Journal of Microbiology (2014) Vol. 52, No. 5, pp. 355–365
DOI 10.1007/s12275-014-3443-3

Variations in 16S rRNA-based Microbiome Profiling between Pyrosequencing Runs and between Pyrosequencing Facilities

Minseok Kim[#] and Zhongtang Yu[*]

Department of Animal Sciences, The Ohio State University, Columbus, OH, USA
[#]Present address: USDA, ARS, USMARC, Clay Center, NE 68933, USA

(Received Aug 21, 2013 / Revised Nov 15, 2013 / Accepted Nov 25, 2013)

Pyrosequencing of 16S rRNA gene amplicons on the 454 FLX Titanium platform has been widely used to analyze microbiomes in various environments. However, different results may stem from variations among sequencing runs or among sequencing facilities. This study aimed to evaluate these variations between different pyrosequencing runs by sequencing 16S rRNA gene amplicon libraries generated from three sets of rumen samples twice each on the 454 FLX Titanium system at two independent sequencing facilities. Similar relative abundances were found for predominant taxa represented by large numbers of sequence reads but not for minor taxa represented by small numbers of sequence reads. The two sequencing facilities revealed different bacterial profiles with respect to both predominant taxa and minor taxa, including the most predominant genus *Prevotella*, the family *Lachnospiraceae*, and the phylum *Proteobacteria*. Differences in primers used to generate amplicon libraries may be a major source of variations in microbiome profiling. Because different primers and regions of 16S rRNA genes are often used by different researchers, significant variations likely exist among studies. Quantitative interpretation for relative abundance of taxa, especially minor taxa, from prevalence of sequence reads and comparisons of results from different studies should be done with caution.

Keywords: 16S rRNA genes, microbiomes, primers, pyrosequencing, sequencing facility

Introduction

Microbes are present in various environments on the Earth and play an important role therein. However, most microbes remain uncultured and unknown, hindering understanding of their ecology such as diversity, community composition, and population dynamics. The use of 16S rRNA gene, in most cases one or a few hypervariable regions, as a marker gene has revolutionized analysis of diversity and composition of microbiomes by potentially allowing detection of all members, both culturable and unculturable, of any microbiome (Pace, 1997). The availability of specialty databases, including the RDP (Wang *et al.*, 2007), Greengenes (DeSantis *et al.*, 2006; Wang *et al.*, 2007), and Silva (Pruesse *et al.*, 2007), has greatly facilitated taxonomic classification of 16S rRNA gene sequences recovered from microbiome samples and then determination of diversity and composition of microbiomes. Initially, construction and analysis of 16S rRNA gene clone libraries were the primary method used to analyze microbiomes, but the depth of coverage affordable by this methodology is limited, making it infeasible to comprehensively characterize the diversity and composition of any complex microbiome.

The next-generation sequencing (NGS) technologies, primarily the Illumina and Roche-454 technologies, have enabled detailed scrutiny of various complex microbiomes (Petrosino *et al.*, 2009). Although the Illumina platform is more cost-effective compared to the Roche-454 platform on a per base basis, the Illumina platforms currently can only read up to 250 bases while the Roche-454 platforms can read up to 1000 bases (Weinstock, 2012). Therefore, the Roche-454 platform can result in better taxonomic classification, and indeed it has been used in more metagenomic studies than other NGS technologies. With either of the NGS technologies, PCR amplicons are prepared from metagenomic DNA using domain-specific primers. Several factors, such as PCR bias (Rajendhran and Gunasekaran, 2011), primers bias (Frank *et al.*, 2008), choices of sequence aligners, and bioinformatic analysis pipelines (Barriuso *et al.*, 2011), have been shown to affect the results of microbiome analysis. However, repeatability of microbiome analysis using NGS technologies has not been systematically examined. The objective of this study was to examine the magnitude of variations between pyrosequencing runs performed at two sequencing facilities using the same type of FLX Titanium system.

Materials and Methods

Sampling and DNA extraction

Rumen samples were obtained for three previous studies (Felix and Loerch, 2011; Mathew *et al.*, 2011; Stiverson *et al.*, 2011) and preserved at -80°C. Metagenomic DNA was extracted from each rumen sample as described previously (Yu and Morrison, 2004). In Study 1 (Felix and Loerch, 2011), rumen samples were collected from Angus beef cattle fed the following four diets: 1) a control diet (designated as "C"), 2) control diet supplemented with haylage (designated as "H"), 3) control diet supplemented with monensin (designated as "M"), and 4) control diet supplemented with both

*For correspondence. E-mail: yu.226@osu.edu; Tel.: +1-614-292-3057; Fax: +1-614-292-2929

monensin and haylage (designated as "MH"). Four composite metagenomic DNA samples were prepared from each of the four dietary groups (four cattle per group). In Study 2 (Mathew *et al.*, 2011), rumen samples were collected from Holstein dairy cattle fed the following three diets: 1) a control diet (designated as "C"), 2) control diet supplemented with Rumensin® (monensin sodium, designated as "R"), and 3) control diet supplemented with both Rumensin® and fat (designated as "RF"). These rumen samples were fractionated into their adherent fractions (Ad) and liquid fractions (Lq) as described previously (Kim *et al.*, 2011a). Six composite metagenomic DNA samples were prepared from the six fractioned samples (two fraction per group, and six cattle per group). In Study 3 (Stiverson *et al.*, 2011), rumen samples were collected from sheep fed the following two diets: 1) orchardgrass hay only (designated as "H") and 2) a mixture of orchardgrass hay and corn (designated as "C"). These rumen samples were also fractionated into their adherent fractions and liquid fractions. Four composite metagenomic DNA samples were prepared from the four fractioned samples (two fractions per group, and four sheep per group).

Pyrosequencing

The V1-V3 region of 16S rRNA genes was amplified from each composite DNA sample using modified universal primers 454BactF (5′-AKRGTTYGATYNTGGCTCAG-3′) and 454BactR (5′-GTNTBACCGCDGCTGCTG-3′) that have increased degeneracy (Nelson, 2011) compared to the original primers 27F and 519R (Lane *et al.*, 1991). A unique barcoded region was tagged to each primer downstream of the sequencing primers A and B. Each PCR reaction mixture (50 μl) contained 100 ng composite DNA, 1X PCR buffer (20 mM Tris-HCl; pH 8.4 and 50 mM KCl), 200 μM deoxynucleoside triphosphates, 100 nM (each) primer, 1.75 mM $MgCl_2$, 670 ng of bovine serum albumin/μl, and 1.25 U of Platinum Taq DNA polymerase (Invitrogen, USA). PCR reaction was performed with 30 thermal cycles (denaturation at 95°C for 30 sec, annealing at 60°C for 30 sec, and extension at 72°C for 60 sec) using a PTC-100 thermocycler (MJ Research, USA). All amplicons were purified using a QIAquick Gel Extraction Kit (QIAGEN, USA) and then quantified using a NanoDrop ND-1000 UV-Vis Spectrophotometer (NanoDrop, USA). The 14 amplicon libraries prepared from the three sets of samples were each diluted with the Elution buffer of the QIAquick Gel Extraction Kit to a concentration of 20 ng/μl and pooled together at an equal amount (also equal molar ratio because they have the same length). This pool was sequenced twice on two different dates at the Plant-Microbe Genomics Facility, The Ohio State University (OSU), using a 454 GS FLX Titanium system (designated as OSU datasets 1 and 2).

Similar V1-V3 amplicon libraries were also prepared from the same 14 composite DNA samples and sequenced twice on two different dates at the Research and Testing Laboratories (RTL) in Texas using a 454 GS FLX Titanium system. An amplicon library of the V1-V3 region was prepared for each sample similarly as described above, except for the primers Gray28F (5′-GAGTTTGATCNTGGCTCAG-3′) and Gray-519R (5′-GTNTTACNGCGGCKGCTG-3′) that have been used at RTL (Dowd *et al.*, 2008; Callaway *et al.*, 2010; Pitta *et al.*, 2010). The HotStarTaq Plus Master Mix Kit (QIAGEN) was used in the PCR with 100 ng composite DNA, and the PCR conditions were 32 cycles of denaturation at 94°C for 30 sec, annealing at 60°C for 40 sec, and extension at 72°C for 60 sec. The resultant two datasets from the two sequencing runs were referred to as RTL datasets 1 and 2. The OSU datasets and the RTL datasets were compared to evaluate intra-facility variations, while the two datasets from each of the two sequencing facilities were also compared to examine inter-facility variations.

Sequence processing and bioinformatics analysis

The programs implemented in the QIIME software package v.1.4.0 (Caporaso *et al.*, 2010) were used to process and analyze all the sequences. From the initial sequence reads returned from both sequencing facilities, the barcodes and primers were trimmed off. Sequences that had a read length shorter than 200 nt or longer than 650 nt, a mean quality score below Q25, or a homopolymer stretch longer than 8 nt were removed. The resultant sequences were denoised using the default method (Reeder and Knight, 2010) in QIIME to identify and remove erroneous sequences. Then non-chimeric sequences were obtained using the UCHIME program (Edgar *et al.*, 2011). Species-level OTUs were clustered at 97% sequence similarity using the uclust algorithm (Edgar, 2010). Taxonomic assignment was conducted using the RDP naïve Bayesian rRNA Classifier (Wang *et al.*, 2007). To avoid impact from different library sizes (Gihring *et al.*, 2012), we used the same number of sequences for all the 56 "samples" (14 composite samples × 2 runs × 2 sequencing facilities) to assess the inter-run and inter-facility variations. As recommended by Benson *et al.* (2010), log-transformed relative abundance (reads of a taxon/ total reads) of each taxon was used to compare the microbiome structure and to examine the distribution of taxa among the datasets using Principal Component Analysis (PCA), which was performed using the XLSTAT's statistical analysis software (Addinsoft, USA). Diversity indices including Chao1, PD_whole_tree, and Shannon diversity index were calculated for each group of samples using the OTU table produced by QIIME.

Statistics

Data were analyzed using one-way ANOVA with the PROC GLM of SAS (SAS Institute Inc., USA). Contrasts were used to compare means of log-transformed relative abundance of each taxon and diversity indices between the OSU datasets and the RTL datasets for each diet/sample group of each sample set. Comparisons of the datasets between the two sequencing facilities irrespective of diet were also performed for each sample set. Significance was declared at $P<0.05$ and tendency was declared at $P<0.10$. Percent coverage was calculated for all the samples by dividing the number of observed OTUs by the asymptote that was predicted from the rarefaction curve using PROC NLIN of SAS as described previously (Larue *et al.*, 2005).

Table 1. Diversity statistics

Study	Sample	Diversity indices (1st run : 2nd run)										
		No. of OTUs (coverage)[a]	Contrast		Chao1	Contrast		PD_whole_tree	Contrast		Shannon	Contrast	
			Trt[b]	Facility[c]		Trt[b]	Facility[c]		Trt[b]	Facility[c]		Trt[b]	Facility[c]
Study 1	C (OSU)	394 (58%) : 293 (70%)	<0.05	<0.001	1420: 762	<0.1	<0.01	30.4 : 25.4	<0.01	<0.001	6.8 : 6.5	<0.05	<0.001
	C (RTL)	186 (92%) : 132 (99%)			299 : 169			19.7 : 15.2			6.1 : 5.7		
	H (OSU)	491 (55%) : 387 (70%)	<0.01		1951 : 1051	<0.05		36.4 : 32.7	<0.01		7.4 : 7.0	<0.05	
	H (RTL)	248 (87%) : 238 (92%)			370 : 425			24.2 : 24.7			6.7 : 6.6		
	M (OSU)	341 (60%) : 266 (80%)	<0.05		1125 : 628	0.14		26.8 : 24.2	<0.01		6.6 : 6.3	<0.01	
	M (RTL)	169 (94%) : 107 (96%)			260 : 123			18.2 : 12.7			5.8 : 5.3		
	MH (OSU)	462 (53%) : 352 (76%)	<0.01		2132 : 1009	<0.05		35.1 : 31.4	<0.01		7.2 : 7.0	<0.01	
	MH (RTL)	212 (93%) : 151 (97%)			326 : 222			21.7 : 17.1			6.2 : 5.8		
Study 2	Lq-C (OSU)	921 (39%) : 746 (55%)	<0.05	<0.001	4860 : 2872	<0.01	<0.001	70.7 : 62.8	<0.001	<0.001	9.3 : 8.9	<0.1	<0.01
	Lq-C (RTL)	607 (75%) : 376 (86%)			1000 : 697			46.8 : 36.5			8.6 : 7.6		
	Lq-R (OSU)	886 (41%) : 678 (55%)	<0.05		4236 : 2651	<0.01		62.4 : 56.4	<0.01		9.2 : 8.5	0.13	
	Lq-R (RTL)	571 (76%) : 367 (92%)			976 : 637			45.9 : 36.3			8.5 : 7.6		
	Lq-RF (OSU)	859 (42%) : 689 (56%)	<0.05		4394 : 2243	<0.01		64.9 : 58.0	<0.001		9.0 : 8.6	<0.1	
	Lq-RF (RTL)	567 (73%) : 368 (87%)			1013 : 529			43.3 : 35.8			8.3 : 7.5		
	Ad-C (OSU)	906 (39%) : 741 (51%)	<0.05		4431 : 3342	<0.01		69.2 : 63.2	<0.001		9.2 : 8.7	0.20	
	Ad-C (RTL)	629 (74%) : 412 (91%)			1213 : 664			46.5 : 38.3			8.7 : 8.0		
	Ad-R (OSU)	897 (39%) : 740 (49%)	0.10		4417 : 2969	<0.05		64.2 : 62.6	<0.01		9.3 : 8.7	0.46	
	Ad-R (RTL)	621 (76%) : 556 (74%)			967 : 1419			43.9 : 47.0			8.8 : 8.4		
	Ad-RF (OSU)	893 (39%) : 724 (50%)	<0.05		4852 : 3258	<0.01		61.4 : 56.9	<0.001		9.2 : 8.6	<0.1	
	Ad-RF (RTL)	550 (77%) : 361 (87%)			842 : 518			38.6 : 31.9			8.4 : 7.7		
Study 3	Lq-C (OSU)	484 (53%) : 379 (68%)	<0.05	<0.001	1741 : 1043	<0.1	<0.001	42.8 : 36.6	<0.01	<0.001	7.1 : 6.8	<0.05	<0.01
	Lq-C (RTL)	234 (85%) : 168 (95%)			423 : 241			23.3 : 18.2			6.1 : 5.6		
	Lq-H (OSU)	727 (38%) : 575 (51%)	<0.01		4465 : 2503	<0.001		63.3 : 54.2	<0.001		8.0 : 7.6	<0.05	
	Lq-H (RTL)	377 (79%) : 254 (92%)			685 : 444			36.0 : 28.6			7.1 : 6.3		
	Ad-C (OSU)	435 (52%) : 371 (61%)	<0.05		1817 : 1296	<0.1		37.9 : 36.8	<0.05		6.5 : 6.3	0.23	
	Ad-C (RTL)	267 (84%) : 189 (91%)			491 : 254			26.8 : 21.8			6.2 : 5.8		
	Ad-H (OSU)	659 (45%) : 573 (51%)	<0.01		2472 : 1951	<0.05		53.0 : 49.6	<0.01		8.1 : 7.7	0.19	
	Ad-H (RTL)	430 (80%) : 298 (92%)			798 : 479			37.3 : 27.9			7.7 : 7.1		

[a] Coverage was calculated by dividing the number of observed OTUs by the asymptote predicted from the rarefaction curve using the nonlinear model procedure of SAS as described previously (Larue *et al.*, 2005)
[b] Contrast analysis for the two sequencing facilities for each diet group.
[c] Contrast analysis for the two sequencing facilities for all the diet groups.

Results

Data summary

A total of 80,640 (1,440 sequences/sample × 56 samples) non-chimeric quality-checked sequences were obtained from the three sets of samples and subjected to OTU clustering and taxonomic classification. Collectively, the 80,640 sequences were assigned to 16 phyla: *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria*, TM7, *Cyanobacteria*, *Spirochaetes*, *Synergistetes*, *Planctomycetes*, SR1, *Chloroflexi*, *Fibrobacteres*, *Lentisphaerae*, *Tenericutes*, *Verrucomicrobia*, and *Fusobacteria*. Of the 16 phyla, *Firmicutes* and *Bacteroidetes* were the first and the second most predominant phyla and accounted for 57% and 34% of the total sequences, respectively. This result supports the previous study on the global bacterial diversity reported for the rumen (Kim *et al.*, 2011b). *Actinobacteria* and *Proteobacteria* were the third and the fourth most predominant phyla and accounted for 1.5% and 1.3% of the total sequences, respectively. These 4 phyla were found in all the 56 samples. However, the remaining 12 phyla accounted for less than 1% of the total sequences, and some of these minor phyla were found only in some of the 56 samples. The total sequences were clustered to a total of 11,831 OTUs across all the 56 samples. Only the OTUs and taxa that were each represented by ≥ 1% of total sequences in at least one sample of each sample set were used to examine intra- and inter-facility variations. For all the three sample sets, the percent coverage was greater in the RTL datasets than in the OSU datasets (Table 1).

Comparison between the sequencing facilities – inter-facility variations

For the set of 16 "samples" (4 composite rumen samples × 2 runs × 2 sequencing facilities) collected in study 1 (Felix and Loerch, 2011), taxa represented by ≥1% of total sequences in at least one sample were assigned to 4 phyla, 5 classes, 5 orders, 7 families, and 8 genera (Table 2). At phylum level, the relative abundance of *Firmicutes* and *Actinobacteria* in two diet groups (H and MH for *Firmicutes*, and C and M for *Actinobacteria*) was significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets, whereas the opposite held true for the relative abundance of *Proteobacteria* in all the four diet groups ($P<0.001$). Collectively, the contrast analysis showed that the relative abundance of

Table 2. **The relative abundance (% of total sequences) of taxa in the samples from study 1 (Felix and Loerch, 2011)**

Rank	Taxon	Facility	Diet (1st run : 2nd run)				Contrast				
			C[a]	H[a]	M[a]	MH[a]	C[b]	H[b]	M[b]	MH[b]	Facility[c]
Phylum	*Firmicutes*	OSU	53.5 : 53.9	60.0 : 62.2	51.0 : 51.2	54.4 : 54.0	0.20	<0.05	0.24	<0.01	<0.05
		RTL	56.9 : 55.6	53.3 : 58.3	51.3 : 46.7	47.2 : 47.4					
	Bacteroidetes	OSU	41.2 : 39.8	35.4 : 32.7	43.8 : 42.1	40.1 : 38.3	<0.1	0.57	0.45	0.57	0.20
		RTL	35.1 : 35.7	36.0 : 29.7	37.6 : 43.8	41.4 : 40.3					
	Proteobacteria	OSU	0.7 : 0.8	0.8 : 0.6	0.8 : 0.9	0.9 : 0.8	< 0.001	<0.001	< 0.001	< 0.001	<0.001
		RTL	5.2 : 5.6	6.9 : 9.5	7.5 : 7.0	8.8 : 9.7					
	Actinobacteria	OSU	2.7 : 3.5	1.7 : 1.9	3.0 : 4.4	2.1 : 4.0	<0.05	<0.1	<0.05	0.12	<0.01
		RTL	1.1 : 1.6	1.4 : 0.6	1.8 : 1.4	1.5 : 1.8					
Class	*Clostridia*	OSU	51.8 : 51.6	58.5 : 60.4	49.4 : 49.9	52.0 : 52.1	0.74	<0.05	0.13	<0.01	<0.01
		RTL	53.1 : 51.6	50.4 : 55.5	49.1 : 44.3	44.8 : 45.4					
	Erysipelotrichi	OSU	1.6 : 2.0	1.3 : 1.4	1.5 : 1.1	1.9 : 1.7	<0.01	<0.01	<0.05	0.39	<0.001
		RTL	3.5 : 3.3	2.5 : 2.6	1.9 : 2.2	2.6 : 1.6					
	Bacteroidia	OSU	39.8 : 39.1	34.6 : 31.32	43.0 : 40.8	39.31 : 37.2	<0.1	0.47	0.57	0.59	0.21
		RTL	34.2 : 34.0	34.6 : 27.9	36.6 : 43.8	39.9 : 39.9					
	γ-Proteobacteria	OSU	0.6 : 0.5	0.6 : 0.4	0.7 : 0.6	0.9 : 0.8	<0.001	<0.001	<0.001	<0.001	<0.001
		RTL	4.9 : 5.5	6.9 : 8.9	7.3 : 6.9	9.6 : 8.5					
	Actinobacteria	OSU	2.7 : 3.5	1.7 : 1.9	3.0 : 4.4	2.1 : 4.0	<0.05	<0.1	<0.05	0.12	<0.01
		RTL	1.1 : 1.6	1.4 : 0.6	1.8 : 1.4	1.8 : 1.5					
Order	*Clostridiales*	OSU	51.8 : 51.6	58.3 : 60.3	49.4 : 49.9	51.9 : 52.1	0.76	<0.05	0.13	<0.01	<0.01
		RTL	53.1 : 51.5	50.3 : 55.4	49.1 : 44.3	45.4 : 44.8					
	Erysipelotrichales	OSU	1.6 : 2.0	1.3 : 1.4	1.5 : 1.1	1.9 : 1.7	<0.01	<0.01	<0.05	0.39	<0.001
		RTL	3.5 : 3.3	2.5 : 2.6	1.9 : 2.2	1.6 : 2.6					
	Bacteroidales	OSU	39.8 : 39.1	34.6 : 31.3	43.0 : 40.8	39.3 : 37.2	<0.1	0.47	0.58	0.59	0.21
		RTL	34.2 : 34.0	34.6 : 27.9	36.6 : 44.0	40.0 : 40.0					
	Aeromonadales	OSU	0.4 : 0.4	0.5 : 0.2	0.5 : 0.5	0.6 : 0.6	<0.001	<0.001	<0.001	<0.001	<0.001
		RTL	4.9 : 5.4	6.9 : 7.9	7.3 : 6.3	8.5 : 8.8					
	Coriobacteriales	OSU	2.6 : 3.5	1.7 : 1.9	2.9 : 4.4	2.0 : 3.9	<0.1	<0.1	<0.1	0.20	<0.01
		RTL	1.1 : 1.6	**1.4 : 0.4**[d]	1.8 : 1.4	1.5 : 1.8					
Family	*Lachnospiraceae*	OSU	31.9 : 29.8	37.5 : 38.6	23.7 ; 25.3	30.5 : 32.4	<0.01	<0.01	<0.01	<0.001	<0.001
		RTL	23.1 : 23.4	27.9 : 23.3	19.9 : 15.9	16.9 : 17.9					
	Ruminococcaceae	OSU	6.3 : 5.5	6.0 : 6.2	2.5 : 3.1	4.5 : 5.4	<0.001	<0.01	<0.01	<0.01	<0.001
		RTL	11.9 : 11.5	8.5 : 11.3	5.5 : 4.0	9.5 : 10.0					
	Veillonellaceae	OSU	10.4 : 13.1	10.2 : 12.4	20.0 : 18.6	13.6 : 12.3	<0.05	<0.1	0.28	0.11	<0.01
		RTL	16.8 : 15.6	11.8 : 18.2	22.0 : 23.3	17.4 : 16.0					
	Erysipelotrichaceae	OSU	1.6 : 2.0	1.3 : 1.4	1.5 : 1.1	1.9 : 1.7	<0.01	<0.01	<0.05	0.39	<0.001
		RTL	3.5 : 3.3	2.5 : 2.6	1.9 : 2.2	1.6 : 2.6					
	Prevotellaceae	OSU	39.0 : 38.3	32.9 : 29.2	41.8 : 39.9	38.3 : 35.9	0.11	0.47	0.58	0.59	0.23
		RTL	33.3 : 33.2	32.9 : 25.7	35.8 : 42.4	38.8 : 38.9					
	Succinivibrionaceae	OSU	0.4 : 0.4	0.5 : 0.2	0.5 : 0.5	0.6 : 0.6	<0.001	<0.001	<0.001	<0.001	<0.001
		RTL	4.9 : 5.4	6.9 : 7.9	7.3 : 6.3	8.5 : 8.8					
	Coriobacteriaceae	OSU	2.6 : 3.5	1.7 : 1.9	2.9 : 4.4	2.0 : 3.9	<0.1	<0.1	<0.1	0.20	<0.01
		RTL	1.1 : 1.6	**1.4 : 0.4**	1.8 : 1.4	1.5 : 1.8					
Genus	*Dialister*	OSU	5.4 : 7.3	3.5 : 5.2	10.1 : 9.9	3.9 : 4.0	0.14	0.21	0.95	0.13	<0.05
		RTL	8.5 : 7.6	4.5 : 6.2	9.8 : 10.0	4.8 : 5.4					
	Syntrophococcus	OSU	1.5 : 2.0	0.8 ; 1.1	1.7 : 2.0	0.9 : 0.6	<0.01	<0.001	<0.01	<0.01	<0.001
		RTL	4.7 : 4.4	3.1 : 2.4	3.7 : 3.6	1.5 : 2.2					
	Butyrivibrio	OSU	1.7 : 1.0	3.1 : 2.9	0.9 : 1.2	1.9 : 2.6	<0.1	0.21	<0.05	<0.1	<0.01
		RTL	**0.0 : 0.5**	**0.1 : 2.3**	0.0 : 0.0	**0.0 : 0.7**					
	Mitsuokella	OSU	1.0 : 0.7	1.6 : 1.6	2.0 : 1.9	2.6 : 2.4	0.77	0.54	<0.1	<0.05	<0.05
		RTL	0.6 : 0.6	**0.4 : 1.9**	**0.8 : 0.1**	**0.6 : 0.0**					
	Succiniclasticum	OSU	0.8 : 1.0	0.6 : 1.1	**0.4 : 0.1**	0.5 : 0.5	<0.05	<0.05	<0.05	<0.01	<0.001
		RTL	2.2 : 2.3	1.6 : 2.3	**0.4 : 0.8**	1.9 : 2.0					
	Prevotella	OSU	26.0 : 26.32	22.8 : 20.5	31.3 : 29.7	21.5 : 19.9	<0.1	0.34	0.43	0.37	0.20
		RTL	23.4 : 18.9	22.3 : 16.2	26.1 : 29.4	24.8 : 21.6					
	Anaerobiospirillum	OSU	0.4 : 0.4	0.2 : 0.1	0.4 : 0.5	0.6 : 0.5	<0.001	<0.001	<0.001	<0.001	<0.001
		RTL	4.1 : 4.3	3.3 : 4.4	6.3 : 6.3	7.0 : 8.3					
	Olsenella	OSU	1.7 : 1.0	1.3 : 1.7	**1.8 : 3.9**	**1.0 : 2.8**	0.25	0.14	0.29	0.59	<0.1
		RTL	3.1 : 1.4	**1.4 : 0.3**	1.5 : 1.3	1.3 : 1.3					

[a] C, control diet; H, control diet supplemented with haylage; M, control diet supplemented with monensin; and MH, control diet supplemented with both monensin and haylage.
[b] Contrast analysis for the two sequencing facilities for each diet group. Log-transformed relative abundance values were used in the contrast analysis.
[c] Contrast analysis for the two sequencing facilities for all the diet groups. Log-transformed relative abundance values were used in the contrast analysis.
[d] Relative abundance values are bolded and underlined if they differ by ≥3-fold between the two sequencing runs at each sequencing facility. This is also applied to Tables 3 and 4.

Table 3. **The relative abundance (% of total sequences) of taxa in the samples from study 2 (Mathew *et al.*, 2011)**

Rank	Taxon	Facility	Fraction[a] (1st run : 2nd run)						Contrast[b]						
			Lq-C[a]	Lq-R[a]	Lq-RF[a]	Ad-C[a]	Ad-R[a]	Ad-RF[a]	Lq-C[b]	Lq-R[b]	Lq-RF[b]	Ad-C[b]	Ad-R[b]	Ad-RF[b]	Facility[c]
Phylum	*Firmicutes*	OSU	47.4 : 47.4	43.1 : 43.2	56.2 : 57.4	63.9 : 61.7	63.1 : 62.0	80.3 : 75.1	<0.01	<0.001	0.50	<0.01	0.70	0.12	0.16
		RTL	43.7 : 37.6	35.0 : 33.8	58.5 : 58.3	69.4 : 73.3	62.0 : 65.1	81.0 : 85.0							
	Bacteroidetes	OSU	43.4 : 41.4	51.3 : 50.2	36.7 : 35.2	20.6 : 20.3	27.2 : 27.9	10.7 : 13.3	<0.05	<0.1	0.86	0.15	0.39	<0.01	0.30
		RTL	47.2 : 54.4	56.6 : 60.1	35.6 : 35.3	18.9 : 17.4	28.2 : 23.5	9.4 : 8.0							
	TM7	OSU	2.1 : 2.8	1.1 : 2.2	1.2 : 0.8	5.8 : 8.9	2.4 : 2.9	1.1 : 2.2	0.28	0.11	<0.01	<0.05	0.12	0.12	<0.001
		RTL	1.6 : 1.7	1.2 : 0.6	0.3 : 0.4	3.8 : 2.4	1.5 : 1.5	1.4 : 0.6							
	Actinobacteria	OSU	0.9 : 0.6	0.2 : 0.3	0.9 : 1.0	2.6 : 1.7	1.0 : 0.4	2.7 : 3.2	0.55	0.73	0.77	<0.001	<0.01	<0.01	<0.001
		RTL	0.8 : 1.0	0.2 : 0.4	0.7 : 1.0	0.4 : 0.3	0.1 : 0.2	0.9 : 0.8							
Class	*Clostridia*	OSU	46.7 : 46.2	41.8 : 42.0	54.5 : 55.1	61.0 : 59.4	61.7 : 59.0	76.9 : 71.5	<0.001	<0.001	0.50	<0.05	0.98	0.16	<0.05
		RTL	41.1 : 36.6	34.0 : 32.6	56.2 : 56.3	65.1 : 69.0	58.3 : 62.3	76.7 : 80.1							
	Bacteroidia	OSU	38.1 : 35.9	1.1 : 2.2	1.2 : 0.8	17.9 : 17.4	24.9 : 24.9	9.8 : 12.4	0.28	0.11	<0.01	0.55	0.63	<0.01	0.98
		RTL	42.8 : 50.0	1.2 : 0.6	0.3 : 0.4	17.4 : 16.3	26.5 : 21.5	8.8 : 7.4							
	Actinobacteria	OSU	0.9 : 0.6	0.2 : 0.3	0.9 : 1.0	2.6 : 1.6	1.0 : 0.4	2.7 : 3.2	0.55	0.73	0.77	<0.001	<0.01	<0.01	<0.001
		RTL	0.8 : 1.0	0.2 : 0.4	0.7 : 1.0	0.4 : 0.3	0.1 : 0.2	0.9 : 0.8							
Order	*Clostridiales*	OSU	46.0 : 45.6	41.3 : 41.8	54.0 : 54.5	59.6 : 58.8	59.5 : 57.7	75.0 : 68.8	<0.001	<0.001	0.44	<0.05	0.95	0.12	<0.05
		RTL	40.8 : 36.4	33.4 : 32.1	55.8 : 56.0	63.4 : 67.2	56.5 : 60.5	74.7 : 78.2							
	Bacteroidales	OSU	38.1 : 35.9	47.1 : 47.2	33.3 : 33.0	17.9 : 17.4	24.9 : 24.9	9.8 12.4	<0.05	<0.1	0.96	0.55	0.63	<0.01	0.98
		RTL	42.8 : 50.2	54.3 : 58.4	33.3 : 33.2	17.4 : 16.3	26.5 : 21.5	8.8 : 7.4							
Family	*Lachnospiraceae*	OSU	25.2 : 26.1	24.9 : 24.8	24.0 : 23.8	32.9 : 29.4	35.1 : 34.4	40.3 : 35.1	<0.001	<0.001	<0.001	<0.1	<0.001	<0.001	<0.001
		RTL	16.7 : 14.7	15.5 : 14.3	18.8 : 16.0	26.8 : 27.8	21.8 : 23.4	29.3 : 26.0							
	Ruminococcaceae	OSU	8.7 : 9.4	6.5 : 8.4	16.5 : 17.6	10.2 : 14.0	9.4 : 10.8	14.0 : 16.9	0.32	0.90	0.19	0.14	<0.05	<0.05	<0.01
		RTL	12.8 : 8.6	8.5 : 6.6	21.1 : 20.8	15.5 : 14.7	13.8 : 15.0	20.9 : 25.8							
	Veillonellaceae	OSU	2.1 : 2.0	3.3 : 2.8	4.8 : 4.9	1.4 : 1.8	1.9 : 1.0	1.4 : 1.3	0.33	0.37	0.26	0.42	<0.05	0.42	<0.05
		RTL	2.2 : 3.8	3.1 : 5.7	5.4 : 10.1	1.4 : 3.2	2.6 : 3.3	1.3 : 2.6							
	Prevotellaceae	OSU	25.7 : 25.9	36.7 : 38.0	25.3 : 26.3	9.7 : 9.4	13.3 : 14.2	4.2 : 7.6	<0.1	0.22	0.48	0.71	0.73	0.15	0.34
		RTL	32.2 : 40.0	44.3 : 48.5	28.0 : 30.2	9.8 : 8.3	17.2 : 12.4	5.2 : 3.7							
	Porphyromonadaceae	OSU	1.8 : 1.7	2.4 : 1.9	1.7 : 1.2	**1.5 : 0.4**	1.9 : 1.3	1.5 : 0.7	0.33	0.45	<0.05	0.41	0.99	0.90	0.19
		RTL	1.1 : 0.9	1.5 : 1.3	0.8 : 0.1	1.7 : 1.0	2.0 : 1.3	0.7 : 1.7							
Genus	*Succiniclasticum*	OSU	1.3 : 1.4	3.0 : 2.5	4.7 : 4.6	1.2 : 1.4	1.7 : 0.8	1.4 : 1.1	0.12	0.30	0.24	0.71	<0.05	0.33	<0.01
		RTL	1.9 : 3.3	3.0 : 5.7	5.4 : 10.1	**0.9 : 7.4**	2.6 : 2.9	1.3 : 2.6							
	Butyrivibrio	OSU	2.4 : 2.0	1.9 : 1.3	1.1 : 0.8	3.4 : 3.0	2.6 : 3.0	1.6 : 1.7	<0.01	<0.001	<0.05	<0.01	<0.05	<0.01	<0.001
		RTL	0.4 : 0.3	0.1 : 0.1	0.6 : 0.1	0.3 : 0.5	0.8 : 0.6	**0.6 : 0.1**							
	Moryella	OSU	**0.1 : 0.7**	1.0 : 1.3	1.2 : 1.7	0.8 : 1.0	1.6 : 2.6	1.9 : 2.2	0.97	0.21	<0.05	0.73	<0.1	0.63	<0.05
		RTL	0.2 : 0.5	0.5 : 0.4	**0.7 : 0.0**	0.5 : 1.0	0.4 : 0.5	1.5 : 1.3							
	Prevotella	OSU	17.3 : 17.8	25.7 : 26.9	17.3 : 18.5	3.9 : 4.3	6.7 : 6.7	1.6 : 3.9	0.14	0.28	0.69	0.75	0.70	0.98	0.22
		RTL	24.2 : 28.8	32.9 : 37.3	19.7 : 20.0	4.7 : 3.1	8.9 : 6.2	3.4 : 1.8							

[a] Lq-C, the liquid fraction in the control diet; Lq-R, the liquid fraction in the control diet supplemented with Rumensin®; Lq-RF, the liquid fraction in the control diet supplemented with both Rumensin® and fat; Ad-C, the adherent fraction in the control diet; Ad-R, the adherent fraction in the control diet supplemented with Rumensin®; and Ad-RF, the adherent fraction in the control diet supplemented with both Rumensin® and fat.
[b] Contrast analysis for the two sequencing facilities for each fraction group. Log-transformed relative abundance values were used in the contrast analysis.
[c] Contrast analysis for the two sequencing facilities for all the fraction groups. Log-transformed relative abundance values were used in the contrast analysis.

these three phyla determined by the two sequencing facilities differed significantly ($P<0.05$) irrespective of diet.

The relative abundance of the classes *Actinobacteria* and *Clostridia*, the order *Clostridiales*, and the family *Lachnospiraceae* in at least two diet groups (C and M for *Actinobacteria*, H and MH for *Clostridia* and *Clostridiales*, and all four diet groups for *Lachnospiraceae*) was significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets (Table 2). The opposite held true for the relative abundance of the classes *Erysipelotrichi* and *Gammaproteobacteria*, the orders *Aeromonadales* and *Erysipelotrichales*, and the families *Erysipelotrichaceae*, *Ruminococcaceae*, *Veillonellaceae*, and *Succinivibrionaceae* in at least one diet group (C, H or M for *Erysipelotrich*, *Erysipelotrichales*, and *Erysipelotrichaceae*; all the four diet groups for *Gammaproteobacteria*, *Aeromonadales*, *Ruminococcaceae*, and *Succinivibrionaceae*; C for *Veillonellaceae*). The relative abundance of the order *Coriobacteriales* and the family *Coriobacteriaceae* in the C, the H, and the M diet groups tended to be greater ($P<0.1$) in the OSU datasets than in the RTL datasets. Based on the contrast analysis, the two sequencing facilities showed significantly different ($P<0.05$) relative abundance for all these classes, orders, and families irrespective of diet.

At genus level, the relative abundance of *Butyrivibrio* and *Mitsuokella* in two diet group (M for *Butyrivibrio*, and MH for *Mitsuokella*) was significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets, whereas that of *Anaerobiospirilum*, *Succiniclasticum*, and *Syntrophococcus* in all the four diet groups was significantly greater ($P<0.05$) in the RTL datasets than in the OSU datasets (Table 2). All these genera had significantly different ($P<0.05$) relative abundance between the OSU datasets and the RTL datasets irrespective of diet. Although the relative abundance of *Dialister* was not consistently different (greater or smaller) between the OSU datasets and the RTL datasets, it was significantly different ($P<0.05$) between the OSU datasets and the RTL datasets irrespective of diet.

For the set of 24 "samples" (6 composite rumen samples × 2 runs × 2 sequencing facilities) collected in study 2 (Mathew *et al.*, 2011), taxa represented each by ≥1% of total sequences in at least one sample were assigned to 4 phyla, 4 classes, 2 orders, 5 families, and 4 genera (Table 3). The relative abundance of *Firmicutes* in two liquid fractions (Lq-C and Lq-R) was significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets, whereas the opposite held true for that of *Firmicutes* in one adherent fraction (Ad-C). On the contrary, the relative abundance of *Bacteroidetes* in one adherent fraction (Ad-RF) was significantly smaller ($P<0.05$) in the OSU datasets than in the RTL datasets, whereas the opposite held true for that of *Bacteroidetes* in one liquid fraction (Lq-C). However, no significant difference ($P>0.1$) in relative abundance of these two phyla was seen across all the samples between the OSU datasets and the RTL datasets irrespective of fraction. The relative abundances of the phyla *Actinobacteria* and TM7 in at least two fraction groups (three Ad fractions for *Actinobacteria*, and Lq-RF and Ad-C for TM7) were significantly greater ($P>0.05$) in the OSU datasets than in the RTL datasets. Based on the contrast analysis, the relative abundance of these two phyla were significantly different ($P<0.05$) between the OSU datasets and the RTL datasets irrespective of fraction.

As shown at phylum level, the class *Clostridia* and the order *Clostridiales* in two liquid fractions (Lq-C and Lq-R) had greater relative abundance in the OSU datasets than in the RTL datasets, whereas the opposite held true for *Clostridia* and *Clostridiales* in one adherent fraction (Ad-C) (Table 3). Similarly, the relative abundance of the class *Bacteroidia* and the order *Bacteroidales* in the one adherent fraction (Ad-RF) was significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets. *Clostridia* and *Clostridiales* were also significantly different ($P<0.05$) in relative abundance between the OSU datasets and the RTL datasets irrespective of fraction. The relative abundance of the families *Lachnospiraceae* and *Porphyromonadaceae* was significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets, except for *Lachnospiraceae* in Ad-C and for *Porphyromonadaceae* in Lq-RF). The opposite held true for the relative abundance of the families *Veillonellaceae* and *Ruminococcaceae* in at least one fraction group (Ad-R for *Veillonellaceae*, and Ad-R and Ad-RF for *Ruminococcaceae*). Except for that of *Porphyromonadaceae*, the contrast analysis indicated that the relative abundances of all these families were found to be significantly different ($P<0.05$) between the OSU datasets and the RTL datasets irrespective of diet or fraction.

At genus level, the relative abundance of *Butyrivibrio* in all the six fraction groups was significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets (Table 3). The relative abundance of *Moryella* in the Lq-RF fraction group was significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets, whereas that of *Succiniclasticum* in the Ad-R fraction group was significantly greater ($P<0.05$) in the RTL datasets than in the OSU datasets. Collectively, the contrast analysis showed that the two sequencing facilities resulted in significantly different ($P<0.05$) relative abundance for all these genera irrespective of diet or fraction.

For the set of 16 "samples" (4 composite rumen samples × 2 runs × 2 sequencing facilities) collected in study 3 (Stiverson *et al.*, 2011), taxa represented by ≥1% of total sequences in at least one sample were assigned to 3 phyla, 3 classes, 3 orders, 5 families, and 4 genera (Table 4). At phylum level, the relative abundance of *Actinobacteria* and *Firmicutes* in one fraction group (Lq-C for *Actinobacteria*; and Ad-C for *Firmicutes*) was significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets, whereas the opposite held true for the relative abundance of *Bacteroidetes* in two Ad fraction groups ($P<0.05$). *Bacteroidetes* had significantly different ($P<0.05$) relative abundance between the OSU dataset and the RTL datasets irrespective of diet or fraction, while *Firmicutes* tended to have different ($P<0.1$) relative abundance between the datasets generated at the two sequencing facilities irrespective of diet or fraction.

The relative abundance of the classes *Actinobacteria* and *Clostridia*, the orders *Clostridiales* and *Coriobacteriales*, and the families *Coriobacteriaceae* and *Lachnospiraceae* in at least one fraction group (Lq-C for *Actinobacteria*; Ad-C for *Clostridia* and *Clostridiales*; Lq-C, Lq-H, and Ad-H for *Coriobacteriales* and *Coriobacteriaceae*; and Lq-C, Lq-H, and Ad-C for *Lachnospiraceae*) was significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets (Table 4). However, the OSU datasets had a significantly smaller ($P<0.05$) rela-

Table 4. **The relative abundance (% of total sequences) of taxa in the samples from study 3 (Stiverson *et al.*, 2011)**

Rank	Taxon	Facility	Fraction[a] (1st run : 2nd run)				Contrast[b]				
			Lq-C[a]	Lq-H[a]	Ad-C[a]	Ad-H[a]	Lq-C[b]	Lq-H[b]	Ad-C[b]	Ad-H[b]	Facility[c]
Phylum	*Firmicutes*	OSU	55.3 : 56.3	38.8 : 37.4	75.8 : 73.8	76.0 : 73.2	0.23	0.28	<0.01	0.77	<0.1
		RTL	53.0 : 51.0	44.8 : 36.7	58.5 : 61.8	75.5 : 76.1					
	Bacteroidetes	OSU	37.6 : 37.8	52.1 : 54.3	20.4 : 21.5	15.2 : 15.7	<0.1	0.42	<0.001	<0.05	<0.01
		RTL	41.4 : 43.8	46.0 : 55.2	28.5 : 32.9	18.1 : 18.6					
	Actinobacteria	OSU	2.4 : 1.7	0.4 : 0.6	1.1 : 2.1	1.2 : 2.0	<0.05	0.80	0.18	<0.1	0.11
		RTL	0.6 : 0.3	0.5 : 0.4	**10.1 : 1.3**	0.5 : 0.4					
Class	*Clostridia*	OSU	53.6 : 54.4	37.1 : 36.5	74.9 : 72.2	75.2 : 72.5	0.15	0.21	<0.01	0.93	<0.1
		RTL	50.0 : 49.0	44.0 : 35.8	56.4 : 59.9	74.0 : 74.4					
	Bacteroidia	OSU	34.3 : 34.9	46.7 : 48.1	19.6 : 20.4	12.9 : 13.5	<0.1	0.36	<0.001	<0.01	<0.001
		RTL	39.1 : 41.4	40.6 : 48.8	27.6 : 32.4	16.4 : 17.0					
	Actinobacteria	OSU	2.4 : 1.7	0.4 : 0.6	1.1 : 2.1	1.2 : 2.0	<0.05	0.80	0.18	<0.1	0.11
		RTL	0.6 : 0.3	0.5 : 0.4	**10.1 : 1.3**	0.5 : 0.4					
Order	*Clostridiales*	OSU	53.2 : 54.0	35.6 : 34.4	74.7 : 71.9	74.4 : 71.1	0.18	<0.1	<0.01	0.87	<0.1
		RTL	49.8 : 48.8	43.3 : 35.1	56.1 : 59.4	72.5 : 71.6					
	Bacteroidales	OSU	34.3 : 34.9	46.7 : 48.1	19.6 : 20.4	12.9 : 13.5	<0.1	0.36	<0.001	<0.01	<0.001
		RTL	39.1 : 41.4	40.6 : 48.8	27.6 : 32.4	16.4 : 17.0					
	Coriobacteriales	OSU	2.4 : 1.6	0.4 : 0.5	1.0 : 1.8	1.0 : 1.9	<0.01	<0.001	0.43	<0.01	
		RTL	0.5 : 0.3	0.1 : 0.0	1.0 : 1.1	0.4 : 0.3					
Family	*Lachnospiraceae*	OSU	23.2 : 22.2	12.0 : 10.8	55.4 : 54.4	43.8 : 39.9	<0.01	<0.01	<0.01	<0.1	<0.001
		RTL	12.4 : 11.5	9.2 : 5.6	32.9 : 35.7	31.3 : 32.9					
	Ruminococcaceae	OSU	16.9 : 20.0	9.6 : 9.5	5.3 : 5.0	17.3 : 19.4	<0.001	0.13	<0.001	0.44	<0.001
		RTL	27.7 : 26.9	11.9 : 9.9	10.2 : 9.6	20.4 : 18.7					
	Veillonellaceae	OSU	2.9 : 2.9	3.5 : 3.0	6.0 : 4.9	0.6 : 1.0	0.80	0.11	0.33	<0.05	1.00
		RTL	2.6 : 2.9	2.6 : 2.4	4.2 : 5.1	1.2 : 1.3					
	Prevotellaceae	OSU	32.4 : 32.6	30.8 : 33.8	18.8 : 19.2	10.4 : 10.9	<0.05	0.11	<0.001	<0.01	<0.001
		RTL	37.6 : 40.5	25.9 : 30.9	26.5 : 31.7	13.7 : 14.9					
	Coriobacteriaceae	OSU	2.4 : 1.6	0.4 : 0.5	1.0 : 1.8	1.0 : 1.9	<0.01	<0.001	0.43	<0.01	<0.001
		RTL	0.5 : 0.3	0.1 : 0.0	1.0 : 1.1	0.4 : 0.3					
Genus	*Butyrivibrio*	OSU	0.1 : 0.1	3.3 : 2.3	1.9 : 1.6	13.3 : 11.3	<0.001	<0.001	<0.001	<0.01	<0.001
		RTL	0.0 : 0.0	1.0 : 0.6	0.1 : 0.0	4.2 : 4.4					
	Ruminococcus	OSU	1.5 : 3.8	0.2 : 0.3	0.6 : 1.0	0.6 : 0.4	0.25	0.54	<0.1	0.14	<0.1
		RTL	4.9 : 4.5	**0.1 : 0.4**	2.0 : 2.5	**1.2 : 0.1**					
	Succiniclasticum	OSU	0.1 : 0.1	1.5 : 1.0	4.1 : 2.9	0.3 : 0.8	0.69	0.58	1.00	<0.1	<0.33
		RTL	0.1 : 0.2	1.5 : 1.6	2.9 : 4.1	0.9 : 1.1					
	Prevotella	OSU	23.3 : 22.9	14.1 : 16.4	15.3 : 14.9	6.0 : 6.9	<0.01	<0.39	<0.001	<0.001	<0.001
		RTL	29.2 : 30.2	14.9 : 17.7	22.9 : 25.2	10.1 : 10.1					

[a] Lq-C, the liquid fraction in the diet of a mixture of orchardgrass hay and corn; Lq-H, the liquid fraction in the diet of orchardgrass hay; Ad-C, the adherent fraction in the diet of a mixture of orchardgrass hay and corn; and Ad-H, the adherent fraction in the diet of orchardgrass hay.
[b] Contrast analysis for the two sequencing facilities for each fraction group. Log-transformed relative abundance values were used in the contrast analysis.
[c] Contrast analysis for the two sequencing facilities for all the fraction groups. Log-transformed relative abundance values were used in the contrast analysis.

tive abundance of the class *Bacteroidia*, the order *Bacteroidales*, the families *Prevotellaceae*, *Veillonellaceae*, and *Ruminococcaceae* than the RTL datasets in at least one fraction group (Ad-C and Ad-H for *Bacteroidia* and *Bacteroidales*; Lq-C, Ad-C, and Ad-H for *Prevotellaceae*; Ad-H for *Veillonellaceae*; and Lq-C and Ad-C for *Ruminococcaceae*). Except *Actinobacteria* and *Veillonellaceae*, each of these taxa was shown to have significantly different ($P<0.05$) or tended to be different ($P<0.1$) relative abundance between the two sequencing facilities irrespective of diet or fraction.

At genus level, the relative abundance of *Butyrivibrio* in all the four fraction groups was significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets, whereas that of *Prevotella* in the Lq-C, the Ad-C, and the Ad-H fraction groups was significantly greater ($P<0.05$) in the RTL datasets than in the OSU datasets (Table 4). The contrast analysis showed that the two sequencing facilities showed significantly different ($P<0.001$) relative abundance for these two genera irrespective of diet or fraction.

Comparison between the two sequencing runs at each sequencing facility – intra-facility variations

In the first set of 16 "samples" collected in study 1 (Felix and Loerch, 2011), the relative abundance of the phyla *Firmicutes* and *Bacteroidetes*, each of which was represented by ≥29% of total sequences, were similar between the two sequencing runs at each sequencing facility (Table 2). The relative abundance of phyla *Proteobacteria* and *Actinobacteria* was each represented by <10% of total sequences in all the 16

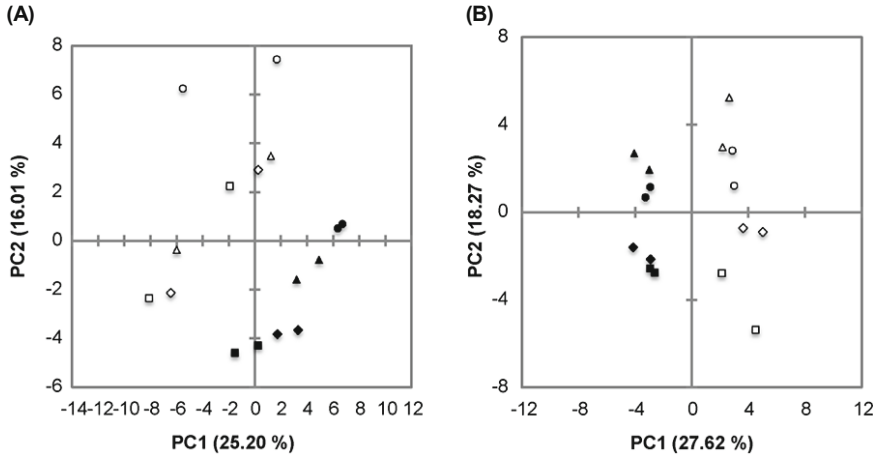

Fig. 1. Principal component analysis of all the taxa (A) or only the predominant taxa observed in at least 50% of the samples (B) collected from study 1 (Felix and Loerch, 2011). Closed symbols, the OSU datasets; open symbols, the RTL datasets. Circles, sample C; triangles, sample M; squares, sample H; diamonds, sample MH.

samples and did not differ by ≥3-fold between the two runs at either sequencing facility. The relative abundance of the order *Coriobacteriales* and the family *Coriobacteriaceae* was each represented by <10% of total sequences in all the 16 samples and differed by ≥3-fold between the two runs at RTL in the H diet group. At genus level, the relative abundance of *Prevotella*, which was represented by ≥20% of total sequences in 10 of the 16 samples, was similar between the two runs at both sequencing facilities. However, the relative abundance of *Butyrivibrio*, *Mitsuokella*, *Succiniclasticum*, and *Olsenella*, each of which was represented by <4% of total sequences in all the 16 samples, differed by ≥3-fold between the two runs at one or both sequencing facilities in at least one diet group.

In the second set of samples collected in study 2 (Mathew *et al.*, 2011), the relative abundance of the phyla *Firmicutes* and *Bacteroidetes* each exceeded 20% of total sequences, and their relative abundances determined by the two sequencing runs at either facility were similar, whereas that of the phyla TM7 and *Actinobacteria*, each of which was represented by <10% of total sequences, differed though not by ≥3-fold between the two runs at both sequencing facilities (Table 3). This also held true for the major classes (*Clostridia* and *Bacteroidia*), major orders (*Clostridiales* and *Bacteroidales*), major families (*Lachnospiraceae* and *Prevotellaceae*). The relative abundance of family *Porphyromonadaceae* was represented by <3% of total sequences in each of the 24 samples and it differed by ≥3-fold between the two runs at OSU in the Ad-C fraction group. At genus level, the relative abundance of *Prevotella*, which was represented by ≥20% of total sequences in 7 of the 24 samples, was similar between the two runs at both sequencing facilities. The relative abundance of minor genera *Succiniclasticum*, *Butyrivibrio*, and *Moryella* differed by ≥3-fold between the two runs at one or both sequencing facilities in at least one fraction group.

In the third set of samples collected in study 3 (Stiverson *et al.*, 2011), the relative abundance of the largest phyla *Firmicutes* and *Bacteroidetes* each exceeded 20% of total sequences in most of the 16 samples (Table 4). The relative abundances of these phyla were similar between the two runs at both sequencing facilities, but that of the third largest phylum *Actinobacteria*, which was represented by <3% of total sequences, differed by ≥3-fold between the two runs at RTL in the Ad-C fraction group. This also held true at class level. At genus level, the relative abundance of the most predominant genus *Prevotella* was represented by 6–29% of total sequences among the samples and it did not differ by ≥3-fold between the two runs at either sequencing facilities. However, the relative abundance of *Ruminococcus*, which was represented by <5% of total sequences in all the 16

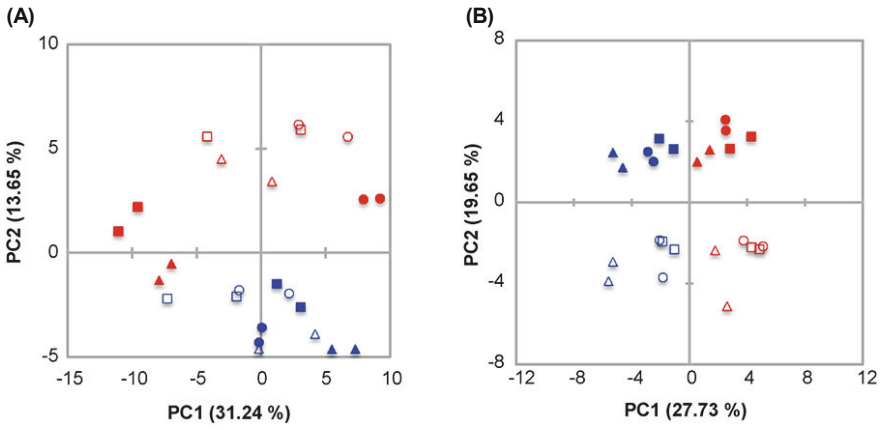

Fig. 2. Principal component analysis of all the taxa (A) or only the predominant taxa observed in at least 50% of the samples (B) collected from study 1 (Mathew *et al.*, 2011). Closed symbols, the OSU datasets; open symbols, the RTL datasets. Red symbols, sample Lq; blue symbols, sample Ad. Circles, sample C; squares, sample R; triangles, sample RF.

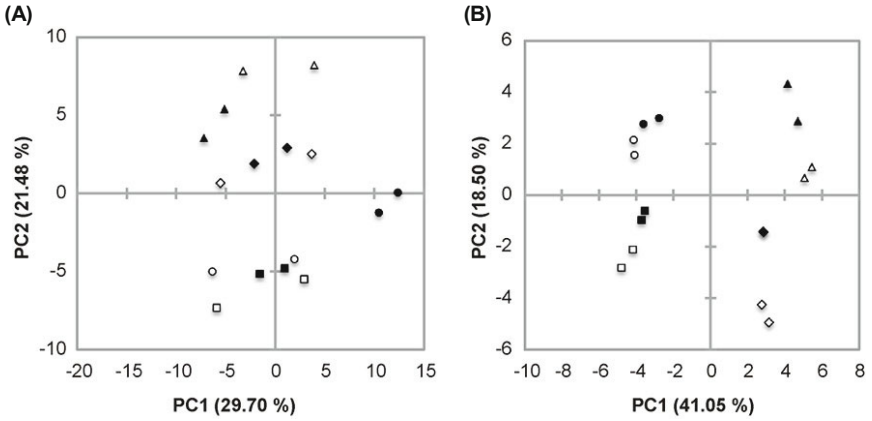

Fig. 3. Principal component analysis of all the taxa (A) or only the predominant taxa observed in at least 50% of the samples (B) collected from study 1 (Stiverson *et al.*, 2011). Closed symbols, the OSU datasets; open symbols, the RTL datasets. Circles, sample Lq-C; triangles, sample Lq-H; squares, sample Ad-C; diamonds, sample Ad-H.

samples, differed by ≥3-fold between the two runs at RTL in the Lq-H and the Ad-H fraction groups.

Comparison of microbiome profiles

The overall intra- and inter-facility variations were assessed using PCA. For the first set of samples collected in study 1, variations of varying degrees were seen between the two sequencing runs, especially the two runs at RTL (Fig. 1). Such variations were generally larger when all the detected taxa were included in PCA than only the predominant taxa detected in ≥50% of the samples were included, indicating that microbiome profiling can be affected to a greater extent by variations in minor taxa than in major taxa. The datasets from the two sequencing facilities were separated along PC1, while the datasets from the two sequencing runs at each sequencing facility and from the different samples tended to be separated along PC2, either when all taxa or only the major taxa were included. Separations between the two sequencing runs and the two sequencing facilities also varied among the samples. The above observations also applied to the two sets of samples collected in studies 2 and 3, but the two sequencing facilities did not produce consistent variations for all the samples and sequencing runs (Figs. 2 and 3).

For all the samples, the OSU datasets generated more observed OTUs, greater Chao1 estimate, PD_whole_tree values, and Shannon's diversity index than the RTL datasets (Table 1). For the first set of 16 samples collected in study 1 (Felix and Loerch, 2011), the numbers of observed OTUs, PD_whole_tree values, and the Shannon's diversity index in all the four diet groups were significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets. These results indicate that microbial diversity was greater in the OSU datasets than in the RTL datasets. All diversity indices were significantly different ($P<0.01$) between the two sequencing facilities irrespective of diet. For the second set of 24 samples collected in study 2 (Mathew *et al.*, 2011), the PD_whole_tree values in all the six fraction groups were also significantly greater ($P<0.01$) in the OSU datasets than in the RTL datasets, while Shannon's diversity index in the Lq-C, the Lq-RF and the Ad-RF fraction groups tended to be greater ($P<0.1$) in the OSU datasets than in the RTL datasets. The contrast analysis showed that all the diversity indices were significantly different ($P<0.01$) between the two sequencing facilities irrespective of diet or fraction of this set of samples. For the third set of 16 samples collected in study 3 (Stiverson *et al.*, 2011), the PD_whole_tree values in all the 4 fraction groups were also significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets, while Shannon's diversity index in the Lq-C and the Lq-H fraction groups was significantly greater ($P<0.05$) in the OSU datasets than in the RTL datasets. Based on the contrast analysis, all diversity indices were significantly different ($P<0.01$) between the two sequencing facilities irrespective of diet or fraction of this sample set. These results again indicate that the OSU datasets captured a greater diversity than the RTL datasets.

Discussion

By cost-effectively sequencing thousands of amplicons of partial 16S rRNA genes, pyrosequencing empowered by the 454 GS FLX Titanium system (Roche) can potentially detail the diversity and composition of complex microbiomes, including minor OTUs that are represented by small numbers of sequence reads. Although it produces significant amounts of artifactual sequences (Gomez-Alvarez *et al.*, 2009; Kunin *et al.*, 2010), bioinformatic tools are being improved to identify and remove such artifactual sequences (Quince *et al.*, 2009; Reeder and Knight, 2010). Powerful software packages or pipelines (e.g., mothur and QIIME) and sequence alignment programs continue to improve in reliability and accuracy. However, variations among different sequencing runs at the same sequencing facility and among different sequencing facilities cannot be readily addressed by any bioinformatic tool.

In the present study, we assessed the variations between two independent pyrosequencing runs and between two sequencing facilities using three sets of rumen samples collected from cattle. We chose to use rumen samples because they have relatively similar diversity and composition so that most of the major and minor taxa (or OTUs) are shared across all the samples and can be included in comparison. The results showed that only the most predominant taxa that are represented by ≥20% of total sequences in at least one sample did not significantly differ in relative abundance between the two runs at both sequencing facilities. However, many

minor taxa differed in relative abundance by ≥3-fold between the two pyrosequencing runs at both sequencing facilities. As demonstrated in the present study, relatively small variations in relative abundance of major taxa can cause several fold variations in relative abundance of minor taxa. Therefore, differences in relative abundance of minor taxa among samples may result from unreliable or unpredictable quantitative repeatability rather than actual differences among samples.

In some studies, taxa or OTUs that were represented by a small number of sequences (e.g., 10 to 50) were excluded from analysis based on assumption that these sequences could be artifactual sequences (Stanley *et al.*, 2012). Benson *et al.* (2010) suggested that quantitative repeatability was not reliable for taxa represented by <30 sequence reads. In our study, however, both intra- and inter-facility variations were found for many taxa that were represented by more than 30 sequences even though each of our samples had a relatively small number of sequences. The reasons for inter-run and inter-facility variations are not known. However, it should also be noted that although PCR can introduce amplification bias, which can potentially affect the results of pyrosequencing analysis, the inter-run variations seen in the present study probably stemmed from other factors because the two sequencing runs done at the OSU facility used the same amplicon libraries that were properly stored at -80°C. Choice of primers used to generate amplicons of partial 16S rRNA genes has been shown to affect results of sequencing-based analysis of microbiomes (Morales and Holben, 2009; Soergel *et al.*, 2012). The primers used to produce the OSU datasets were different from those used to produce the RTL datasets and also had a greater degeneracy, especially the primer 454BactF (modified from primer 27F). Therefore, the differential diversity and composition reveled between the two sequencing facilities might largely result from differences in the primers used. The differences in the primers might also explain the higher greater Shannon's diversity index and the lower percent coverage seen in the OSU datasets than in the RTL datasets. The higher relative abundance of *Lachnospiraceae* in the OSU datasets than in the RTL datasets across all the three sets of samples might also be attributed to the differences in the primers. However, in a post-hoc analysis using the Probe Match function at RDP, we found that the primers used to create the OSU datasets and the primer used to generate the RTL datasets can match similar numbers of the *Lachnospiraceae* sequences archived in RDP (21,478 vs. 22,606). Therefore, aspects of primers other than degeneracy, such as annealing locations and sequences, might also contribute to the observed differences in relative abundance of *Lachnospiraceae* between the two sequencing facilities. The same explanation may also apply to the greater relative abundance of the phylum *Proteobacteria* in the OSU datasets than in the RTL datasets among the samples of study 1. Therefore, choice of primers and their degeneracy need to be taken into consideration when interpreting differential relative abundances observed between studies even if the sequencing and data analysis were done using the same methodology. Primers that have maximal inclusiveness should be used in future studies to minimize biased results.

The power of NGS technologies hinges upon their ability to detect minor OTUs or taxa present in microbiomes. Consistent with a previous study (Kauserud *et al.*, 2012), OTU recovery was consistent between sequencing runs on the same machine. However, quantitative interpretation of relative abundances of OTUs or taxa from prevalence of sequencing reads, which has been practiced commonly in most of the published studies, may not be reliable as previously thought, especially for minor OTUs or taxa that are represented by a relatively low percentage of total sequences. Because different numbers of sequences are always produced from different samples, a specific percentage rather than a specific number of sequences may be a better cutoff value for their exclusion from analysis. In addition, although rarely done, multiple sequence datasets generated from different runs may be used to minimize biased results. Furthermore, a standardized set of parameters and procedures, such as primers, PCR conditions, sequence quality check and exclusion, may improve consistency of analysis results and facilitate comparisons among studies.

The present study used only rumen microbiome samples to examine variations between pyrosequencing runs and between different sequencing facilities. Inter-run variations likely occur when other types of microbiomes are analyzed, but actual assessment is needed to assess the magnitude of variations. Therefore, caution needs to be exercised when comparing the results from one's study to those of previous studies.

Acknowledgements

The work was partially supported by an OARDC ward (2010-007).

References

Barriuso, J., Valverde, J.R., and Mellado, R.P. 2011. Estimation of bacterial diversity using next generation sequencing of 16S rDNA: a comparison of different workflows. *BMC Bioinformatics* **12**, 473.

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K., and *et al.* 2010. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc. Natl. Acad. Sci. USA* **107**, 18933–18938.

Callaway, T.R., Dowd, S.E., Edrington, T.S., Anderson, R.C., Krueger, N., Bauer, N., Kononoff, P.J., and Nisbet, D.J. 2010. Evaluation of bacterial diversity in the rumen and feces of cattle fed different levels of dried distillers grains plus solubles using bacterial tag-encoded FLX amplicon pyrosequencing. *J. Anim. Sci.* **88**, 3977–3983.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., and *et al.* 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072.

Dowd, S.E., Callaway, T.R., Wolcott, R.D., Sun, Y., McKeehan, T., Hagevoort, R.G., and Edrington, T.S. 2008. Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial

tag-encoded FLX amplicon pyrosequencing (bTEFAP). *BMC Microbiol.* **8**, 125.

Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461.

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.

Felix, T.L. and Loerch, S.C. 2011. Effects of haylage and monensin supplementation on performance, carcass characteristics, and ruminal metabolism of feedlot cattle fed diets containing 60% dried distillers grains. *J. Anim. Sci.* **89**, 2614–2623.

Frank, J.A., Reich, C.I., Sharma, S., Weisbaum, J.S., Wilson, B.A., and Olsen, G.J. 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl. Environ. Microbiol.* **74**, 2461–2470.

Gihring, T.M., Green, S.J., and Schadt, C.W. 2012. Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environ. Microbiol.* **14**, 285–290.

Gomez-Alvarez, V., Teal, T.K., and Schmidt, T.M. 2009. Systematic artifacts in metagenomes from complex microbial communities. *ISME J.* **3**, 1314–1317.

Kauserud, H., Kumar, S., Brysting, A.K., Norden, J., and Carlsen, T. 2012. High consistency between replicate 454 pyrosequencing analyses of ectomycorrhizal plant root samples. *Mycorrhiza* **22**, 309–315.

Kim, M., Morrison, M., and Yu, Z. 2011a. Phylogenetic diversity of bacterial communities in bovine rumen as affected by diets and microenvironments. *Folia Microbiol (Praha)* **56**, 453–458.

Kim, M., Morrison, M., and Yu, Z. 2011b. Status of the phylogenetic diversity census of ruminal microbiomes. *FEMS Microbiol. Ecol.* **76**, 49–63.

Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* **12**, 118–123.

Lane, D.J. 1991. 16S/23S rRNA sequencing, pp. 115–175. *In* Stackebrandt, E. and Goodfellow, M.D. (ed.), Nucleic acid techniques in bacterial systematics. John Wiley and Sons, New York, N.Y., USA.

Larue, R., Yu, Z., Parisi, V.A., Egan, A.R., and Morrison, M. 2005. Novel microbial diversity adherent to plant biomass in the herbivore gastrointestinal tract, as revealed by ribosomal intergenic spacer analysis and *rrs* gene sequencing. *Environ. Microbiol.* **7**, 530–543.

Mathew, B., Eastridge, M.L., Oelker, E.R., Firkins, J.L., and Karnati, S.K. 2011. Interactions of monensin with dietary fat and carbohydrate components on ruminal fermentation and production responses by dairy cows. *J. Dairy Sci.* **94**, 396–409.

Morales, S.E. and Holben, W.E. 2009. Empirical testing of 16S rRNA gene PCR primer pairs reveals variance in target specificity and efficacy not suggested by in silico analysis. *Appl. Environ. Microbiol.* **75**, 2677–2683.

Nelson, M.C. 2011. An integrated investigation of the microbial communities underpinning biogas production in anaerobic digestion systems. Ph. D. Dissertation, The Ohio State University, Columbus, Ohio, USA.

Pace, N.R. 1997. A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740.

Petrosino, J.F., Highlander, S., Luna, R.A., Gibbs, R.A., and Versalovic, J. 2009. Metagenomic pyrosequencing and microbial identification. *Clin. Chem.* **55**, 856–866.

Pitta, D.W., Pinchak, E., Dowd, S.E., Osterstock, J., Gontcharova, V., Youn, E., Dorton, K., Yoon, I., Min, B.R., Fulford, J.D., and *et al.* 2010. Rumen bacterial diversity dynamics associated with changing from bermudagrass hay to grazed winter wheat diets. *Microb. Ecol.* **59**, 511–522.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glockner, F.O. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196.

Quince, C., Lanzen, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M., Read, L.F., and Sloan, W.T. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods* **6**, 639–641.

Rajendhran, J. and Gunasekaran, P. 2011. Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. *Microbiol. Res.* **166**, 99–110.

Reeder, J. and Knight, R. 2010. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat. Methods* **7**, 668–669.

Soergel, D.A., Dey, N., Knight, R., and Brenner, S.E. 2012. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J.* **6**, 1440–1444.

Stanley, D., Denman, S.E., Hughes, R.J., Geier, M.S., Crowley, T.M., Chen, H., Haring, V.R., and Moore, R.J. 2012. Intestinal microbiota associated with differential feed conversion efficiency in chickens. *Appl. Microbiol. Biotechnol.* **96**, 1361–1369.

Stiverson, J., Morrison, M., and Yu, Z. 2011. Populations of select cultured and uncultured bacteria in the rumen of sheep and the effect of diets and ruminal fractions. *Int. J. Microbiol.* **2011**, 750613.

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.

Weinstock, G.M. 2012. Genomic approaches to studying the human microbiota. *Nature* **489**, 250–256.

Yu, Z. and Morrison, M. 2004. Improved extraction of PCR-quality community DNA from digesta and fecal samples. *Biotechniques* **36**, 808–812.