HHS Public Access

Author manuscript

J Microbiol Methods. Author manuscript; available in PMC 2021 September 01.

Published in final edited form as:

J Microbiol Methods. 2020 September ; 176: 106033. doi:10.1016/j.mimet.2020.106033.

Influence of PCR cycle number on 16S rRNA gene amplicon sequencing of low biomass samples

Monica C. Witzke[a], **Alexis Gullic**[b], **Peggy Yang**[c], **Nathan J. Bivens**[d], **Pamela R. F. Adkins**[a], **Aaron C. Ericsson**[e,*]

[a]Department of Veterinary Medicine and Surgery, College of Veterinary Medicine, University of Missouri, Columbia, MO, USA

[b]University of Missouri, College of Veterinary Medicine, University of Missouri, Columbia, MO, USA

[c]University of Michigan, Unit for Laboratory Animal Medicine, Ann Arbor, MI, USA

[d]DNA Core Facility, University of Missouri, Columbia, MO, USA

[e]University of Missouri Metagenomics Center (MUMC), Department of Veterinary Pathobiology, College of Veterinary Medicine, University of Missouri, Columbia, MO, USA

Abstract

The objective of this study was to evaluate the effects of increased PCR cycle number on sequencing results from samples with low microbial biomass, including bovine milk, and murine pelage and blood. We hypothesized that subjecting DNA from such samples to higher PCR cycle numbers would increase 16S rRNA sequencing coverage. DNA was extracted from matched samples of each type and multiple PCR cycle numbers were evaluated to generate a total of 96 libraries from 24 milk samples, 46 libraries from 23 pelage samples, and 170 libraries from 85 blood samples. 16S rRNA sequencing was performed on the Illumina MiSeq platform, and the coverage per sample, detected richness, and beta-diversity were evaluated. Across all sample types, higher PCR cycle numbers were associated with increased coverage. Surprisingly however, while higher PCR cycle numbers resulted in greater number of useable datapoints, no differences were detected in metrics of richness or beta-diversity. While reagent controls amplified for 40 cycles yielded similarly increased coverage, control and experimental samples were clearly

*Corresponding author at: 4011 Discovery Drive, S123B, Columbia, MO, 65201, ericssona@missouri.edu (Aaron C. Ericsson).

Author Statement

Monica Witzke: Conceptualization, Investigation, Data curation, Writing – original draft; Alexis Gullic: Investigation, Data curation; Peggy Yang: Investigation, Data curation; Nathan Bivens: Investigation, Resources; Pamela Adkins: Investigation, Resources, Supervision, Writing – review and editing; Aaron Ericsson: Conceptualization, Data curation, Formal analysis, Resources, Visualization, Writing – review and editing, Project administration

DECLARATION OF COMPETING INTEREST

None.

DATA AVAILABILITY

All 16S rRNA sequencing data are available at the NCBI Sequence Read Archive under BioProject accession PRJNA628962.

differentiated based on beta-diversity. The results from this study support the use of higher PCR cycle numbers to evaluate samples with low microbial biomass.

Keywords

Illumina Sequencing; Dairy Cattle; Mouse; Milk; Microbiome; Low Biomass; Pelage; Blood

1. INTRODUCTION

Next-generation sequencing of targeted amplicon libraries (e.g., 16S rRNA or internal transcribed spacer [ITS] libraries) allows researchers to characterize the taxonomic composition of complex bacterial or fungal communities. As these libraries rely on PCR to amplify a variable region of the microbial genome, they can be performed with relatively small amounts of total bacterial biomass. When extracting DNA from feces, for example, the vast majority of the sample weight is bacterial biomass and one-quarter of a mouse fecal pellet provides more than enough DNA. There is however an emerging interest in the bacterial communities present in other biological samples, such as blood and milk (Young et al., 2015; Metzger et al., 2018b). Such samples are often difficult to sequence using next-generation sequencing technologies due to low microbial biomass and excessive contamination with host cells. While there are commercially available kits designed to separate eukaryotic and prokaryotic DNA, they are not practical in situations where there is also relatively low bacterial biomass, because of the large quantity of starting material required to achieve enough bacterial biomass.

During library generation using fecal DNA, most laboratories use relatively low PCR cycle numbers, most commonly 25 cycles (Ericsson et al., 2015; Young et al., 2015). Previous studies have evaluated the effects of PCR cycle number on next-generation sequencing results (Wu et al., 2010; Sze and Schloss, 2019). However, these studies commonly used moderate to high microbial biomass samples such as soil and feces, or mock communities. Studies that evaluated samples with low microbial biomass, such as milk or blood, often used much higher PCR cycle numbers such as 35 (Andrews et al., 2019; Dahlberg et al., 2019) and 40 (Young et al., 2015; Metzger et al., 2018a; Metzger et al., 2018b), but did not show direct comparisons of libraries created at different cycle numbers from matched samples. Higher cycle numbers have been shown to lead to decreased data quality when working with high biomass samples (Sze and Schloss 2019), however low biomass samples often return uninterpretable data (Bjerre et al., 2019) due to low coverage or a complete absence of PCR amplification. For low biomass samples, the potential benefit of increased coverage may outweigh concerns of read quality, which can be filtered out during the bio-informatics stage of analysis.

The objective of this study was to evaluate the effects of PCR cycle number on the sequencing results from three different sample types, which are known to contain low microbial biomass and variable host DNA content. We hypothesized that an increased PCR cycle number would allow for the successful sequencing of low microbial biomass samples. DNA was extracted from matched milk, furred pelage, and blood samples, amplified using

different PCR cycle numbers (25, 30, 35, or 40 cycles in the case of milk; 25 and 40 in the case of pelage and blood), and then sequenced on the Illumina MiSeq platform to compare data quality and agreement between libraries created using differing numbers of PCR cycles.

2. MATERIALS AND METHODS

2.1. Sample collection

Lactating Holstein dairy cows were selected for inclusion in the study based on somatic cell count (SCC, < 100,000 cells/mL). Quarter level milk samples were collected from cows (n = 10) at the University of Missouri's Foremost Dairy Research and Teaching farm just before milking. Milk samples were collected aseptically, according to the National Mastitis Council recommendations (Middleton et al., 2017). Briefly, the cows' teats were pre-dipped with an iodine-based teat dip, wiped with an individual cloth, and foremilk was stripped. Then, teat ends were cleaned with 70% Isopropyl alcohol, 2–3 streams of milk were discarded, and approximately 30 mL of milk was collected into 50 mL sterile conical tubes (Falcon, Corning, New York). The samples were frozen at −20°C until further processing.

Full thickness dorsal pelage samples, measuring approximately 3 × 2 cm, were collected post-mortem from 23 adult outbred CD-1 mice of both sexes. Mice were bred and group-housed under barrier conditions in an Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) international-accredited facility, in individually ventilated microisolator cages, on a 14:10 light:dark cycle, and received autoclaved standard chow (Purina LabDiet 5058) and autoclaved, acidified water *ad libitum*. Samples were collected using aseptic technique on a downdraft table, immediately placed in sterile Eppendorf tubes, and frozen at −20°C until further processing.

Blood samples were collected post mortem via cardiac puncture, placed in anti-coagulant (EDTA)-tubes and frozen at −20°C until further processing. While these mice were from a different cohort than that used for the pelage samples, the generation and maintenance of mice was exactly as described above.

2.2. Milk culture

Milk samples for culture and sequencing were thawed at room temperature (approximately 25°C) and five aliquots of 1.5 mL were prepared using sterile 2 mL tubes (Thermo Scientific, Waltham, MA). One aliquot of each milk sample was cultured with traditional methods to determine if an intramammary infection was present (Middleton et al., 2017). Culturing was performed using a sterile cotton-tipped swab to spread approximately 10 μL of milk onto one-half of a Columbia Blood Agar plate (5% Sheep Blood, Remel, Lenexa, KS). The plates were incubated at 37°C and read at 24 and 48 hours. Only samples that had no bacterial growth at the 48-hour reading were used for downstream sequencing (n = 24).

2.3. DNA Extraction

2.3.1. Milk sample preparation—The remaining four aliquots of 1.5 mL each, were centrifuged at 13,000 × g for 20 minutes to separate the fat layer from the supernatant. After centrifugation, a sterile cotton tipped swab was used to remove the layer of fat. Samples

were then vortexed to resuspend the pellet. Of the four aliquots of each milk sample (n = 96), 800 uL of milk was used as the starting material for DNA extraction.

2.3.2. DNA Extraction and quantification—DNA was extracted from the milk samples (800 uL/sample), full thickness pelage, and blood (800 uL/sample) with a PowerFecal DNA Isolation Kit (Qiagen, Germantown, MD) following the manufacturer's instructions, with the exception that samples were initially subjected to mechanical cell lysis with a TissueLyser II (10 min at 30 Hz, Qiagen), rather than using the vortex adaptor described in the protocol. After extraction, DNA was quantified via fluorometry (Qubit 2.0) using 10 μL of sample DNA and the quant-iT Broad Range dsDNA assay (Life Technologies, Carlsbad, CA).

2.4. Library preparation and sequencing of the 16S rRNA gene amplicons

For milk, 96 libraries were generated using four matched DNA samples from 24 quarter level milk samples. There were 24 libraries generated using each of the four PCR cycle numbers tested, including 25, 30, 35, and 40, all of which were sequenced on a single flow cell. For pelage, 46 libraries were generated using two matched DNA samples from 23 mice (23 generated using 25 PCR cycles, and 23 generated using 40 cycles) on a single flow cell. For blood, 170 libraries were generated, using two matched DNA samples from 85 mice (85 generated using 25 cycles, an 85 generated using 40 cycles); only the latter were sequenced, again on a single flow cell. Library preparation and Illumina sequencing were performed at the University of Missouri DNA Core Facility. The V4 region of the 16S rRNA gene was amplified with the use of previously published universal primers (U515F/806R) (Caporaso et al. 2011) and flanked with Illumina adapter sequences (Walters et al., 2011). Dual-indexed forward and reverse primers were used in all reactions. PCR was performed in 50 μL reactions containing 100 ng metagenomic DNA, primers (0.2 μM each), dNTPs (200 μM each), and Phusion high-fidelity DNA polymerase (1U). Amplification parameters were $98°C^{(3:00)}$ + [$98°C^{(0:15)}$ + $50°C^{(0:30)}$ + $72°C^{(0:30)}$] × 25 to 40 cycles (as indicated) +$72°C^{(7:00)}$. Amplicon pools (5 μL/reaction) were combined, thoroughly mixed, and then purified by addition of Axygen Axyprep MagPCR clean-up beads to an equal volume of 50 μL of amplicons and incubated for 15 minutes at room temperature. Products were then washed multiple times with 80% ethanol and the dried pellet then resuspended in 32.5 μL EB buffer, incubated for two minutes at room temperature, and then placed on the magnetic stand for five minutes. The final amplicon pool was evaluated using the Advanced Analytical Fragment Analyzer automated electrophoresis system, quantified using quant-iT HS dsDNA reagent kits, and diluted according to Illumina's standard protocol for sequencing on the MiSeq instrument. All libraries, except those using blood DNA and amplified for 25 cycles, were sequenced on the Illumina MiSeq platform following the manufacturer's protocol.

2.5. Bioinformatics

The DNA sequences were assembled and annotated at the MU Informatics Research Core Facility. Primers were designed to match the 5' ends of the forward and reverse reads. Cutadapt (version 2.6; https://github.com/marcelm/cutadapt) was used to remove the primer from the 5′ end of the forward read. If found, the reverse complement of the primer to the

reverse read was then removed from the forward read as were all bases downstream. Thus, a forward read could be trimmed at both ends if the insert was shorter than the amplicon length. The same approach was used on the reverse read, but with the primers in the opposite roles. Read pairs were rejected if one read or the other did not match a 5′ primer, and an error-rate of 0.1 was allowed. Two passes were made over each read to ensure removal of the second primer. A minimal overlap of 3 with the 3′ end of the primer sequence was required for removal.

The Qiime2 dada2 plugin (version 1.10.0) was used to denoise, de-replicate, and count ASVs (amplicon sequence variants), incorporating the following parameters: 1) forward and reverse reads were truncated to 150 bases, 2) forward and reverse reads with number of expected errors higher than 2.0 were discarded, and 3) Chimeras were detected using the "consensus" method and removed. R version 3.5.1 and Biom version 2.1.7 were used in Qiime2. Taxonomies were assigned to final sequences using the Silva.v132 database, using the classify-sklearn procedure.

2.6. Statistical methods

The number of detected sequences and OTUs were rank-transformed due to lack of normality as determined by the Shapiro-Wilks test, and a repeated measures ANOVA was performed, using SigmaPlot 14.0. Beta diversity was determined with the weighted Bray-Curtis and unweighted Jaccard similarities in PAST (Hammer et al., 2001), analyzed with a permutational analysis of variance (PERMANOVA), and displayed in a Principal Coordinate Analysis (PCoA) based on Bray-Curtis distances. The proportion of changes in coverage (i.e., the number of high-quality sequences retained) per subject with increasing cycle numbers, relative to 25 PCR cycles, was also evaluated. Statistical significance was determined at $p < 0.05$.

3. RESULTS

Of the aliquoted milk samples used for sequencing (n = 96), only eight had detectable DNA yields (Median = 5.12 ng/ml; Range: 4.08–6.20 ng/ml) based on fluorometric values obtained using broad-range reagents, which provide a lower limit of detection of 2 ng of DNA. Among all 96 samples, a total of 2,168,824 high-quality sequences were obtained with 25, 30, 35, and 40 cycles of PCR. Despite the paucity of DNA based on fluorometric values, the mean (± standard deviation) read count was 22592 ± 456199 sequences with a minimum of 120 and a maximum of 513,457 sequences retained from libraries generated at 25 and 40 cycles, respectively. Of note, while the proportion of filtered reads increased in libraries amplified for 35 and 40 cycles, the greater total number of raw (i.e., unfiltered) sequences was sufficiently increased due to increased cycle number to offset the increased filtering (Figure 1A). Supplementary Table 1 summarizes the total number, and proportion of raw reads represented by that number, remaining in each group of samples at each stage of filtering.

The overall number of sequences detected in the milk samples increased ($p < 0.001$) with higher PCR cycle numbers (Figure 1B). Samples that were subjected to 40 cycles returned the greatest number of sequences followed by 35, 30, and 25 cycles.

For each set of four replicate samples from 24 quarter level milk samples, the highest mean read count was obtained with 40 PCR cycles followed by 35, 30 and 25 cycles (Table 1). However, 30 and 35 cycles of PCR yielded higher read counts than 40 PCR cycles in 6/24 and 4/24 of samples, respectively.

To determine if the increase in coverage was associated with an increase in detected richness, the number of detected OTUs were compared. When data from milk samples were subsampled to a uniform read count, there were no differences in detected richness among samples achieving the subsampling threshold. Thus, while the use of higher PCR cycle numbers increased the number of useable data points (Figures 1C–D), no differences in detected richness were found regardless of the subsampling threshold.

Lastly, beta-diversity was assessed to determine how increased cycle number affected overall community composition. Based on Jaccard similarities, PERMANOVA detected no difference ($p = 0.114$) across PCR cycle number (Figure 2A). Similarly, comparison based on Bray-Curtis similarities (Figure 2B) detected no significant difference between cycle numbers ($p = 0.311$). This suggests that while additional rare taxa were detected with higher cycle numbers, the rare taxa did not alter beta-diversity substantially. Thus, data from samples processed at 25 cycles still accurately profiles the dominant taxa and general beta-diversity, while libraries generated using increased cycle numbers provided more comprehensive profiles.

To determine if increased cycle number could similarly benefit amplification and sequencing of other low biomass sampled, additional samples were processed including portions of furred pelage, and aliquots of blood, collected from healthy mice maintained in an SPF facility. To conserve resources, libraries were generated from matched DNA using only 25 or 40 PCR cycles for these sample types.

DNA yields from pelage were below the limit of detection of our broad-range fluorometric assay for all samples. Comparison of the total number of sequences and extent of filtering revealed a similar proportion of sequences were retained from the pelage libraries generated at 25 cycles (43%) and 40 cycles (46%) (Figure 3A, Supplementary Table 1). However, as the number of unfiltered reads was so much greater in the pelage libraries amplified for 40 cycles, the final number of retained sequences was also higher in those libraries ($p = 9.2\times10^{-5}$, paired t-test; Figure 3B). As with the milk sample data, there was no difference in the detected richness of pelage samples in libraries generated with 25 or 40 cycles, regardless of how data were subsampled. All 23 samples met the lower threshold of 1028 sequences per sample (Figure 3C), but at the higher subsampling threshold of 10165 sequences per sample, three libraries amplified for 25 cycles and one library amplified for 40 cycle were not included (Figure 3D).

As with data generated from milk samples, comparisons of β-diversity between mouse pelage libraries amplified with 25 or 40 cycle numbers failed to detect significant differences, regardless of the similarity metric used and despite the relatively high number of samples ($n = 23$/method) (Figure 4).

Lastly, to evaluate the effect of increased cycle number on samples with extremely low bacterial biomass, murine blood from was used as a source of DNA. While there is not an actual microbial community colonizing peripheral circulation of healthy animals, microbial signatures can be retrieved from blood samples, likely reflecting bacterial DNA within phagocytic leukocytes (e.g., macrophages). Blood samples from 85 healthy mice with a mean (±SD) volume of 391 μL (± 131 μL) yielded 2.47 μg (± 1.09 μg) total DNA, presumably of leukocyte origin. Following library generation at either 25 or 40 cycles using matched DNA, libraries were pooled and analyzed on a Fragment Analyzer to evaluate overall amplification. Analysis of all 85 amplicon libraries generated using 25 cycles produced no characteristic peak at 439 bp, representing the expected amplicon size. Due to the lack of detectable amplification, this pool was not sequenced. In contrast, libraries generated using 40 cycles produced a small peak at the expected size (Figure 5A), and were thus sequenced on a single flow cell. Recognizing the increased potential for amplification of contaminating DNA with increased PCR cycle numbers, we also amplified and sequenced blank reagents (n = 10), as negative technical controls, on a separate flow cell.

While the blood samples generated a total of over 130,000 raw sequences, filtering resulted in removal of all but 1.1% of those reads (Figure 5B, Supplementary Table 1). Nonetheless, data were generated for all 85 samples when amplified for 40 cycles, with a mean read count of 1377 (± 432) sequences per samples. Notably however, reagent controls amplified for 40 cycles also yielded surprisingly high sequence counts (Figure 5B), bringing into question the veracity of other sequence data generated using 40 PCR cycles. Moreover, subsampling of data from blood and negative controls suggested similar richness, albeit substantially less than that detected in pelage and some milk samples.

Lastly, considering the negative control data, beta-diversity among all experimental and control samples was compared using data subsampled to a uniform depth of 1028 sequences per sample to maximize the number of samples included in the analysis. Regardless of the similarity metric used, PCoA demonstrated clear separation of all sample groups, with control samples clustering distinctly from all experimental groups, particularly with regard to the milk and pelage samples (Figure 6A–B). As seen previously, libraries from matched sample types showed no separation based on cycle number. Collectively, we interpret these data as support for the use of increased PCR cycle number with low biomass samples, provided reagent controls are processed alongside experimental samples and taken into consideration in downstream analyses.

4. DISCUSSION

The current study explored the impact of PCR cycle number (25, 30, 35, and 40 cycles) on 16S rRNA sequencing results of healthy bovine milk samples. In this study, bovine milk samples were used as an example of a sample type known to contain low microbial biomass. Low, medium, and high biomass samples have been defined as samples with 10, 1,000, and 100,000 16S rRNA copies per microliter (Bender et al., 2018). Samples that are of low microbial biomass are generally those that were once considered ‘sterile’ and include body sites such as milk and blood, and environmental sites, such as glacial ice, air, rocks. These low biomass samples often fail sequencing using next-generation sequencing technologies

(Eisenhofer et al., 2019) due to low bacterial biomass. There are concerns with using a PCR cycle number greater than 25 when working with high biomass samples due to effects on read quality. However, the results of the current study indicate that, while read quality may or may not be adversely affected by an increased PCR cycle number, the increase in pre-filtering read counts more than compensates for any increased in losses due to filtering.

Previous studies have determined that using PCR cycle numbers greater than 25 can result in increased error rate and chimera formation (Sze and Schloss, 2019). However, the previously mentioned study was conducted using fecal DNA at the standard concentration for library preparation, likely resulting in loss of efficiency in PCR at higher cycles due to saturation of reagents. Chimeras can result from biological similarity, sequencer error, or poor-quality alignment (Edgar et al., 2011) and can cause non-existent organisms to be identified (Wang and Wang, 1996). However, chimera removal standardly occurs after sequencing and before sequence annotation, therefore minimizing this concern. Another apprehension with using PCR cycles greater than 25 is that PCR error could be higher, which may result in inflated alpha-diversity measures that do not accurately represent the true population (Bokulich et al., 2013). Therefore, beta diversity measures should be calculated and reported in combination with alpha diversity to evaluate changes in overall community composition. Moreover, subsampling at multiple depths will greatly mitigate these concerns.

All of the above notwithstanding, the optimized PCR parameter of 25 cycles is preferred when working with high microbial biomass samples like feces or soil, where μg-level masses of DNA are available, and a precise amount of template DNA can be used during library prep. The current study was performed using extremely low microbial biomass milk samples where error rate and chimeras are of less concern than complete sample failure or sequencing to such a low depth that data are uninterpretable.

The present study and others (Young et al., 2015; Dahlberg et al., 2019) have evaluated milk samples from cows with low SCC (< 100,000 cells/mL). Additionally, in the current study, only samples that had no growth of bacterial colonies with traditional culture methods were sequenced. Normal milk appearance, low SCC, and negative culture results indicate that these samples were from cows without an intramammary infection and free of subclinical or clinical mastitis. Previous studies evaluating the milk microbiome have reported higher (Metzger et al., 2018b) and lower (Young et al., 2015; Andrews et al., 2019; Dahlberg et al., 2019) read counts than the study at hand. However, the previously mentioned studies evaluated healthy and infected quarters (Metzger et al., 2018b; Andrews et al., 2019), potentially contributing to successful amplification.

Results from the present study suggest that a higher PCR cycle number allows for increased sample coverage and thus the possibility of increased group sample sizes, without altering the overall community composition noticeably. Ultimately, however, the goals of a specific project, or convenience in combining multiple sample types on one plate for library preparation, might dictate whether or not an increased PCR cycle number is warranted. Coverage can also theoretically be increased by reducing the number of libraries sequenced on a single flow cell, although this brings a concomitant increase in the cost per sample. It

should also be noted that reduced multiplexing on a single flow cell would be expected to result in similar increases in read counts from negative controls.

In the present study, reagents controls yielded high read counts, supporting the introduction of contaminating bacterial DNA at some stage of sample processing or sequencing. While these data certainly raise questions regarding experimental data generated using the same PCR parameters, the fact that detected richness was substantially higher in many experimental samples (even at lower coverage), and that beta-diversity was consistent within experimental groups yet different from the controls, we believe that a significant portion of the experimental data are valid. While beyond the scope of the current study, multiple approaches are available allowing for removal of potential contaminants found in control samples from experimental data, and data generated from low biomass samples should always be analyzed in the context of control data, particularly when extending the PCR parameters.

Based on the results of this study, using PCR cycle numbers greater than 25, such as 35 or 40 cycles, for samples with low microbial biomass will improve 16S rRNA sequencing by allowing for the detection of more rare taxa while not influencing overall bacterial composition. These results contribute to our current understanding of the influences of PCR parameters on the results of 16S rRNA sequencing and provide support for the use of high cycle numbers for samples that contain low microbial biomass.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors would like to thank the Foremost Dairy Farm staff and the MU DNA Core Facility.

FUNDING

This study was funded by Mizzou Advantage and through a K01 from NIH awarded to A. C. Ericsson (OD019924).

REFERENCES

Andrews T, Neher DA, Weicht TR, Barlow JW. 2019 Mammary microbiome of lactating organic dairy cows varies by time, tissue site, and infection status. PLoS ONE 14, e0225001. doi:10.1371/journal.pone.0225001. [PubMed: 31725757]

Bender JM, Li F, Adisetiyo H, Lee D, Zabih S, Hung L, Wilkinson TA, Pannaraj PS, She RC, Bard JD, Tobin NH, and Aldrovandi GM. 2018 Quantification of variation and the impact of biomass in targeted 16S rRNA gene sequencing studies. Microbiome 6, pp. 155. doi:10.1186/s40168-018-0543-z. [PubMed: 30201048]

Bjerre RD, Hugerth LW, Boulund F, Seifert M, Johansen JD, and Engstrand L. 2019 Effects of sampling strategy and DNA extraction on human skin microbiome investigations. Sci. Rep 9:17287. doi:10.1038/s41598-019-53599-z. [PubMed: 31754146]

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, and Caporaso JG. 2013 Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat. Methods 10, pp. 57–59. doi:10.1038/nmeth.2276. [PubMed: 23202435]

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA,

McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, and Knight R. 2010 QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, pp. 335–336. doi:10.1038/nmeth.f.303. [PubMed: 20383131]

Dahlberg J, Sun L, Waller KP, Östensson K, McGuire M, Agenäa S, and Dicksved J. 2019 Microbiota data from low biomass milk samples is markedly affected by laboratory and reagent contamination. PLoS One 14, e0128257. doi:10.1371/journal.pone.0218257.

Edgar RC, Haas BJ, Clemente JC, Quince C, and Knight R. 2011 UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, pp. 2194–2200. doi:10.1093/bioinformatics/btr381. [PubMed: 21700674]

Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. 2019 Contamination in low microbial biomass microbiome studies: issues and recommendations. Cell Press. 27, pp. 105–117. doi:10.1016/j.tim.2018.11.003.

Ericsson AC, Davis JW, Spollen W, Bivens N, Givan S, Hagan CE, McIntosh M, and Franklin CL. 2015 Effects of vendor and genetic background on the composition of the fecal microbiota of inbred mice. PLoS One 10, e0116704. doi:10.1371/journal.pone.0116704. [PubMed: 25675094]

Hammer Ø, Harper DA, and Ryan PD. 2001 PAST: paleontological statistics software package for education and data analysis. Palaeontologia electronica 4, pp. 9. doi:10.1038/s41598-019-48546-x.

McGinnis S and Madden TL. 2004 BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 32, pp. 20–25. doi:10.1093/nar/gkh435.

Metzger SA, Hernandez LL, Skarlupka JH, Suen G, Walker TM, and Ruegg PL. 2018a Influence of sampling technique and bedding type on the milk microbiota: Results of a pilot study. J. Dairy Sci 101, pp. 6346–6356. doi:10.3168/jds.2017-14212. [PubMed: 29680645]

Metzger SA, Hernandez LL, Skarlupka JH, Walker TM, Suen G, and Ruegg PL. 2018b A Cohort Study of the milk microbiota of healthy and inflamed Bovine mammary glands from dryoff through 150 days in milk. Front. Vet. Sci 5, pp. 247. doi:10.3389/fvets.2018.00247. [PubMed: 30356776]

Middleton JR, Fox LK, Pighetti G, and Petersson-Wolfe C. 2017 Laboratory handbook on Bovine mastitis. National Mastitis Council, Inc., New Prague, Minnesota.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, and Glöckner FO. 2013 The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, pp. 590–596. doi:10.1093/nar/gks1219.

Sze MA and Schloss PD. 2019 The impact of DNA polymerase and number of rounds of amplification in PCR on 16S rRNA gene sequence data. mSphere 4, e00163–00119. doi:10.1128/mSphere.00163-19. [PubMed: 31118299]

Walters WA, Caporaso JG, Lauber CL, Berg-Lyons D, Fierer N, and Knight R. 2011 PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. Bioinformatics 27, pp. 1159–1161. doi:10.1093/bioinformatics/btr087. [PubMed: 21349862]

Wang GC and Wang Y. 1996 The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiology 142, pp. 1107–1114. doi:10.1099/13500872-142-5-1107. [PubMed: 8704952]

Wu J-Y, Jiang X-T, Jiang Y-X, Lu S-Y, Zou F, and Zhou H-W. 2010 Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. BMC Microbiology. 10, pp. 255. doi:10.1186/1471-2180-10-255. [PubMed: 20937143]

Young W, Hine BC, Wallace OA, Callaghan M, and Bibiloni R. 2015 Transfer of intestinal bacterial components to mammary secretions in the cow. PeerJ 3, e888. doi:10.7717/peerj.888. [PubMed: 25922791]

Highlights

- Increased PCR cycle number during generation of bacterial 16S rRNA gene amplicon libraries from low biomass samples is associated with greater depth of sequencing.
- Increased PCR cycle number is also associated with increased error rate during library preparation, but this loss is offset by the increased size in amplicon pool.
- While negative reagent controls amplified for an extended cycle number reveal contaminating sequences, controls and experimental samples are clearly differentiated based on beta-diversity.

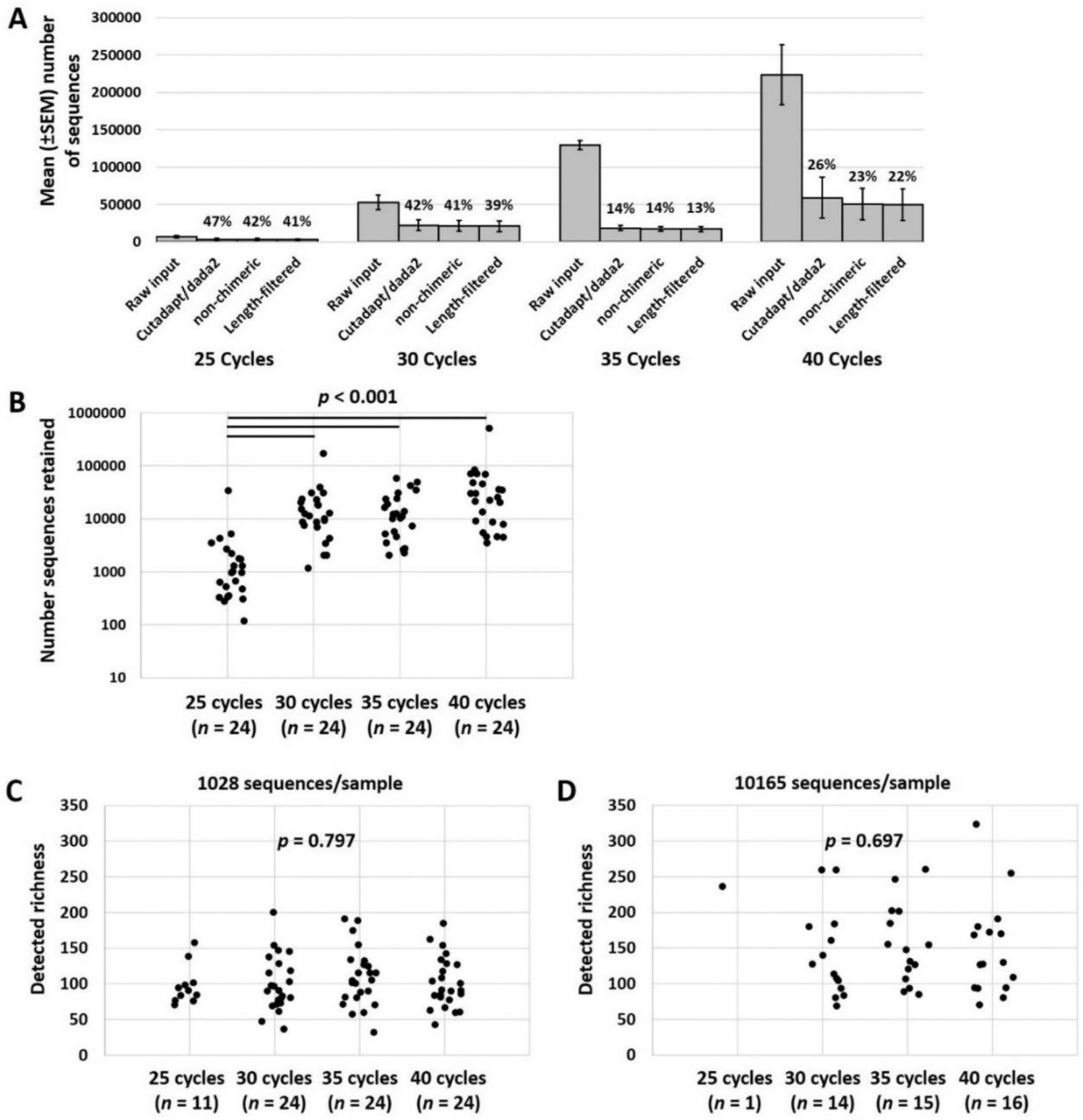

Figure 1.

Histograms showing the number of sequences retained at each level of data filtering from DNA extracted from matched milk samples ($n = 24$), and amplified for 25, 30, 35, or 40 PCR cycles (A). Dot plots show the total number of sequences retained at each cycle number (B), and the detected richness in datasets subsampled to a uniform depth of 1028 sequences (C) or 10165 sequences (D) per sample. Bars indicate significant ($p < 0.001$) differences (Friedman's repeated measures ANOVA on ranks due to lack of normality.

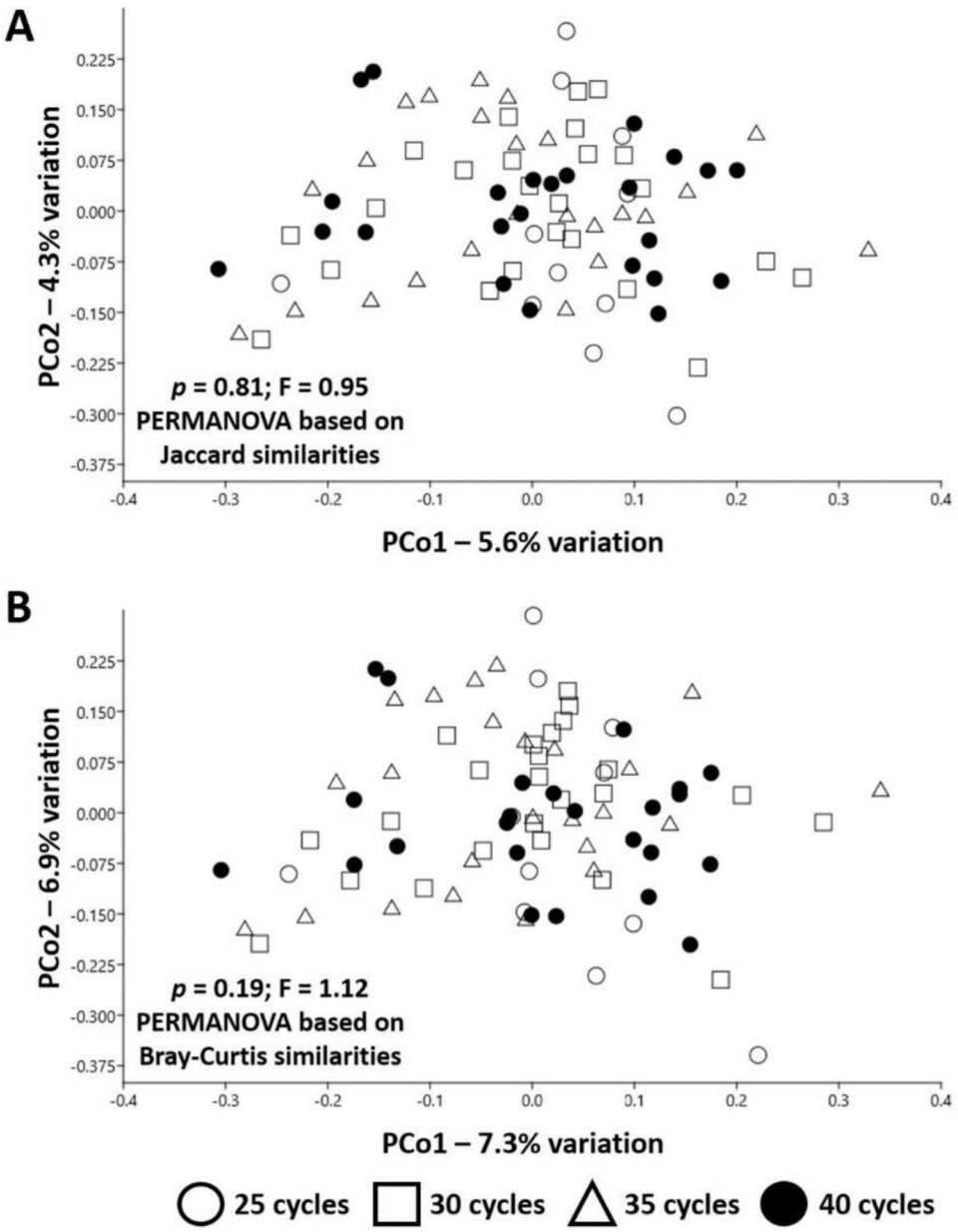

Figure 2.

Comparisons of beta-diversity, based on Jaccard (A) and Bray-Curtis (B) similarities via principal coordinate analysis, of libraries prepared from matched milk samples (n = 24) and amplified for 25, 30, 35, or 40 cycles (legend at bottom). Results of one-factor PERMANOVA are provided on each plot.

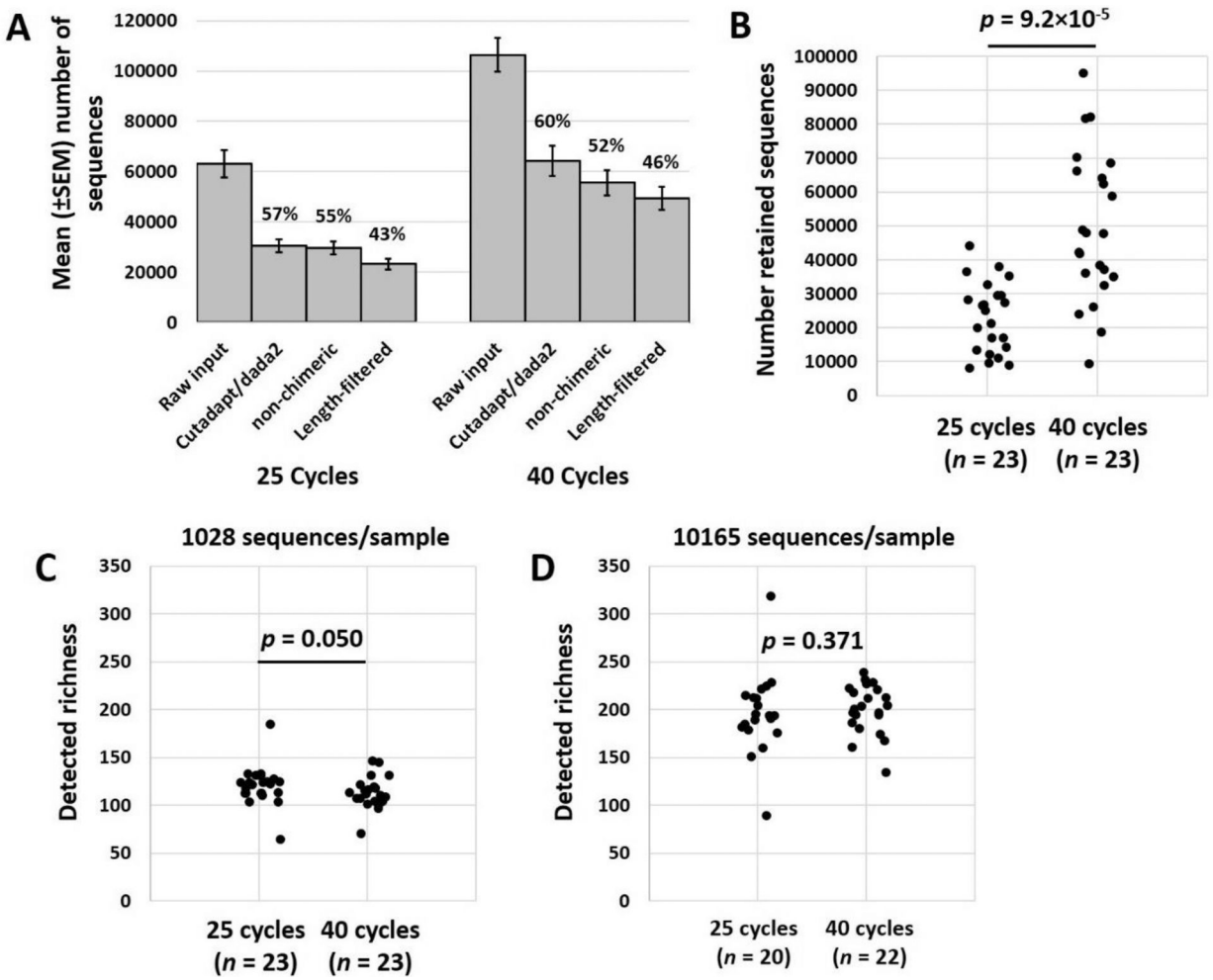

Figure 3.

Histograms showing the number of sequences retained at each level of data filtering from DNA extracted from matched pelage samples (n = 23), and amplified for 25 or 40 PCR cycles (A). Dot plots show the total number of sequences retained at each cycle number (B), and the detected richness in datasets subsampled to a uniform depth of 1028 sequences (C) or 10165 sequences (D) per sample.

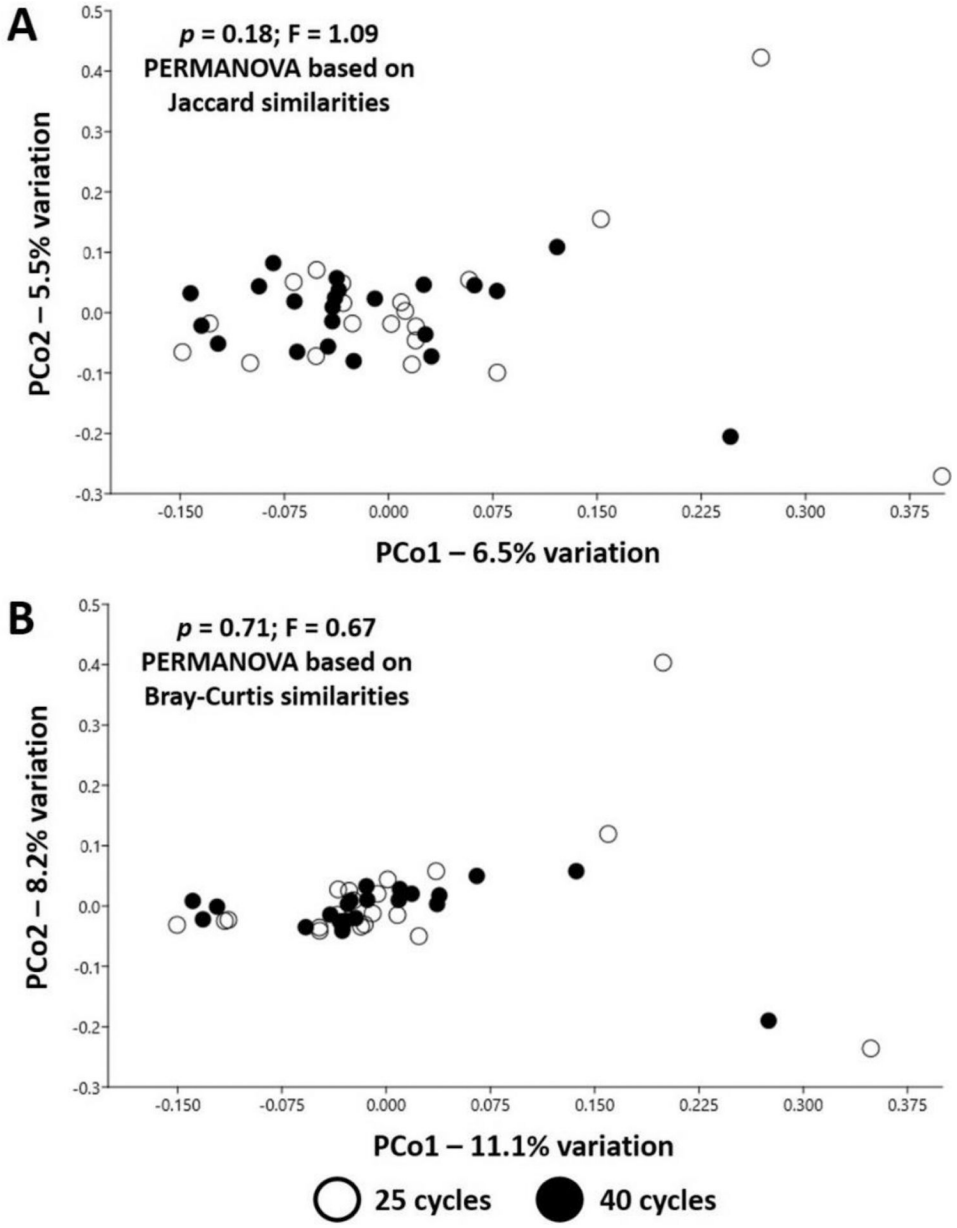

Figure 4.

Comparisons of beta-diversity, based on Jaccard (A) and Bray-Curtis (B) similarities via principal coordinate analysis, of libraries prepared from matched pelage samples (n = 23) and amplified for 25 or 40 cycles (legend at bottom). Results of one-factor PERMANOVA are provided on each plot.

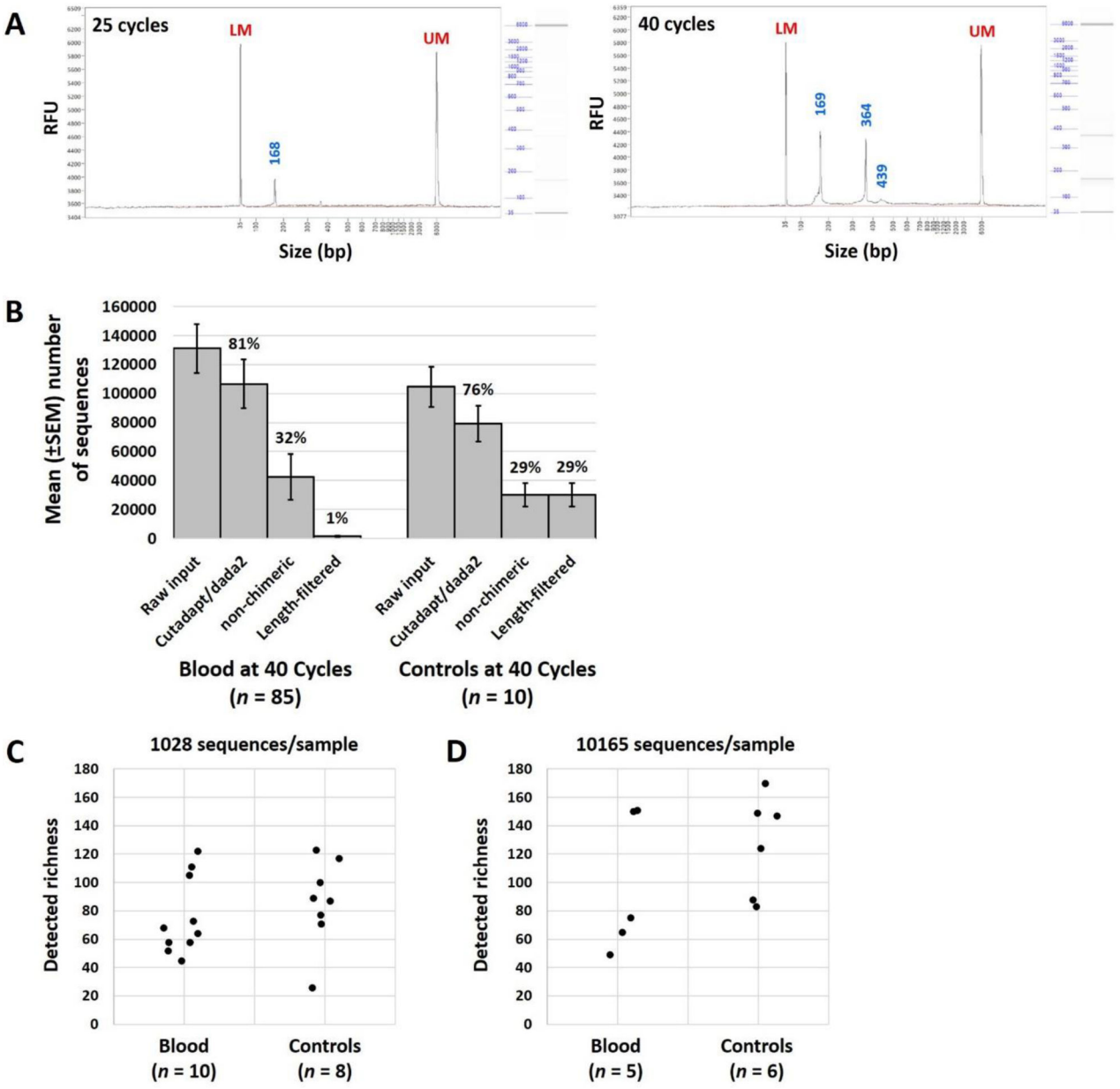

Figure 5.

Quantification of amplified DNA at 439 bp in pooled 16S rRNA libraries generated from murine blood (n = 85) amplified for 25 or 40 cycles (A); histograms showing the number of sequences retained at each level of data filtering from DNA extracted from blood samples (n =853) and reagent controls (n = 10), and amplified for 40 PCR cycles (B). Dot plots show the detected richness in datasets subsampled to a uniform depth of 1028 sequences (C) or 10165 sequences (D) per sample.

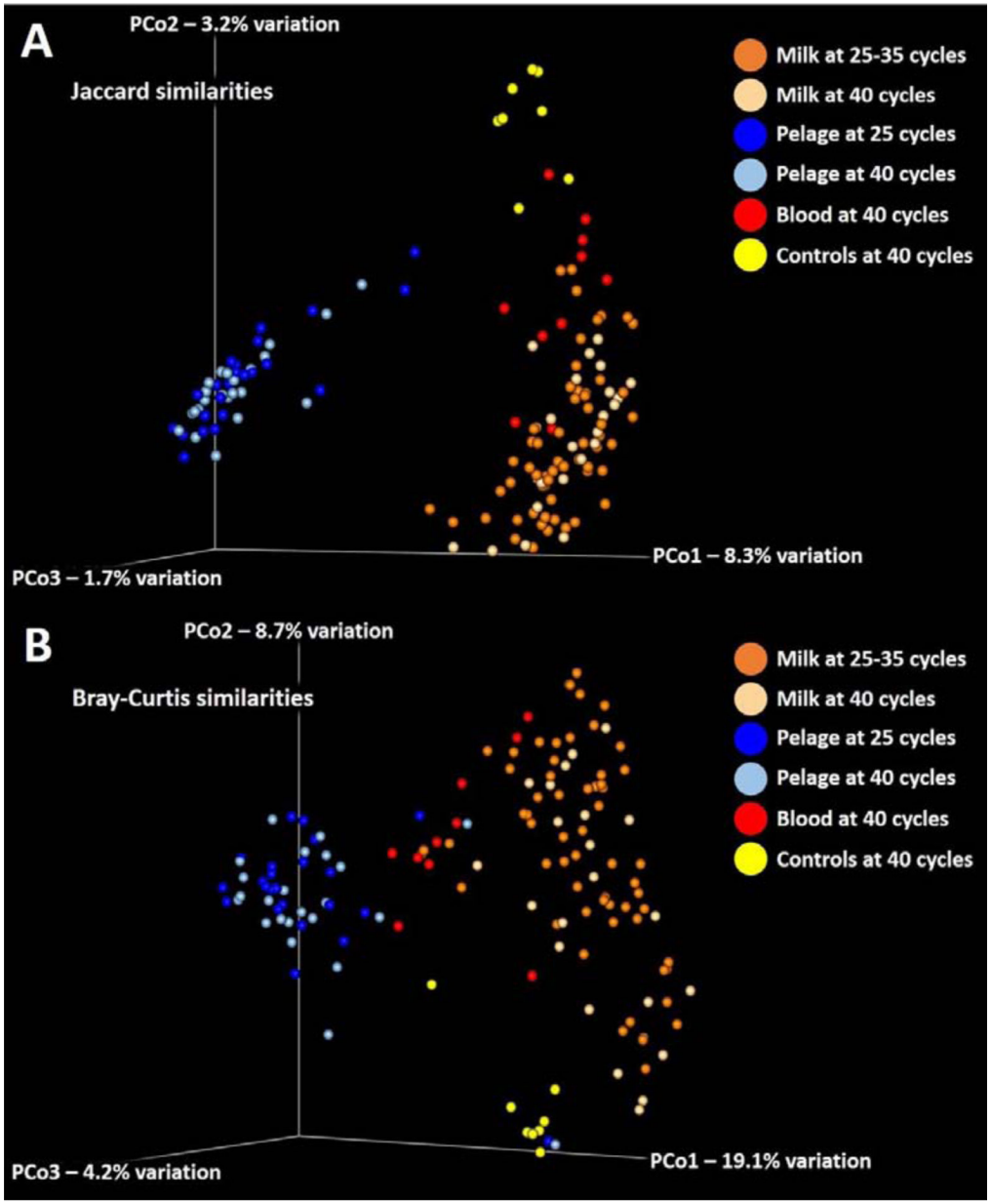

Figure 6.

Three-dimensional PCoA showing beta-diversity among all amplified libraries, subsampled to a uniform depth of 1028 sequences per sample, and ordinated according to Jaccard (A) or Bray-Curtis (B) similarities; legend at right.

Table 1.

The highest proportion of read counts obtained with each cycle threshold.

PCR Cycle Number	[a]Highest read count (%)	[b]Read count (±SEM)
25	0/24 (0.0)	2783 (1410)
30	6/24 (25.0)	20838 (6995)
35	4/24 (16.7)	17042 (3242)
40	14/24 (58.3)	49704 (20776)

[a]The proportion of samples with the highest read count for a given cycle number.

[b]The average read count for each PCR cycle threshold (± standard error of the mean).