



Evaluating Established Methods for Rumen 16S rRNA Amplicon Sequencing With Mock Microbial Populations

Emily McGovern^{1,2}, Sinéad M. Waters¹, Gordon Blackshields¹ and Matthew S. McCabe^{1*}

¹ Animal and Bioscience Research Department, Animal and Grassland Research and Innovation Centre, Teagasc, Carlow, Ireland, ² UCD College of Health and Agricultural Sciences, University College Dublin, Dublin, Ireland

OPEN ACCESS

Edited by:

Stuart Edward Denman,
Commonwealth Scientific
and Industrial Research Organisation
(CSIRO), Australia

Reviewed by:

Marcell Nikolausz,
Helmholtz-Zentrum für
Umweltforschung (UFZ), Germany
Alex V. Chaves,
The University of Sydney, Australia

*Correspondence:

Matthew S. McCabe
matthew.mccabe@teagasc.ie

Specialty section:

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

Received: 12 February 2018

Accepted: 05 June 2018

Published: 25 June 2018

Citation:

McGovern E, Waters SM,
Blackshields G and McCabe MS
(2018) Evaluating Established
Methods for Rumen 16S rRNA
Amplicon Sequencing With Mock
Microbial Populations.
Front. Microbiol. 9:1365.
doi: 10.3389/fmicb.2018.01365

The rumen microbiome scientific community has utilized amplicon sequencing as an aid in identifying potential community compositional trends that could be used as an estimation of various production and performance traits including methane emission, animal protein production efficiency, and ruminant health status. In order to translate rumen microbiome studies into executable application, there is a need for experimental and analytical concordance within the community. The objective of this study was to assess these factors in relation to selected currently established methods for 16S phylogenetic community analysis on a microbial community standard (MC) and a DNA standard (DS; ZymoBIOMICS™). DNA was extracted from MC using the RBBC method commonly used for microbial DNA extraction from rumen digesta samples. 16S rRNA amplicon libraries were generated for the MC and DS using primers routinely used for rumen bacterial and archaeal community analysis. The primers targeted the V4 and V3–V4 region of the 16S rRNA gene and samples were subjected to both 20 and 28 polymerase chain reaction (PCR) cycles under identical cycle conditions. Sequencing was conducted using the Illumina MiSeq platform. As the bacteria contained in the microbial mock community were well-classified species, and for ease of explanation, we used the results of the Basic Local Alignment Search Tool classification to assess the DNA, PCR cycle number, and primer type. Sequence classification methodology was assessed independently. Spearman's correlation analysis indicated that utilizing the repeated bead beating and column method for DNA extraction in combination with primers targeting the 16S rRNA gene using 20 first-round PCR cycles was sufficient for amplicon sequencing to generate a relatively accurate depiction of the bacterial communities present in rumen samples. These results also emphasize the requirement to develop and utilize positive mock community controls for all rumen microbiomic studies in order to discern errors which may arise at any step during a next-generation sequencing protocol.

Keywords: mock community, rumen, amplicon, PCR, microbiota, phylogenetic analysis

INTRODUCTION

The rumen microbiota enables the ruminant to digest cellulose-rich feedstuff and to convert them into a wide range of metabolites that are used by both microbes for their own proliferation and also for the supply of nutrients to the host animal. This symbiotic relationship between the rumen microbiota and host is vital to the ruminant's survival and performance (Saleem et al., 2012). The diverse microbial community adapts to a wide array of dietary feedstuffs and management strategies (Knapp et al., 2014). Elucidating the rumen microbiome composition, adaptation, and function is of international interest due to its implications, ranging from climate change to applied animal production and animal health (McCann et al., 2014).

Classical knowledge of rumen microbiology was previously based on anaerobic, culture-dependent methods. However, in recent years, the development of next-generation sequencing (NGS) has permitted in-depth sequencing and data analysis of various types of environmental samples (Stulberg et al., 2016). Amplicon sequencing targeting the 16S rRNA in bacteria and archaea remains the most cost-effective and facile tool to provide valuable phylogenetic information for the comparison of microbial diversity in large numbers of environmental samples. Numerous research groups investigating the rumen microbiome have utilized amplicon sequencing as an aid in identifying potential community compositional trends that could be used as an estimation of various performance traits, including methane emission, animal protein production, and ruminant health status (Hook et al., 2011; Petri et al., 2013; Shi et al., 2014; Roehe et al., 2016; Shabat et al., 2016). It has been identified that there is no standardization between amplicon studies (Sinha et al., 2017) and it has been emphasized that this lack of standardization is causing issues with research replication as well as problems translating research into applications (Sinha et al., 2015). The assessment of US microbiome research also identified a need for standardized protocols between the microbiome research communities in order to facilitate comparisons between studies (Stulberg et al., 2016).

The method utilized for DNA extraction can lead to dramatic differences in microbial output composition (Henderson et al., 2013; Brooks et al., 2015). Hence, validating extraction methods with a mock microbial community is an absolute necessity to ensure accurate representation of microbial communities in samples. The extraction method assessed in the current study is the repeated bead beating and column (RBBC) method (Yu and Morrison, 2004) as it is a common method utilized for extraction of DNA from rumen contents (Henderson et al., 2013; Wallace et al., 2014; McCabe et al., 2015). However, this DNA extraction method has not been validated with a mock microbial community standard (MC) to verify its efficiency at extracting accurate representative quantities of DNA from both Gram-positive and Gram-negative bacteria.

Comparisons of the rumen bacterial and archaeal community compositions are typically investigated by sequencing one or more of the nine different hypervariable regions of the 16S rRNA gene, due to conservation across the domains of bacteria

and archaea (Pace, 1997; Tringe and Hugenholtz, 2008). Most NGS amplicon studies in recent years have been conducted on the Illumina MiSeq as this platform gives the highest number of reads at the lowest cost. However, the length of reliable high-quality reads remains at 250 bp for the forward read and approximately 230 bp for the reverse read. Consequently, this allows sequencing of only one or two 16S rRNA variable regions which limits the analysis to genus level identification at best. Despite this limited resolution, amplicon sequencing on the Illumina MiSeq is a very useful approach for screening large numbers of complex microbial community samples to identify samples that show major perturbations in the composition of microbial genera. The choice of a specific hyper variable region can affect both the accuracy and specificity of phylogenetic assignment and the community diversity estimation (Liu et al., 2008; Youssef et al., 2009). In the current study, we tested two commonly used primer pairs – 515F/806R (Caporaso et al., 2011) and Pro341F/Pro805R (Takahashi et al., 2014) – that both simultaneously amplify bacteria and archaea. It has been previously shown that simultaneous amplification of archaea and bacteria using multi-domain specific universal primers may not be as accurate as separate analysis of the archaeal and bacterial domain (Klindworth et al., 2013; Fischer et al., 2016). However, simultaneous amplification of archaea and bacteria is particularly appropriate for rumen microbiomic diversity studies as there are major efforts to develop strategies in cattle and sheep production that reduce the numbers of rumen archaea relative to bacteria (Roehe et al., 2016). It also facilitates investigation into relationships between microbial taxa from both archaeal and bacterial domains in a convenient, one step-approach, for example, McCabe et al. (2015) previously deciphered an inverse relationship between a member of *Methanobrevibacter gottschalkii* clade and uncharacterized Succinivibrionaceae species utilizing this method. The primer pair Pro341F/Pro805R targets V3–V4 and is a subsequent modification of the primers reported by Klindworth et al. (2013). Pro341F/Pro805R was reported to generate microbial profiles that were more accurate than those obtained with other commonly used bacterial and archaeal primer pairs (Takahashi et al., 2014). The primer pair 515F/806R targets the V4 region of the 16S rRNA gene and was previously validated with a mock community. However, Caporaso et al. (2011) utilized an alternative amplification protocol with a single round of polymerase chain reaction (PCR) rather than the two-step PCR protocol used in this investigation.

Polymerase chain reaction cycle conditions are another contentious issue that can cause misrepresentation of microbial community composition. A two-step PCR protocol is routinely used to initially amplify the hypervariable region of interest and, second, to barcode samples in order to facilitate multiplex sequencing (Illumina, 2013). It was our aim to assess our current methodology (McCabe et al., 2015) against a lower number of PCR cycle conditions for the first amplicon generation step. Superfluous PCR cycles can lead to the formation of PCR by-products (Qiu et al., 2001) such as chimeras (Wang and Wang, 1997), heteroduplexes, and single-stranded DNA

molecules (Jensen et al., 1993). The classification of sequences is yet another element of consideration for 16S rRNA phylogenetic analysis of bacterial and archaeal communities. For the present study, we used multiple sequence alignment with Basic Local Alignment Search Tool (BLAST) to identify species from DNA that were extracted from mock community cells. This BLAST approach was successful as the mock community is composed of only eight well-characterized bacterial species and two well-characterized yeast species. We also tested the Ribosomal Database Project (RDP) classifier (Wang et al., 2007), which is a popular naïve Bayesian classifier (Puniya et al., 2015) for taxonomic assignment of rumen bacterial and archaeal species (McCabe et al., 2015; Meale et al., 2016; Popova et al., 2017). The taxonomic classifier can be used independently and can be encapsulated into other programs and pipelines. The overall objective of the study was to evaluate standard methodology frequently utilized for phylogenetic surveys of the rumen bacterial and archaeal community and assess whether it is giving an accurate representation of microbial communities using a commercially available mock MC and a DNA standard (DS). The mock microbial community utilized for this evaluation is commercially available and is composed of five Gram-positive bacteria (i.e., difficult to lyse): *Lactobacillus fermentum*, *Enterococcus faecalis*, *Staphylococcus aureus*, *Listeria monocytogenes*, and *Bacillus subtilis*, three Gram-negative bacteria: *Escherichia coli*, *Salmonella enterica*, and *Pseudomonas aeruginosa*, and two yeast species: *Saccharomyces cerevisiae* and *Cryptococcus neoformans*.

MATERIALS AND METHODS

DNA Extraction

In order to replicate the DNA extraction protocol identically to our rumen extraction protocol, the ZymoBIOMICS™ Microbial Community Standard (Zymo Research Corp., Irvine, CA, United States) was diluted in a 2-mL microfuge tube (DNA/RNA/DNase/RNase-free; Merck KGaA, Darmstadt, Germany) with molecular grade phosphate buffered saline (PBS; Merck KGaA, Darmstadt, Germany), and the tube and its contents were snap frozen in liquid nitrogen. The frozen tube contents were then homogenized to a fine frozen powder under liquid nitrogen using a pestle and mortar and the frozen powder was stored in a 2-mL microfuge tube at -80°C . Approximately 250 mg homogenized frozen powder was used for DNA extraction from the mock community. DNA was also extracted from 250 mg of homogenized frozen rumen solid digesta powder sample and molecular grade water as a positive and a negative control (NC), respectively. DNA was extracted using the RBBC method (Yu and Morrison, 2004). The DNA quality (i.e., approximation of high molecular weight DNA:degraded DNA and presence/absence of RNA) was assessed on an agarose gel at a normalized concentration of 100 ng/ μL . DNA concentration and purity were assessed with two consecutive readings on the Nanodrop 1000 spectrophotometer.

16S rRNA Gene Amplicon Library Preparation

Five replicate amplicon libraries were generated for both the ZymoBIOMICS™ Microbial Community Standard cells (MC; Zymo Research Corp., Irvine, CA, United States), which were extracted using the RBBC extraction method (Yu and Morrison (2004), and the ZymoBIOMICS™ Microbial Community Standard DNA (DS; Zymo Research Corp., Irvine, CA, United States). The ZymoBIOMICS™ Microbial Community Standard DNA (DS) was generated at Zymo by extracting DNA from pure colonies of each of eight bacteria species and two yeast strains and combining the resulting 10 DNA extractions in carefully defined amounts (12% for each of the bacterial DNA extracts and 2% for each of the two yeast DNA extracts). Libraries were also generated for positive (rumen digesta) and negative extraction (molecular grade water) controls, respectively.

Polymerase chain reactions were performed for amplicon libraries with 20 ng of rumen DNA 515F/806R (V4; Caporaso et al., 2011) and Pro341F/Pro805R (V3–V4) primers (Takahashi et al., 2014) and designed with Nextera overhang adapters, using 1X KAPA HiFi HotStart ReadyMix DNA polymerase (Roche Diagnostics, West Sussex, United Kingdom). Both 515F/806R and Pro341F/Pro805R matched to all bacterial DNA targets within MC and DS with no mismatch. Two sets of cycle conditions were applied for the initial PCR (PCR1) stage:

- (a) 95°C for 3 min, 20 cycles; 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and then 72°C for 5 min
- (b) 95°C for 3 min, 28 cycles; 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and then 72°C for 5 min.

Amplicons were purified using the Qiaquick PCR Purification Kit (Qiagen, Manchester, United Kingdom). A second PCR step was used to attach dual indices and Illumina sequencing adapters using Nextera XT index kit (Illumina, San Diego, CA, United States). Cycle conditions were 95°C for 3 min, 8 cycles; 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and then 72°C for 5 min. Amplicon generation was validated through visualization on a 2% (w/v) agarose gel (Yu and Morrison, 2004). Amplicons were pooled in equal concentrations and gel purified to remove adapter dimers using the Qiagen Gel Extraction Kit (Qiagen, Manchester, United Kingdom). An extra purification with the QIAquick purification kit (Qiagen, Manchester, United Kingdom) was used to remove residual agarose. The pooled purified libraries were measured for purity and quantity on the Nanodrop 1000 and further quantified using the KAPA SYBR FAST universal kit with Illumina Primer Premix (Roche Diagnostics, West Sussex, United Kingdom). The library pool was then diluted and denatured as recommended by the Illumina MiSeq library preparation guide. Sequencing was conducted using 500 cycle MiSeq reagent kits (Illumina, San Diego, CA, United States).

Zymo also provided the fastq files (ZE) containing 16S rRNA sequencing data from their extraction of the microbial mock community with the ZymoBIOMICS™ DNA Mini Kit (Zymo Research Corp., Irvine, CA, United States). These data had been generated with primers targeting V3–4 region and the amplicons

were sequenced on Illumina MiSeq (2 × 250 bp; Illumina, San Diego, CA, United States).

Sequence Analysis

Read 1 and Read 2 of 16S rRNA gene from all amplicons were quality checked with FASTQC (Andrews, 2010)¹ and overlapping reads (>Q30) were merged using BBMerge from the BBDNA package². The merged sequence outputs were processed and classified using four different methods:

- (1) A high throughput NCBI BLAST (version 2.5) of individual sequences of merged reads was conducted against the NCBI 16S ribosomal RNA database.
- (2) Merged reads from all samples were also subsequently combined into a single dataset for processing with QIIME (version 1.9). Chimeric sequences were removed via USEARCH using the ChimeraSlayer GOLD database (Edgar et al., 2011). A combination of *de novo* and reference-based operational taxonomic unit (OTU) identification was carried out using the open reference calling method implemented within QIIME. Sequences were clustered at 97% identity into individual OTUs using UCLUST (Edgar, 2010). A single representative sequence from each clustered OTU was aligned to the Greengenes (version gg_13_8), RDP (version 16), SILVA (Silva_128_release), and NCBI 16S ribosomal RNA database independently. Taxonomic classification for each OTU was determined using the RDP classifier (version 2.12).
- (3) Identical sequences were identified from merged sequence reads and a unique representative sequence of the original sequences was used for sequence classification. Classification was conducted with the R package; microclass, using the taxMachine tool and the contax_trim database. Merged ZE sequence reads were also classified using the RDP classifier (version 2.9) in R, with the contax_trim database.

Statistical Analysis

Sequence data from amplicon libraries ($n = 48$) were compared to the theoretical composition of the Zymo standards. 16S rRNA gene abundance was calculated from theoretical genomic DNA composition with the following formula: 16S copy number = total genomic DNA (g) × unit conversion constant (bp/g)/genome size (bp) × 16S copy number per genome (Zymo, 2016). Relative abundance was calculated for identified sequences in each sample at genus level. The experiment was conducted as a 2 × 2 factorial design. MC and DS replicates for each primer set and both cycle conditions examined were analyzed as an experimental unit for statistical analysis ($n = 5$ per unit). Group means were calculated for MC and DS replicates R Studio (v3.1). Spearman's correlation analysis was conducted with hmisc package R Studio (v3.1; Harrell, 2014). A power analysis was conducted prior to experimentation to ensure that technical replication

was sufficient to deduce a significant correlation between the technical treatments examined in the current experiment and the theoretical composition of the mock community supplied by Zymo. This utilized a significance level of 0.05, an expected correlation of 0.7, and a power of 0.2 (Champely, 2015). Similarity was declared as significant if $P < 0.05$ and as a trend if $0.05 > P \leq 0.1$.

Summary of Methods

DNA was extracted from MC using the RBBC method commonly used for microbial DNA extraction from rumen digesta samples. 16S rRNA amplicon libraries were generated for the MC and DS using primers routinely used for rumen bacterial and archaeal community analysis. The primers targeted the V4 and V3–V4 region of the 16S rRNA gene and samples were subjected to both 20 and 28 PCR cycles under identical cycle conditions. Sequencing was conducted using the Illumina MiSeq platform. As the bacteria contained in the microbial mock community were well-classified species, and for ease of explanation, we used the results of the BLAST classification to assess the DNA, PCR cycle number, and primer type. Sequence classification methodology was assessed independently.

RESULTS

DNA Extraction

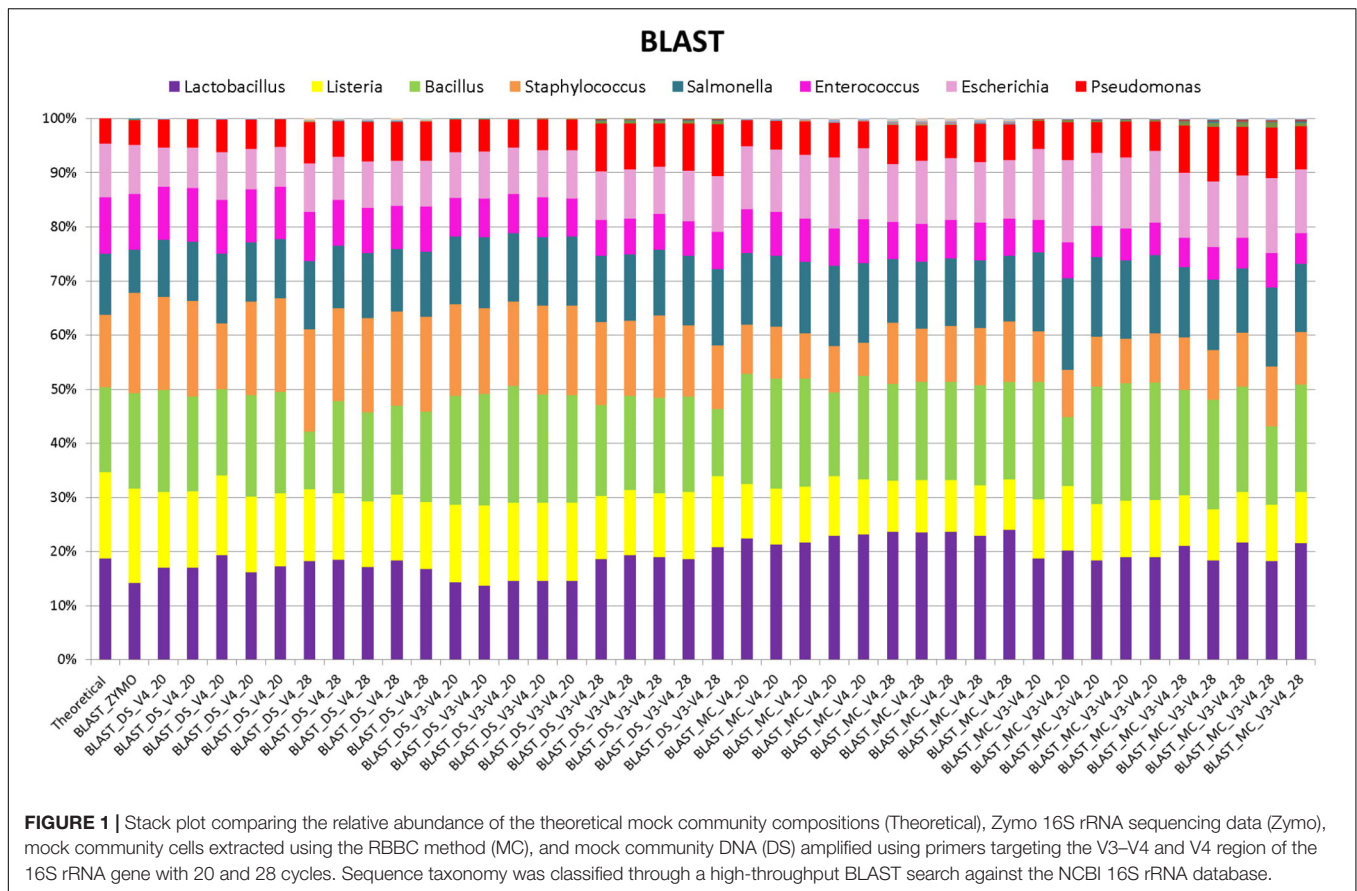
When BLAST analysis was used to analyze the MiSeq amplicon sequence data, all eight expected bacterial genera were detected for all combinations of DNA extraction/primer pair/PCR cycle number (**Figure 1**). The largest effect on accuracy of genera relative abundances was DNA extraction method. As expected, DS (DNA that was pre-extracted by Zymo from each species then combined in equal amounts) showed significant ($P < 0.05$) Spearman correlation with the theoretical mock community relative abundances for all combinations of PCR cycle numbers and primers. MC (DNA extracted using RBBC method) showed non-significant Spearman correlations with theoretical mock community relative abundances ($P > 0.05$; **Figure 2**). The best result achieved for correlations between MC and the theoretical composition was a tendency toward significance (Spearman = 0.67, $P = 0.07$) when 20 PCR cycles were used in combination with the V4 primers.

Interestingly, correlation of ZE (FASTQ files from a MiSeq amplicon run conducted at Zymo using DNA they extracted from the mock community with their Zymo kit) with the theoretical mock community relative abundance values showed a Spearman correlation of 0.69 ($P = 0.058$) which was similar to the correlation between MC/20 PCR cycles/V4 primers and the theoretical mock community (Spearman correlation of 0.67, $P = 0.071$; **Figure 2**). This shows that the RBBC DNA extraction method, amplicon library generation, and MiSeq sequencing that we conducted produced bacterial relative abundance profiles that were equivalent to those produced with the procedures that were conducted by Zymo.

Background amplification was determined to be the presence of any genera other than the eight expected to be present in the

¹<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

²<https://sourceforge.net/projects/bbmap/>



mock community. Background amplification was present in all samples including ZE (Figure 3). Background amplification was increased in the MC samples in comparison to the DS samples ($P < 0.05$; Table 1 and Figure 3).

Primer Selection

Two primer pairs were investigated in this study; 515F/806R and Pro341F/Pro805R which targeted the V4 region and the V3–V4 region of the 16S rRNA gene, respectively. As expected, both primer sets detected eight bacterial genera (Figure 1) and both archaeal and bacterial species were detected in our positive control samples. The relative abundances of the genera present in the DS and the MC amplified using the two primer pairs were evaluated via correlation analysis with the theoretical mock community. The DS bacterial relative abundance profiles for V4 and V3–V4 showed strong positive correlations with the theoretical mock community profiles ($P < 0.05$; Figure 2). Both DS and MC amplified with V3–V4 primers showed a positive correlation to their counterpart amplified with V4 primers ($P < 0.01$; Figure 2).

At 28 PCR cycle, there was an increase in background amplification of both MC and DS amplified with V3–V4 primers compared to MC and DS amplified with V4 primers ($P < 0.01$; Table 2). Background amplification was comparable between MC and DS amplified with V3–V4 primers and MC and DS amplified with V4 primers at 20 PCR cycles ($P > 0.5$; Table 2).

PCR Cycle Number

Two rounds of PCR were used for library preparation. The first round of PCR used primers that amplified 16S rRNA gene and the second round added the sequencing adapters and barcodes. Too many PCR cycles are thought to generate amplicon libraries that do not accurately represent microbiome composition. We tested the effect of increasing the number of PCR cycles in the first round of PCR (PCR1) from 20 PCR1 cycles to 28 PCR1 cycles. Relative abundances of taxa derived from 16S rRNA amplicon sequences from DS and MC amplified with 20 and 28 PCR cycles were compared against those of the theoretical mock community composition. The DS amplified at both 20 and 28 PCR1 cycles were significantly correlated with the theoretical mock community composition regardless of which primer pair was used ($P < 0.05$; Figure 2). DS and MC amplified at 20 PCR1 cycles were significantly correlated with their 28 PCR1 cycle counterpart ($P < 0.01$; Figure 2). There was an increase in detection of unexpected taxa at 28 PCR1 cycles in comparison to samples amplified at 20 PCR1 cycles for both DS and MC ($P < 0.01$; Figure 3 and Table 3).

Detection of Unexpected Taxa

Unexpected taxa were detected in all samples at relative abundances that varied between 0.12% and 1.6% (Figure 3). A higher number of PCR1 cycles (i.e., 28 PCR1 cycles versus 20 PCR1 cycles) clearly caused higher relative abundances of

BLAST

Spearman's Correlation Coefficient

	Theoretical	ZYMO	DS_V4_20	DS_V4_28	DS_V3.V4_20	DS_V3.V4_28	MC_V4_20	MC_V4_28	MC_V3.V4_20	MC_V3.V4_28
Theoretical	1	0.69	0.88	0.88	0.76	0.79	0.67	0.62	0.62	0.57
ZYMO	0.69	1	0.81	0.76	0.88	0.57	0.26	0.31	0.31	0.24
DS_V4_20	0.88	0.81	1	0.90	0.90	0.88	0.69	0.71	0.71	0.67
DS_V4_28	0.88	0.76	0.90	1	0.83	0.93	0.69	0.74	0.64	0.71
DS_V3.V4_20	0.76	0.88	0.90	0.83	1	0.79	0.55	0.57	0.62	0.55
DS_V3.V4_28	0.79	0.57	0.88	0.93	0.79	1	0.81	0.86	0.79	0.88
MC_V4_20	0.67	0.26	0.69	0.69	0.55	0.81	1	0.98	0.98	0.95
MC_V4_28	0.62	0.31	0.71	0.74	0.57	0.86	0.98	1	0.95	0.98
MC_V3.V4_20	0.62	0.31	0.71	0.64	0.62	0.79	0.98	0.95	1	0.93
MC_V3.V4_28	0.57	0.24	0.67	0.71	0.55	0.88	0.95	0.98	0.93	1

P value

	Theoretical	ZYMO	DS_V4_20	DS_V4_28	DS_V3.V4_20	DS_V3.V4_28	MC_V4_20	MC_V4_28	MC_V3.V4_20	MC_V3.V4_28
Theoretical	NA	0.06	0.00	0.00	0.03	0.02	0.07	0.10	0.10	0.14
ZYMO	0.06	NA	0.01	0.03	0.00	0.14	0.53	0.46	0.46	0.57
DS_V4_20	0.00	0.01	NA	0.00	0.00	0.00	0.06	0.05	0.05	0.07
DS_V4_28	0.00	0.03	0.00	NA	0.01	0.00	0.06	0.04	0.09	0.05
DS_V3.V4_20	0.03	0.00	0.00	0.01	NA	0.02	0.16	0.14	0.10	0.16
DS_V3.V4_28	0.02	0.14	0.00	0.00	0.02	NA	0.01	0.01	0.02	0.00
MC_V4_20	0.07	0.53	0.06	0.06	0.16	0.01	NA	0.00	0.00	0.00
MC_V4_28	0.10	0.46	0.05	0.04	0.14	0.01	0.00	NA	0.00	0.00
MC_V3.V4_20	0.10	0.46	0.05	0.09	0.10	0.02	0.00	0.00	NA	0.00
MC_V3.V4_28	0.14	0.57	0.07	0.05	0.16	0.00	0.00	0.00	0.00	NA

FIGURE 2 | Correlation matrix of Spearman correlation coefficient values between the theoretical mock community compositions (Theoretical), Zymo 16S rRNA sequencing data (Zymo), mock community cells extracted using the RBBC method (MC), and mock community DNA (DS) amplified using primers targeting the V3–V4 and V4 region of the 16S rRNA gene with 20 and 28 cycles. Sequence taxonomy was classified through a high throughput BLAST search against the NCBI 16S rRNA database. Values with a positive significant correlation are illustrated with shaded color while insignificant correlations are blank.

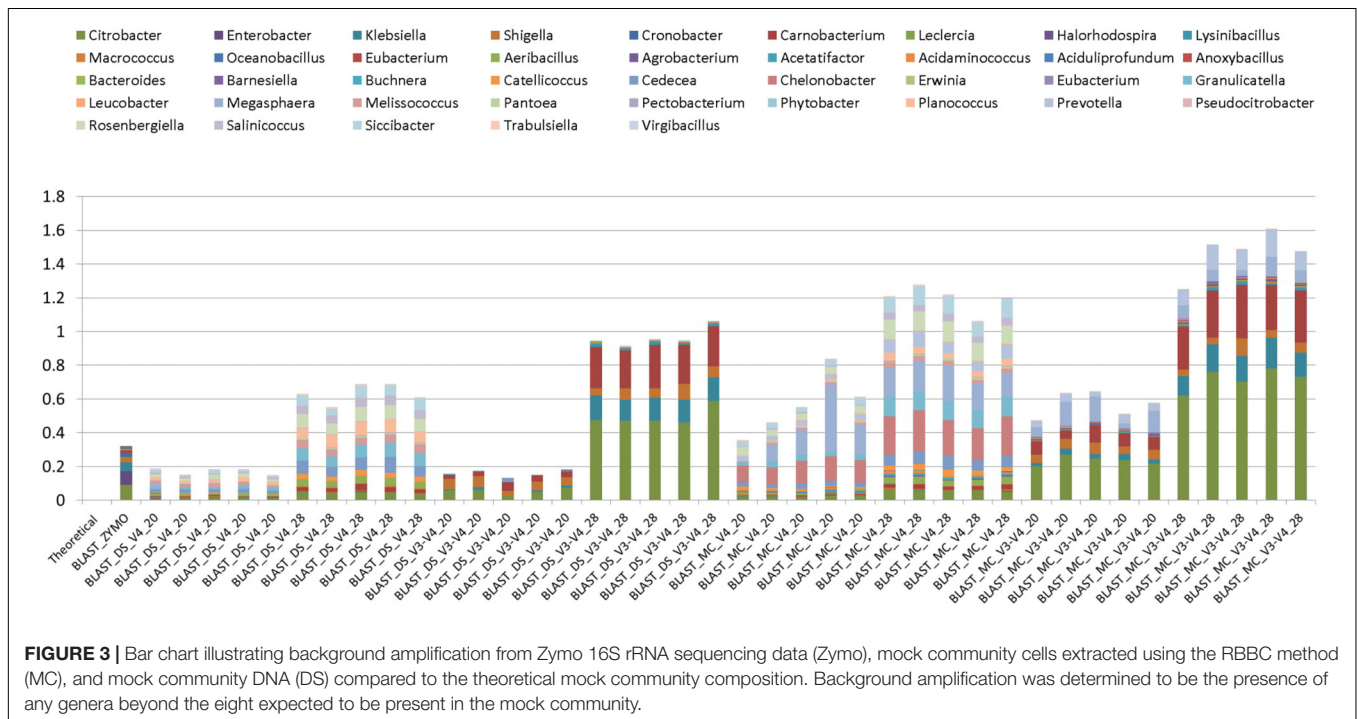


FIGURE 3 | Bar chart illustrating background amplification from Zymo 16S rRNA sequencing data (Zymo), mock community cells extracted using the RBBC method (MC), and mock community DNA (DS) compared to the theoretical mock community composition. Background amplification was determined to be the presence of any genera beyond the eight expected to be present in the mock community.

TABLE 1 | Mean relative abundance of background amplification from mock community cells extracted using the RBBC method for DNA extraction (MC) in comparison to mock community DNA (DS) amplified with two primer sets targeting the V3–V4 and V4 region of the 16S rRNA gene at both 20 and 28 cycles.

	DNA extraction				P-value
	DS	SEM	MC	SEM	
BLAST_V3–V4_20	0.16	0.01	0.57	0.03	<0.05
BLAST_V3–V4_28	0.96	0.03	1.47	0.06	<0.05
BLAST_V4_20	0.17	0.01	0.56	0.08	<0.05
BLAST_V4_28	0.63	0.03	1.19	0.04	<0.05

P-values were calculated using a Mann–Whitney U-test to evaluate background amplification from each treatment group ($n = 5$) resulting from DNA extraction method.

TABLE 2 | Mean relative abundance of amplification of unexpected bacteria which results from two primer sets targeting the V3–V4 and V4 region of the 16S rRNA gene which were used for amplification of mock community cells extracted with the RBBC (MC) and mock community DNA amplified at both 20 and 28 cycles.

	Primer comparison				P-value
	V3–V4	SEM	V4	SEM	
BLAST_DS_20	0.16	0.01	0.17	0.01	NS
BLAST_DS_28	0.96	0.03	0.63	0.03	<0.05
BLAST_MC_20	0.57	0.03	0.56	0.08	NS
BLAST_MC_28	1.47	0.06	1.19	0.04	<0.05

P-values were calculated using a Mann–Whitney U-test to compare background amplification from each treatment group ($n = 5$) which results from the primer set utilized.

TABLE 3 | Mean relative abundance of background amplification resulting from the number of PCR cycles during library preparation; 20 and 28 PCR cycles, constructed from DNA extracted from mock community cells extracted using the RBBC method (MC) and mock community DNA (DS) amplified with two primer sets targeting the V3–V4 and V4 region of the 16S rRNA gene.

	PCR cycle number				P-value
	20 PCR1	SEM	28 PCR1	SEM	
BLAST_DS_V4	0.17	0.01	0.63	0.03	<0.05
BLAST_DS_V3–V4	0.16	0.01	0.96	0.03	<0.05
BLAST_MC_V4	0.56	0.08	1.19	0.04	<0.05
BLAST_MC_V3–V4	0.57	0.03	1.47	0.06	<0.05

P-values were calculated using a Mann–Whitney U-test to evaluate background amplification from each treatment group ($n = 5$) resulting from PCR cycle number.

unexpected taxa (Figure 3). The highest and lowest unexpected taxa detection was in the MC_V3–V4_28 and DS_V4_20, respectively. Unexpected genera detection was one to three times higher in MC than DS in all cases when the same primer/PCR cycle number combinations were compared. This indicates that this “background noise” was influenced by the RBBC extraction method. Primers also influenced unexpected taxa detection. At 28 PCR1 cycles (but not 20 cycles), there was a significant ($P < 0.01$) increase in detection of unexpected genera in both MC and DS amplified with V3–V4 primers compared to MC and

DS amplified with V4 primers. V4 primers amplified different unexpected taxa than V3–V4 (Figure 3).

Sequence Classifier

Basic Local Alignment Search Tool analysis of amplicon sequences against the NCBI 16S rRNA database identified the eight bacterial genera present in the mock community from DS, MC, and ZE (Figure 1). The ZE, DS, and MC amplified with the V4 primers all showed significant correlation to the theoretical mock community (Figure 2). Classification was consistent with the theoretical composition with minimal background amplification detected (<2%; Figure 3). Although suitable for this experiment due to the well-defined nature and convenient number of the mock community components, BLAST in conjunction with the NCBI database is seldom used for 16S rRNA phylogenetic analysis of less well-defined communities, such as the rumen.

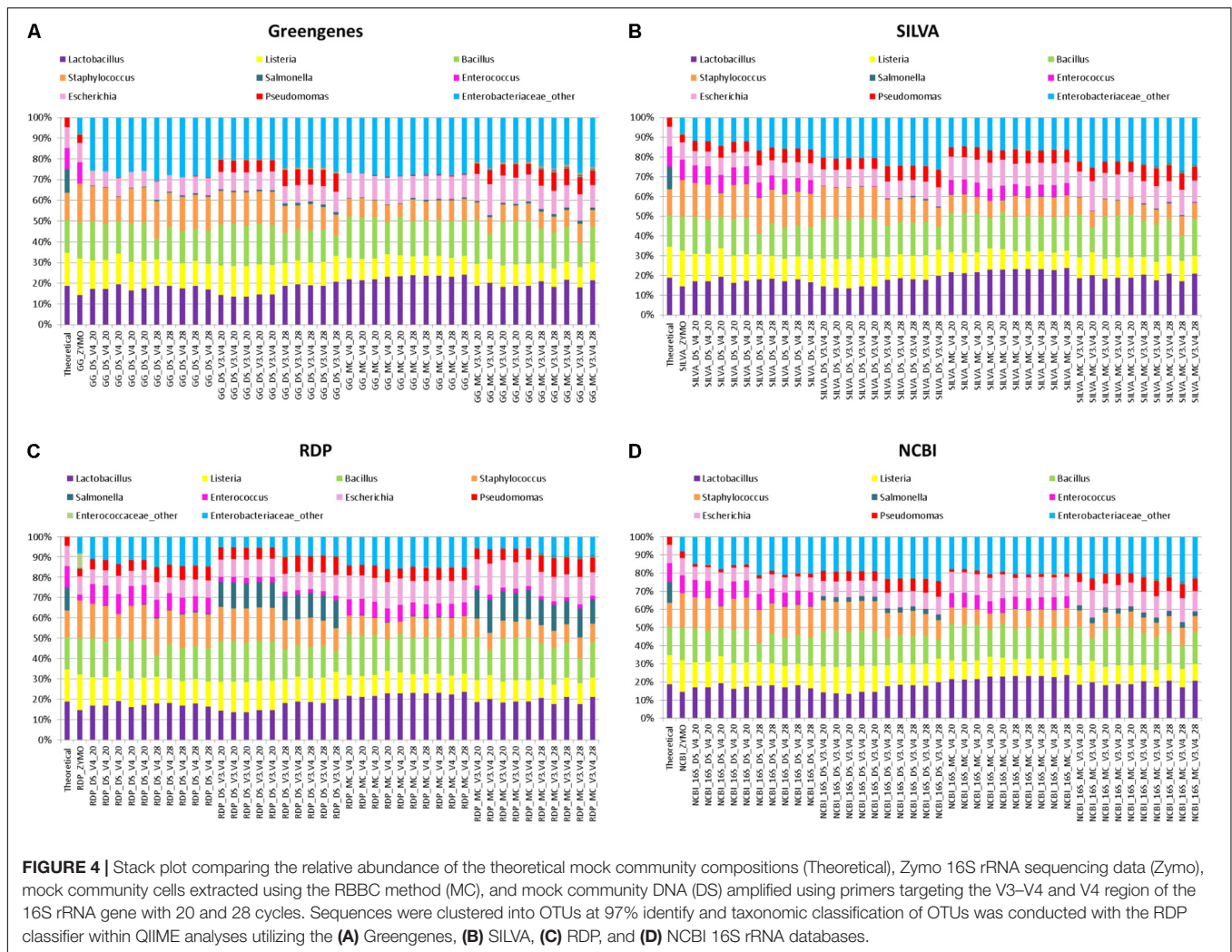
OTU-Based Approach

More commonly used for taxonomic classification of 16S rRNA amplicon sequences in rumen phylogenetic studies is the generation of OTUs from sequences that share 97% identity and classification of the OTUs with the RDP classifier. This approach was investigated independently in combination with four databases: Greengenes, NCBI-16S-rRNA, RDP, and SILVA. The RDP classifier identified all eight bacterial genera from DS, MC, and ZE (Figure 4). However, a large proportion of the 16S rRNA OTUs that were classified as members of the Enterobacteriaceae family were not classified as far as genus level in DS, MC, and the ZE using the RDP classifier. These OTUs were assigned to “Enterobacteriaceae_other” (Figure 4). There were two genera from this family within the mock community: *Escherichia* and *Salmonella*. Representation of these two genera was inaccurate in all of the relative abundance profiles that were generated from OTUs.

Salmonella was identified using the RDP database for the MC (~13%) and DS (~12%) at 20 and 28 PCR cycles when amplified with V3–V4 primers. However, *Salmonella* was not identified for ZE (V3–V4 amplicon sequence files supplied by Zymo) using the RDP database for sequence identification despite the same region (V3–V4) of the 16S rRNA gene being amplified (Figure 4C). The majority of OTUs from DS and MC amplified with the V4 primers which were identified as belonging to the Enterobacteriaceae family with the RDP database were not classified to genus level, with less than 0.2% of sequences identified as *Salmonella* (Figure 4C).

Utilizing the Greengenes and NCBI databases also showed that OTUs spanning the V3–V4 region of the 16S rRNA gene increased *Salmonella* identification in comparison to the V4 sequences from MC and DS but this increase was only small and the *Salmonella* genus was severely underrepresented with both V4 and V3–V4 primers for Greengenes and NCBI (Figures 4A,D). *Salmonella* was equally underrepresented in all samples (<0.1%) regardless of OTU sequence length when the SILVA database was used (Figure 4B).

Enterococcus remained almost absent from the MC and DS samples identified using the Greengenes database irrespective of



primer type used for amplification (Figure 4A). *Enterococcus* was identified in the correct proportions from the ZE (Figure 4A) when classified with the Greengenes database (Figure 4A). *Enterococcus* was also underrepresented in DS and MC samples amplified with V3–V4 primers when identified with the SILVA, RDP, and NCBI databases in comparison to DS and MC samples amplified with V4 primers and ZE (Figures 4B–D). *Pseudomonas* was underrepresented from DS and MC samples amplified with V4 primers when classified with the Greengenes database (<1%; Figure 4A).

Spearman’s rank correlation analysis was used to compare the DS, MC, and ZE sequences classified with the RDP classifier and Greengenes, RDP, SILVA, and the NCBI 16S rRNA database (Figure 5). Zymo amplified their fastq files with primers targeting the V3–V4 region (which were designed by Zymo) of the 16S rRNA gene; a similar correlation between this amplified community and the theoretical mock community ($r = 0.64$, $P < 0.1$) is observed for Greengenes and SILVA databases (Figure 5). ZE classified with the RDP and NCBI databases gave a reduced correlation to the theoretical mock community; $r = 0.55$ and $r = 0.62$, respectively, no tendency toward the theoretical

mock community was observed (Figure 5). The Zymo mock community gave a similar community structure regardless of database used for sequence classification (Figure 5). *Salmonella* was underrepresented in all of the ZE. Approximately 7% of the sequences belonging to the Enterococcaceae family were left unclassified at the genus level when classified with the RDP database (Figure 4).

The DS amplified with V3–V4 primers with 28 PCR1 cycles showed a similar correlation for the Greengenes, SILVA, NCBI, and RDP databases and the community compositions tended toward the theoretical mock community ($P > 0.05$; Figure 5). This correlation extended for the DS amplified with V3–V4 primers with 20 PCR1 cycles when sequences were classified with the RDP database ($P < 0.05$; Figure 5C) but not for the Greengenes, SILVA, and NCBI databases ($P > 0.05$; Figures 5A,B,D). All four databases showed a significant correlation for the MC and DS amplified with the V4 primers at 20 and 28 PCR1 cycles (Figure 5).

DNA extracted from mock community cells amplified with V4 primers with 20 PCR cycles and classified with the Greengenes database were the only condition where a tendency toward the

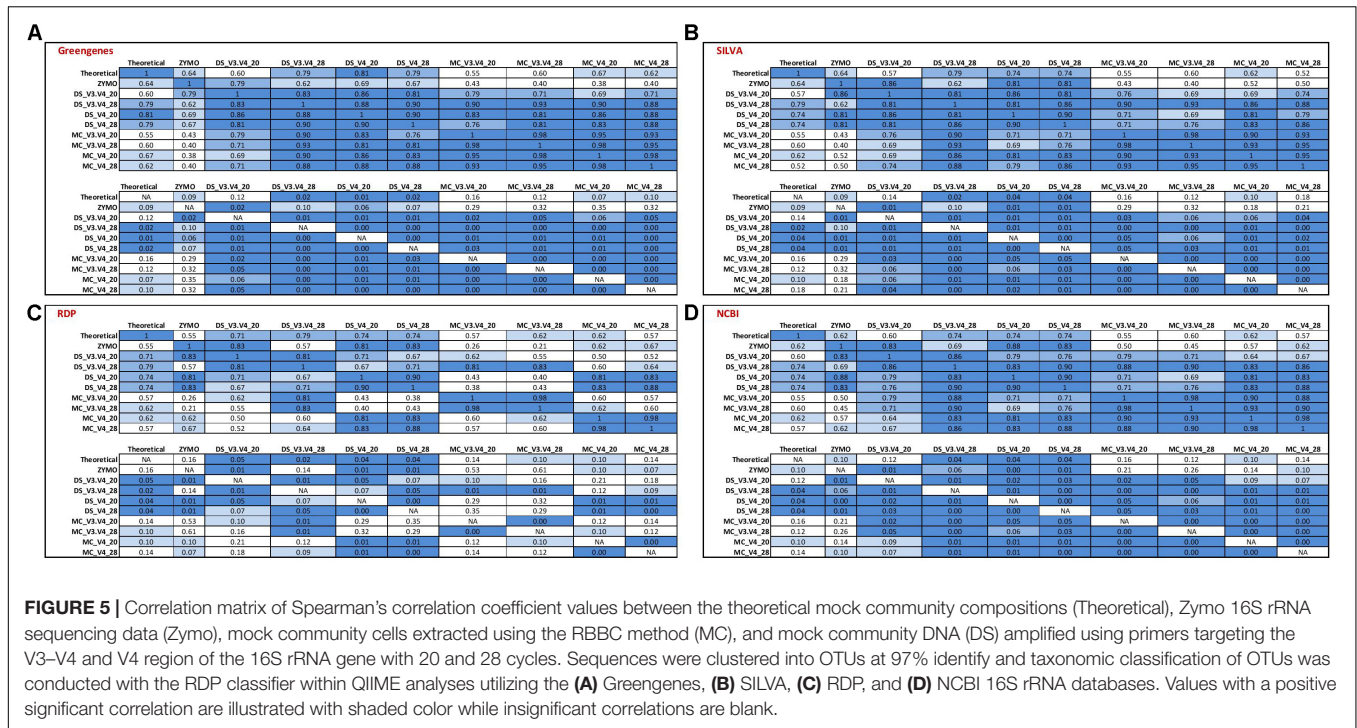


FIGURE 5 | Correlation matrix of Spearman's correlation coefficient values between the theoretical mock community compositions (Theoretical), Zymo 16S rRNA sequencing data (Zymo), mock community cells extracted using the RBBC method (MC), and mock community DNA (DS) amplified using primers targeting the V3–V4 and V4 region of the 16S rRNA gene with 20 and 28 cycles. Sequences were clustered into OTUs at 97% identity and taxonomic classification of OTUs was conducted with the RDP classifier within QIIME analyses utilizing the (A) Greengenes, (B) SILVA, (C) RDP, and (D) NCBI 16S rRNA databases. Values with a positive significant correlation are illustrated with shaded color while insignificant correlations are blank.

theoretical community for sequenced MC samples was observed ($P < 0.1$). The V4 sequences amplified from DS and MC samples at 20 and 28 PCR1 cycles were significantly correlated to their V3–V4 counterparts ($P < 0.05$) when classified with using the Greengenes, SILVA, and NCBI databases (Figures 5A,B,D). This correlation was not apparent in the MC samples classified with the RDP database ($P < 0.05$; Figure 5C). For Greengenes, SILVA, and NCBI databases, DS samples were correlated with their MC counterpart amplified with identical primers with the same PCR1 cycle conditions ($P < 0.05$; Figures 5A,B,D). This was also observed for MC and DS samples classified with the RDP databases at 28 PCR1 cycles for both primers and for samples amplified with V4 primers at 20 PCR1 cycles ($P < 0.05$; Figure 5C). A significant correlation was not observed between MC and DS both amplified with V3–V4 primers at 20 PCR1 cycles when classified with the RDP database ($P > 0.5$; Figure 5C).

taxMachine

Taxonomic classification of unique sequences from DS, MC, and ZE with taxMachine identified all eight bacterial genera (Figure 6). Similarly to the BLAST classification, the classification was consistent with the theoretical composition. With taxMachine as taxonomic classifier, the Spearman's correlation coefficients (Figure 7) were similar to that of the BLAST classification (Figure 2). ZE tended toward the theoretical mock community ($r = 0.69$, $P < 0.1$). The DS samples amplified with the V3–V4 and V4 at 20 and 28 PCR1 cycles were all significantly correlated with the theoretical mock community. Samples derived from DNA extracted from mock community cells showed a lower correlation to the

theoretical mock community with primer choice and PCR1 cycle number influencing this correlation (Figure 7). MC amplified with 20 PCR1 cycles showed increased correlation to the theoretical mock community over MC amplified with 28 PCR1 cycles. DNA extracted from mock community cells and DS amplified with V4 primers showed increased correlation to the theoretical mock community over MC and DS amplified with V3–V4 primers. DNA extracted from mock community cells amplified with V4 primers with 20 PCR1 cycles tended toward the theoretical mock community.

As a comparative tool, unique sequences from ZE were also classified with a version of the RDP classifier within R package rRDP. Spearman's correlation coefficients for unique sequences classified with RDP were significantly correlated for the theoretical mock community ($r = 0.71$, $P < 0.05$).

Negative Controls

DNA was extracted from water as a NC and was subjected to 20 and 28 PCR1 cycles with primers targeting the V3–V4 and V4 region of the 16S rRNA gene (Figure 3). Compared to DS and MC, read numbers were negligible (<1000 sequences) for NC amplified with V4 primers at 20 and 28 PCR1 cycles and NC amplified with V3–V4 at 20 PCR1 cycles. One NC sample (amplified V3–V4 at 28 PCR1 cycles) had ~14,500 merged sequences after quality filtering. Although this was much higher than the other NC samples, it was less than a 16th of the average number of sequences obtained for DS and MC community samples. NC amplified with 28 PCR1 cycles had a higher number of merged sequence reads when amplified with both V4 and V3–V4 primers.

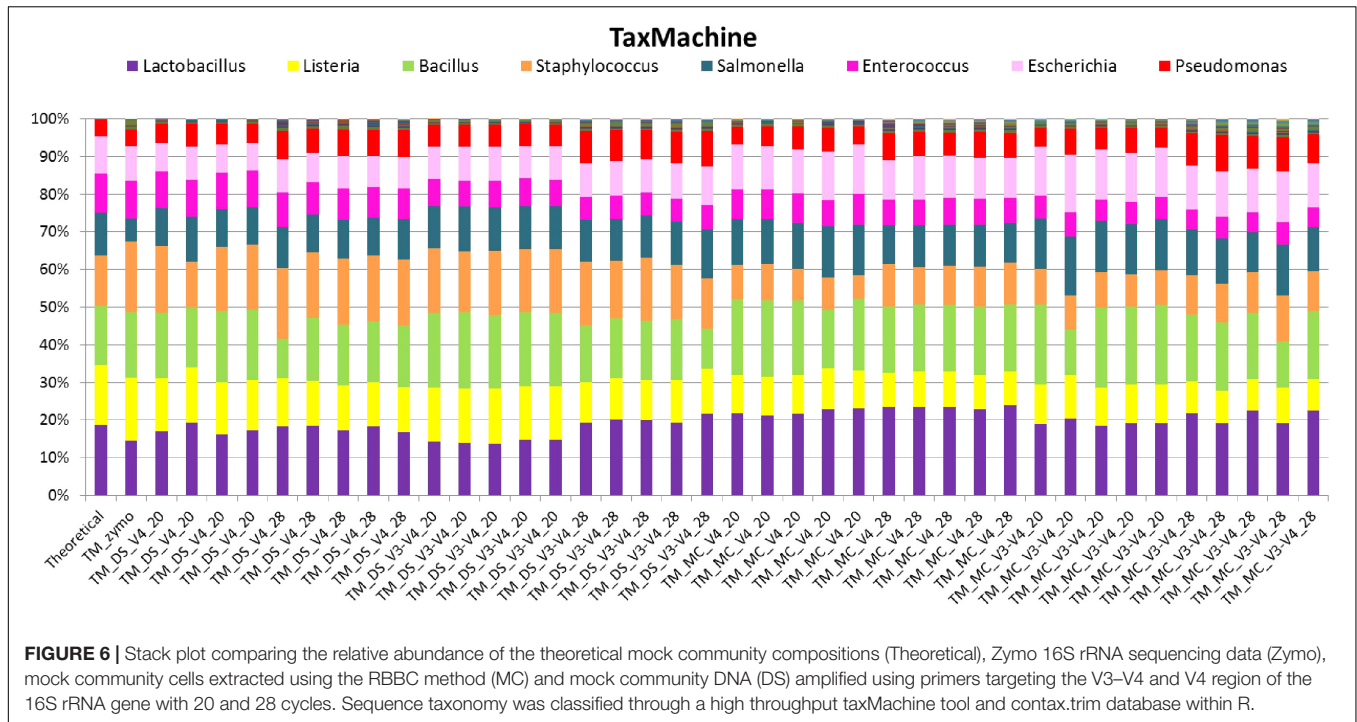


FIGURE 6 | Stack plot comparing the relative abundance of the theoretical mock community compositions (Theoretical), Zymo 16S rRNA sequencing data (Zymo), mock community cells extracted using the RBBC method (MC) and mock community DNA (DS) amplified using primers targeting the V3–V4 and V4 region of the 16S rRNA gene with 20 and 28 cycles. Sequence taxonomy was classified through a high throughput taxMachine tool and contax.trim database within R.

taxMachine

Spearman's Correlation Coefficient

	Theoretical	ZYMO	DS_V4_20	DS_V4_28	DS_V3.V4_20	DS_V3.V4_28	MC_V4_20	MC_V4_28	MC_V3.V4_20	MC_V3.V4_28
Theoretical	1	0.69	0.88	0.90	0.76	0.76	0.67	0.57	0.62	0.38
ZYMO	0.69	1	0.81	0.79	0.88	0.60	0.26	0.33	0.31	0.14
DS_V4_20	0.88	0.81	1	0.93	0.90	0.83	0.69	0.67	0.71	0.52
DS_V4_28	0.90	0.79	0.93	1	0.81	0.90	0.62	0.62	0.57	0.48
DS_V3.V4_20	0.76	0.88	0.90	0.81	1	0.76	0.55	0.55	0.62	0.40
DS_V3.V4_28	0.76	0.60	0.83	0.90	0.76	1	0.71	0.74	0.67	0.71
MC_V4_20	0.67	0.26	0.69	0.62	0.55	0.71	1	0.95	0.98	0.86
MC_V4_28	0.57	0.33	0.67	0.62	0.55	0.74	0.95	1	0.93	0.93
MC_V3.V4_20	0.62	0.31	0.71	0.57	0.62	0.67	0.98	0.93	1	0.83
MC_V3.V4_28	0.38	0.14	0.52	0.48	0.40	0.71	0.86	0.93	0.83	1

P value

	Theoretical	ZYMO	DS_V4_20	DS_V4_28	DS_V3.V4_20	DS_V3.V4_28	MC_V4_20	MC_V4_28	MC_V3.V4_20	MC_V3.V4_28
Theoretical	NA	0.06	0.00	0.00	0.03	0.03	0.07	0.14	0.10	0.35
ZYMO	0.06	NA	0.01	0.02	0.00	0.12	0.53	0.42	0.46	0.74
DS_V4_20	0.00	0.01	NA	0.00	0.00	0.01	0.06	0.07	0.05	0.18
DS_V4_28	0.00	0.02	0.00	NA	0.01	0.00	0.10	0.10	0.14	0.23
DS_V3.V4_20	0.03	0.00	0.00	0.01	NA	0.03	0.16	0.16	0.10	0.32
DS_V3.V4_28	0.03	0.12	0.01	0.00	0.03	NA	0.05	0.04	0.07	0.05
MC_V4_20	0.07	0.53	0.06	0.10	0.16	0.05	NA	0.00	0.00	0.01
MC_V4_28	0.14	0.42	0.07	0.10	0.16	0.04	0.00	NA	0.00	0.00
MC_V3.V4_20	0.10	0.46	0.05	0.14	0.10	0.07	0.00	0.00	NA	0.01
MC_V3.V4_28	0.35	0.74	0.18	0.23	0.32	0.05	0.01	0.00	0.01	NA

FIGURE 7 | Correlation matrix of Spearman correlation coefficient values between the theoretical mock community compositions (Theoretical), Zymo 16S rRNA sequencing data (Zymo), mock community cells extracted using the RBBC method (MC), and mock community DNA (DS) amplified using primers targeting the V3–V4 and V4 region of the 16S rRNA gene with 20 and 28 cycles. Sequences were assigned taxonomic classification using the taxMachine tool and the contax.trim database within R. Values with a positive significant correlation are illustrated with shaded color while insignificant correlations are blank.

DISCUSSION

Ruminants are unique in their ability to convert cellulose in plant cell walls into high-quality meat and milk protein suitable for human consumption. Despite their importance in animal protein production, ruminants are a major contributor to global anthropogenic greenhouse gas emissions by producing methane via enteric fermentation. In consideration of this, the majority of current rumen microbiomic studies aim to elucidate the microbial community present in the rumen in order to increase nutrient utilization and reduce methane emissions. It is crucial that these studies are giving an accurate representation of the rumen environment and that the studies are reproducible and suitable for meta-analysis. Currently, there is a lack of concordance between rumen microbiomic studies with research groups, including our own, failing to include negative extraction controls and positive controls in their analysis (McCabe et al., 2015; Myer et al., 2015; Meale et al., 2016; Shabat et al., 2016; Popova et al., 2017). Among the human microbiome community, concerns have been raised in response to the recent uncertainty surrounding the accuracy, consistency, and reliability of 16S rRNA phylogenetic studies (Sinha et al., 2015, 2017). In our present study, we subjected some of the methods that are currently used for investigation of the rumen microbiota using mock community cells and DSs. Power analysis inferred that technical replication of five was sufficient to deduce a significant correlation between the technical treatments examined in the current experiment and the theoretical composition of the mock community supplied by Zymo.

Zymo maintain that they provide an unbiased approach for DNA extraction. The fastq files provided by Zymo containing 16S rRNA sequencing data from their extraction of the microbial mock community with the ZymoBIOMCS™ DNA Mini Kit were used initially to confirm the method of taxonomic classification. All eight genera were detected from ZE using BLAST as a method of classification, confirming that the method of classification was suitable for methodological comparative analysis. The results obtained for the ZE showed a compositional trend toward that of the theoretical mock community. 16S rRNA sequences amplified from the DS under the amplification conditions regularly utilized in rumen microbiomic studies showed better correlation to the theoretical mock community than ZE. Results from the sequenced mock community DNA indicated that both the V3–V4 and V4 primer set in this study accurately identified the genera present in the mock microbial community. All eight genera were detected and in relatively accurate proportions. The DS amplified with primers targeting the V4 region of the 16S rRNA gene showed an improved correlation to the mock microbial community than the DS samples amplified with primers targeting the V3–V4 region of the 16S rRNA gene. This may have been due to the increased length of the V3–V4 primer sequence. The base quality of the sequencing deteriorates at the end of the read (Schirmer et al., 2015). The increased length of the V3–V4 sequence allows only around 20 bp of overlap between sequence reads when a 250-cycle MiSeq reagent kit is used (Illumina had serious QC issues with the 300-cycle kits at the time of writing) and requires utilization of lower quality read ends. This may

have contributed to the marginally lower correlation coefficient values discerned for the V3–V4 sequences relative to the V4 sequences. The correlation of DS to the mock community did not differ greatly between 20 and 28 PCR1 cycles. However, non-specific background amplification was increased for samples generated with 28 PCR1 cycles in comparison to amplicons generated with 20 PCR1 cycles indicating that 20 PCR1 cycles was favorable for 16S rRNA gene amplicon generation. The results provide confirmation that both primer types are suitable for amplification of bacteria for phylogenetic community analysis and that 20 PCR1 cycles was the preferable option for the amplicon amplification when utilizing these primer sets.

Having confirmed that the amplification process was satisfactory, we examined the DNA extraction methodology commonly utilized for extraction of microbial DNA from rumen microbiota. All eight genera were produced from the sequences output from the MC, inferring that the extraction protocol is suitable for both Gram-positive and Gram-negative bacteria. The MC did not reflect the theoretical mock community as precisely as the DS; however, the correlation coefficient obtained for the MC samples was similar to that obtained when comparing the ZE to the theoretical mock community. Zymo deemed this result to be an accurate representation of the theoretical mock community profile. The increase in background amplification in MC in comparison to DS may have been a contributing factor in the reduced correlation with the theoretical mock community. Addition of any steps to the full NGS protocol is invariably going to contribute to non-specific amplification, due to the sensitivity of “universal” bacterial primers (Caporaso et al., 2011; Takahashi et al., 2014) and possible residual bacterial DNA present in extraction reagents. Non-specific amplification could have been generated during the PCR step of the protocol due to the KAPA Taq DNA polymerase which is purified from *E. coli*. However, this is not likely as the number of sequence reads in the NCs was low and there was a high correlation to the theoretical mock community in the DS samples. Therefore, the reduced correlation coefficient for MC in comparison to DS is likely to be due to minor extraction inefficiencies. 16S rRNA gene amplification of DNA extraction from microbial cells would inevitably be less efficient than amplification from DNA directly; however, the DS and MC did show strong correlations to each under the same conditions highlighting the minor variation between the two amplified communities. It was also noted that MC followed the same pattern as DS, with samples amplified with V4 primers showing a higher correlation to the theoretical composition of the mock microbial community than samples amplified with primers targeting the V3–V4 region of the 16S rRNA gene, further validating the results from both the sampled MC and DS. The results provided confirmation that both primer types are suitable for amplification of bacteria for phylogenetic community analysis and that 20 PCR1 cycles was the preferable option for the amplicon amplification when utilizing these primer sets in this two PCR step protocol. The results indicate that MC provided the best simulation of the theoretical mock community when amplified with primers targeting the V4 region of the 16S rRNA gene at 20 PCR1 cycles.

The major area of concern arose not with the DNA extraction or amplicon generation process but with the taxonomic classification of the 16S rRNA sequences. NCBI BLAST in conjunction with the NCBI 16S database was chosen as a method for this assessment as the Zymo mock community being examined contained well-characterized bacteria. However, BLAST is computationally expensive, slow and not refined for the phylogenetic assessment of rumen microbiota which are largely uncharacterized to species or even family level. A more common method for handling of 16S rRNA gene sequences from rumen phylogenetic studies is the publicly available pipelines such as QIIME (Caporaso et al., 2010). QIIME is a wrapper for applications such as quality filtering, OTU picking, and taxonomic assignment. A general approach to dealing with rumen 16S rRNA gene sequences is to cluster sequences at 97% identity utilizing QIIME's open reference-based OTU picking approach and use the RDP classifier algorithm for taxonomic classification of 16S rRNA amplicons (Wang et al., 2007). Representative OTUs are then assigned to bacterial taxonomies using RDP classifier via QIIME. This taxonomic-independent, OTU-based approach was developed to overcome the underrepresentation of 16S rRNA sequences originating in the publicly available 16S rRNA databases. This solution to cluster sequences at 97% identity was proposed by Stackebrandt and Goebel (1994) when few full-length 16S rRNA sequences were available and sequencing studies were still in their infancy. Due to the generic nature of the mock microbial community, this taxonomic-independent approach was assumed not to be a necessity but was undertaken in order to assess rumen microbiomic approaches to phylogenetic assessment. Our results indicated that the RDP classifier struggled to identify 16S rRNA V3–V4 and V4 OTUs to genus level, using the default confidence threshold of 80%. For all the databases we assessed, the RDP classifier failed to discern between the members of the Enterobacteriaceae family within ZE, DS, and MC samples when sequences were clustered into OTUs at 97% identity. It should be noted that, at a species level, the sequences of *E. coli* and *S. enterica* are two of the most well-characterized bacterial species with numerous genomic assemblies for each on the NCBI database (Edgar, 2018); reducing these sequences to representative OTU sequences may not necessarily emulate the diversity of the original sequences. It is therefore possible the confidence threshold was too high for accurate identification at a genus level.

Although not the initial aim of our study, taxonomic classification became a concern. Liland et al. (2017) highlight that while the RDP classifier tool is a good method for classification, it was not perfect and that they had found that other methods outperformed the RDP classifier (Vinje et al., 2015). A new ready-to-use tool for taxonomic classification method for 16S rRNA sequences, taxMachine, within the microclass R-package, became available during the course of our study (Liland et al., 2017). This was used to classify samples sequenced within the experiment as an alternative tool to the standard QIIME pipeline. The taxMachine tool used an internal optimized

database contax.trim database for sequence classification. Liland et al. (2017) emphasized that there is no comprehensive gold standard databases available for taxonomic classification of 16S rRNA sequences due to our growing knowledge of the microbial community. This variability and lack of comprehensiveness were evident in database coverage and were visible in our own results of the RDP classification with the Greengenes, SILVA, RDP, and NCBI databases, giving varied correlation coefficient values for the comparison of ZE, DS, and MC to the theoretical mock community. The microcontax dataset was designed based on consensus taxonomy assignment from several data repositories and was highlighted to be the closest to a gold standard currently available (Liland et al., 2017). For this element of the investigation, we moved away from the convention of clustering sequences at 97% identity (Edgar, 2018). There was improvement in the community profiles from amplicon sequences derived from ZE, DS, and MC when identical sequences were aligned and classified with the taxMachine tool. Interestingly, the same result was observed when the OTU clustering step was removed and identical ZE sequences were classified with the RDP classifier within the rRDP package in R (Hahsler and Nagar, 2014). This indicates that the standard conventional pipeline we have been utilizing for 16S rRNA may not be providing the most accurate results. Clustering sequencing into OTUs of 97% identity provided a facile tool for management of the vast amount of sequence data obtained from rumen amplicon studies but it may be time to move away from this concept; however, this will need to be explored further with a mock community which more closely resembles the diversity and genera present in the rumen.

Despite the limitations in taxonomic classification highlighted in the current study, it is acknowledged that global efforts are ongoing to improve our interpretation of rumen microbial communities (Seshadri et al., 2018; Stewart et al., 2018).

CONCLUSION/FUTURE WORK

Methods for phylogenetic analysis are fundamental for translation of research findings into application. This study highlights that we are following the right path for accurate community analysis; however, still have improvements to make. The analysis indicated that utilizing the RBBC method for DNA extraction in combination with primers targeting the 16S rRNA gene using 20 PCR cycles was sufficient for amplicon sequencing to generate a relatively accurate depiction of the bacterial communities present in rumen samples. However, this study was conducted using a low diversity bacterial mock community. In order to obtain a more tangible representation of the efficiency of our DNA extraction and PCR methodology, it is important to conduct a similar study using a mock community which simulates the microbiota present in the rumen. However, this is currently unavailable and common pipelines for bioinformatics analysis will also need to be appraised. Although beyond the scope of this current study, we identified some major flaws in the rumen microbiomic pipelines we and others had been utilizing. This study also highlights the need for positive mock community controls within all rumen microbiomic studies in

order to discern errors which may arise at any step during an NGS protocol.

AUTHOR CONTRIBUTIONS

MM, EM, and SW conceived and designed the experiments. EM and MM performed the experiments and wrote the paper. EM, MM, and GB analyzed the data. EM, MM, SW, and GB contributed reagents, materials, and analysis tools, interpreted the results.

REFERENCES

- Andrews, S. (2010). *FastQC: A Quality Control Tool for High throughput Sequence Data*. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., et al. (2015). The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol.* 15:66. doi: 10.1186/s12866-015-0351-6
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl. 1), 4516–4522. doi: 10.1073/pnas.1000080107
- Champely, S. (2015). *pwr: Basic Functions for Power Analysis. R Package Version 1*. Available at: <https://cran.r-project.org/package=pwr>
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Edgar, R. C. (2018). Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics*. doi: 10.1093/bioinformatics/bty113 [Epub ahead of print].
- 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
- Fischer, M. A., Güllert, S., Neulinger, S. C., Streit, W. R., and Schmitz, R. A. (2016). Evaluation of 16S rRNA gene primer pairs for monitoring microbial community structures showed high reproducibility within and low comparability between datasets generated with multiple archaeal and bacterial primer pairs. *Front. Microbiol.* 7:1297. doi: 10.3389/fmicb.2016.01297
- Hahsler, M., and Nagar, A. (2014). *rRDP: Interface to the RDP Classifier*. Available at: <http://sourceforge.net/projects/rdp-classifier/>
- Harrell, F. E. Jr. (2014). *Hmisc Package Version 3.14–16*. Available at: <http://cran.r-project.org/package=Hmisc>
- 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
- Hook, S. E., Steele, M. A., Northwood, K. S., Dijkstra, J., France, J., and Wright, A. G. (2011). Impact of subacute ruminal acidosis (SARA) adaptation and recovery on the density and diversity of bacteria in the rumen of dairy cows. *FEMS Microbiol. Ecol.* 78, 275–284. doi: 10.1111/j.1574-6941.2011.01154.x
- Illumina (2013). *16S Metagenomic Sequencing Library Preparation*. Available at: http://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf [accessed January 8, 2017].
- Jensen, M. A., Webster, J. A., and Straus, N. (1993). Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. *Appl. Environ. Microbiol.* 59, 945–952.
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical

FUNDING

The study was funded by a FACCE-JPI “Rumen Stability” grant (2014). EM was funded by a Walsh Fellowship award (Ref. No. 2014231).

ACKNOWLEDGMENTS

The authors acknowledge Paul Cormican for the bioinformatic advice.

- and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1. doi: 10.1093/nar/gks808
- Knapp, J., Laur, G., Vadas, P., Weiss, W., and Tricarico, J. (2014). Invited review: enteric methane in dairy cattle production: quantifying the opportunities and impact of reducing emissions. *J. Dairy Sci.* 97, 3231–3261. doi: 10.3168/jds.2013-7234
- Liland, K. H., Vinje, H., and Snipen, L. (2017). microclass: an R-package for 16S taxonomy classification. *BMC Bioinformatics* 18:172. doi: 10.1186/s12859-017-1583-2
- Liu, Z., DeSantis, T. Z., Andersen, G. L., and Knight, R. (2008). Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res.* 36:e120. doi: 10.1093/nar/gkn491
- McCabe, M. S., Cormican, P., Keogh, K., O’Connor, A., O’Hara, E., Palladino, R. A., et al. (2015). Illumina MiSeq phylogenetic amplicon sequencing shows a large reduction of an uncharacterised Succinivibrionaceae and an increase of the *Methanobrevibacter gottschalkii* clade in feed restricted cattle. *PLoS One* 10:e0133234. doi: 10.1371/journal.pone.0133234
- McCann, J. C., Wiley, L. M., Forbes, T. D., Rouquette, F. M. Jr., and Tedeschi, L. O. (2014). Relationship between the rumen microbiome and residual feed intake-efficiency of Brahman bulls stocked on bermudagrass pastures. *PLoS One* 9:e91864. doi: 10.1371/journal.pone.0091864
- Meale, S. J., Li, S., Azevedo, P., Derakhshani, H., Plaizier, J. C., Khafipour, E., et al. (2016). Development of ruminal and fecal microbiomes are affected by weaning but not weaning strategy in dairy calves. *Front. Microbiol.* 7:582. doi: 10.3389/fmicb.2016.00582
- Myer, P. R., Smith, T. P., Wells, J. E., Kuehn, L. A., and Freetly, H. C. (2015). Rumen microbiome from steers differing in feed efficiency. *PLoS One* 10:e0129174. doi: 10.1371/journal.pone.0129174
- Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740. doi: 10.1126/science.276.5313.734
- Petri, R. M., Schwaiger, T., Penner, G. B., Beauchemin, K. A., Forster, R. J., and McKinnon, J. J. (2013). Changes in the rumen epimural bacterial diversity of beef cattle as affected by diet and induced ruminal acidosis. *Appl. Environ. Microbiol.* 79, 3744–3755. doi: 10.1128/aem.03983-12
- Popova, M., McGovern, E., McCabe, M. S., Martin, C., Doreau, M., Arbre, M., et al. (2017). The structural and functional capacity of ruminal and cecal microbiota in growing cattle was unaffected by dietary supplementation of linseed oil and nitrate. *Front. Microbiol.* 8:937. doi: 10.3389/fmicb.2017.00937
- Puniya, A. K., Singh, R., and Kamra, D. N. (2015). *Rumen Microbiology: From Evolution to Revolution*. Berlin: Springer. doi: 10.1007/978-81-322-2401-3
- Qiu, X., Wu, L., Huang, H., McDonel, P. E., Palumbo, A. V., Tiedje, J. M., et al. (2001). Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Appl. Environ. Microbiol.* 67, 880–887. doi: 10.1128/aem.67.2.880-887.2001
- Roehle, R., Dewhurst, R. J., Duthie, C.-A., Rooke, J. A., McKain, N., Ross, D. W., et al. (2016). Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. *PLoS Genet.* 12:e1005846. doi: 10.1371/journal.pgen.1005846
- Saleem, F., Ametaj, B. N., Bouatra, S., Mandal, R., Zebeli, Q., Dunn, S. M., et al. (2012). A metabolomics approach to uncover the effects of grain diets on rumen health in dairy cows. *J. Dairy Sci.* 95, 6606–6623. doi: 10.3168/jds.2012-5403

- Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., and Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* 43:e37. doi: 10.1093/nar/gku1341
- Seshadri, R., Leahy, S. C., Attwood, G. T., Teh, K. H., Lambie, S. C., Cookson, A. L., et al. (2018). Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat. Biotechnol.* 36, 359–367. doi: 10.1038/nbt.4110
- Shabat, S. K. B., Sasson, G., Doron-Faigenboim, A., Durman, T., Yaacoby, S., Miller, M. E. B., et al. (2016). Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *ISME J.* 10, 2958–2972. doi: 10.1038/ismej.2016.62
- Shi, W., Moon, C. D., Leahy, S. C., Kang, D., Froula, J., Kittelmann, S., et al. (2014). Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res.* 24, 1517–1525. doi: 10.1101/gr.168245.113
- Sinha, R., Abnet, C. C., White, O., Knight, R., and Huttenhower, C. (2015). The microbiome quality control project: baseline study design and future directions. *Genome Biol.* 16:276. doi: 10.1186/s13059-015-0841-8
- Sinha, R., Abu-Ali, G., Vogtmann, E., Fodor, A. A., Ren, B., Amir, A., et al. (2017). Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat. Biotechnol.* 35, 1077–1086. doi: 10.1038/nbt.3981
- Stackebrandt, E., and Goebel, B. M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 44, 846–849. doi: 10.1099/00207713-44-4-846
- Stewart, R. D., Auffret, M. D., Warr, A., Wisner, A. H., Press, M. O., Langford, K. W., et al. (2018). Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* 9:870. doi: 10.1038/s41467-018-03317-6
- Stulberg, E., Fravel, D., Proctor, L. M., Murray, D. M., LoTempio, J., Chrisey, L., et al. (2016). An assessment of US microbiome research. *Nat. Microbiol.* 1:15015. doi: 10.1038/nmicrobiol.2015.15
- Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., and Nishijima, M. (2014). Development of a prokaryotic universal primer for simultaneous analysis of *Bacteria* and *Archaea* using next-generation sequencing. *PLoS One* 9:e105592. doi: 10.1371/journal.pone.0105592
- Tringe, S. G., and Hugenholtz, P. (2008). A renaissance for the pioneering 16S rRNA gene. *Curr. Opin. Microbiol.* 11, 442–446. doi: 10.1016/j.mib.2008.09.011
- Vinje, H., Liland, K. H., Almøy, T., and Snipen, L. (2015). Comparing K-mer based methods for improved classification of 16S sequences. *BMC Bioinformatics* 16:205. doi: 10.1186/s12859-015-0647-4
- Wallace, R. J., Rooke, J. A., Duthie, C.-A., Hyslop, J. J., Ross, D. W., McKain, N., et al. (2014). Archaeal abundance in post-mortem ruminal digesta may help predict methane emissions from beef cattle. *Sci. Rep.* 4:5892. doi: 10.1038/srep05892
- Wang, G. C., and Wang, Y. (1997). Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Appl. Environ. Microbiol.* 63, 4645–4650.
- Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive 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
- Youssef, N., Sheik, C. S., Krumholz, L. R., Najjar, F. Z., Roe, B. A., and Elshahed, M. S. (2009). Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Appl. Environ. Microbiol.* 75, 5227–5236. doi: 10.1128/aem.00592-09
- Yu, Z., and Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* 36, 808–812. doi: 10.2144/04365ST04
- Zymo (2016). *ZymoBIOMICS™ Microbial Community DNA Standard*. Irvine, CA: Zymo Research.

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 © 2018 McGovern, Waters, Blackshields and McCabe. 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) and the copyright owner 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.