



# Evaluation of Host Depletion and Extraction Methods for Shotgun Metagenomic Analysis of Bovine Vaginal Samples

 Chian Teng Ong,<sup>a</sup> Gry Boe-Hansen,<sup>b</sup> Elizabeth M. Ross,<sup>a</sup> Patrick J. Blackall,<sup>a</sup> Conny Turni,<sup>a</sup> Ben J. Hayes,<sup>a</sup> Ala E. Tabor<sup>a,c</sup>

<sup>a</sup>Queensland Alliance for Agriculture and Food Innovation, Centre for Animal Science, The University of Queensland, Brisbane, Queensland, Australia

<sup>b</sup>School of Veterinary Science, Faculty of Science, The University of Queensland, Brisbane, Queensland, Australia

<sup>c</sup>School of Chemistry and Molecular Bioscience, Faculty of Science, The University of Queensland, Brisbane, Queensland, Australia

**ABSTRACT** The reproductive tract metagenome plays a significant role in the various reproductive system functions, including reproductive cycles, health, and fertility. One of the major challenges in bovine vaginal metagenome studies is host DNA contamination, which limits the sequencing capacity for metagenomic content and reduces the accuracy of untargeted shotgun metagenomic profiling. This is the first study comparing the effectiveness of different host depletion and DNA extraction methods for bovine vaginal metagenomic samples. The host depletion methods evaluated were slow centrifugation (Soft-spin), NEBNext Microbiome DNA Enrichment kit (NEBNext), and propidium monoazide (PMA) treatment, while the extraction methods were DNeasy Blood and Tissue extraction (DNeasy) and QIAamp DNA Microbiome extraction (QIAamp). Soft-spin and QIAamp were the most effective host depletion method and extraction methods, respectively, in reducing the number of cattle genomic content in bovine vaginal samples. The reduced host-to-microbe ratio in the extracted DNA increased the sequencing depth for microbial reads in untargeted shotgun sequencing. Bovine vaginal samples extracted with QIAamp presented taxonomical profiles which closely resembled the mock microbial composition, especially for the recovery of Gram-positive bacteria. Additionally, samples extracted with QIAamp presented extensive functional profiles with deep coverage. Overall, a combination of Soft-spin and QIAamp provided the most robust representation of the vaginal microbial community in cattle while minimizing host DNA contamination.

**IMPORTANCE** In addition to the host tissue collected during the sampling process, bovine vaginal samples are saturated with large amounts of extracellular DNA and secreted proteins that are essential for physiological purposes, including the reproductive cycle and immune defense. Due to the high host-to-microbe genome ratio, which hampers the sequencing efficacy for metagenome samples and the recovery of the actual metagenomic profiles, bovine vaginal samples cannot benefit from the full potential of shotgun sequencing. This is the first investigation on the most effective host depletion and extraction methods for bovine vaginal metagenomic samples. This study demonstrated an effective combination of host depletion and extraction methods, which harvested higher percentages of 16S rRNA genes and microbial reads, which subsequently led to a taxonomical profile that resembled the actual community and a functional profile with deeper coverage. A representative metagenomic profile is essential for investigating the role of the bovine vaginal metagenome for both reproductive function and susceptibility to infections.

**KEYWORDS** cattle, metagenomics, microbiome, shotgun, vagina, veterinary microbiology

The term metagenome refers to the genomes and genes of the microorganisms present in a defined environment (1). Metagenomic studies can bring forth a breakthrough to microbial discovery and ecology studies. For instance, the discovery of nonculturable and new microbial species has been challenging with traditional culture-based and molecular

**Editor** Kang Ning, Huazhong University of Science and Technology

**Copyright** © 2022 Ong et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Ala E. Tabor, a.tabor@uq.edu.au.

The authors declare no conflict of interest.

**Received** 26 May 2021

**Accepted** 21 March 2022

**Published** 11 April 2022

approaches and is now commonly reported by various metagenomic sequencing projects (2, 3). The entire microbial community can be accessed without extensive labor and a stringent need to customize a suitable growth environment or specific markers for each microbial species (4, 5). Profiling the microbial community in host systems has revealed a greater coverage of microbial inhabitation in host systems, including those which were long presumed to be sterile, for example, urine (6), milk (7), fetal fluid (8), and uterus (9). The microbiome was demonstrated to play a vital role in host physiology (10, 11), nutrition (12, 13), and health (14, 15). A balanced microbiome is beneficial to healthy biological processes such as digestion, immune system maturation, toxin degradation, and pathogen defense (16). However, imbalanced microbiomes have been associated with disease development, such as urinary tract infection (17) and reproductive diseases (18). In addition to the interrogation of the microbial community profile, diversity, and abundances, the untargeted shotgun metagenomic approach also provides insights into the functional characteristics of the microbiome in the complex host-microbial relationships under different circumstances (19–21).

Increasing research efforts have been directed to bovine reproductive tract metagenomic studies in the past 10 years. Metagenomes of bovine reproductive tracts have been associated with bovine reproductive functions, including reproductive cycle (22), gestation (23), and reproduction health (24, 25). Microbial infiltration may occur during calving or mating and lead to undesirable shifts of the metagenome in the bovine reproductive tract (26–28). Disturbed commensal microflora have been associated with the development of reproductive diseases, including metritis and endometritis (29, 30).

One of the challenges in host-derived metagenomic studies is host DNA contamination, which has been commonly observed in most host-derived microbiome studies (31–34). In addition to the vaginal epithelial tissue collected during the sampling process, metagenome samples from the bovine reproductive tract are saturated with large amounts of extracellular DNA and secreted proteins essential for various physiological purposes, including reproductive cycle and host defense (35, 36). The vast amount of host genetic content in the metagenome sample dominates the sequencing capacity, and, in turn, reduces the sequencing depth for the metagenomic content (37). Additionally, the host genome is larger than the average microbial genome. Hence, host genomic information could easily dominate the sequencing capacity even with a small number of host cells (38). For instance, the cattle genome is 2.7 Gb, approximately 750 times larger than the average bacterial genome (~3.6 Mb) (39, 40). Importantly, the ratio of host-to-microbe cells varies at different sampling sites. For example, fecal samples typically yield much less host genomic material than saliva, mucus, skin, and vaginal swabs. Additionally, metagenome samples constitute a mixture of microorganisms, including bacteria, fungi, archaea, viruses, and protists, which represent different cell structures. It is challenging to efficiently extract a mixture of microbial DNA using one extraction method (37, 38). While increasing sequencing depth can recover representative metagenomic profiles from metagenome samples with a large amount of host genomic materials (34), this increases the costs and requires extensive computational effort. Therefore, host depletion is pivotal for cost-effective metagenomic studies of host-derived metagenome samples.

In most of the bovine vaginal metagenomic studies, DNeasy Blood and Tissue extraction kit (DNeasy) was incorporated for metagenome DNA extraction (41). However, the benefits of effective host depletion for metagenomic studies were reported, primarily for human clinical samples (42–45). In general, these methods manipulate the differences between mammalian cells and microbial cells, for example, cell size, cell wall structures, and DNA methylation pattern, to achieve the separation either before or after DNA extraction. Nonetheless, the different complexity and chemistry matrix of the host-derived samples pose different degrees of impact on the performances of the host depletion and extraction methods (38, 46). Therefore, we conducted the first study to evaluate the efficacies and potential bias of the different host depletion and extraction methods on bovine vaginal samples intended for metagenomic studies. In total, three host depletion methods and two extraction methods were evaluated from different aspects, including the percentage of 16S ribosomal (r)

**TABLE 1** Details of the bacteria used to assess the effectiveness of the depletion and DNA extraction methods

Bacterial species	Strain	Gram stain	Aerotolerance	Median GC (%)	Genome size (Mb)	Spiked concentration (CFU/mL × 10 <sup>6</sup> )
<i>Aliarcobacter cryaerophilus</i>	CCUG17804	Gram-negative	Aerobic	27.4	2.09	4.55
<i>Campylobacter fetus</i> subsp. <i>venerealis</i>	ATCC 19438	Gram-negative	Microaerophilic	33.1	1.82	4.55
<i>Enterococcus faecalis</i>	BR1200	Gram-positive	Facultative anaerobe	37.4	2.97	2.22
<i>Pseudomonas aeruginosa</i>	ATCC 27853	Gram-negative	Aerobic	66.2	6.6	9.91
<i>Staphylococcus aureus</i>	ATCC 29213	Gram-positive	Anaerobic	32.7	2.84	1.82

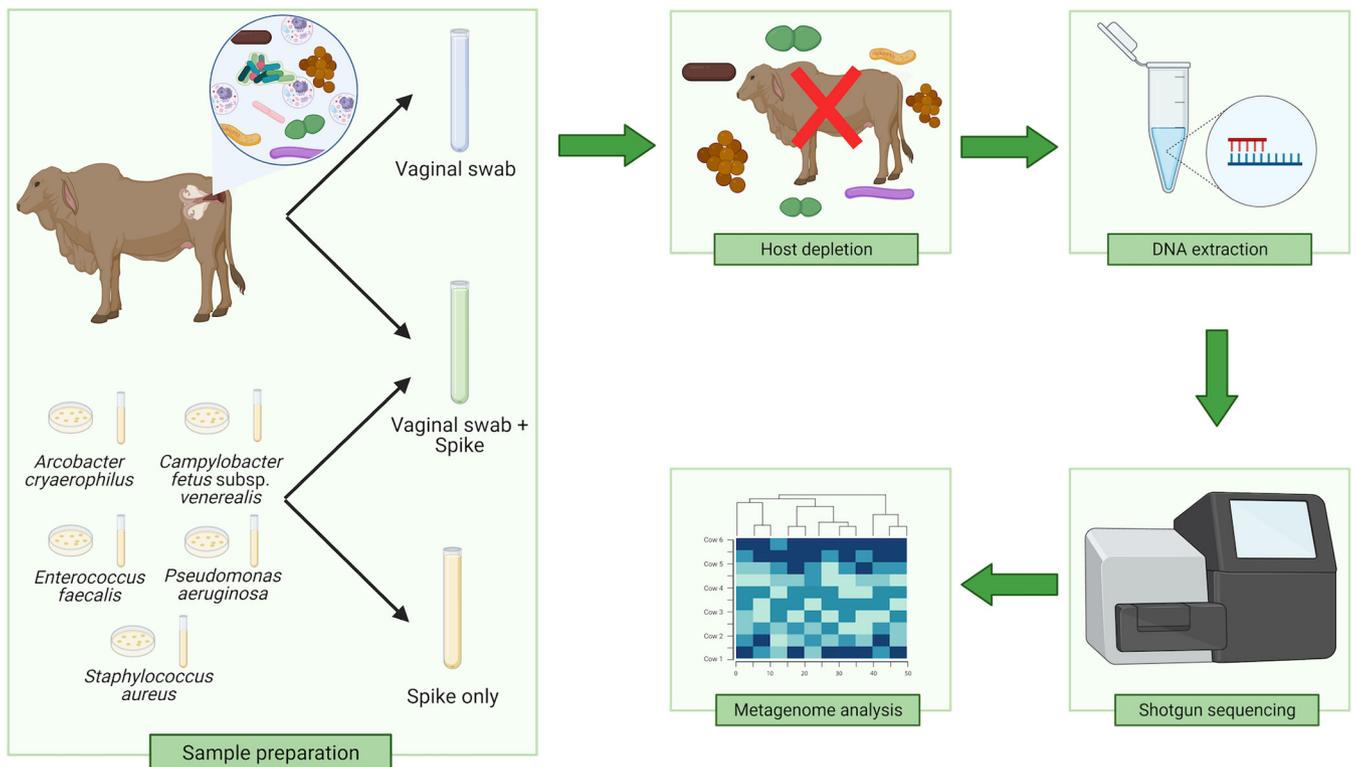
RNA genes in the extracted samples, the percentage of microbial reads, the coverage and accuracy of the metagenomic profiles.

**RESULTS**

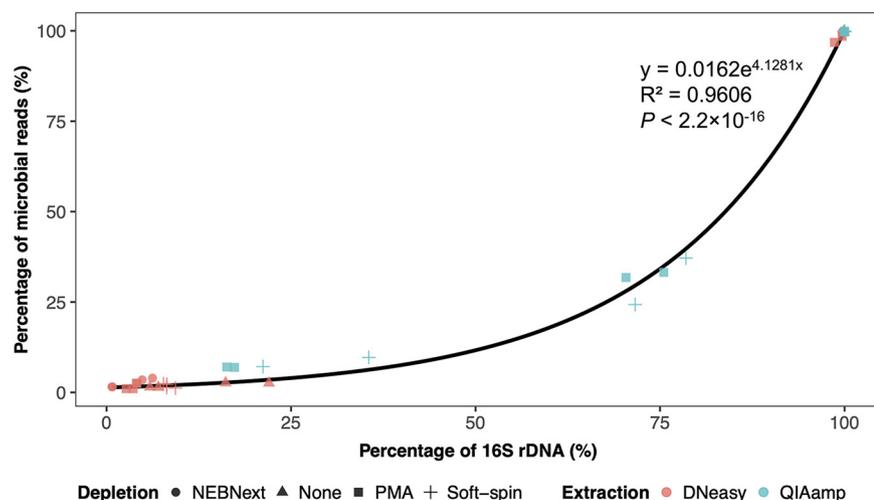
**Spiked suspension.** The total amount of spiked bacteria in each of the vaginal swabs, 14 mock samples, and 14 positive-control samples was  $2.3 \times 10^7$  CFU/mL (Table 1). According to culture plate counting, *Pseudomonas aeruginosa* constituted the highest percentage (43.01%), followed by *Campylobacter fetus* subsp. *venerealis* (19.73%) and *Aliarcobacter cryaerophilus* (19.73%), *Enterococcus faecalis* (9.63%), and *Staphylococcus aureus* (7.89%).

**Sample preparation and extraction.** In total, 42 samples were prepared, including 14 vaginal swabs, 14 mock samples, and 14 positive-control samples (Fig. 1).

**Shotgun sequencing.** Samples processed using the combination of NEBNext and QIAamp extraction were eliminated from downstream analyses because the DNA quantities were too low and were not eligible for shotgun sequencing. Whole-metagenome shotgun sequencing of the samples ( $n = 36$ ) returned an average of 6.47 Gbp of data and an average of 51,956,238 raw paired-end reads per sample. An average of 48,965,875 paired-end reads per sample remained after trimming (Text S1).



**FIG 1** Experimental design. Three sample types were used to examine the efficacies of depletion and extraction methods. Vaginal swabs were vaginal swabs without spiked bacteria. The mock sample contained vaginal samples and five bacterial suspensions. Positive-control samples contained only the five bacterial suspensions. The samples were treated with different depletion and extraction methods to obtain the nucleic acids for shotgun sequencing and subsequent metagenomic analysis.



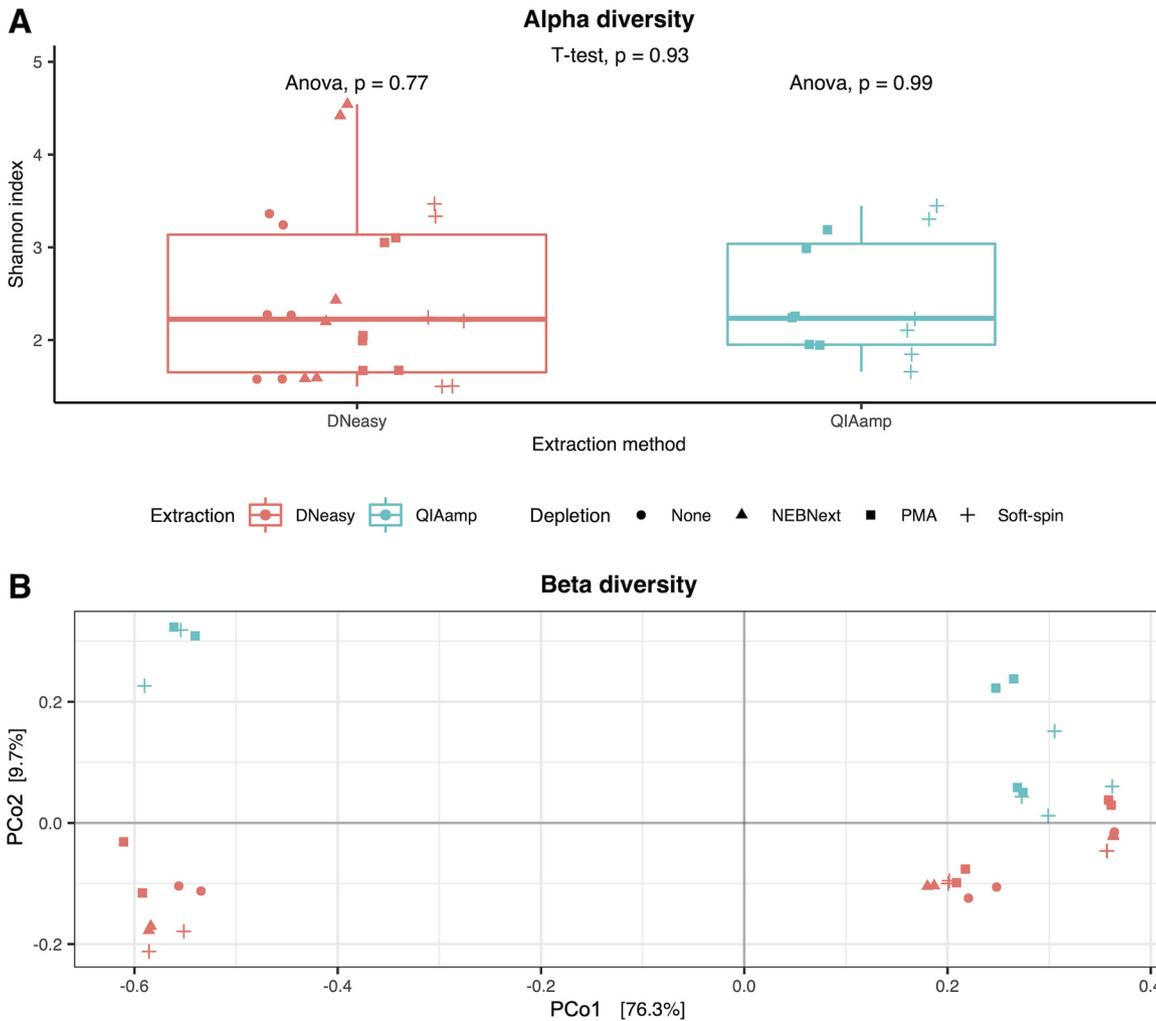
**FIG 2** The relationship between the percentage of 16S rRNA genes and the percentage of microbial reads identified in the bovine vaginal samples. The markers represent all the samples involved in this study ( $n = 36$ ), including vaginal swabs, mock samples, and positive-controls. Samples were treated with different depletion methods, including None (circle), NEBNext (triangle), PMA (square), and Soft-spin (cross). Red represents samples extracted using DNeasy and blue represents samples extracted using QIAamp.

**Metagenome assembly.** The paired-end reads were assembled into contigs, and the accuracy of contig-based metagenomic profiles was investigated. The number of contigs ranged from 335 to 19,711 and the average contig length ranged from 484 to 48,911 (Text S1). The contig coverage was examined by realigning the reads back to the contigs, and the mapped percentage ranged from 15.03 to 99.80% (Text S1). The mapped percentage of the assembled contigs evaluated using MetaQUAST ranged from 24.06 to 99.88% (Text S1).

**Relationship between the percentage of 16S rRNA genes and the percentage of microbial reads in the samples.** To analyze the impacts of host DNA depletion in the bovine vaginal samples, the relationship between the percentage of the 16S rRNA genes and the percentage of microbial reads identified in the samples ( $n = 36$ ) was investigated (Fig. 2). The result indicated the proportion of microbial reads sequenced by shotgun sequencing was exponentially proportional ( $P < 0.001$ ) to the percentage of 16S rRNA genes in the bovine vaginal samples. The impacts of depletion and extraction methods on the proportion of 16S rRNA genes and microbial reads were not significantly different. However, Soft-spin and QIAamp recovered the highest percentages of 16S rRNA genes (mean = 53.3% and 65.5%) and microbial reads (mean = 40.4% and 46.4%), respectively.

**Impacts on the taxonomic profile.** Alpha diversity of each sample ( $n = 36$ ) was examined using the Shannon index. In general, there was a higher alpha diversity in the vaginal swab (mean = 3.45) followed by a mock sample (mean = 2.21) and positive-control (mean = 1.67). However, the alpha diversities between samples treated with different depletion methods (ANOVA,  $P = 0.77$  and  $0.99$ ) and different extraction methods ( $t$  test,  $P = 0.93$ ) were not significantly different (Fig. 3A). The Bray-Curtis dissimilarity demonstrated that the dissimilarity caused by depletion methods was not significant (permutational multivariate analysis of variance [PERMANOVA],  $P = 0.58$ ) while the dissimilarity resulting from the extraction methods was significant (PERMANOVA,  $P = 0.005$ ) (Fig. 3B).

The abundances of the bacteria spiked into the mock sample and positive-control were examined to assess the impacts of depletion and extraction methods on the recovery of taxonomic profile. The abundances of the spiked bacteria in the negative controls (vaginal swab) samples were less than 2% (Fig. 4). The impacts of depletion methods on the abundances of the spiked bacteria were not significantly different from each other (Table 2). Contrarily, the impacts of the extraction method were insignificant on the abundances of *A. cryaerophilus*, *C. fetus*, and *P. aeruginosa* but were significant for *E. faecalis* (ANOVA,  $P < 0.001$ ) and *S. aureus* (ANOVA,  $P = 0.003$ ). *Enterococcus faecalis* and *S. aureus* were detected at lower abundances (mean = 0.404% and 0.34%) in samples treated with DNeasy compared to QIAamp (mean = 8.47% and 1.96%). *Enterococcus faecalis* and *S. aureus* constituted 9.63%



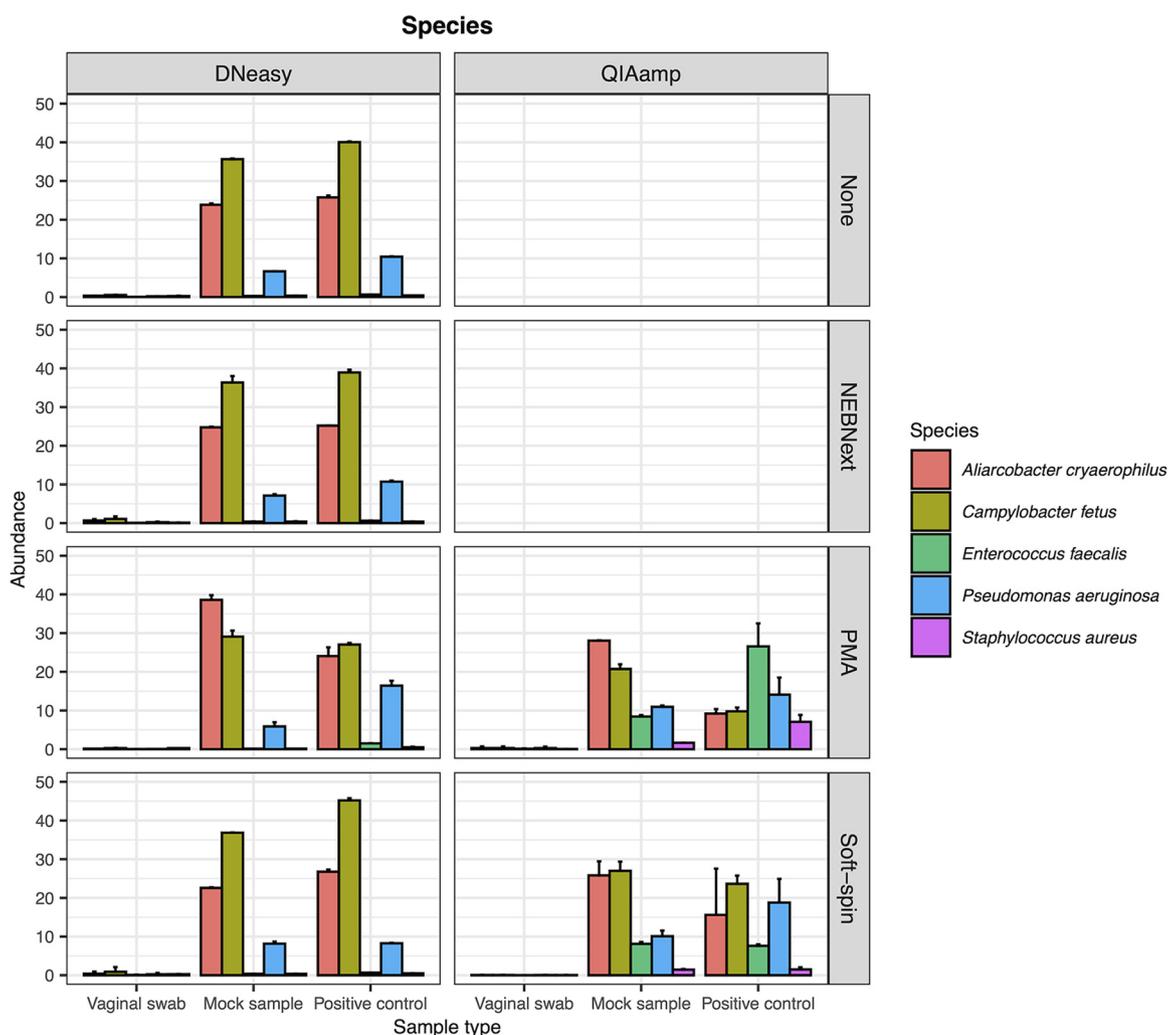
**FIG 3** Alpha (A) and Beta (B) diversity of the samples involved in this study ( $n = 36$ ), including vaginal swab, mock sample, and positive-control. Samples were treated with different depletion methods, including None (circle), NEBNext (triangle), PMA (square), and Soft-spin (cross). Red represents samples extracted using DNeasy and blue represents samples extracted using QIAamp. ANOVA test was applied to examine the significance of differences between the alpha diversity of samples treated with different depletion methods while the  $t$  test was applied to examine the significance of differences between the alpha diversity of samples processed with different extraction methods. Beta diversity was represented by principal coordinate analysis ordination of the Bray-Curtis dissimilarity matrix.

and 7.89% in the spiked suspension (Table 1). Hence, QIAamp was demonstrated to recover percentages of bacteria that were closer to the spiked proportion.

**Impacts on functional classification.** To study the impacts of host depletion and extraction methods on the contig-based functional classification of shotgun metagenomic data, the total functional annotations were reported by eight databases, including Enzyme, Gene Orthology, KEGG Orthology, KEGG pathways, Metabolic-hmm, MetaCyc pathways, Pfam, and Tigrfam. The results indicated that the extraction methods contributed significant impacts on the number of functional annotations reported ( $P < 0.001$ ) (Fig. 5).

**DISCUSSION**

This is the first study that has evaluated multiple host depletion and metagenome extraction methods on bovine vaginal metagenome samples. The exponential relationship between the percentage of 16S rRNA genes and microbial reads depicted in this study suggested the importance of host depletion for efficient metagenome sequencing. Host DNA contamination reduced the sequencing depth for metagenome content and subsequently led to reduced sensitivity of metagenomic studies performed with untargeted shotgun sequencing (47). The results indicated that Soft-spin and QIAamp



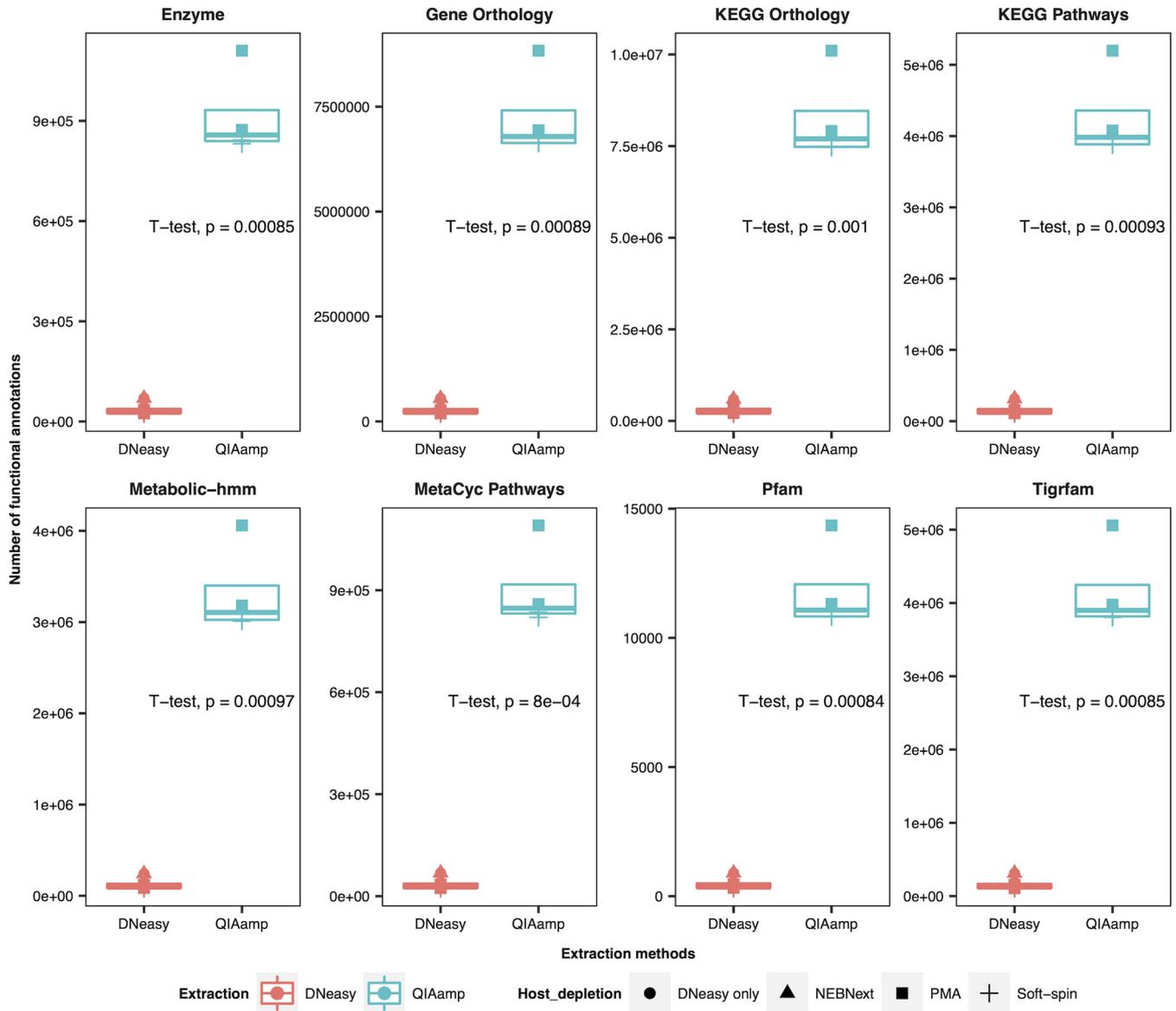
**FIG 4** Abundances of spiked bacteria, including *Aliarcobacter cryaerophilus*, *Campylobacter fetus*, *Enterococcus faecalis*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus*, in the samples involved in this study ( $n = 36$ ), including vaginal swabs, mock samples, and positive-controls.

outperformed the other host depletion and extraction methods, including the commonly used extraction method “DNeasy only,” in removing the cattle DNA from bovine vaginal samples. Consequently, samples extracted with QIAamp resulted in taxonomic profiles that more resembled the spiked community and functional profiles with deeper coverage.

The efficacy of PMA treatment was dependent on the microbial content in the metagenome samples. Propidium monoazide unbiasedly digests the exposed DNA. Hence, the selective impacts of PMA-based host depletion treatments depend greatly on the lysis methods and the permeability of microbial cell membrane (38, 48, 49). Potentially, the integrity of the microbial cells in vaginal swab samples was not preserved in vaginal mucus during the sample

**TABLE 2** Significance of different, calculated using ANOVA test, between the abundances of spiked bacteria in the samples ( $n = 36$ ) processed with different depletion and extraction methods

Species	ANOVA, <i>P</i> value	
	Depletion	Extraction
<i>Aliarcobacter cryaerophilus</i>	0.991	0.32
<i>Campylobacter fetus</i>	0.444	0.065
<i>Enterococcus faecalis</i>	0.195	0.000156
<i>Pseudomonas aeruginosa</i>	0.861	0.189
<i>Staphylococcus aureus</i>	0.27	0.003



**FIG 5** Number of functional annotations identified by the different functional and pathway databases. A *t* test was applied to examine the significance of differences between the number of functional annotations in samples processed with different extraction methods.

collection and transfer. Hence, the exposed microbial DNA content was digested together with the cattle genomic materials before extraction. Metagenome samples treated with NEBNext recovered fewer 16S rRNA genes and microbial sequences than samples extracted without a host depletion step, indicating the low efficacy of NEBNext in depleting cattle genomic materials from vaginal samples. This is potentially due to the low molecular weight genomic DNA extracted using DNeasy and QIAamp. NEBNext is a postextraction method that requires >15 kb high-molecular-weight DNA, which is typically achieved by conventional extraction methods for effective capture and subsequent host removal (37). The low throughput of conventional extraction methods renders them less ideal for quantitative metagenomic studies with multiple samples (50, 51). Extraction methods used in this study were all based on the vigorous lysis required for bacterial cells, which is likely to shear the long fragment DNA in most cases, making the NEBNext protocol unsuitable for this use. Of the tested host depletion methods (Soft-spin, PMA treatment, and NEBNext), Soft-spin was the most efficient and consistent at depleting the cattle DNA from the bovine vaginal samples. It was demonstrated that Soft-spin resulted in higher percentages of 16S rRNA genes and microbial sequences. Soft-spin effectively

separated the cattle tissues in the bovine vaginal samples and allowed a higher percentage of microbial DNA to be extracted.

DNeasy has been commonly incorporated in bovine reproductive tract metagenome studies (52–57), while QIAamp is relatively less common (29). However, our results indicated that QIAamp was more effective than DNeasy in maximizing the 16S rRNA genes and microbial sequences in both the actual and mock bovine vaginal samples. Unlike DNeasy, which relied solely on chemical lysis for nucleic acid extraction, QIAamp implemented two separate lysis steps. The first step was the host enzymatic lysis and degradation with benzonase. The efficacy of benzonase in nucleic acid hydrolyzation is well documented and has been widely used in biopharmaceutical production (46, 58, 59). The second step was the bead-beating mechanical lysis and proteinase K lysis of microbial cell walls. Therefore, metagenome samples extracted with QIAamp extraction recovered a higher percentage of microbial DNA and sequences, particularly *E. faecalis* and *S. aureus*, which were significantly less captured by the DNeasy extraction. Gram-positive bacteria like *E. faecalis* and *S. aureus* possess a thick outer cell wall made up of peptidoglycan, allowing Gram-positive bacteria to be more resistant to heat and chemical stresses (60, 61). Bead-beating mechanical lysis helps to break the thick microbial cell walls and release the encapsulated nucleic acid, consequently improving the yield and quality of microbial DNA (62, 63). The effectiveness of QIAamp in increasing the microbial DNA and the subsequent effects on the metagenomic profiles were also reported in other host-derived metagenomic studies, which investigated fluid collected from infected respiratory tract and biopsy specimen samples from diabetic foot infections (64, 65).

The vaginal samples used in this study were collected from acyclic heifers, which were often the control group for comparative bovine reproductive tract metagenomic studies (66–69). The mock sample simulated the bovine vaginal samples collected from mated cows or cows that develop reproductive diseases, which often have richer microbial content (70). Vaginal swabs are one of the most challenging sample types for metagenomic investigation due to the high (>90%) host DNA content (38, 47). Our findings demonstrated that Soft-spin and QIAamp were effective in extracting the metagenomic content from bovine vaginal samples, either with low or enriched microbial contents and subsequently increased the sequencing efficiency and recovered metagenome profiles with higher accuracy and coverage. This study provides a beneficial reference for other host-derived metagenomic studies, especially samples with a similar chemistry matrix and low microbial content as bovine vaginal swabs.

One of the greatest advantages of sequencing the metagenome samples with untargeted shotgun methods is the functional profiling of the metagenome. The assembled contigs can be mapped to the protein and pathway databases for functional annotations to investigate the functional focus of the metagenome under different conditions (21, 71). Our results demonstrated that there was a significantly higher recovery rate of the functional annotations in the samples extracted using QIAamp. The increased number of functional annotations identified, and the relative coverage of each annotation essentially improved the functional profiles of metagenome samples by QIAamp. An extensive functional profile with deep coverage provides information regarding the metagenome's functional capacity, including the virulence factors, antibiotic resistance, and metabolic pathways (72). Nonetheless, the accuracy of the functional profile identified by QIAamp was not explored in this study.

Because samples in this study were collected from healthy heifers, the impacts of infected host tissues and the influx of inflammatory host cells on the extracted metagenomic content remain unexplored. Further studies shall investigate the host-to-microbe ratio in samples collected from cattle diagnosed with reproductive diseases and its impacts on the sequencing efficiency and recovery of metagenomic profiles. There were only 2 replicates used in the method tested for each sample, this potentially undermined the credibility of the result. Additionally, the proportion of each spiked bacteria in the extracted DNA shall be verified using an individual quantification test.

Our results demonstrated that Soft-spin was more efficient than NEBNext and PMA in reducing the host-to-microbe ratio in bovine reproductive tract metagenome samples. The relationship between 16S rRNA genes and microbial reads detected signified that the shotgun metagenomic sequencing was increasingly more efficient as the percentage of 16S

rRNA genes in the metagenome samples increased. QIAamp extraction outperformed DNeasy by improving the microbe-to-host DNA ratio, providing more accurate taxonomic profiles, and increasing the sensitivity in deciphering the functional profiles of bovine reproductive tract metagenome samples. This study provides optimal metagenomic conditions for the application in future bovine vaginal reproductive microbiome research.

## MATERIALS AND METHODS

**Sample collection and transportation.** Samples were collected from 30 Droughtmaster heifers from a cattle farm in Northern Queensland under UQ Animal Ethics Approval AE30009. An experienced veterinarian conducted cattle health assessments and sample collection. The heifers were between 12 and 18 months old during sample collection. Using transrectal ultrasound, the heifers were defined as not pregnant and without the presence of a corpus luteum, but with the presence of either small or large follicles. The heifers were defined as being acyclic, likely in the late prepubertal period. The heifers' average live weight, hip height, and body score were  $316.5 \pm 30.7$  kg,  $193 \pm 43$  cm, and  $2.8 \pm 0.25$  (scale 0 to 5), respectively. Vaginal samples from the heifers were collected using the Tricamper (DAF Queensland, Australia) sampling tool following the manufacturer's protocol. Briefly, the Tricamper was inserted into the cow's vagina with the leading edge in contact with the dorsal wall of the vagina. The Tricamper was moved back and forth in the vagina to collect the swab. Upon removing the Tricamper from the vagina, the other end of the Tricamper was blocked to prevent spillage. The swab sample was immediately preserved in a 10 mL tube preloaded with 5 mL phosphate-buffered saline (PBS) by excising the head of the Tricamper device. The samples were kept on ice during delivery and were processed within 6 h upon arrival to the laboratory. Each sample was first vortexed for 15 s and followed by an additional 15 s of vortex after the Tricamper head was removed from the tube. The vaginal mucus samples were then transferred into a new sterile tube.

**Spiked bacteria and mock bacterial community.** To monitor the recovery of the metagenomic profiles, five bacteria were used for spiking, *Pseudomonas aeruginosa* (ATCC 27853), *Staphylococcus aureus* (ATCC 29213), *Campylobacter fetus* subsp. *venerealis* (ATCC 19438), *Aliarcobacter* previously *Arcobacter cryaerophilus* (312), and *Enterococcus faecalis* (BR1200). *Aliarcobacter cryaerophilus* and *C. fetus* subsp. *venerealis* are pathogens associated with abortion and infertility in cattle (73, 74). *Enterococcus faecalis*, *P. aeruginosa*, and *S. aureus* are commonly isolated from bovine reproductive tract samples but are regarded as normal flora or environmental organisms (75–77). Additionally, these bacteria formed a range of characteristics in terms of genome sizes, Gram stain, GC content, and aerotolerance, which were hypothesized to affect the efficiencies of host depletion methods and DNA extraction methods differently. *Aliarcobacter cryaerophilus*, *E. faecalis*, *P. aeruginosa*, and *S. aureus* were cultured on the BBL *Brucella* agar (Becton, Dickinson) under aerobic conditions at 37°C for 24 h. *Campylobacter fetus* subsp. *venerealis* were cultured on the tryptone soya agar (TSA) supplemented with 5% defibrinated sheep blood (Oxoid) under microaerophilic conditions at 37°C for 48 h. Colonies of each bacterial isolate were resuspended in sterile PBS to reach an optical density measured at wavelength 600 nm ( $OD_{600}$ ) for yielding colony count of approximately  $1 \times 10^8$  CFU/mL.

**The mock bacterial community in heifer vaginal swab samples.** The 30 vaginal swab samples were combined and homogenized in a sterile bottle. Five milliliters of the homogenized vaginal swab samples were then redistributed into 28 new sterile tubes. The vaginal swab with no spiked bacteria was labeled as "vaginal swab". The "mock samples" were prepared by adding 0.1 mL of each bacterial suspension into 5 mL homogenized vaginal swab samples. "Positive controls" were prepared by adding 0.1 mL of each bacterial suspension to 5 mL of sterile water. All samples were prepared in duplicates.

The bacterial load was monitored by the plate count method. Serial dilutions were made from the bacterial suspensions, and 0.1 mL of each dilution was plated onto TSA supplemented with 5% defibrinated sheep blood. For *P. aeruginosa*, *S. aureus*, and *E. faecalis*, the number of colonies was counted after 24 h. For *A. cryaerophilus* and *C. fetus* subsp. *venerealis*, the number of colonies was counted after 48 h.

**Host depletion and extraction methods.** Two of each sample type, including vaginal swabs, mock samples, and positive-controls, were processed with one of the depletion methods, including "None," "NEBNext," "Soft-spin" and "PMA". After being treated with the depletion method, the samples were extracted with either "DNeasy" or "QIAamp".

**Slow and short centrifugation (Soft-spin).** Slow and short centrifugation (Soft-spin) was performed for respective samples before extraction. The samples were centrifuged at  $1000 \times g$  for 1 min at 4°C. Without disturbing the pellet, the supernatant was collected into a new tube for DNA extraction.

**NEBNext Microbiome DNA Enrichment kit (NEBNext).** NEBNext is a postextraction host depletion method. Extracted DNA was processed with the NEBNext Microbiome DNA Enrichment kit (NEBNext) according to the manufacturer's protocol. Briefly, one microgram of the extracted DNA was mixed with magnetic beads bound to the methyl-CpG binding domain (MBD) to capture the methylated host DNA. A magnetic rack was used to separate the captured host DNA. The supernatant, which contained the microbial DNA was precipitated and diluted in TE buffer (10 mM Tris/0.1 mM EDTA, pH 8).

**Osmosis and propidium monoazide treatment (PMA).** PMA is a preextraction host depletion method. Samples were processed according to a published protocol (38). Briefly, vaginal samples were centrifuged at  $4000 \times g$  for 10 min to obtain the cell pellet. The pellet was resuspended in 0.2 mL of sterile water by pipetting and brief vortexing. The suspension was left at room temperature for 5 min to allow mammalian cell lysis by osmosis. A 10  $\mu$ L aliquot of 0.2 mM PMA dye (Biotium) solution was added to each of the samples. The mixtures were briefly vortexed to ensure thorough mixing and were then

**TABLE 3** List of primers used in quantitative PCR for bacterial and cattle DNA quantification

Target gene	Primer ID	Primer sequence	Amplicon size	Annealing temp	Reference
B2M	B2M-F	ACCTGAACCTGCTATGTGTATGG	134 bp	58°C	(79)
	BRM-R	GTGGGACAGCAGGTAGAAAAG			
16S rRNA	764-F	CAAACAGGATTAGATACCC	143 bp	54°C	(110, 111)

incubated in the dark for 5 min at room temperature. Subsequently, the samples were placed on ice within 25 cm of a light source for 25 min to allow light activation of the PMA molecule. The samples were briefly vortexed at 5-min intervals with light exposure.

**Qiagen DNeasy blood and tissue kit (DNeasy).** The sample was centrifuged at  $4000 \times g$  for 15 min to obtain a cell pellet. The supernatant was discarded, and DNA was extracted from the cell pellet according to the manufacturer's instructions for Gram-positive bacteria. Briefly, the pellet was treated with enzymatic lysis buffer and proteinase K digestion before column precipitation. Precipitated DNA was washed and eluted in 60  $\mu$ L of TE buffer.

**Qiagen QIAamp DNA Microbiome kit (QIAamp).** DNA was extracted according to the manufacturer's instructions. First, the sample was centrifuged at  $4000 \times g$  for 15 min to obtain the cell pellet. Briefly, the pellet was subjected to lysis buffer, benzonase enzyme, bead beating, and pathogen lysis buffer before column precipitation. The precipitated DNA was eluted in 60  $\mu$ L of TE buffer.

**Quantitative PCR (qPCR) assay and cycling conditions.** To quantify the bacterial copy number, qPCR was performed on the extracted DNA with PowerUp SYBR Green Master Mix (Applied Biosystems) using 16S rRNA gene-specific primers, 764F and 907R, for bacteria (78). Primers encompassing the beta2-microglobulin (B2M) gene were utilized for estimating the cattle copy number (79) (Table 3). Thermal cycling was initiated with the activation of Uracil-DNA Glycosylase at 50°C for 2 min followed by incubation of the DNA polymerase at 95°C for 2 min and 40 cycles of denaturing and annealing steps, which were 95°C for 15 s and primer-pair specific annealing temperature for 15 s, respectively. The qPCR was finished with an extension at 72°C for 1 min. Serial dilutions of cattle blood DNA and cultured bacterial samples were amplified alongside to generate the standard curves. All qPCR assays, including the nontemplate controls, were performed in duplicates. Results were generated using a CFX96 real-time PCR detection system (Bio-Rad) and data were analyzed using CFX Maestro software (Bio-Rad) with the Cq thresholds determined by the software.

The amount of amplicon was calculated from the standard curves generated with known concentrations of 16S rRNA genes and cattle blood DNA. The estimated copy number of cattle and bacteria was calculated as previously described (80). Briefly, the amount of amplicon (SQ) was calculated from the standard curves generated with known concentrations of bacterial DNA and cattle blood DNA. The estimated copy number of cattle and bacteria were derived from the ratio of SQ in gram multiplied by the Avogadro's number  $N_A$  ( $6.0221 \times 10^{23} \text{ mol}^{-1}$ ) and length of amplicon multiplied by the mean molar mass of a base pair  $M_{bp}$  ( $660 \text{ g mol}^{-1}$ ). The estimated copy number of cattle and bacteria were determined to compute the percentage of 16S rRNA genes.

**Shotgun sequencing, quality filtering, and contamination read removal.** The quantity of extracted DNA was measured using Qubit 4 fluorometer (Invitrogen) with either Qubit dsDNA Broad Range assay kit (Invitrogen) or Qubit 1x dsDNA High Sensitivity kit (Invitrogen). Extracted DNA was sent for shotgun sequencing to the Australian Centre of Ecogenomics (ACE, University of Queensland). Sequencing was performed on the NextSeq500 platform using NextSeq 500/550 High Output v2 2x150bp paired-end chemistry with 5 Gbp sequencing coverage to yield at least 50 million paired-end reads per metagenome sample.

The quality of the reads was assessed using FastQC 0.11.4 (81). The paired-end reads were trimmed with Trimmomatic 0.39.1 using the paired-end mode to remove the Nextera adapters, leading and trailing N bases, leading and trailing bases below quality score 15, bases with an average quality score below 15 in every 4-base wide sliding window and reads below 35 bases in length (82). The quality paired-end reads were mapped to the ARS-UCD1.2 *Bos taurus* genome (GCA\_002263795.2) (39) using Bowtie2 2.3.4.3 (83) to determine the number of cattle reads in each sample. SAMtools 1.3 (84) were applied to retrieve the alignments and BEDTools 2.25.0 (85) were used to extract paired-end reads of noncattle sequences. The microbial read percentage was calculated by getting the percentage of noncattle sequences in the total reads generated. The statistical details of the alignments were retrieved from the Binary Alignment Mapping (BAM) output files using SAMtools (84). The same procedure was used to remove human-read contamination by replacing the ARS-UCD1.2 *Bos taurus* genome with the GRCh38 *Homo sapiens* genome (GCA\_000001405.15). The number of microbial and bovine reads was determined from the BAM output to compute the percentage of microbial reads.

**Read-based taxonomic classifications.** After quality was filtered and cleaned, the metagenomic reads were annotated taxonomically with Kraken2 (86). Kraken2 examined the *k*-mers information in the metagenomic reads and query the *k*-mers information against the database. The abundances, at the species level, of the organisms identified in the metagenomic samples, were computed using the Bayesian Reestimation of Abundance with Kraken (Bracken) (87). Bioinformatic analysis and visualization were conducted on R studio (88) with R packages, including vegan 2.5.7 (89), phyloseq 1.34.0 (90), dplyr (91), and ggplot2 (92).

**Metagenomic assembly.** The quality paired-end reads of each sample were assembled into contigs using MEGAHIT 1.2.9 (93, 94), which incorporated *mercy-kmers* function to recover species with low abundances and were sequenced in low depth in a metagenome sample. The minimum contig length was limited to 200 bp. Additionally, the paired-end reads from all the samples were pooled to perform a coassembly using MEGAHIT 1.2.9. The reads were aligned back to the contigs using BMap 38.84 (95) to determine the quality of the assemblies. The quality of the assembled contigs was also evaluated with MetaQUAST 5.0.2 (96). The coassembly was first analyzed using MetaQUAST against the SILVA 16S rRNA

database (97) to identify the overall genome content of the metagenome samples. The identified genomes were downloaded from the NCBI database (98) to serve as the reference metagenome when evaluating the independent assemblies on MetaQUAST.

**Contig-based taxonomic and functional classifications.** The assembled contigs were annotated using an integrated pipeline MetaErg 1.2.0 (99). MetaErg performed HMMs profile similarity searches against several databases, including Pfam-A (100), TIGRFAM (101), FOAM (102), metabolic-hmms (103), and *casgenes.hmm* (104). MetaErg also performed DIAMOND (double index alignment of next-generation sequencing data) searches against Swiss-Prot (105) and the MetaErg in-built database GenomeDB. Mapping files generated from searches against Swiss-Prot, FOAM, and TIGRFAMs databases were incorporated in MinPath (106) to infer to KEGG (107) and MetaCyc (108) metabolic pathways. To weigh the relative abundances of the taxonomic, functional, and pathway compositions of the metagenome samples, individual BAM files were generated using BMap (95) by aligning the metagenome reads of each sample to the coassembled contigs. A depth file was constructed with the *jgi\_summarize\_bam\_contig\_depths* script in MetaBat 2 (109) using the BAM files.

**Data availability.** The datasets generated during the current study are available in the NCBI sequence read archive (SRA) database under BioProject [PRJNA786360](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA786360) and BioSamples [SRR17146641](https://www.ncbi.nlm.nih.gov/biosamples/SRR17146641) to [SRR17146676](https://www.ncbi.nlm.nih.gov/biosamples/SRR17146676).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 0.3 MB.

## ACKNOWLEDGMENTS

We gratefully acknowledge the support provided by Meat and Livestock Australia Donor Company project P.PSH.799.

We also thank Geoffry Fordyce from Queensland Alliance for Agricultural and Food Innovation (QAAFI) at The University of Queensland for conducting health assessments for the Droughmaster heifers and the sample collection to enable these experiments.

We declare no conflict of interest.

## REFERENCES

- Marchesi JR, Ravel J. 2015. The vocabulary of microbiome research: a proposal. *Microbiome* 3:31. <https://doi.org/10.1186/s40168-015-0094-5>.
- Wilkins LGE, Ettinger CL, Jospin G, Eisen JA. 2019. Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. *Sci Rep* 9:3059. <https://doi.org/10.1038/s41598-019-39576-6>.
- Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, Lawley TD, Finn RD. 2019. A new genomic blueprint of the human gut microbiota. *Nature* 568:499–504. <https://doi.org/10.1038/s41586-019-0965-1>.
- Forbes JD, Knox NC, Ronholm J, Pagotto F, Reimer A. 2017. Metagenomics: the next culture-independent game changer. *Front Microbiol* 8:1069–1069. <https://doi.org/10.3389/fmicb.2017.01069>.
- Hilton SK, Castro-Nallar E, Pérez-Losada M, Toma I, McCaffrey TA, Hoffman EP, Siegel MO, Simon GL, Johnson WE, Crandall KA. 2016. Metatransomic and metagenomic approaches vs. culture-based techniques for clinical pathology. *Front Microbiol* 7:484–484. <https://doi.org/10.3389/fmicb.2016.00484>.
- Mueller ER, Wolfe AJ, Brubaker L. 2017. Female urinary microbiota. *Curr Opin Urol* 27:282–286. <https://doi.org/10.1097/MOU.0000000000000396>.
- Stower H. 2019. The breast milk microbiome. *Nat Med* 25:541–541. <https://doi.org/10.1038/s41591-019-0427-1>.
- Guzman CE, Wood JL, Egidi E, White-Monsant AC, Semencic L, Grommen SVH, Hill-Yardin EL, De Groef B, Franks AE. 2020. A pioneer calf foetus microbiome. *Sci Rep* 10:17712. <https://doi.org/10.1038/s41598-020-74677-7>.
- Li F, Chen C, Wei W, Wang Z, Dai J, Hao L, Song L, Zhang X, Zeng L, Du H, Tang H, Liu N, Yang H, Wang J, Madsen L, Brix S, Kristiansen K, Xu X, Li J, Wu R, Jia H. 2018. The metagenome of the female upper reproductive tract. *GigaScience* 7:1–8. <https://doi.org/10.1093/gigascience/giy107>.
- Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UME, Zhong X, Koenig SSK, Fu L, Ma ZS, Zhou X, Abdo Z, Forney LJ, Ravel J. 2012. Temporal dynamics of the human vaginal microbiota. *Sci Transl Med* 4:132ra52–132ra52.
- Moore SG, Ericsson AC, Poock SE, Melendez P, Lucy MC. 2017. Hot topic: 16S rRNA gene sequencing reveals the microbiome of the virgin and pregnant bovine uterus. *J Dairy Sci* 100:4953–4960. <https://doi.org/10.3168/jds.2017-12592>.
- McCann JC, Wickersham TA, Loor JJ. 2014. High-throughput methods redefine the rumen microbiome and its relationship with nutrition and metabolism. *Bioinform Biol Insights* 8:109–125. <https://doi.org/10.4137/BBI.S15389>.
- Loor JJ, Elolimy AA, McCann JC. 2016. Dietary impacts on rumen microbiota in beef and dairy production. *Anim Front* 6:22–29. <https://doi.org/10.2527/af.2016-0030>.
- Rodrigues NF, Kastle J, Coutinho TJ, Amorim AT, Campos GB, Santos VM, Marques LM, Timenetsky J, de Farias ST. 2015. Qualitative analysis of the vaginal microbiota of healthy cattle and cattle with genital-tract disease. *Genet Mol Res* 14:6518–6528. <https://doi.org/10.4238/2015.June.12.4>.
- Wang Y, Wang J, Li H, Fu K, Pang B, Yang Y, Liu Y, Tian W, Cao R. 2018. Characterization of the cervical bacterial community in dairy cows with metritis and during different physiological phases. *Theriogenology* 108:306–313. <https://doi.org/10.1016/j.theriogenology.2017.12.028>.
- Fujimura KE, Slusher NA, Cabana MD, Lynch SV. 2010. Role of the gut microbiota in defining human health. *Expert Rev Anti Infect Ther* 8:435–454. <https://doi.org/10.1586/eri.10.14>.
- Magruder M, Sholi AN, Gong C, Zhang L, Edusei E, Huang J, Albakry S, Satlin MJ, Westblade LF, Crawford C, Dadhania DM, Lubetzky M, Taur Y, Littman E, Ling L, Burnham P, De Vlaminck I, Pamer E, Suthanthiran M, Lee JR. 2019. Gut uropathogen abundance is a risk factor for development of bacteriuria and urinary tract infection. *Nat Commun* 10:5521. <https://doi.org/10.1038/s41467-019-13467-w>.
- Jeon SJ, Vieira-Neto A, Gobikrushanth M, Daetz R, Mingoti RD, Parize ACB, de Freitas SL, da Costa ANL, Bicalho RC, Lima S, Jeong KC, Galvão KN. 2015. Uterine microbiota progression from calving until establishment of metritis in dairy cows. *Appl Environ Microbiol* 81:6324–6332. <https://doi.org/10.1128/AEM.01753-15>.
- Handelsman J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68:669–685. <https://doi.org/10.1128/MMBR.68.4.669-685.2004>.
- Tessler M, Neumann JS, Afshinnekoo E, Pineda M, Hersch R, Velho LFM, Segovia BT, Lansac-Toha FA, Lemke M, DeSalle R, Mason CE, Brugler MR. 2017. Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Sci Rep* 7:6589. <https://doi.org/10.1038/s41598-017-06665-3>.
- Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 35:833–844. <https://doi.org/10.1038/nbt.3935>.

22. Quereda JJ, Barba M, Mocé ML, Gomis J, Jiménez-Trigos E, García-Muñoz Á, Gómez-Martín Á, González-Torres P, Carbonetto B, García-Roselló E. 2020. Vaginal microbiota changes during estrous cycle in dairy heifers. *Front Vet Sci* 7:371–371. <https://doi.org/10.3389/fvets.2020.00371>.
23. Chen SY, Deng F, Zhang M, Jia X, Lai SJ. 2020. Characterization of vaginal microbiota associated with pregnancy outcomes of artificial insemination in dairy cows. *J Microbiol Biotechnol* 30:804–810. <https://doi.org/10.4014/jmb.2002.02010>.
24. Moore SG, Ericsson AC, Behura SK, Lamberson WR, Evans TJ, McCabe MS, Poock SE, Lucy MC. 2019. Concurrent and long-term associations between the endometrial microbiota and endometrial transcriptome in postpartum dairy cows. *BMC Genomics* 20:405. <https://doi.org/10.1186/s12864-019-5797-8>.
25. Santos TMA, Gilbert RO, Bicalho RC. 2011. Metagenomic analysis of the uterine bacterial microbiota in healthy and metritic postpartum dairy cows. *J Dairy Sci* 94:291–302. <https://doi.org/10.3168/jds.2010-3668>.
26. Noakes DE, Parkinson TJ, England GCW, Arthur GH. 2001. Chapter 22 - Infertility in the cow: structural and functional abnormalities, management deficiencies and non-specific infections, p 383–472. *In* Noakes DE, Parkinson TJ, England GCW, Arthur GH (ed), *Arthur's Veterinary Reproduction and Obstetrics* (Eighth Edition). W.B. Saunders, Philadelphia, PA, USA.
27. Sheldon IM, Noakes DE, Rycroft AN, Pfeiffer DU, Dobson H. 2002. Influence of uterine bacterial contamination after parturition on ovarian dominant follicle selection and follicle growth and function in cattle. *Reproduction* 123:837–845. <https://doi.org/10.1530/rep.0.1230837>.
28. Bondurant RH. 1999. Inflammation in the bovine female reproductive tract. *J Anim Sci* 77 Suppl 2:101–110. [https://doi.org/10.2527/1999.77suppl\\_2101x](https://doi.org/10.2527/1999.77suppl_2101x).
29. Pascottini OB, Van Schyndel SJ, Spricigo JFW, Rousseau J, Weese JS, LeBlanc SJ. 2020. Dynamics of uterine microbiota in postpartum dairy cows with clinical or subclinical endometritis. *Sci Rep* 10:12353. <https://doi.org/10.1038/s41598-020-69317-z>.
30. Knudsen LRV, Karstrup CC, Pedersen HG, Angen Ø, Agerholm JS, Rasmussen EL, Jensen TK, Klitgaard K. 2016. An investigation of the microbiota in uterine flush samples and endometrial biopsies from dairy cows during the first 7 weeks postpartum. *Theriogenology* 86:642–650. <https://doi.org/10.1016/j.theriogenology.2016.02.016>.
31. Feigelman R, Kahlert CR, Baty F, Rassouli F, Kleiner RL, Kohler P, Brutsche MH, von Mering C. 2017. Sputum DNA sequencing in cystic fibrosis: non-invasive access to the lung microbiome and to pathogen details. *Microbiome* 5:20. <https://doi.org/10.1186/s40168-017-0234-1>.
32. Bicalho MLS, Machado VS, Higgins CH, Lima FS, Bicalho RC. 2017. Genetic and functional analysis of the bovine uterine microbiota. part I: metritis versus healthy cows. *J Dairy Sci* 100:3850–3862. <https://doi.org/10.3168/jds.2016-12058>.
33. Bicalho MLS, Lima S, Higgins CH, Machado VS, Lima FS, Bicalho RC. 2017. Genetic and functional analysis of the bovine uterine microbiota. Part II: purulent vaginal discharge versus healthy cows. *J Dairy Sci* 100:3863–3874. <https://doi.org/10.3168/jds.2016-12061>.
34. Robbins SJ, Singleton CM, Chan CX, Messer LF, Geers AU, Ying H, Baker A, Bell SC, Morrow KM, Ragan MA, Miller DJ, Forêt S, Ball E, Beeden R, Berumen M, Aranda M, Ravasi T, Bongaerts P, Hoegh-Guldberg O, Cooke I, Leggat B, Sprungala S, Fitzgerald A, Shang C, Lundgren P, Yfffe T, Rubino F, van Oppen M, Weynberg K, Robbins SJ, Singleton CM, Xin Chan C, Messer LF, Geers AU, Ying H, Baker A, Bell SC, Morrow KM, Ragan MA, Miller DJ, Foret S, Voolstra CR, Tyson GW, Bourne DG, Voolstra CR, Tyson GW, Bourne DG, ReFuGe C, ReFuGe2020 Consortium. 2019. A genomic view of the reef-building coral *Porites lutea* and its microbial symbionts. *Nat Microbiol* 4:2090–2100. <https://doi.org/10.1038/s41564-019-0532-4>.
35. Echternkamp SE, Hansel W. 1973. Concurrent changes in bovine plasma hormone levels prior to and during the first postpartum estrous cycle. *J Anim Sci* 37:1362–1370. <https://doi.org/10.2527/jas1973.3761362x>.
36. Paiano RB, Gonçalves CGP, Mendes JPG, Bonilla J, Birgel DB, Birgel Junior EH. 2019. Comparative biochemical profiles, production and reproduction status of the post-partum dairy cows with and without purulent vaginal discharge. *Reprod Dom Anim* 54:1188–1194. <https://doi.org/10.1111/rda.13496>.
37. Feehery GR, Yigit E, Oyola SO, Langhorst BW, Schmidt VT, Stewart FJ, Dimalanta ET, Amaral-Zettler LA, Davis T, Quail MA, Pradhan S. 2013. A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One* 8:e76096. <https://doi.org/10.1371/journal.pone.0076096>.
38. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. 2018. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* 6:42. <https://doi.org/10.1186/s40168-018-0426-3>.
39. Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elisk CG, Tseng E, Rowan TN, Low WY, Zimin A, Coudrey C, Hall R, Li W, Rhie A, Ghurye J, McKay SD, Thibaud-Nissen F, Hoffman J, Murdoch BM, Snelling WM, McDanel TG, Hammond JA, Schwartz JC, Nandolo W, Hagen DE, Dreischer C, Schultheiss SJ, Schroeder SG, Phillippy AM, Cole JB, Van Tassel CP, Liu G, Smith TPL, Medrano JF. 2020. *De novo* assembly of the cattle reference genome with single-molecule sequencing. *GigaScience* 9:giaa021. <https://doi.org/10.1093/gigascience/giaa021>.
40. diCenzo GC, Finan TM. 2017. The divided bacterial genome: structure, function, and evolution. *Microbiol Mol Biol Rev* 81:e00019-17. <https://doi.org/10.1128/MMBR.00019-17>.
41. Ong CT, Turni C, Blackall PJ, Boe-Hansen G, Hayes BJ, Tabor AE. 2021. Interrogating the bovine reproductive tract metagenomes using culture-independent approaches: a systematic review. *Anim Microbiome* 3:41. <https://doi.org/10.1186/s42523-021-00106-3>.
42. Horz H-P, Scheer S, Vianna ME, Conrads G. 2010. New methods for selective isolation of bacterial DNA from human clinical specimens. *Anaerobe* 16:47–53. <https://doi.org/10.1016/j.anaerobe.2009.04.009>.
43. Zhou L, Pollard AJ. 2012. A novel method of selective removal of human DNA improves PCR sensitivity for detection of *Salmonella typhi* in blood samples. *BMC Infect Dis* 12:164. <https://doi.org/10.1186/1471-2334-12-164>.
44. Lim YW, Haynes M, Furlan M, Robertson CE, Harris JK, Rohwer F. 2014. Purifying the impure: sequencing metagenomes and metatranscriptomes from complex animal-associated samples. *J Vis Exp* 94:e52117. <https://doi.org/10.3791/52117>.
45. Hasan MR, Rawat A, Tang P, Jithesh PV, Thomas E, Tan R, Tilley P. 2016. Depletion of human DNA in spiked clinical specimens for improvement of sensitivity of pathogen detection by next-generation sequencing. *J Clin Microbiol* 54:919–927. <https://doi.org/10.1128/JCM.03050-15>.
46. Nelson MT, Pope CE, Marsh RL, Wolter DJ, Weiss EJ, Hager KR, Vo AT, Brittnacher MJ, Radey MC, Hayden HS, Eng A, Miller SI, Borenstein E, Hoffman LR. 2019. Human and extracellular DNA depletion for metagenomic analysis of complex clinical infection samples yields optimized viable microbiome profiles. *Cell Rep* 26:2227–2240. <https://doi.org/10.1016/j.celrep.2019.01.091>.
47. Pereira-Marques J, Hout A, Ferreira RM, Weber M, Pinto-Ribeiro I, van Doorn L-J, Knecht CW, Figueiredo C. 2019. Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. *Front Microbiol* 10. <https://doi.org/10.3389/fmicb.2019.01277>.
48. Ganda E, Beck KL, Haiminen N, Kawas B, Cronk B, Anderson RR, Goodman LB, Wiedmann M. 2020. DNA extraction and host depletion methods significantly impact and potentially bias bacterial detection in a biological fluid. *bioRxiv*. <https://www.doi.org/10.1101/2020.08.21.262337>.
49. Wood JM. 2015. Bacterial responses to osmotic challenges. *J Gen Physiol* 145:381–388. <https://doi.org/10.1085/jgp.201411296>.
50. Casals-Pascual C, González A, Vázquez-Baeza Y, Song SJ, Jiang L, Knight R. 2020. Microbial diversity in clinical microbiome studies: sample size and statistical power considerations. *Gastroenterology* 158:1524–1528. <https://doi.org/10.1053/j.gastro.2019.11.305>.
51. Mitra S. 2019. Multiple data analyses and statistical approaches for analyzing data from metagenomic studies and clinical trials, p 605–634. *In* Anisimova M (ed), *Evolutionary Genomics: Statistical and Computational Methods*. Springer, New York, NY, USA.
52. Ault TB, Clemmons BA, Reese ST, Dantas FG, Franco GA, Smith TPL, Edwards JL, Myer PR, Pohler KG. 2019. Bacterial taxonomic composition of the postpartum cow uterus and vagina prior to artificial insemination. *J Anim Sci* 97:4305–4313. <https://doi.org/10.1093/jas/skz212>.
53. Ault TB, Clemmons BA, Reese ST, Dantas FG, Franco GA, Smith TPL, Edwards JL, Myer PR, Pohler KG. 2019. Uterine and vaginal bacterial community diversity prior to artificial insemination between pregnant and nonpregnant postpartum cows. *J Anim Sci* 97:4298–4304. <https://doi.org/10.1093/jas/skz210>.
54. Wang Y, Ametaj BN, Ambrose DJ, Gänzle MG. 2013. Characterisation of the bacterial microbiota of the vagina of dairy cows and isolation of pediocin-producing *Pediococcus acidilactici*. *BMC Microbiol* 13:19. <https://doi.org/10.1186/1471-2180-13-19>.
55. Nguyen TT, Miyake A, Tran TTM, Tsuruta T, Nishino N. 2019. The relationship between uterine, fecal, bedding, and airborne dust microbiota from dairy cows and their environment: a pilot study. *Animals* 9:1007. <https://doi.org/10.3390/ani9121007>.
56. Laguardia-Nascimento M, Branco KMGR, Gasparini MR, Giannattasio-Ferraz S, Leite LR, Araujo FMG, Salim ACdM, Nicoli JR, de Oliveira GC, Barbosa-Stancioli EF. 2015. Vaginal microbiome characterization of

- Nellore cattle using metagenomic analysis. *PLoS One* 10:e0143294. <https://doi.org/10.1371/journal.pone.0143294>.
57. Clemmons BA, Reese ST, Dantas FG, Franco GA, Smith TPL, Adeyosoye OI, Pohler KG, Myer PR. 2017. Vaginal and uterine bacterial communities in postpartum lactating cows. *Front Microbiol* 8:1047. <https://doi.org/10.3389/fmicb.2017.01047>.
  58. Zhu Y, Li M, Chen W, Peters A. 2013. The smart solution for DNA removal in biopharmaceutical production by benzonase endonuclease. *J Appl Virol* 2:25–33.
  59. Moreno JM, Sanchez-Montero JM, Sinisterra JV, Nielsen LB. 1991. Contribution to the study of the enzymatic activity of benzonase. *J Mol Catal* 69:419–427. [https://doi.org/10.1016/0304-5102\(91\)80120-R](https://doi.org/10.1016/0304-5102(91)80120-R).
  60. Silhavy TJ, Kahne D, Walker S. 2010. The bacterial cell envelope. *Cold Spring Harb Perspect Biol* 2:a000414. <https://doi.org/10.1101/cshperspect.a000414>.
  61. Jordan S, Hutchings MI, Mascher T. 2008. Cell envelope stress response in Gram-positive bacteria. *FEMS Microbiol Rev* 32:107–146. <https://doi.org/10.1111/j.1574-6976.2007.00091.x>.
  62. de Boer R, Peters R, Gierveld S, Schuurman T, Kooistra-Smid M, Savelkoul P. 2010. Improved detection of microbial DNA after bead-beating before DNA isolation. *J Microbiol Methods* 80:209–211. <https://doi.org/10.1016/j.mimet.2009.11.009>.
  63. Ketchum RN, Smith EG, Vaughan GO, Phippen BL, McParland D, Al-Mansoori N, Carrier TJ, Burt JA, Reitzel AM. 2018. DNA extraction method plays a significant role when defining bacterial community composition in the marine invertebrate *Echinometra mathaei*. *Front Mar Sci* 5:255. <https://doi.org/10.3389/fmars.2018.00255>.
  64. Heravi FS, Zakrzewski M, Vickery K, Hu H. 2020. Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. *J Microbiol Methods* 170:105856. <https://doi.org/10.1016/j.mimet.2020.105856>.
  65. Wen Y, Xiao F, Wang C, Wang Z. 2016. The impact of different methods of DNA extraction on microbial community measures of BALF samples based on metagenomic data. *Am J Transl Res* 8:1412–1425.
  66. Gonzalez Moreno C, Fontana C, Cocconcelli PS, Callegari ML, Otero MC. 2016. Vaginal microbial communities from synchronized heifers and cows with reproductive disorders. *J Appl Microbiol* 121:1232–1241. <https://doi.org/10.1111/jam.13239>.
  67. Messman RD, Contreras-Correa ZE, Paz HA, Perry G, Lemley CO. 2020. Vaginal bacterial community composition and concentrations of estradiol at the time of artificial insemination in Brangus heifers. *J Anim Sci* 98:skaa178. <https://doi.org/10.1093/jas/skaa178>.
  68. Miranda-CasoLuengo R, Lu J, Williams EJ, Miranda-CasoLuengo AA, Carrington SD, Evans ACO, Meijer WG. 2019. Delayed differentiation of vaginal and uterine microbiomes in dairy cows developing postpartum endometritis. *PLoS One* 14:e0200974. <https://doi.org/10.1371/journal.pone.0200974>.
  69. Wagener K, Grunert T, Prunner I, Ehling-Schulz M, Drillich M. 2014. Dynamics of uterine infections with *Escherichia coli*, *Streptococcus uberis* and *Trueperella pyogenes* in post-partum dairy cows and their association with clinical endometritis. *Vet J* 202:527–532. <https://doi.org/10.1016/j.tvjl.2014.08.023>.
  70. Jeon SJ, Galvão KN. 2018. An advanced understanding of uterine microbial ecology associated with metritis in dairy cows. *Genomics Inform* 16:e21. <https://doi.org/10.5808/GI.2018.16.4.e21>.
  71. Franzosa EA, Hsu T, Sirota-Madi A, Shafquat A, Abu-Ali G, Morgan XC, Huttenhower C. 2015. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat Rev Microbiol* 13:360–372. <https://doi.org/10.1038/nrmicro3451>.
  72. Rath S, Rud T, Karch A, Pieper DH, Vital M. 2018. Pathogenic functions of host microbiota. *Microbiome* 6:174. <https://doi.org/10.1186/s40168-018-0542-0>.
  73. Di Blasio A, Traversa A, Giacometti F, Chiesa F, Piva S, Decastelli L, Dondo A, Gallina S, Zoppi S. 2019. Isolation of *Arcobacter* species and other neglected opportunistic agents from aborted bovine and caprine fetuses. *BMC Vet Res* 15:257. <https://doi.org/10.1186/s12917-019-2009-3>.
  74. Veron M, Chatelain R. 1973. Taxonomic study of the genus *Campylobacter* Sebald and Veron and designation of the neotype strain for the type species. *Campylobacter fetus* (Smith and Taylor) Sebald and Veron. *Int J Syst Evol Microbiol* 23:122–134. <https://doi.org/10.1099/00207713-23-2-122>.
  75. Mushin R, Ziv G. 1973. An epidemiological study of *Pseudomonas aeruginosa* in cattle and other animals by pyocine typing. *J Hyg (Lond)* 71:113–122. <https://doi.org/10.1017/s0022172400046271>.
  76. Roberson JR, Fox LK, Hancock DD, Gay JM, Besser TE. 1994. Ecology of *Staphylococcus aureus* isolated from various sites on dairy farms. *J Dairy Sci* 77:3354–3364. [https://doi.org/10.3168/jds.S0022-0302\(94\)77277-5](https://doi.org/10.3168/jds.S0022-0302(94)77277-5).
  77. Chingwaru W, Mpuchane SF, Gashe BA. 2003. *Enterococcus faecalis* and *Enterococcus faecium* isolates from milk, beef, and chicken and their antibiotic resistance. *J Food Prot* 66:931–936. <https://doi.org/10.4315/0362-028x-66.6.931>.
  78. Hiergeist A, Reischl U, Gessner A, Priority Program 1656 Intestinal Microbiota Consortium/quality assessment participants. 2016. Multicenter quality assessment of 16S ribosomal DNA-sequencing for microbiome analyses reveals high inter-center variability. *Int J Med Microbiol* 306:334–342. <https://doi.org/10.1016/j.ijmm.2016.03.005>.
  79. Weller M, Fortes MRS, Marcondes MI, Rotta PP, Gionbeli TRS, Valadares Filho SC, Campos MM, Silva FF, Silva W, Moore S, Guimarães SEF. 2016. Effect of maternal nutrition and days of gestation on pituitary gland and gonadal gene expression in cattle. *J Dairy Sci* 99:3056–3071. <https://doi.org/10.3168/jds.2015-9673>.
  80. Ritalahti KM, Amos BK, Sung Y, Wu Q, Koenigsberg SS, Löffler FE. 2006. Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple *Dehalococcoides* strains. *Appl Environ Microbiol* 72:2765–2774. <https://doi.org/10.1128/AEM.72.4.2765-2774.2006>.
  81. Andrew S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
  82. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
  83. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
  84. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
  85. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
  86. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>.
  87. Lu J, Breitwieser FP, Thielen P, Salzberg SL. 2017. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput Sci* 3:e104. <https://doi.org/10.7717/peerj-cs.104>.
  88. RStudio Team. 2020. RStudio: Integrated development environment for R. RStudio, PBC, Boston, MA. <http://www.rstudio.com/>.
  89. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens HH, Szoecs E, Wagner H. 2020. vegan: community ecology package. R package version 2.5–7. <https://CRAN.R-project.org/package=vegan>.
  90. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>.
  91. Hadley WR, L Henry K, Müller 2020. dplyr: A grammar of data manipulation. <https://CRAN.R-project.org/package=dplyr>.
  92. Wickham H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York USA. <https://ggplot2.tidyverse.org>.
  93. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
  94. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW. 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102:3–11. <https://doi.org/10.1016/j.ymeth.2016.02.020>.
  95. Bushnell B. 2020. BBMap short-read aligner, and other bioinformatics tools, v38.84.
  96. Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* 32:1088–1090. <https://doi.org/10.1093/bioinformatics/btv697>.
  97. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>.
  98. Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 33:D501–D504.
  99. Dong X, Strous M. 2019. An integrated pipeline for annotation and visualization of metagenomic contigs. *Front Genet* 10:999. <https://doi.org/10.3389/fgene.2019.00999>.
  100. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44:D279–D285. <https://doi.org/10.1093/nar/gkv1344>.

101. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. 2013. TIGRFAMs and genome properties in 2013. *Nucleic Acids Res* 41:D387–D395. <https://doi.org/10.1093/nar/gks1234>.
102. Prestat E, David MM, Hultman J, Taş N, Lamendella R, Dvornik J, Mackelprang R, Myrold DD, Jumpponen A, Tringe SG, Holman E, Mavromatis K, Jansson JK. 2014. FOAM (functional ontology assignments for metagenomes): a hidden Markov model (HMM) database with environmental focus. *Nucleic Acids Res* 42:e145. <https://doi.org/10.1093/nar/gku702>.
103. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS, Banfield JF. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* 7:13219. <https://doi.org/10.1038/ncomms13219>.
104. Burstein D, Harrington LB, Strutt SC, Probst AJ, Anantharaman K, Thomas BC, Doudna JA, Banfield JF. 2017. New CRISPR–Cas systems from uncultivated microbes. *Nature* 542:237–241. <https://doi.org/10.1038/nature21059>.
105. Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* 28:45–48. <https://doi.org/10.1093/nar/28.1.45>.
106. Ye Y, Doak TG. 2009. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput Biol* 5:e1000465. <https://doi.org/10.1371/journal.pcbi.1000465>.
107. Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30. <https://doi.org/10.1093/nar/28.1.27>.
108. Karp PD, Riley M, Paley SM, Pellegrini-Toole A. 2002. The MetaCyc database. *Nucleic Acids Res* 30:59–61. <https://doi.org/10.1093/nar/30.1.59>.
109. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>.
110. Imase M, Watanabe K, Aoyagi H, Tanaka H. 2008. Construction of an artificial symbiotic community using a *Chlorella*-symbiont association as a model. *FEMS Microbiol Ecol* 63:273–282. <https://doi.org/10.1111/j.1574-6941.2007.00434.x>.
111. Lane D. 1991. 16S/23S rRNA sequencing, p 115–175. *In* Stackebrandt E, Goodfellow M. (ed), *Nucleic acid techniques in bacterial systematics*. John Wiley & Sons, New York, USA.