



Taxonomic Assessment of Rumen Microbiota Using Total RNA and Targeted Amplicon Sequencing Approaches

Fuyong Li¹, Gemma Henderson², Xu Sun¹, Faith Cox², Peter H. Janssen² and Le Luo Guan^{1*}

¹ Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada, ² AgResearch Ltd., Grasslands Research Centre, Palmerston North, New Zealand

OPEN ACCESS

Edited by:

David Berry,
University of Vienna, Austria

Reviewed by:

Tim Urich,
University of Greifswald, Germany
David William Waite,
Ministry for Primary Industries,
New Zealand
Ilias Lagkouravdos,
Technical University of Munich,
Germany

*Correspondence:

Le Luo Guan
lguan@ualberta.ca

Specialty section:

This article was submitted to
Microbial Symbioses,
a section of the journal
Frontiers in Microbiology

Received: 18 February 2016

Accepted: 08 June 2016

Published: 22 June 2016

Citation:

Li F, Henderson G, Sun X, Cox F, Janssen PH and Guan LL (2016) Taxonomic Assessment of Rumen Microbiota Using Total RNA and Targeted Amplicon Sequencing Approaches. *Front. Microbiol.* 7:987. doi: 10.3389/fmicb.2016.00987

Taxonomic characterization of active gastrointestinal microbiota is essential to detect shifts in microbial communities and functions under various conditions. This study aimed to identify and quantify potentially active rumen microbiota using total RNA sequencing and to compare the outcomes of this approach with the widely used targeted RNA/DNA amplicon sequencing technique. Total RNA isolated from rumen digesta samples from five beef steers was subjected to Illumina paired-end sequencing (RNA-seq), and bacterial and archaeal amplicons of partial 16S rRNA/rDNA were subjected to 454 pyrosequencing (RNA/DNA Amplicon-seq). Taxonomic assessments of the RNA-seq, RNA Amplicon-seq, and DNA Amplicon-seq datasets were performed using a pipeline developed in house. The detected major microbial phylotypes were common among the three datasets, with seven bacterial phyla, fifteen bacterial families, and five archaeal taxa commonly identified across all datasets. There were also unique microbial taxa detected in each dataset. *Elusimicrobia* and *Verrucomicrobia* phyla; *Desulfovibrionaceae*, *Elusimicrobiaceae*, and *Sphaerochaetaceae* families; and *Methanobrevibacter woesei* were only detected in the RNA-Seq and RNA Amplicon-seq datasets, whereas *Streptococcaceae* was only detected in the DNA Amplicon-seq dataset. In addition, the relative abundances of four bacterial phyla, eight bacterial families and one archaeal taxon were different among the three datasets. This is the first study to compare the outcomes of rumen microbiota profiling between RNA-seq and RNA/DNA Amplicon-seq datasets. Our results illustrate the differences between these methods in characterizing microbiota both qualitatively and quantitatively for the same sample, and so caution must be exercised when comparing data.

Keywords: total RNA sequencing, targeted amplicon sequencing, rumen microbiota, bacteria, archaea

INTRODUCTION

Microbiota play essential roles in many ecosystems, including the animal gastrointestinal tract, and have attracted much attention in the past decade due to the understanding of their functions in host productivity and health (Holmes et al., 2012; Million et al., 2013; Yeoman and White, 2014). Numerous studies have shown that changes in gastrointestinal microbiota at the taxonomic and/or functional levels are associated with host dysfunction and metabolic diseases (Marchesi et al., 2016; Ojeda et al., 2016), highlighting the importance of studying the interactions that exist between

gastrointestinal microbiota and host animals. Therefore, an accurate assessment of the composition and diversity of rumen microbiota is essential to link microbiota changes to host performance under different conditions.

To date, culture-independent molecular-based taxonomic assessment of microbiota has primarily relied on the sequencing of PCR amplicons of targeted microbial genes at the DNA level (DNA Amplicon-seq). Although Amplicon-seq has been widely used, it can be biased due to primer selection (Hong et al., 2009) and/or amplification cycling conditions (Huber et al., 2009). It is also limited in discovering novel microbial phylotypes because the associated primers are designed based on known sequences (Urich et al., 2008; Ross et al., 2012). In addition, to study different groups of microbes within the same microbiota, a wide range of primers is needed (Kittelmann et al., 2013). Total DNA sequencing (metagenomics) has also been widely used to study microbiota without PCR amplification, and provides information on the presence and absence of phylotypes, but it mainly offers insight in microbial functions through studying microbiota-associated genes. Although a couple of studies have assessed microbial profiles based on 16S rDNA sequences generated in metagenomics datasets (Ellison et al., 2014; Logares et al., 2014), most metagenomic studies rely on parallel DNA Amplicon-seq to characterize microbial communities (Mason et al., 2014; Rooks et al., 2014) due to the low fraction of 16S rDNA reads present in metagenomics datasets (Logares et al., 2014). Meanwhile, DNA-based methods do not directly measure the activity of the microbiota because they cannot distinguish the presence of genes that stem from active cells, inactive but alive cells, dead cells, or lysed cells (Gaidos et al., 2011).

To overcome these limitations of DNA-based approaches, recent improvements in RNA sequencing have created a great opportunity to study potentially active microbiota. However, RNA sequencing has mainly been applied to elucidate the functions of microbiota through mRNA enrichment (de Menezes et al., 2012; Franzosa et al., 2014) and to study active phylotypes through 16S rRNA amplicon sequencing (RNA Amplicon-seq; Gaidos et al., 2011; Kang et al., 2013). Total RNA sequencing (RNA-seq) has been explored for taxonomic assessment of microbial communities in a number of environments, including soil (Urich et al., 2008; Tveit et al., 2014), hydrothermal vents (Lanzen et al., 2011), and the animal gut (Poulsen et al., 2013; Schwab et al., 2014). However, most of these studies did not compare outcomes between DNA- and RNA-based methods for the same samples and did not compare RNA-seq vs. Amplicon-seq (DNA/RNA), except Berry et al. (2012), who used DNA Amplicon-seq and RNA-seq to study shifts in murine gut microbiota in dextran sodium sulfate (DSS)-induced colitis, and Lanzen et al. (2011), who explored microbial communities at both the DNA and RNA levels in the hydrothermal vents. To date, it is not conclusive which method is the most reliable to assess animal gastrointestinal microbial communities because the different outcomes of these methods have not yet been comprehensively compared.

In this study, we compared bacterial and archaeal community profiles in rumen digesta samples using RNA-seq and RNA/DNA Amplicon-seq with standard protocols and a pipeline developed

in house. The rumen microbial community is complex and includes bacteria, archaea, protozoa and fungi (Edwards et al., 2004). Although Poulsen et al. (2013) used RNA-seq to study rumen microbiota, they mainly focused on methanogens and did not analyze bacteria or compare RNA-seq with Amplicon-seq (DNA and RNA). In the current study, our aim was to gain a better understanding of the differences between the techniques using different genetic materials (DNA vs. RNA) and how they affect interpretation of microbiota-associated data.

MATERIALS AND METHODS

Animals and Sampling

Rumen digesta samples were collected from five 10-month-old crossbred beef steers, which were raised under feedlot conditions on a high-energy finishing diet, as previously described (Hernandez-Sanabria et al., 2013) and followed the guidelines of the Canadian Council on Animal Care (Olfert et al., 1993). The animal protocol was approved by the Animal Care and Use Committee of University of Alberta (protocol no. Moore-2006-55). Animals were not starved before the sampling, and were slaughtered before feeding. For each animal, ~3 g of rumen digesta were collected at slaughter and stored in RNA_{later} (Ambion, Carlsbad, CA, USA) at -20°C for further analysis.

Nucleic Acid Extractions

Total RNA was extracted from rumen digesta using a modified procedure based on the acid guanidinium-phenol-chloroform method (Chomczynski and Sacchi, 1987; Béra-Maillet et al., 2009). Specifically, for ~200 mg of rumen digesta sample, 1.5 ml of TRIzol reagent (Invitrogen, Carlsbad, CA, USA), 0.4 ml of chloroform, 0.3 ml of isopropanol, and 0.3 ml of high salt solution (1.2 M sodium acetate, 0.8 M NaCl) were used. RNA quality and quantity was determined with the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA), respectively. RNA samples with the RNA integrity number (RIN) higher than 7.0 were used for downstream analysis. DNA was extracted from 25 to 30 mg of freeze-dried and ground rumen digesta according to the PCQI method (modified phenol-chloroform with bead beating and QIAquick PCR purification kit; Rius et al., 2012; Henderson et al., 2013).

RNA Library Construction and Sequencing (RNA-seq)

Total RNA (5 μl of 20 ng/ μl) from each sample was used to construct an RNA library following the TruSeq RNA sample Prep v2 LS protocol (Illumina, San Diego, CA, USA), without the mRNA enrichment (rRNA removal) step. The quality and concentration of cDNA fragments (~260 bp) containing artificial sequences (adapters, index sequences, and primers; ~120 bp) and inserted cDNA sequences (~140 bp) were assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies) and a Qubit 2.0 fluorometer (Invitrogen), respectively, before sequencing. RNA libraries were paired-end sequenced (2×100 bp) using an Illumina HiSeq2000 platform (McGill University and Génome Québec Innovation Centre, QC, Canada).

Amplicon-seq of 16S rRNA/rDNA Using Pyrosequencing (RNA/DNA Amplicon-seq)

For the DNA Amplicon-seq, partial bacterial and archaeal 16S rRNA genes (the V1-V3 region for bacteria and the V6-V8 region for archaea) were amplified as previously described by Kittelmann et al. (2013) and sequenced using 454 GS FLX Titanium chemistry at Eurofins MWG Operon (Ebersberg, Germany). For the RNA Amplicon-seq, total RNA was first reverse-transcribed into cDNA using SuperScript II reverse transcriptase (Invitrogen) with random primers following procedures for first-strand cDNA synthesis. Then, partial 16S rRNA amplicons of bacteria and archaea were generated using the same primers as for the DNA Amplicon-seq and sequenced using a 454 pyrosequencing platform at McGill University and Génome Québec Innovation Centre (Montreal, QC, Canada).

Analysis of the RNA-seq Dataset

The sequence data quality was checked using the FastQC program (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The program Trimmomatic (version 0.32; Bolger et al., 2014) was used to trim residual artificial sequences, cut bases with quality scores below 20, and remove reads shorter than 50 bp. The filtered reads were then sorted to enrich 16S rRNA fragments for taxonomic identification and mRNA reads for functional analysis (not reported in this study) using SortMeRNA (version 1.9; Kopylova et al., 2012) by aligning with the rRNA reference databases SILVA_SSU (release 119), SILVA_LSU (release 119; Quast et al., 2013), and Rfam 11.0 (Burge et al., 2013). After the 16S rRNA sequences were enriched, downstream analyses were performed using the mothur program (version 1.31.2; Schloss et al., 2009) according to the procedures (http://www.mothur.org/wiki/MiSeq_SOP) described by Kozich et al. (2013), with modifications. For taxonomic identification, regionally enriched reference datasets were built for bacteria and archaea (mothur command: pcr.seqs). Specifically, sequences belonging to the V1-V3 region (mean length: 466 bp) were extracted from the aligned Greengenes 16S rRNA gene database (version gg_13_5_99 accessed May 2013; DeSantis et al., 2006) for bacteria. For archaea, sequences belonging to the V6-V8 region (mean length: 456 bp) were extracted from the aligned rumen-specific archaeal 16S rRNA gene database derived from a previous study (Janssen and Kirs, 2008). The starting and ending positions of the targeted regions were located based on the alignment of Amplicon-seq reads to the reference databases (Figure 1) because these amplicons were generated using designed primers for known regions.

The overall RNA-seq data analysis pipeline is illustrated in Figure 1. Briefly, sorted paired-end reads belonging to bacterial and archaeal 16S rRNA were joined (mothur command: make.contigs) to increase the length by combining the forward and reverse sequences. Joined sequences (mean length: 140 bp) with ambiguous bases or longer than 200 bp were discarded (mothur command: screen.seqs) to remove sequences without overlapped regions. Identical sequences were then binned to generate a set of unique sequences to facilitate the counting

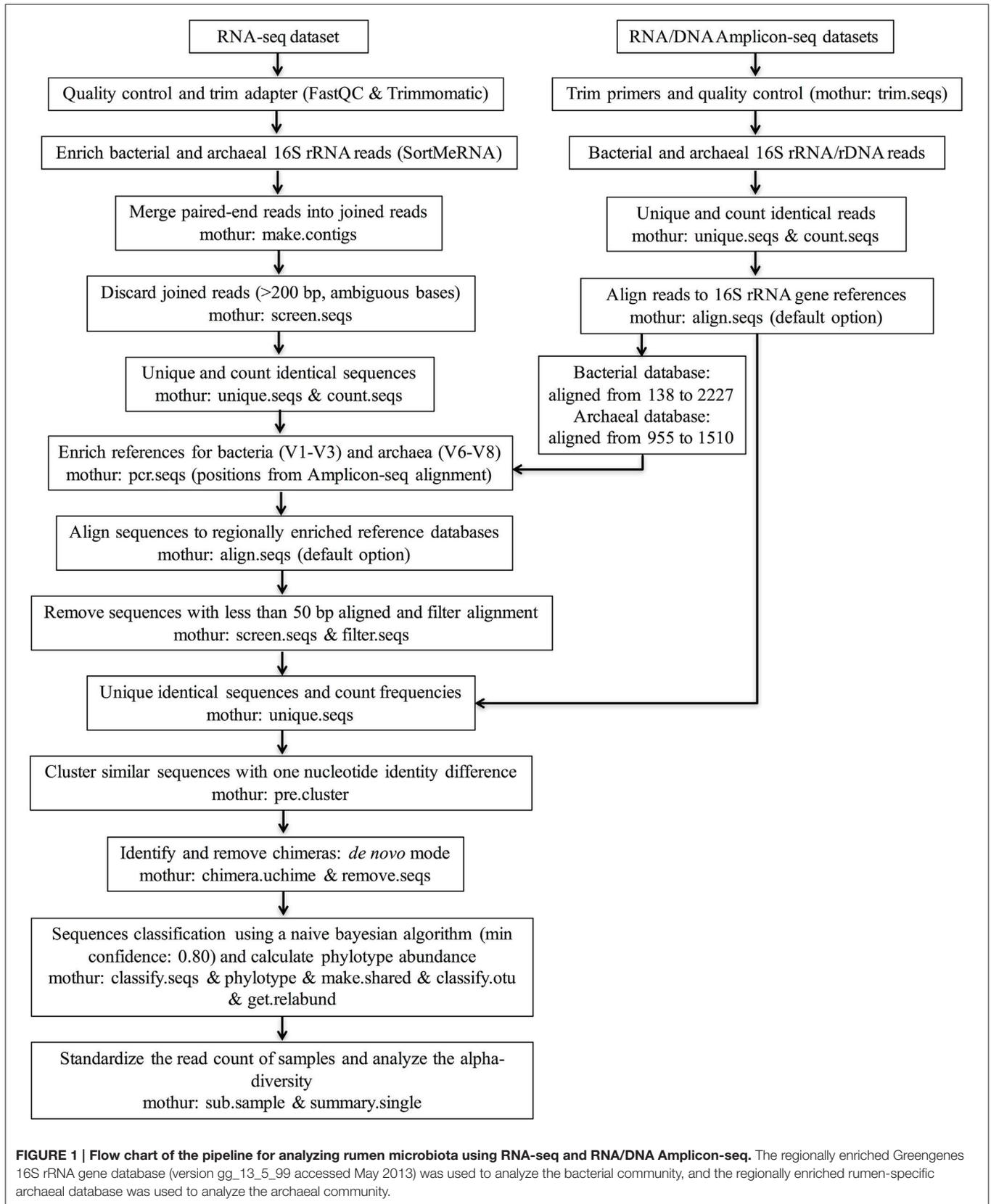
of their frequencies in each sample (mothur commands: unique.seqs and count.seqs). Next, the bacterial and archaeal 16S rRNA sequences in the sample datasets were aligned to the regionally enriched bacterial (the V1-V3 region) and archaeal (the V6-V8 region) references (see above), respectively, (mothur command: align.seqs with default option). Sequences for which less than 50 bp aligned to the reference datasets were culled (mothur command: screen.seqs). After alignment filtering (mothur command: filter.seqs), combining the identical sequences, and counting the frequencies (mothur command: unique.seqs), pre-clustering was performed to decrease the complexity of our sample datasets by clustering highly similar sequences with one nucleotide identity difference (mothur command: pre.cluster). UCHIME (Edgar et al., 2011) in *de novo* mode and with default settings was applied to identify and remove chimeras (mothur command: chimera.uchime & remove.seqs). Finally, chimera-depleted sequences were subject to taxonomic assignment to different phylotypes using a naive Bayesian algorithm (Wang et al., 2007) with a minimum confidence of 0.8 (mothur command: classify.seqs), and then the taxonomic rank and the relative abundance for each phylotype was calculated (mothur commands: phylotype, make.shared, classify.otu, and get.relabund). To make alpha-diversity estimators comparable among samples and different methods, all samples were standardized to the same number of sequences (the smallest sampling size) by randomly selecting sequences from the chimera-depleted datasets (mothur command: sub.sample). Alpha-diversity analysis was conducted at the bacterial family level and archaeal species level. The Good's coverage, the Shannon index, the inverse Simpson index, the number of observed phylotypes, and the Chao estimator were calculated based on the normalized samples (mothur command: summary.single).

Analysis of RNA/DNA Amplicon-seq Datasets

The procedures were similar to those described for the RNA-seq dataset (Figure 1). Briefly, after trimming primers and screening homopolymer runs (maximum length: 6), only sequences over 200 bp in length with an average quality score over 25 and with less than 6 ambiguous bases were included in the analysis. This step was performed using mothur (version 1.31.2; Schloss et al., 2009) with the command *trim.seqs*. After clustering similar sequences, the chimeras were checked and discarded from reads with good quality. The chimera-depleted reads were used for taxonomic identification and to calculate the relative abundance of each phylotype. The alpha-diversity was analyzed using the standardized chimera-depleted sequences according to the lowest number of reads.

Validation of Microbial Relative Abundances Using qRT-PCR and qPCR

Quantitative reverse transcription PCR (qRT-PCR) and quantitative PCR (qPCR) were further performed to validate the relative abundance data obtained from RNA-seq, RNA Amplicon-seq, and DNA Amplicon-seq. Three primer pairs



(Supplementary Table 2) were used to enumerate the total bacteria, *Bacteroidetes*, and *Gammaproteobacteria* in each rumen sample. Standard curves were constructed using serial dilutions of plasmid DNA from clones identified as *Butyrivibrio hungatei* (for total bacteria, using an initial concentration of 8.50×10^8 mol/ μ l), *Prevotella* sp. (for *Bacteroidetes*, using an initial concentration of 2.89×10^8 mol/ μ l), and *Tolomonas auensis* (for *Gammaproteobacteria*, using an initial concentration of 2.68×10^8 mol/ μ l). The copy numbers for each standard curve were calculated as described previously (Li et al., 2009). For qRT-PCR, cDNA was first reverse-transcribed from 20 ng of total RNA using iScript reverse transcription supermix for RT-qPCR (Bio-Rad, Hercules, CA, USA) and then diluted 5 times. One microliter of diluted cDNA was subjected to a qRT-PCR reaction using SYBR Green chemistry (Fast SYBRH Green Master Mix; Applied Biosystems) in a StepOnePlus Real-Time PCR System (Applied Biosystems). The relative abundances of *Bacteroidetes* and *Gammaproteobacteria* compared to total bacteria were calculated according to the following equation: relative abundance = Q_{Target}/Q_{U2} , where Q_{Target} was the quantity of each target, and Q_{U2} was the total bacteria quantity. Concurrently, we performed qPCR using total DNA (1 μ l \times 10 ng/ μ l total DNA per reaction) and followed the same procedures mentioned above to verify DNA Amplicon-seq results.

Statistical Analysis

In this study, only taxa with a relative abundance $> 0.1\%$ in at least two samples within the RNA-seq, RNA Amplicon-seq, and DNA Amplicon-seq datasets were defined as detectable and subjected to downstream statistical analysis. Statistical summaries (mean and SEM) of the detected taxa, and Pearson's correlation analysis for qRT-PCR/qPCR validation were all performed using R 3.1.2 (R Core Team, 2014). Principal coordinate analysis (PCoA) of the microbial profiles generated from the different datasets was conducted based on the Bray-Curtis dissimilarity matrix. The microbial relative abundance was arcsine-square-root transformed (Franzosa et al., 2014), and then Repeated Measures ANOVA was performed to compare the differences among three datasets. *P* values were adjusted into FDR using Benjamini-Hochberg method (Benjamini and Hochberg, 1995), and a threshold of $FDR < 0.15$ was applied to determine the significance (Korpela et al., 2016). Co-occurrence analysis was performed for the bacterial families and archaeal taxa detected in both the RNA-seq and DNA Amplicon-seq datasets (with relative abundance $> 0.1\%$ in all five samples) based on Spearman's rank correlation (Barberan et al., 2012). An association network was constructed using CoNet (Faust et al., 2012) and displayed using Cytoscape 3.2.1 (Shannon et al., 2003; Faust et al., 2012). Effective alpha-diversity estimators (Jost, 2007) were compared among the RNA-seq, RNA Amplicon-seq, and DNA Amplicon-seq datasets using paired Wilcoxon signed rank test.

Data Submission

RNA-seq and RNA Amplicon-seq datasets were submitted into the NCBI Sequence Read Archive (SRA) under the accession number PRJNA275012, and DNA Amplicon-seq sequences were

also placed in the NCBI SRA under the accession number PRJNA273417.

RESULTS AND DISCUSSION

Analyzing Rumen Microbiota Using RNA-seq and RNA/DNA Amplicon-seq

This study assessed active rumen microbial communities using RNA-seq and is the first study to compare RNA-seq outcomes with the well-accepted Amplicon-seq methods to evaluate rumen microbiota. It has been reported that rRNA levels directly relate to the protein synthesis potential of microorganisms (Blazewicz et al., 2013) and are correlated with activity (Poulsen et al., 1993; Bremer and Dennis, 2008). Therefore, rRNA abundance data obtained from total RNA sequencing could potentially be used as one of the indices to taxonomically assess potentially active microbes within a sample. To explore the possibility of taxonomic profiling using total RNA-seq, we firstly enriched 16S rRNA sequences from the RNA-seq dataset (Figure 1). In total, an average of $38,496,238 \pm 2,037,011$ (mean \pm SEM) reads per sample ($192,481,188$ reads in total) were obtained after quality control filtration. Among them, $92.9 \pm 1.1\%$ belonged to small and large subunit rRNA, with $13.7 \pm 5.6\%$ and $0.2 \pm 0.0\%$ being bacterial and archaeal 16S rRNA, respectively, (Supplementary Table 1). It is notable that a large fraction of rRNA was classified as eukaryotic 18S ($22.1 \pm 5.9\%$) and 28S ($32.2 \pm 8.2\%$) rRNA (Supplementary Table 1). Although these reads were not analyzed in the current study, the high number of these sequences indicates the possibility of assessing rumen eukaryotic microbiota using RNA-seq in future studies. After combining paired-end reads and removing non-overlapping sections, $10,782,833$ bacterial and $152,585$ archaeal joint 16S rRNA sequences were obtained.

In this study, we included regionally enriched 16S rRNA gene reference datasets for taxonomic analysis rather than aligning the sequences to the full-length 16S rRNA gene database directly. Because the length of Illumina RNA-seq reads is short, 16S rRNA sequences could be randomly aligned to different regions of the 16S rRNA gene. It is known that different hypervariable regions of the 16S rRNA gene can affect diversity estimation and taxonomic classification (Logares et al., 2014). If these short 16S rRNA sequences are directly mapped to full-length 16S rRNA gene references for taxonomic analysis, as has been performed in previous studies (Urich et al., 2008; Tveit et al., 2014), it could lead to a mixed taxonomic profile as well as an overestimation of diversity. To avoid such potential bias, two regionally enriched reference datasets were generated from the Greengenes 16S rRNA gene database (version gg_13_5_99 accessed May 2013) (the V1-V3 region for bacteria) and the rumen-specific archaeal 16S rRNA database (Janssen and Kirs, 2008; the V6-V8 region for archaea; see details in the Materials and Methods Section). These regions were chosen because the V2-V3 region is the most efficient region for assessing bacterial community (Chakravorty et al., 2007), while the V6-V8 region is the most efficient region for identifying archaea and estimating the archaeal community diversity (Snelling et al., 2014). After identifying sequences belonging to the bacterial V1-V3 region and the archaeal V6-V8 region, sequences aligned with more

than 50 bp of the reference datasets were then subjected to chimeric sequence detection (2,423,139 bacterial and 25,451 archaeal). After the removal of 9353 bacterial sequences (0.4%) and 139 archaeal sequences (0.6%) through chimera checking, 2,413,786 of the V1-V3 region-enriched bacterial sequences (mean length: 124 bp) and 25,312 of the V6-V8 region-enriched archaeal sequences (mean length: 133 bp) were subject to further taxonomic analysis (Table 1).

RNA and DNA Amplicon-seq of the same rumen samples generated 37,105 (7421 ± 506; mean ± SEM; RNA Amplicon-Seq dataset) and 31,031 (6206 ± 645; DNA Amplicon-Seq dataset) bacterial reads, respectively, as well as 8303 (1661 ± 20 for RNA Amplicon-Seq) and 6663 (1333 ± 95 for DNA Amplicon-Seq) archaeal reads, respectively, (Supplementary Table 1). From these two datasets, 6461/5505 bacterial reads (17.4/17.7%) and 562/662 archaeal reads (6.8/9.3%) were detected and removed as chimeric sequences. In total, 30,644/25,526 bacterial reads (mean length: 476/487 bp) and 7741/6041 archaeal reads (mean length: 451/462 bp) were used for taxonomic identification and quantification (Table 1).

In this study, we used the *de novo* (database-independent) mode to determine chimeras rather than reference-based chimera detection. This is because the existing chimera reference databases only contain sequences from cultured organisms (Haas et al., 2011) and are not suitable for real samples that contain uncultured bacteria and archaea. Notably, higher percentages of chimeric sequences (17.4/17.7% of the bacterial and 6.8/9.3% of the archaeal sequences) were removed from the RNA/DNA Amplicon-seq datasets than from the RNA-seq dataset, which had only 0.4% of their bacterial and 0.6% of their archaeal sequences removed due to the presence of chimeras. In Amplicon-Seq datasets, chimeras are produced during PCR amplification, and they can lead to biased estimation of the diversity and/or the identification of differences between microbial communities (Edgar et al., 2011), while in RNA-seq datasets, chimeras might stem from the cDNA synthesis and/or fragment enrichment procedures used during the library construction. Our results indicate that RNA-seq was less affected by chimera formation than Amplicon-seq.

Microbial Taxa Detected from RNA-seq and RNA/DNA Amplicon-seq

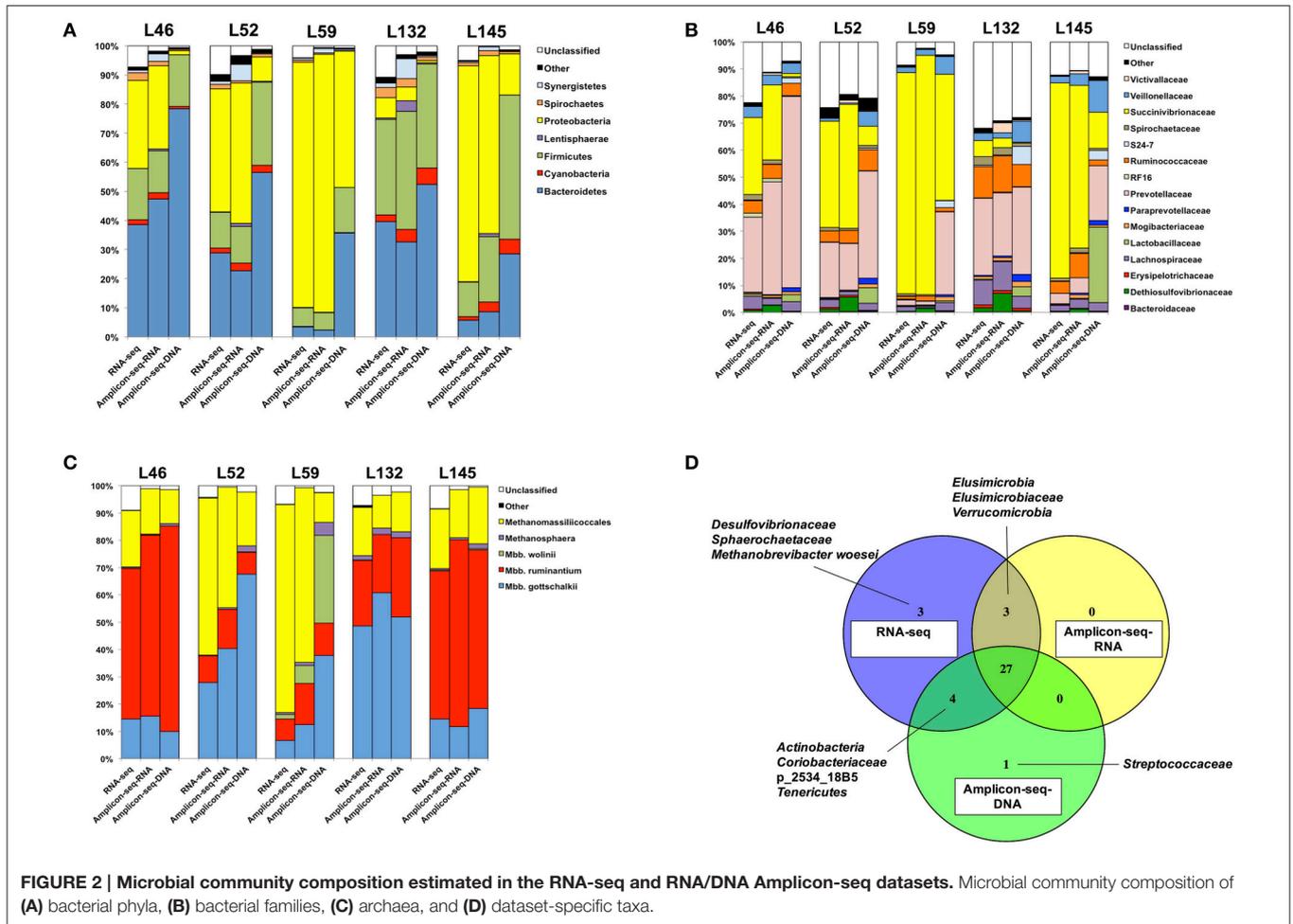
From the RNA-seq dataset, 94.6 and 86.9% of the bacterial V1-V3 region-enriched sequences were classified at the phylum and family level, respectively (Table 1). Due to their short sequence lengths, 86.2% of bacterial sequences from the RNA-seq dataset

could not be classified at the genus level. Thus, only bacterial taxa at the phylum and family levels were retained for further analysis. For archaea, 98.0% of the V6-V8 region-enriched sequences were classified in a mixed taxonomic rank scheme (Table 1). We classified archaeal sequences at different taxonomic levels because most of the predominant archaeal phylotypes (such as *Methanobrevibacter ruminantium* and *Methanobrevibacter gottschalkii*) have well-studied 16S rRNA genes for use as references (Janssen and Kirs, 2008), and even short reads could be classified at the species level. However, for the poorly studied groups, such as *Methanomassiliicoccales*, reads could be only classified at the order level based on this rumen-specific archaeal database (Janssen and Kirs, 2008). From RNA/DNA Amplicon-seq datasets, 98.2/98.8%, 86.8/87.1%, and 98.6/99.7% of the total reads were assigned at the bacterial phylum, bacterial family and archaeal mixed taxa levels, respectively (Table 1).

The bacterial and archaeal taxa detected in the three datasets were generally similar, with a total of 11 bacterial phyla, 21 bacterial families and six archaeal taxa identified. Of these, seven bacterial phyla, 15 bacterial families and five archaeal taxa were commonly detected across the three datasets (Figure 2). Notably, there were unique bacterial and archaeal taxa identified in each dataset (Figure 2D). Firstly, two bacterial families and one archaeal taxon (*Desulfovibrionaceae*, *Sphaerochaetaceae*, and *Methanobrevibacter woesei*) were detected only in the RNA-seq dataset and not in the RNA/DNA Amplicon-seq datasets. Henderson et al. (2015) also reported the absence of *Desulfovibrionaceae* and *Methanobrevibacter woesei*, and the low abundance ($\leq 0.1\%$) of *Sphaerochaetaceae* in the rumen digesta when using DNA Amplicon-seq with the same PCR primers, suggesting that the amplicon-based approach and/or the primers used may have masked the detection of these taxa in the rumen. In addition, lower sequencing depth of RNA/DNA Amplicon-seq could also lead to the missing detection of these taxa. As shown in Table 1, the number of bacterial and archaeal reads from the RNA-seq dataset was about 80–90 times higher and 3–4 times higher than that from RNA/DNA Amplicon-seq datasets, respectively. And thus, increasing sequencing depth could probably enhance the detection of these taxa in the Amplicon-seq datasets. Secondly, two bacterial phyla (*Elusimicrobia* and *Verrucomicrobia*) and one bacterial family (*Elusimicrobiaceae*) were detected only in the RNA-based datasets (RNA-seq and RNA Amplicon-seq) and not in the DNA Amplicon-seq dataset. Our results suggest that these two phyla may be more active in the rumen, and they may be underestimated based on the DNA Amplicon-seq dataset. Moreover, the absence of

TABLE 1 | Summary of sequences used for taxonomic analysis from the chimera-depleted RNA-seq and Amplicon-seq datasets.

	RNA-seq			RNA Amplicon-seq			DNA Amplicon-seq		
	No. of reads	Classified (%)	Unclassified (%)	No. of reads	Classified (%)	Unclassified (%)	No. of reads	Classified (%)	Unclassified (%)
Bacteria (phylum)	2,413,786	94.6	5.4	30,644	98.2	1.8	25,526	98.8	1.2
Bacteria (family)		86.9	13.1		86.8	13.2		87.1	12.9
Archaea (mixed)	25,312	98.0	2.0	7,741	98.6	1.4	6,041	99.7	0.3

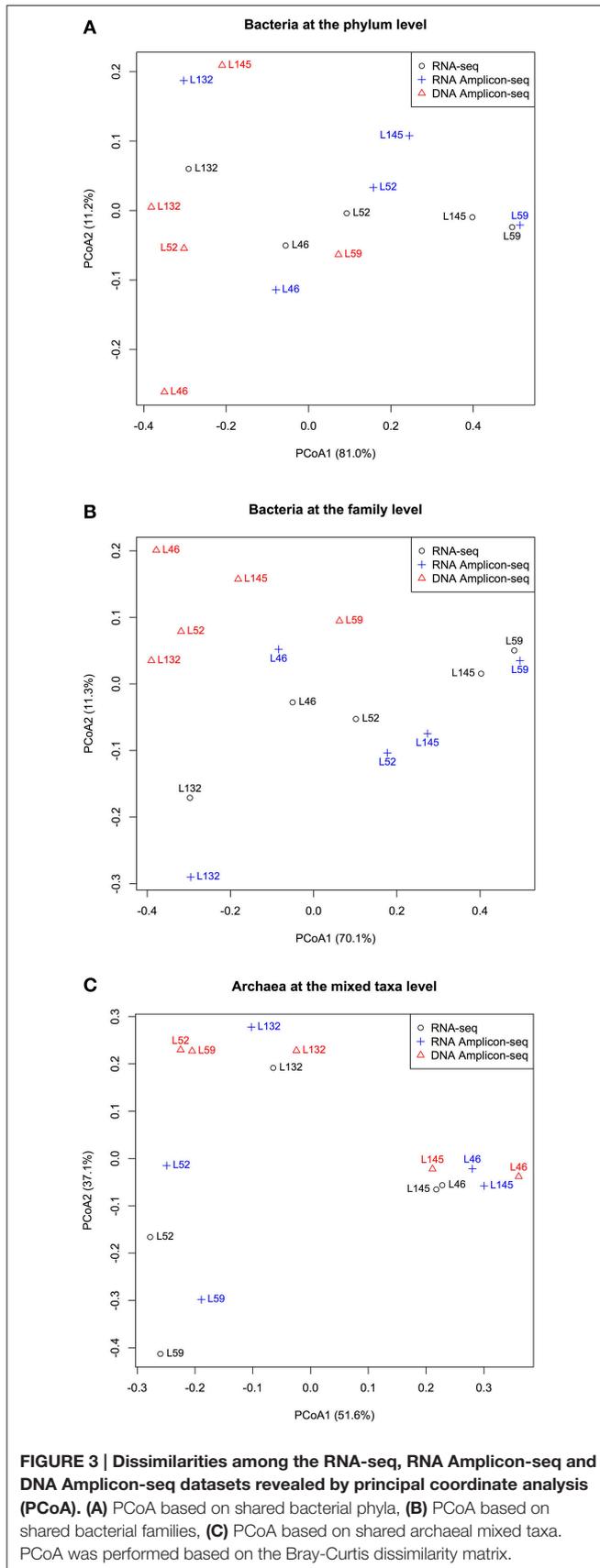


Elusimicrobia and *Verrucomicrobia* in the DNA Amplicon-seq dataset may be due to the unsuccessful isolation of their DNA, and it has been reported that various DNA extraction methods could impact the taxonomic outcomes of rumen microbiota assessments (Henderson et al., 2013). Finally, the bacterial family *Streptococcaceae* was detected only in the DNA Amplicon-seq dataset with a low relative abundance of $0.1 \pm 0.1\%$. Previous studies on the bacterial profiles of rumen digesta from the same cattle used in this study (Hernandez-Sanabria et al., 2012) and different cattle (Petri et al., 2013; Xia et al., 2015) have also reported the absence of *Streptococcaceae*. In a recent study based on a large number ($n = 742$) of rumen and foregut digesta samples and DNA-based Amplicon-seq, *Streptococcaceae* also showed low prevalence in all animals (Henderson et al., 2015). These suggest that *Streptococcaceae* may have low cellular abundance and even lower activities in samples assessed in the current study.

Microbial Relative Abundances Estimated from RNA-seq and RNA/DNA Amplicon-seq

Principal coordinate analysis (PCoA) of the relative abundances of commonly detected bacterial phyla, bacterial families, and

archaeal taxa revealed dissimilarities in microbial profiles between RNA- and DNA-based approaches (Figure 3). For each animal, the RNA-seq and RNA Amplicon-seq generated similar rumen bacterial profiles (at both the phylum and family levels), which generally separated with that assessed using DNA Amplicon-seq (Figures 3A,B). However, for each animal, three datasets displayed generally similar archaeal profiles (Figure 3C); the RNA-based assessment outcomes of L52 and L59 were distinct from their DNA-based profiles because these two samples had a high abundance of *Methanomassiliococcales* in the RNA-based datasets (Figure 2C). Among the shared taxa, four bacterial phyla (*Bacteroidetes*, *Lentisphaerae*, *Proteobacteria*, and *Synergistetes*), eight bacterial families (*Dethiosulfovibrionaceae*, *Lactobacillaceae*, *Mogibacteriaceae*, *Paraprevotellaceae*, *Prevotellaceae*, *S24-7*, *Succinivibrionaceae*, and *Victivallaceae*), and one archaeal taxon (*Methanomassiliococcales*) had significantly different relative abundances among the three datasets (FDR < 0.15; Table 2). Lanzen et al. (2011) reported that the dominant taxa within hydrothermal vent field microbiota showed similar outcomes on the RNA and DNA levels based on their Amplicon-seq results, which is not consistent with our findings and is probably an extremely environment-specific case. In addition, such



discrepancy may also be due to different targeted amplicon regions (the V5-V6 region in Lanzen et al. (2011) vs. the V1-V3/V6-V8 regions in the current study) as well as different reference databases used (the Silva SSURef in Lanzen et al. (2011) vs. the regionally enriched Greengenes/rumen-specific archaeal databases in our study). Therefore, different methods and strategies should be carefully considered for samples from various environmental conditions.

The dominant bacterial phyla detected by DNA Amplicon-seq were *Bacteroidetes* ($50.3 \pm 8.7\%$; mean \pm SEM), *Firmicutes* ($29.4 \pm 6.2\%$), and *Proteobacteria* ($14.3 \pm 8.5\%$), which is consistent with previous studies using DNA-based methods. For example, *Bacteroidetes* (range, 8.0–60.1%) and *Firmicutes* (range, 33.6–85.0%) were reported as the most abundant phyla, and *Proteobacteria* was commonly detected but less abundant (range, 0.6–20.1%) in the rumen (Jami and Mizrahi, 2012; Petri et al., 2013; Kim and Yu, 2014). However, the predominant bacterial phylum detected by RNA-based approaches (RNA-seq and RNA Amplicon-seq) was *Proteobacteria* (47.6 ± 14.2 and $46.3 \pm 14.3\%$, respectively), followed by *Bacteroidetes* (23.3 ± 7.9 and $22.7 \pm 8.1\%$), and *Firmicutes* (16.2 ± 4.5 and $19.2 \pm 6.0\%$). The higher proportion of *Proteobacteria* in the RNA-based datasets confirmed similar findings by Kang et al. (2013) and Kang et al. (2009), who applied RNA amplicon-based sequencing and rRNA-based clone libraries, respectively, to study the rumen microbiota. At the bacterial family level, the most abundant bacterial family was *Succinivibrionaceae* (belonging to the phylum *Proteobacteria*) in the RNA-seq ($45.6 \pm 14.0\%$) and RNA Amplicon-seq ($45.1 \pm 14.4\%$) datasets, while it was *Prevotellaceae* (belonging to the *Bacteroidetes*) in the DNA Amplicon-seq dataset ($38.7 \pm 8.6\%$). *Succinivibrionaceae* was an abundant family at the DNA level when ruminants were fed high-energy diets (Hernandez-Sanabria et al., 2012; Petri et al., 2013; Henderson et al., 2015), and the significance of *Succinivibrionaceae* may be underestimated using DNA Amplicon-seq. The predominance of *Prevotellaceae* detected in the DNA Amplicon-seq dataset is similar to that observed in previously studies using DNA-based approaches (Kittelman et al., 2013; Petri et al., 2013; Henderson et al., 2015). However, its abundance, as estimated in the RNA-seq ($16.4 \pm 5.7\%$) and RNA Amplicon-seq ($17.9 \pm 7.2\%$) datasets, was significantly lower (FDR < 0.15). The family *Lactobacillaceae*, belonging to the phylum *Firmicutes*, also had a lower abundance in the RNA-based datasets ($0.3 \pm 0.1\%$ using RNA-seq and $0.2 \pm 0.1\%$ using RNA Amplicon-seq) than in the DNA Amplicon-seq dataset ($8.1 \pm 5.0\%$, FDR < 0.15). These two families probably had higher cellular abundances but relatively lower activities in the bovine rumen.

From the RNA/DNA Amplicon-seq datasets, *Methanobrevibacter gottschalkii* ($28.2 \pm 9.7/37.1 \pm 10.6\%$), *Methanobrevibacter ruminantium* ($37.1 \pm 12.4/36.4 \pm 13.1\%$), and *Methanomassiliicoccales* ($30.9 \pm 10.0/15.7 \pm 2.0\%$) were dominant but with different rankings. The DNA Amplicon-seq outcomes are generally consistent with previous studies that used the same approaches (Kittelman et al., 2013; Henderson et al., 2015). The archaeal taxon *Methanomassiliicoccales*, which has been previously referred to as Rumen Cluster C or

TABLE 2 | Differential taxa among the RNA-seq and RNA/DNA Amplicon-seq datasets¹.

Taxonomy level	Classification	RNA-seq (%) (Mean ± SEM)	RNA Amplicon-seq (%)	DNA Amplicon-seq (%)	FDR ²
BACTERIA					
Phyla	<i>Bacteroidetes</i>	23.3 ± 7.9 ^a	22.7 ± 8.1 ^a	50.3 ± 8.7 ^b	0.09
Family	<i>Paraprevotellaceae</i>	0.2 ± 0.1 ^a	0.3 ± 0.1 ^a	1.8 ± 0.3 ^b	0.09
Family	<i>Prevotellaceae</i>	16.4 ± 5.7 ^a	17.9 ± 7.2 ^a	38.7 ± 8.6 ^b	0.14
Family	S24-7	0.2 ± 0.1 ^a	0.2 ± 0.1 ^a	3.1 ± 1.0 ^b	0.09
Phyla	<i>Firmicutes</i> ³	16.2 ± 4.5	19.2 ± 6.0	29.4 ± 6.2	0.35
Family	<i>Lactobacillaceae</i>	0.3 ± 0.1 ^a	0.2 ± 0.1 ^a	8.1 ± 5.0 ^b	0.09
Family	<i>Mogibacteriaceae</i>	0.5 ± 0.1 ^a	0.8 ± 0.2 ^{ab}	1.3 ± 0.2 ^b	0.09
Phyla	<i>Lentisphaerae</i>	0.2 ± 0.1 ^a	1.3 ± 0.6 ^b	0.1 ± 0.1 ^a	0.09
Family	<i>Victivallaceae</i>	0.2 ± 0.1 ^a	1.3 ± 0.6 ^b	0.1 ± 0.1 ^a	0.09
Phyla	<i>Proteobacteria</i>	47.6 ± 14.2 ^a	46.3 ± 14.3 ^a	14.3 ± 8.5 ^b	0.09
Family	<i>Succinivibrionaceae</i>	45.6 ± 14.0 ^a	45.1 ± 14.4 ^a	13.8 ± 8.6 ^b	0.09
Phyla	<i>Synergistetes</i>	0.9 ± 0.2 ^a	3.6 ± 1.1 ^b	0.3 ± 0.1 ^c	0.12
Family	<i>Dethiosulfovibrionaceae</i>	0.8 ± 0.2 ^a	3.6 ± 1.1 ^b	0.3 ± 0.1 ^c	0.12
ARCHAEA					
Mixed	<i>Methanomassiliicoccales</i>	38.8 ± 11.9 ^a	30.9 ± 10.0 ^b	15.7 ± 2.0 ^{ab}	0.09

¹Only commonly detected phylotypes among three datasets were compared.

²*P* values were obtained using Repeated Measures ANOVA based on the arcsine-square-root transformed proportion values, and then were adjusted into FDR using Benjamini-Hochberg method (Benjamini and Hochberg, 1995). A threshold of FDR < 0.15 was applied to determine the significance. Within a row, means with different superscript are significantly different.

³The relative abundance of the bacterial phylum Firmicutes did not show differences, but two families belonged to this phylum were different among datasets.

Thermoplasmatales (Janssen and Kirs, 2008; Poulsen et al., 2013; Gaci et al., 2014), was predominant in the RNA-seq dataset (38.8 ± 11.9%), followed by *Methanobrevibacter ruminantium* (30.2 ± 10.4%), and *Methanobrevibacter gottschalkii* (22.4 ± 7.4%). The high proportion of *Methanomassiliicoccales* from the RNA-based datasets supports the hypothesis that they are more active in the rumen, as many studies have suggested (Ohene-Adjei et al., 2007; Wright et al., 2007; Janssen and Kirs, 2008; Williams et al., 2009; Jeyanathan et al., 2011).

qRT-PCR and qPCR Validation of Microbial Relative Abundances among Datasets

qRT-PCR and qPCR were performed to estimate the relative abundances of two predominant phyla (*Bacteroidetes* and *Proteobacteria*) among all three datasets. *Gammaproteobacteria* was selected to represent *Proteobacteria* because 95.8, 98.0, and 96.5% of the *Proteobacteria* reads from the three datasets belonged to the class *Gammaproteobacteria*. The qRT-PCR results for *Bacteroidetes* were in agreement with the RNA-seq (Pearson's correlation coefficient [*r*] = 0.97, *P* < 0.05) and the RNA Amplicon-seq (*r* = 0.97, *P* < 0.05) results, and the qPCR results for *Bacteroidetes* were also consistent with the DNA Amplicon-seq results (*r* = 0.88, *P* = 0.05; Supplementary Figure S1). The relative abundance of *Gammaproteobacteria* estimated using qRT-PCR was correlated with that from the RNA-seq dataset (*r* = 0.97, *P* < 0.05) and the RNA Amplicon-seq dataset (*r* = 0.99, *P* < 0.05), and there was also a high degree of correlation between qPCR and DNA Amplicon-seq (*r* = 0.99, *P* < 0.05) for *Gammaproteobacteria*. The overall consistent trends between the RNA-based approaches and qRT-PCR (as well as between

DNA Amplicon-seq and qPCR) confirm the different relative abundance detected in the three datasets.

Alpha-Diversity Estimators in the RNA-seq and RNA/DNA Amplicon-seq Datasets

In the present study, alpha-diversity indices were estimated based on the observed phylotypes at the family level for bacteria and at the species level for archaea. To avoid potential differences caused by different sequencing depths among the three datasets, the samples were randomly normalized according to the lowest number of reads (3476 bacterial sequences and 1074 archaeal sequences per sample after normalization). The values of Good's coverage were all above 99% for the bacterial and archaeal data from the three datasets, indicating that the numbers of reads after normalization were sufficient to represent the microbial communities. The Shannon index and the inverse Simpson index were not significantly different (*P* > 0.1, the paired Wilcoxon signed rank test) among the three datasets (Table 3). The number of observed phylotypes and the Chao estimator tended to be higher in the RNA-seq dataset than in the RNA/DNA Amplicon-seq datasets for bacteria and archaea (*P* < 0.1, the paired Wilcoxon signed rank test; Table 3), which was further confirmed using rarefaction analysis (Supplementary Figure S2). These results suggest that more microbial taxa could be detected using RNA-seq than using RNA/DNA Amplicon-seq. In the Amplicon-seq datasets, some phylotypes were probably overlooked due to the bias of primers and/or amplification conditions during the PCR process, which may explain the difference in richness among the three datasets.

TABLE 3 | A comparison of alpha-diversity estimators among the RNA-seq and RNA/DNA Amplicon-seq datasets.

	Bacteria			Archaea		
	RNA-seq	RNA Amplicon-seq (Mean ± SEM)	DNA Amplicon-seq	RNA-seq	RNA Amplicon-seq (Mean ± SEM)	DNA Amplicon-seq
Number of observed phylotypes	45.4 ± 2.8 ¹	33.2 ± 2.9 [#]	35.6 ± 3.1 [#]	11.8 ± 0.4 ^a	6.2 ± 0.2 ^b	7.6 ± 0.5 ^c
Chao	52.1 ± 3.8 [*]	43.0 ± 7.3 ^{*#}	40.4 ± 3.3 [#]	12.7 ± 0.5 ^a	6.2 ± 0.2 ^b	8.4 ± 1.0 ^b
Shannon ²	1.9 ± 0.3	1.8 ± 0.3	2.0 ± 0.2	1.2 ± 0.1	1.0 ± 0.0	1.1 ± 0.1
Inverse Simpson	4.3 ± 1.2	4.2 ± 1.3	4.8 ± 1.0	2.5 ± 0.2	2.3 ± 0.1	2.5 ± 0.3
Good's coverage	99.7%	99.8%	99.8%	99.8%	100%	99.9%

¹ Within a row, means with different superscript tend to be different at $P < 0.1$. Comparison was conducted using paired Wilcoxon signed rank test for bacterial and archaeal communities separately, and thus estimators between bacterial and archaeal groups are not comparable.

² Shannon indices showed in the table are the raw values, and the comparison of Shannon indices among datasets was based on the exponentially transformed values (Jost, 2007) using paired Wilcoxon signed rank test.

Co-Occurrence Analysis of Abundant Microbial Taxa Detected from RNA-seq and DNA Amplicon-seq

The relationships that exist among microbial taxa could be a determining factor for microbial community composition (Prosser et al., 2007). To explore the relationships among different taxa in our samples, Spearman's rank correlation was used to identify the co-occurrence patterns of different microbial groups in both the RNA-seq and DNA Amplicon-seq datasets (Figure 4). The RNA Amplicon-seq dataset was not included due to its similar outcomes to the RNA-seq dataset, with high correlation for all samples (Spearman's rank correlation coefficient [ρ] = 0.88 – 0.98, $P < 0.0001$; Figures 2, 3). As shown in Figure 4, the microbial taxa identified using RNA-seq were more closely correlated than those identified using DNA Amplicon-seq. There may be stronger interactions among microbes within a microbiota at the transcriptional level than that at the genomic level, which has been suggested recently by Inceoğlu et al. (2015). Previous studies have revealed associations among microbes (*Ruminococcaceae* and *Mbb. gottschalkii*, *Succinivibrionaceae* and *Methanomassiliicoccales*, *Mbb. gottschalkii* and *Mbb. ruminantium*, and *Methanomassiliicoccales* and *Methanosphaera*, etc) in rumen using DNA Amplicon-seq (Kittelmann et al., 2013; Henderson et al., 2015). In this study, we also detected a negative relationship between *Mbb. gottschalkii* and *Mbb. ruminantium* ($\rho = -0.9$, $P < 0.1$; Figure 4) in the DNA Amplicon-seq dataset. In the RNA-seq dataset, we confirmed that *Succinivibrionaceae* was positively correlated with *Methanomassiliicoccales* ($\rho = 0.9$, $P < 0.1$; Figure 4). The bacterial families *Lachnospiraceae*, *Mogibacteriaceae*, *Prevotellaceae*, *Ruminococcaceae*, and *Spirochaetaceae* were positively correlated with each other ($P < 0.1$), and all of them were negatively correlated with *Succinivibrionaceae* and *Methanomassiliicoccales* ($P < 0.1$) in the RNA-seq dataset (Figure 4B). However, the negative correlations between *Succinivibrionaceae* and the other 4 bacterial families, and between *Mbb. gottschalkii* and *Mbb. ruminantium*, are possibly because they displayed the arithmetic replacement effect (shifts in abundance of predominant phylotypes will have effects on others when analyzing proportion data) as

suggested by Henderson et al. (2015). Meanwhile, only two taxa, *Succinivibrionaceae* and *Mbb. ruminantium*, showed significantly positive correlations between the RNA-seq and DNA Amplicon-seq datasets ($\rho = 1.0$, $P < 0.05$ and $\rho = 0.9$, $P < 0.1$, respectively). No other taxa exhibited strong consistency between the RNA-seq and DNA Amplicon-seq datasets, indicating that cellular abundance did not correspond to the activities of most rumen taxa.

Methodological Caveats of this Study

This study has limitations that should be taken into account. Firstly, the sampling timing may have more of an effect on RNA-based than DNA-based approaches of profiling microbial communities. If assessing the activities of rumen microbiota is the main study objective, the sampling timing should be carefully considered. In this study, rumen digesta samples were collected before the feeding, which probably resulted in different RNA profiles than in digesta samples collected after feeding. However, the same rumen digesta sample was used for RNA and DNA extraction in our study, so the detected differences between RNA- and DNA-based analyses are valid and not biased due to different sampling times. Second, during RNA isolation processes, RNA yield may differ according to extraction method (such as between physical, mechanical, enzymatic, and chemical methods; Stark et al., 2014). Meanwhile, not all microbes can be lysed with equal efficiency, and notably, RNA yields from Gram-positive bacteria are generally lower than those from Gram-negative bacteria (Stark et al., 2014). For instance, members of *Proteobacteria* and *Bacteroidetes* are Gram-negative, while most *Firmicutes* members in the rumen are Gram-positive, which may explain the higher abundances of *Proteobacteria* and *Bacteroidetes*, and the lower abundance of *Firmicutes* in the RNA-seq dataset. Moreover, significant differences in microbial community structures were also demonstrated to correspond to different DNA extraction methods in a report by Henderson et al. (2013). Third, because methods (Griffiths et al., 2000; Leininger et al., 2006) for co-extraction of RNA and DNA could not generate the high quality RNA for RNA-seq of our samples (RNA integrity number < 3.0), the RNA and DNA extractions were conducted separately using two independent

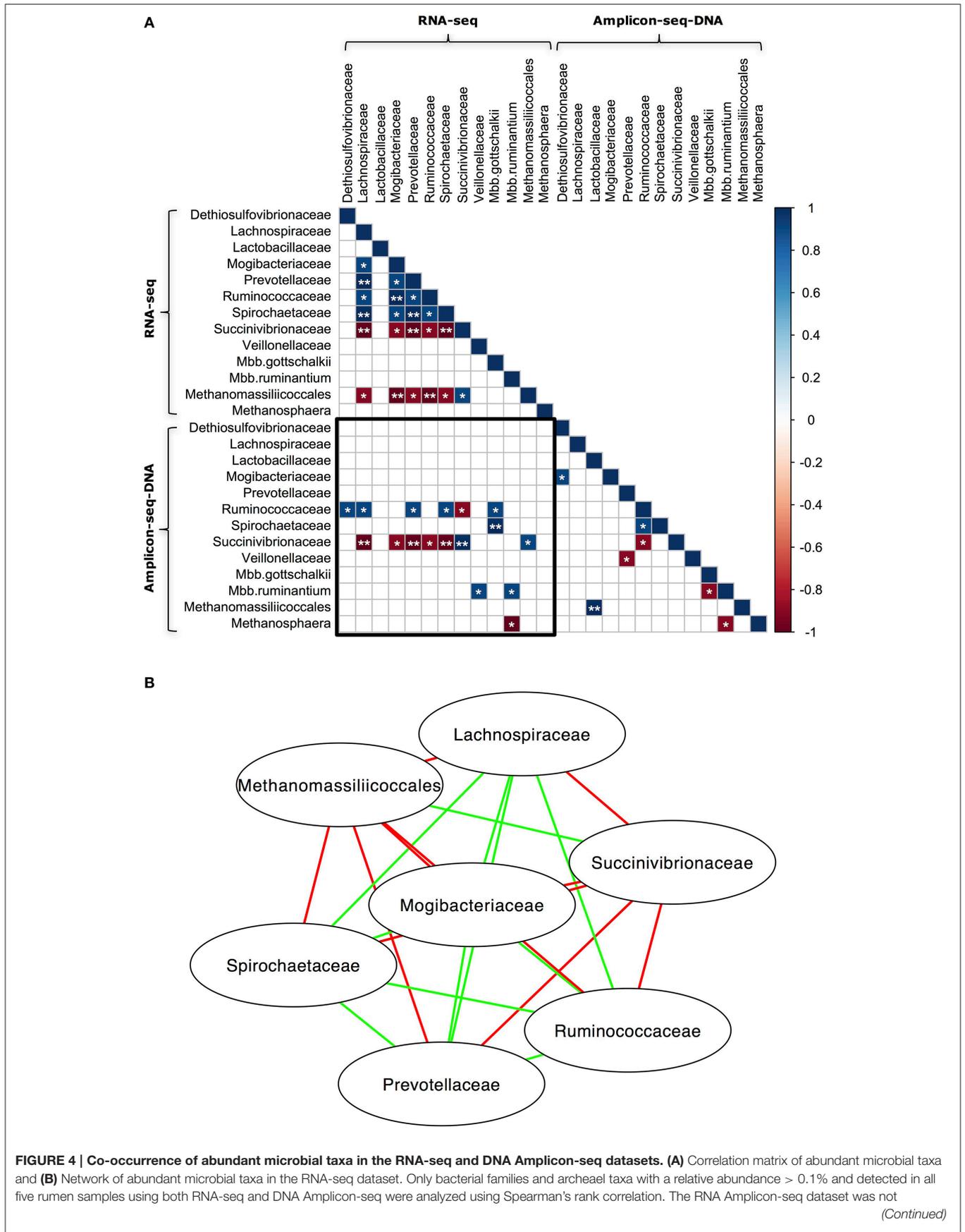


FIGURE 4 | Continued

included into the analysis, because its bacterial and archaeal profiles were similar to profiles from the RNA-seq dataset. In **(A)**, the sub-matrix surrounded by the black square exhibits correlations between taxa in the RNA-seq and DNA Amplicon-seq datasets. Strong correlations (Spearman's rank correlation coefficient [ρ] ≥ 0.9 or ≤ -0.9) were displayed with * ($0.05 < P < 0.1$) and ** ($P < 0.05$), while the other correlations were showed as blank. In **(B)**, a connection with a green/red line means a strong positive/negative correlation ($\rho \geq 0.9$ or ≤ -0.9 and $P < 0.1$).

protocols to ensure the high quality RNA and DNA in the current study, which could potentially lead to differences between RNA- and DNA-based methods. Furthermore, the RNA was transcribed to cDNA using the random primers before making the RNA-seq and RNA Amplicon-seq libraries, while the DNA Amplicon-seq was performed using region-specific primers to amplify DNA template directly, which could also contribute to differences among three datasets. Fourth, because rRNA content per cell varies between different microbial phylotypes (Medlin and Simon, 1998; Sievert et al., 2000), such intrinsic differences in rRNA content could also influence the relative abundance determined using RNA-seq. Future experiments to globally compare the rRNA content per cell among different microbial phylotypes in the rumen microbiota and normalize rRNA concentrations from different phylotypes can improve the accuracy of microbial community profiling using RNA-seq. Fifth, the quantification of total and/or species-specific rRNA is a valid and well-accepted approach to estimate the microbial activity, which has been applied in more than 100 studies (Blazewicz et al., 2013). However, the use of rRNA as an indicator of specific microbial functional activity in complex environmental samples still needs to be further validated by correlating them with the mRNA information within the same RNA-seq dataset. Sixth, in the current study, the RNA-seq dataset was generated on an Illumina HiSeq2000 platform, whereas RNA/DNA Amplicon-seq was performed using a 454 pyrosequencing platform. It has been demonstrated that there are different features between these two platforms, such as read length, accuracy, and throughput (Liu et al., 2012). However, previous studies have revealed the consistency of microbial community profiles generated across sequencing platforms (Caporaso et al., 2012; Nelson et al., 2014), but differences in sequencing depth could have an impact on the detection of low-abundant taxa.

CONCLUSION

A comparison of the microbial profiles generated from RNA-seq and RNA/DNA Amplicon-seq revealed the generation of different taxonomic profiles of the same rumen microbiota between these methods, and thus, their results could not be simply combined. The RNA-based methods could more robustly detect microbial phylotypes with potentially metabolic activities in the rumen and also detect more interactions among these phylotypes than DNA Amplicon-seq. In addition, compared

to RNA/DNA Amplicon-seq, the RNA-seq approach showed more diversity and could detect more bacterial and archaeal phylotypes in the rumen. Although the RNA-seq approach has the advantage of simultaneously identifying and quantifying active microorganisms within a microbiota, the data are not conclusive on which method is the best for analyzing animal gastrointestinal microbiota due to the different technologies and constraints of DNA vs. RNA and due to differences in nucleotide extraction, sequencing, and analysis protocols. Nevertheless, this is the first study to compare RNA-seq and RNA/DNA Amplicon-seq for the taxonomic assessment of rumen microbiota, and differences among these methods should be carefully considered to accurately assess gastrointestinal microbiota in future studies.

AUTHOR CONTRIBUTIONS

FL: experimental design, data generation, data analysis and interpretation, manuscript writing; GH: DNA Amplicon-seq, data interpretation and manuscript writing; XS: protocol optimization of RNA-seq library construction; FC: DNA isolation and DNA Amplicon-seq; PHJ: data interpretation and manuscript writing; LLG: experimental design, data analysis and interpretation, manuscript writing.

ACKNOWLEDGMENTS

We thank staff at Kinsella Research Station, E. Hernandez-Sanabria and M. Zhou for help with the sampling; and G. Liang, Y. Chen, and N. Malmuthuge for experiment assistance and advice. This work was funded by Alberta Livestock and Meat Agency (Edmonton, Canada) (2013R029R), and by the New Zealand Government to support the objectives of the Livestock Research Group of the Global Research Alliance on Agricultural Greenhouse Gases. We also acknowledge the Alberta Innovates-Technology Futures Graduate Student Scholarship and the Global Research Alliance Senior Scientist Award for the financial support to FL and LLG, respectively.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmich.2016.00987>

REFERENCES

- Barberan, A., Bates, S. T., Casamayor, E. O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 6, 343–351. doi: 10.1038/ismej.2011.119
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B* 57, 289–300.
- Béra-Maillet, C., Mosoni, P., Kwasiborski, A., Suau, F., Ribot, Y., and Forano, E. (2009). Development of a RT-qPCR method for the quantification of *Fibrobacter succinogenes* S85 glycoside hydrolase transcripts in the rumen content of gnotobiotic and conventional sheep. *J. Microbiol. Methods* 77, 8–16. doi: 10.1016/j.mimet.2008.11.009
- Berry, D., Schwab, C., Milinovich, G., Reichert, J., Ben Mahfoudh, K., Decker, T., et al. (2012). Phylotype-level 16S rRNA analysis reveals new bacterial indicators of health state in acute murine colitis. *ISME J.* 6, 2091–2106. doi: 10.1038/ismej.2012.39
- Blazewicz, S. J., Barnard, R. L., Daly, R. A., and Firestone, M. K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J.* 7, 2061–2068. doi: 10.1038/ismej.2013.102
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bremer, H., and Dennis, P. P. (2008). Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus* 3. doi: 10.1128/ecosal.5.2.3
- Burge, S. W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E. P., et al. (2013). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* 41, D226–D232. doi: 10.1093/nar/gks1005
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8
- Chakravorty, S., Helb, D., Burday, M., Connell, N., and Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiol. Methods* 69, 330–339. doi: 10.1016/j.mimet.2007.02.005
- Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162, 156–159. doi: 10.1016/0003-2697(87)90021-2
- de Menezes, A., Clipson, N., and Doyle, E. (2012). Comparative metatranscriptomics reveals widespread community responses during phenanthrene degradation in soil. *Environ. Microbiol.* 14, 2577–2588. doi: 10.1111/j.1462-2920.2012.02781.x
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381
- Edwards, J. E., McEwan, N. R., Travis, A. J., and Wallace, R. J. (2004). 16S rDNA library-based analysis of ruminal bacterial diversity. *Antonie Van Leeuwenhoek* 86, 263–281. doi: 10.1023/B:ANTO.0000047942.69033.24
- Ellison, M. J., Conant, G. C., Cockrum, R. R., Austin, K. J., Truong, H., Becchi, M., et al. (2014). Diet alters both the structure and taxonomy of the ovine gut microbial ecosystem. *DNA Res.* 21, 115–125. doi: 10.1093/dnares/dst044
- Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., et al. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* 8:e1002606. doi: 10.1371/journal.pcbi.1002606
- Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., et al. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2329–E2338. doi: 10.1073/pnas.1319284111
- Gaci, N., Borrel, G., Tottey, W., O'Toole, P. W., and Brugère, J. F. (2014). Archaea and the human gut: new beginning of an old story. *World J. Gastroenterol.* 20, 16062–16078. doi: 10.3748/wjg.v20.i43.16062
- Gaidos, E., Rusch, A., and Ilardo, M. (2011). Ribosomal tag pyrosequencing of DNA and RNA from benthic coral reef microbiota: community spatial structure, rare members and nitrogen-cycling guilds. *Environ. Microbiol.* 13, 1138–1152. doi: 10.1111/j.1462-2920.2010.02392.x
- Griffiths, R. L., Whiteley, A. S., O'Donnell, A. G., and Bailey, M. J. (2000). Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl. Environ. Microbiol.* 66, 5488–5491. doi: 10.1128/AEM.66.12.5488-5491.2000
- Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., et al. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 21, 494–504. doi: 10.1101/gr.112730.110
- Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., and Janssen, P. H. (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. doi: 10.1038/srep14567
- Henderson, G., Cox, F., Kittelmann, S., Miri, V. H., Zethof, M., Noel, S. J., et al. (2013). Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS ONE* 8:e74787. doi: 10.1371/journal.pone.0074787
- Hernandez-Sanabria, E., Goonewardene, L. A., Wang, Z., Durunna, O. N., Moore, S. S., and Guan, L. L. (2012). Impact of feed efficiency and diet on adaptive variations in the bacterial community in the rumen fluid of cattle. *Appl. Environ. Microbiol.* 78, 1203–1214. doi: 10.1128/AEM.05114-11
- Hernandez-Sanabria, E., Goonewardene, L. A., Wang, Z., Zhou, M., Moore, S. S., and Guan, L. L. (2013). Influence of sire breed on the interplay among rumen microbial populations inhabiting the rumen liquid of the progeny in beef cattle. *PLoS ONE* 8:e58461. doi: 10.1371/journal.pone.0058461
- Holmes, E., Li, J. V., Marchesi, J. R., and Nicholson, J. K. (2012). Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. *Cell Metab.* 16, 559–564. doi: 10.1016/j.cmet.2012.10.007
- Hong, S., Bunge, J., Leslin, C., Jeon, S., and Epstein, S. S. (2009). Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J.* 3, 1365–1373. doi: 10.1038/ismej.2009.89
- Huber, J. A., Morrison, H. G., Huse, S. M., Neal, P. R., Sogin, M. L., and Mark Welch, D. B. (2009). Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environ. Microbiol.* 11, 1292–1302. doi: 10.1111/j.1462-2920.2008.01857.x
- İnceoğlu, Ö., Lliros, M., Crowe, S. A., García-Armisen, T., Morana, C., Darchambeau, F., et al. (2015). Vertical distribution of functional potential and active microbial communities in meromictic lake Kivu. *Microb. Ecol.* 70, 596–611. doi: 10.1007/s00248-015-0612-9
- Jami, E., and Mizrahi, I. (2012). Composition and similarity of bovine rumen microbiota across individual animals. *PLoS ONE* 7:e33306. doi: 10.1371/journal.pone.0033306
- Janssen, P. H., and Kirs, M. (2008). Structure of the archaeal community of the rumen. *Appl. Environ. Microbiol.* 74, 3619–3625. doi: 10.1128/AEM.02812-07
- Jeyanathan, J., Kirs, M., Ronimus, R. S., Hoskin, S. O., and Janssen, P. H. (2011). Methanogen community structure in the rumens of farmed sheep, cattle and red deer fed different diets. *FEMS Microbiol. Ecol.* 76, 311–326. doi: 10.1111/j.1574-6941.2011.01056.x
- Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology* 88, 2427–2439. doi: 10.1890/06-1736.1
- Kang, S., Denman, S. E., Morrison, M., Yu, Z., and McSweeney, C. S. (2009). An efficient RNA extraction method for estimating gut microbial diversity by polymerase chain reaction. *Curr. Microbiol.* 58, 464–471. doi: 10.1007/s00284-008-9345-z
- Kang, S. H., Evans, P., Morrison, M., and McSweeney, C. (2013). Identification of metabolically active proteobacterial and archaeal communities in the rumen by DNA- and RNA-derived 16S rRNA gene. *J. Appl. Microbiol.* 115, 644–653. doi: 10.1111/jam.12270
- Kim, M., and Yu, Z. (2014). Variations in 16S rRNA-based microbiome profiling between pyrosequencing runs and between pyrosequencing facilities. *J. Microbiol.* 52, 355–365. doi: 10.1007/s12275-014-3443-3

- Kittlmann, S., Seedorf, H., Walters, W. A., Clemente, J. C., Knight, R., Gordon, J. I., et al. (2013). Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLoS ONE* 8:e47879. doi: 10.1371/journal.pone.0047879
- Kopylova, E., Noe, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28, 3211–3217. doi: 10.1093/bioinformatics/bts611
- Korpela, K., Salonen, A., Virta, L. J., Kekkonen, R. A., Forslund, K., Bork, P., et al. (2016). Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. *Nat. Commun.* 7:10410. doi: 10.1038/ncomms10410
- Kozich, J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (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. doi: 10.1128/AEM.01043-13
- Lanzen, A., Jorgensen, S. L., Bengtsson, M. M., Jonassen, I., Ovreas, L., and Urich, T. (2011). Exploring the composition and diversity of microbial communities at the Jan Mayen hydrothermal vent field using RNA and DNA. *FEMS Microbiol. Ecol.* 77, 577–589. doi: 10.1111/j.1574-6941.2011.01138.x
- Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., et al. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442, 806–809. doi: 10.1038/nature04983
- Li, M., Penner, G. B., Hernandez-Sanabria, E., Oba, M., and Guan, L. L. (2009). Effects of sampling location and time, and host animal on assessment of bacterial diversity and fermentation parameters in the bovine rumen. *J. Appl. Microbiol.* 107, 1924–1934. doi: 10.1111/j.1365-2672.2009.04376.x
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., et al. (2012). Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2012:251364. doi: 10.1155/2012/251364
- Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F. M., Ferrera, I., Sarmiento, H., et al. (2014). Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ. Microbiol.* 16, 2659–2671. doi: 10.1111/1462-2920.12250
- Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D., Hirschfield, G. M., Hold, G., et al. (2016). *The gut microbiota and host health: a new clinical frontier.* *Gut* 65, 330–339. doi: 10.1136/gutjnl-2015-309990
- Mason, O. U., Scott, N. M., Gonzalez, A., Robbins-Pianka, A., Baelum, J., Kimbrel, J., et al. (2014). Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. *ISME J.* 8, 1464–1475. doi: 10.1038/ismej.2013.254
- Medlin, L., and Simon, N. (1998). “Phylogenetic analysis of marine phytoplankton,” in *Molecular Approaches to the Study of the Ocean*, ed K. Cooksey (Dordrecht: Springer), 161–186.
- Million, M., Lagier, J. C., Yahav, D., and Paul, M. (2013). Gut bacterial microbiota and obesity. *Clin. Microbiol. Infect.* 19, 305–313. doi: 10.1111/1469-0691.12172
- Nelson, M. C., Morrison, H. G., Benjamino, J., Grim, S. L., and Graf, J. (2014). Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. *PLoS ONE* 9:e94249. doi: 10.1371/journal.pone.0094249
- Ohene-Adjei, S., Teather, R. M., Ivan, M., and Forster, R. J. (2007). Postinoculation protozoan establishment and association patterns of methanogenic archaea in the ovine rumen. *Appl. Environ. Microbiol.* 73, 4609–4618. doi: 10.1128/AEM.02687-06
- Ojeda, P., Bobe, A., Dolan, K., Leone, V., and Martinez, K. (2016). Nutritional modulation of gut microbiota - the impact on metabolic disease pathophysiology. *J. Nutr. Biochem.* 28, 191–200. doi: 10.1016/j.jnutbio.2015.08.013
- Olfert, E. D., Cross, B. M., and McWilliams, A. A. (1993). *Guide to the Care and Use of Experimental Steers.* Ottawa, ON: Canadian Council on Animal Care.
- Petri, R. M., Schwaiger, T., Penner, G. B., Beauchemin, K. A., Forster, R. J., McKinnon, J. J., et al. (2013). Characterization of the core rumen microbiome in cattle during transition from forage to concentrate as well as during and after an acidotic challenge. *PLoS ONE* 8:e83424. doi: 10.1371/journal.pone.0083424
- Poulsen, L. K., Ballard, G., and Stahl, D. A. (1993). Use of rRNA fluorescence *in situ* hybridization for measuring the activity of single cells in young and established biofilms. *Appl. Environ. Microbiol.* 59, 1354–1360.
- Poulsen, M., Schwab, C., Borg Jensen, B., Engberg, R. M., Spang, A., Canibe, N., et al. (2013). Methylophilic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat. Commun.* 4:1428. doi: 10.1038/ncomms2432
- Prosser, J. I., Bohannon, B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Freckleton, R. P., et al. (2007). The role of ecological theory in microbial ecology. *Nat. Rev. Microbiol.* 5, 384–392. doi: 10.1038/nrmicro1643
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219
- R Core Team (2014). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.
- Rius, A. G., Kittlmann, S., Macdonald, K. A., Waghorn, G. C., Janssen, P. H., and Sikkema, E. (2012). Nitrogen metabolism and rumen microbial enumeration in lactating cows with divergent residual feed intake fed high-digestibility pasture. *J. Dairy Sci.* 95, 5024–5034. doi: 10.3168/jds.2012-5392
- Rooks, M. G., Veiga, P., Wardwell-Scott, L. H., Tickle, T., Segata, N., Michaud, M., et al. (2014). Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. *ISME J.* 8, 1403–1417. doi: 10.1038/ismej.2014.3
- Ross, E. M., Moate, P. J., Bath, C. R., Davidson, S. E., Sawbridge, T. I., Guthridge, K. M., et al. (2012). High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. *BMC Genet.* 13:53. doi: 10.1186/1471-2156-13-53
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Schwab, C., Berry, D., Rauch, I., Rennisch, I., Ramesmayer, J., Hainzl, E., et al. (2014). Longitudinal study of murine microbiota activity and interactions with the host during acute inflammation and recovery. *ISME J.* 8, 1101–1114. doi: 10.1038/ismej.2013.223
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sievert, S. M., Ziebis, W., Kuever, J., and Sahn, K. (2000). Relative abundance of Archaea and Bacteria along a thermal gradient of a shallow-water hydrothermal vent quantified by rRNA slot-blot hybridization. *Microbiology* 146, 1287–1293. doi: 10.1099/00221287-146-6-1287
- Snelling, T. J., Genç, B., McKain, N., Watson, M., Waters, S. M., Creevey, C. J., et al. (2014). Diversity and community composition of methanogenic archaea in the rumen of scottish upland sheep assessed by different methods. *PLoS ONE* 9:e106491. doi: 10.1371/journal.pone.0106491
- Stark, L., Giersch, T., and Wünschiers, R. (2014). Efficiency of RNA extraction from selected bacteria in the context of biogas production and metatranscriptomics. *Anaerobe* 29, 85–90. doi: 10.1016/j.anaerobe.2013.09.007
- Tveit, A. T., Urich, T., and Svenning, M. M. (2014). Metatranscriptomic analysis of arctic peat soil microbiota. *Appl. Environ. Microbiol.* 80, 5761–5772. doi: 10.1128/AEM.01030-14
- Urich, T., Lanzén, A., Qi, J., Huson, D. H., Schleper, C., and Schuster, S. C. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS ONE* 3:e2527. doi: 10.1371/journal.pone.0002527
- Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07
- Williams, Y. J., Popovski, S., Rea, S. M., Skillman, L. C., Toovey, A. F., Northwood, K. S., et al. (2009). A vaccine against rumen methanogens can alter the composition of archaeal populations. *Appl. Environ. Microbiol.* 75, 1860–1866. doi: 10.1128/AEM.02453-08

- Wright, A. D. G., Auckland, C. H., and Lynn, D. H. (2007). Molecular diversity of methanogens in feedlot cattle from Ontario and Prince Edward Island, Canada. *Appl. Environ. Microbiol.* 73, 4206–4210. doi: 10.1128/AEM.00103-07
- Xia, Y., Kong, Y., Seviour, R., Forster, R., Vasanthan, T., and McAllister, T. (2015). *In situ* identification and quantification of starch-hydrolyzing bacteria attached to barley and corn grain in the rumen of cows fed barley-based diets. *FEMS Microbiol. Ecol.* 91:fiv077. doi: 10.1093/femsec/fiv077
- Yeoman, C. J., and White, B. A. (2014). Gastrointestinal tract microbiota and probiotics in production animals. *Annu. Rev. Anim. Biosci.* 2, 469–486. doi: 10.1146/annurev-animal-022513-114149

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Li, Henderson, Sun, Cox, Janssen and Guan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.