

# Similarity in milk microbiota in replicates

Josef Dahlberg<sup>1</sup>  | Erik Pelve<sup>2</sup> | Johan Dicksved<sup>3</sup>

<sup>1</sup>Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

<sup>2</sup>Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, Uppsala, Sweden

<sup>3</sup>Department of Animal Nutrition and Management, Swedish University of Agricultural Sciences, Uppsala, Sweden

## Correspondence

Josef Dahlberg, Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden.

Email: [josef.dahlberg@slu.se](mailto:josef.dahlberg@slu.se)

## Funding information

Swedish Research Council Formas, Grant/Award Numbers: 20121365-2272942, 221-2012-1098; SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science; Knut and Alice Wallenberg Foundation; Science for Life Laboratory

## Abstract

Receiving the same results from repeated analysis of the same sample is a basic principle in science. The inability to reproduce previously published results has led to discussions of a reproducibility crisis within science. For studies of microbial communities, the problem of reproducibility is more pronounced and has, in some fields, led to a discussion on the very existence of a constantly present microbiota. In this study, DNA from 44 bovine milk samples were extracted twice and the V3–V4 region of the 16S rRNA gene was sequenced in two separate runs. The FASTQ files from the two data sets were run through the same bioinformatics pipeline using the same settings and results from the two data sets were compared. Milk samples collected maximally 2 h apart were used as replicates and permitted comparisons to be made within the same run. Results show a significant difference in species richness between the two sequencing runs although Shannon and Simpson's diversity was the same. Multivariate analyses of all samples demonstrate that the sequencing run was a driver for variation. Direct comparison of similarity between samples and sequencing run showed an average similarity of 42%–45% depending on whether binary or abundance-based similarity indices were used. Within-run comparisons of milk samples collected maximally 2 h apart showed an average similarity of 39%–47% depending on the similarity index used and that similarity differed significantly between runs. We conclude that repeated DNA extraction and sequencing significantly can affect the results of a low microbial biomass microbiota study.

## KEYWORDS

microbiota, milk, mock community, repeatability, reproducibility

## 1 | BACKGROUND

The ability to get the same, or similar, results from repeated experiments is one of the pillars that science is founded on. Technical replicates, that is, repeated measurements from the same sample, are used to show variation or similarity between measuring equipment

and protocols, while biological replicates generally are defined as measurements of biological distinct samples that show biological variation.

In the past decade, sequencing of amplicons generated from the 16S rRNA gene has become the most commonly used technique to describe bacterial communities in various environments.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *MicrobiologyOpen* published by John Wiley & Sons Ltd.

A technical challenge with 16S amplicon-sequencing is that the technique is prone to the introduction of biases (see Pollock et al., 2018 or Nearing et al., 2021, for review), results can vary with the type of primers used (Clooney et al., 2016) and the bioinformatics can substantially affect the results (O'Sullivan et al., 2021). Laboratory and reagent contamination has been shown to largely affect the results in microbiota studies where the bacterial biomass is low (Dahlberg et al., 2019; de Goffau et al., 2019; Lauder et al., 2016; Salter et al., 2014). As a consequence, methods to identify and cure data from contamination have been developed (Alipour et al., 2018; Davis et al., 2018; Karstens et al., 2019; Łukasik et al., 2017). The ability to get the same results from sample replicates has been examined (Kennedy et al., 2014; Marotz et al., 2019; Schwenker et al., 2022; Wen et al., 2017) and it has been shown that samples with low bacterial biomass have a lower reproducibility (Kennedy et al., 2014). Together these factors contribute to the reported low repeatability and reproducibility in microbiota studies.

In this study, we describe the similarity in microbiota composition between replicates of the same milk samples (prepared with similar but not identical lab protocols) from two separate 16S rRNA gene sequencing runs, further referred to as type 1 replicates. A bacterial-based mock community was included as a control and compared in the same manner. We also compared milk samples collected from the same quarter and within a short time frame in the same sequencing run, referred to as type 2 replicates. The results from this study add

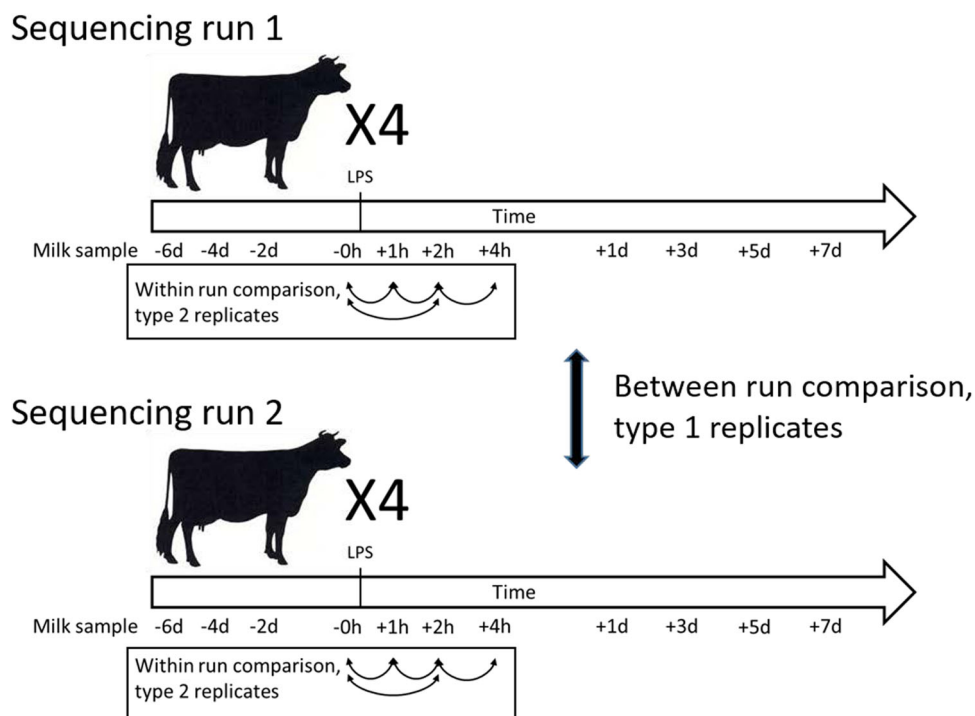
to the list of biases that can affect the results of studies of microbiota, in general, and milk in particular.

## 2 | METHODS

Milk samples came from an experiment that examined the effect of lipopolysaccharide (LPS) infusion into the mammary gland of healthy lactating dairy cows, previously described by Johnzon et al. (2018) and Dahlberg et al. (2023). In short, nine healthy lactating dairy cows were infused with LPS in one quarter and the inflammation processes were followed through repeated blood and milk sampling. Equally many cows were infused with sterile saline for comparison. Eleven milk samples were taken from each infused quarter during the experiment; from each sampling point, 5 mL of milk was aliquoted and stored until DNA extraction. Forty-four milk samples from four LPS-infused cows were subjected to DNA extraction and sequencing in two separate runs. A schematic illustration of the study layout is shown in Figure 1.

### 2.1 | DNA preparation and sequencing run 1

Sample processing for sequencing run 1 has been described by Dahlberg et al. (2023), for clarity and comparison between sequencing runs; 1 mL of room temperature, well-mixed milk was used for



**FIGURE 1** Schematic illustration of study design. Four lactating dairy cows were infused with *E. coli* lipopolysaccharide (LPS) in one udder quarter; milk samples were collected before and after infusion as indicated. Aliquots of milk samples were subjected to DNA extraction, sample processing, and sequencing in two separate runs. The similarity in microbiota was compared between runs (i.e., type 1 replicates) and for selected samples within the same run (i.e., type 2 replicates).

DNA extraction. After centrifugation (at 13,000g, 5 min), DNA was extracted from the cell pellet and fat layer using the DNeasy PowerFood Microbial kit (Qiagen; Lot no. 157017245); bead beating (Precellys24; Bertin Technologies;  $2 \times 45$  s at 6500 rpm with 1 min pause) was included for cell lysis. For each round of DNA extraction, an empty vial into which the first reagent was added was used as a no-template DNA extraction control (NTC) and processed as the milk samples. A bacterial mock community was included for method evaluation as previously described (Dahlberg et al., 2019). The mock community was created with equal numbers of cells and prepared in two different dilutions ( $10^7$  and  $10^5$  cells per mL of each bacterial species) and DNA was extracted and amplified in duplicates using the same protocol as the milk samples.

Extracted DNA from milk samples, NTCs, and the bacterial mock communities were subjected to polymerase chain reaction (PCR) amplification and used to prepare an Illumina MiSeq sequencing library in a two-step manner with primers previously described by Hugerth et al. (2014). In the first step, the V3–V4 region of the 16S rRNA gene was amplified, and in the second step, specific barcodes and Illumina adaptors were attached. One barcode combination per sample (i.e., milk, NTC, and bacterial-based mock community samples) was used. DreamTaq PCR master mix (2 $\times$ ) (Thermo Fisher Scientific) was used in both PCR reactions, in the first PCR reaction 35 cycles were run with 3  $\mu$ L of DNA as a template, and in the second 10 cycles and 5  $\mu$ L of DNA was used. PCR products from the first and second PCR reactions were purified with Ampure Beads (Beckman Coulter) before continuation, 0.9 and 0.7 volumes of beads per volume of PCR product in the first and second PCR, respectively. All DNA extraction and first PCR preparations were performed in a laminar air-flow hood cleaned with 10% bleach, 70% ethanol, and UV-irradiated for 30 min before execution. After the second PCR and cleaning, DNA was quantified with a Qubit 3.0 Fluorometer (Life Technologies) and thereafter pooled into equimolar amounts. The DNA pool was concentrated using Ampure Beads and elution in a smaller volume and thereafter cleaned through gel extraction (GeneJET, Gel Extraction Kit; Thermo Fisher Scientific) to ensure DNA strand length. Sequencing was performed on an Illumina MiSeq sequencer with v3 sequencing chemistry,  $2 \times 300$  bp with 10% PhiX (Illumina Inc.) at the Science for Life Laboratory.

## 2.2 | DNA preparation and sequencing run 2

For the second sequencing run, DNA was prepared as follows: 1 mL of milk was centrifuged at 13,000g for 5 min; the supernatant and the fat layer were removed and DNA was extracted from the cell pellet using the PowerFood Microbial DNA isolation kit (MO BIO Laboratories Inc., Lot no PF14F2 later renamed to Qiagen DNeasy PowerFood Microbial kit). DNA extraction was performed according to the manufacturer's instructions except that Precellys24 (Bertin Technologies,  $2 \times 45$  s at 6500 rpm with 1 min pause) was used for cell lysis. A blank sample with no starting material was included for each round of DNA extraction; these NTCs were PCR-amplified as the milk

samples and barcoded with one barcode combination. The same bacterial-based mock community as previously described was included in two different dilution levels ( $10^7$  and  $10^5$  cells per mL of each bacterial species). DNA extraction and first PCR preparations were performed in a laminar air-flow hood cleaned with 10% bleach, 70% ethanol, and UV-irradiated for 30 min before execution. PCR amplification and attachment of barcodes were performed same as for sequencing run 1 with the following exceptions: (a) Phusion high-fidelity polymerase (Life Technologies) was used, (b) 0.5 or 5  $\mu$ L of DNA was used as template in the first PCR, (c) 10  $\mu$ L of DNA was used as template in second PCR, (d) 0.8 volumes of Ampure beads were used for DNA cleaning, and (e) no gel extraction was performed. Thermocycling conditions were: initial denaturation at 98°C for 30 s, thereafter denaturation at 98°C for 10 s, annealing at 60°C for 30 s, and elongation at 72°C for 7 s, and final elongation at 72°C for 2 min. Thirty-five cycles were run in the first PCR creating amplicons and 10 cycles were run in the second PCR. Samples were pooled into equimolar amounts and sequenced on an Illumina MiSeq sequencer with v3 sequencing chemistry (Illumina Inc.) at the Science for Life Laboratory.

## 2.3 | Illumina sequence data analysis

Data from the two sequencing runs were processed in parallel with the same settings as previously described (Iversen et al., 2022). In short, the DADA2 pipeline was used to denoise, dereplicate reads, merge pair-end reads, and remove chimeras from the raw demultiplexed reads (Callahan et al., 2016). Amplicon sequence variants (ASVs) were assigned to reference sequences using the assign-Taxonomy command (Wang et al., 2007) against the 132 release of the SILVA rRNA database (Quast et al., 2013), formatted for DADA2 by B. Callahan (<https://benjjneb.github.io/dada2/training.html>, accessed on 19 October 2019). The Phyloseq R package was used to construct ASV frequency tables for subsequent statistical analysis (McMurdie & Holmes, 2013). The data analysis was performed and alpha diversity was calculated on relative abundance data at the genus level. Descriptive analysis on sequencing results, statistical calculations, and multivariate analyses (Principal Coordinates Analysis [PCoA's]) were performed using Microsoft Excel 2016, Paleontological Statistics program (PAST, ver 4.05) (Hammer et al., 2001) and R (ver 3.5.3) (R Core Team, 2019). *T*-test was used when two groups were compared and for statistical analysis, values are presented as mean  $\pm$  standard deviation. For statistical analysis of the PCoA, an analysis of similarity (ANOSIM) was used. Statistical significance was set at the level  $p < 0.05$ .

## 2.4 | Data analysis

The data set contained in total of 3.4 million reads, the number of reads per sample was higher in sequencing run 1 ( $56,930 \pm 17,496$ ) compared to run 2 ( $19,955 \pm 11,104$ ). Data from the 44 milk samples

included in sequencing run 1 and 2 were merged in R. The merged data set was used to analyze similarities and differences between the separate sequence runs. For type 1 replicates, the clustering of samples was analyzed by PCoA and ANOSIM using Bray Curtis and Dice similarity indexes. Two similarity matrices, based on the same mentioned similarity indexes, were created in PAST. The similarity of the type 1 replicates between sequencing runs was then measured directly from the similarity matrix. Replicates of the bacterial mock community from sequencing run 1 and 2 were also merged into the data set and compared for the direct similarity between runs.

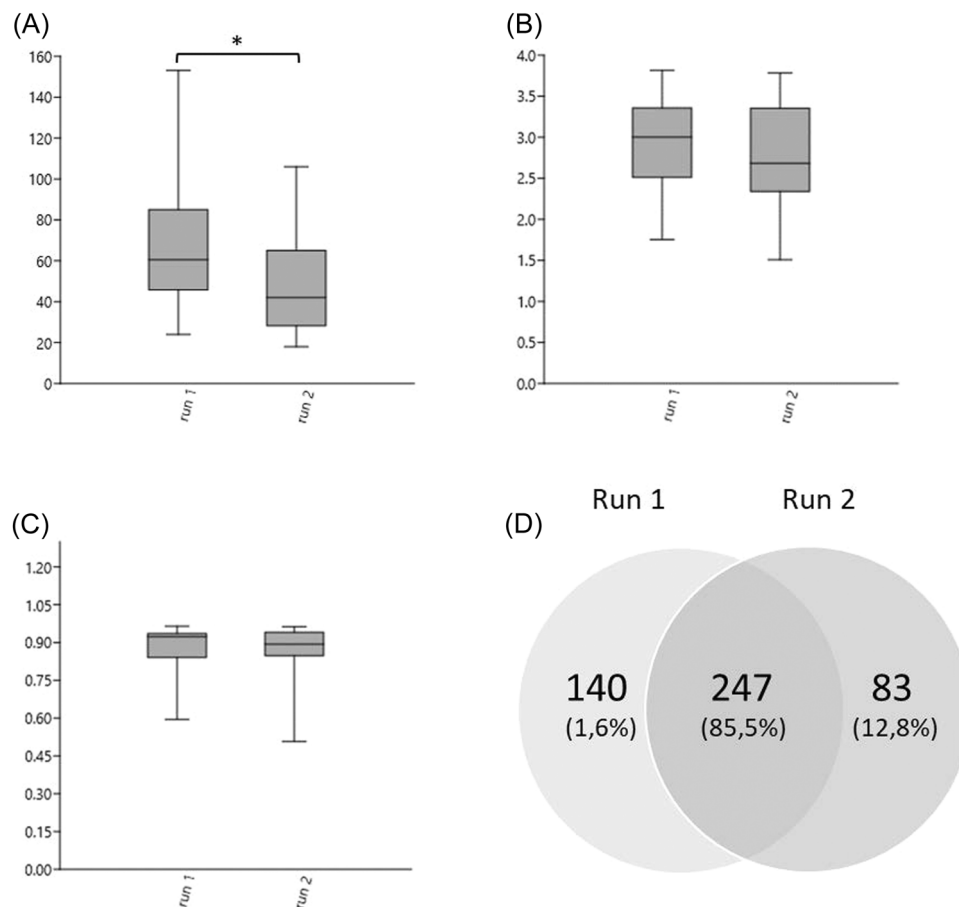
Sample similarity within the sequencing run was measured from a similarity matrix in a comparable manner as for between runs. The two similarity matrices, based on Bray Curtis and Dice similarity indexes created in PAST were used. The relevant comparisons were samples collected from the same individual (udder quarter) 1–2 h apart. More specifically: 1 h postintervention was compared to 0 h before the intervention, 2 h postintervention was compared to 0 h before intervention, 2 h postintervention was compared to 1 h postintervention, and 4 h

postintervention was compared to 2 h postintervention. These samples were biological replicates but were simultaneously expected to have a high similarity in their bacterial composition (further elaborated in Section 4).

### 3 | RESULTS

#### 3.1 | Difference between runs

The 44 milk samples that were subjected to DNA extraction and sequencing twice contained in total of 470 different bacterial taxa. The milk samples sequenced in run 1 had significantly more bacterial taxa, at the genus level, per sample (Chao-1) compared to run 2 ( $65.7 \pm 29.1$  vs.  $48.3 \pm 24.7$ ,  $p < 0.01$  t-test) while there was no difference in Shannon or Simpson diversity (Figure 2). A total of 247 bacterial taxa were shared between the two sequencing runs, 140 bacterial taxa were only identified in run 1, and 83 bacterial taxa were only identified in run 2 (Figure 2). The accumulated data (i.e., reads)



**FIGURE 2** Boxplots of bacterial richness and diversity in type 1 replicate milk samples in different runs. (A) Chao-1 richness. (B) Shannon diversity. (C) Simpson diversity. Brackets with stars above boxes indicate that they differ significantly. For boxes, the 25%–75% quartiles are in the box, and max and min are shown as whiskers. (D) Venn diagram showing the number of shared genera within two sequencing runs of the same 44 samples. The number within parenthesis represents the proportion of reads within each group.

from bacterial taxa that were shared between run 1 and 2 constituted a larger proportion of the data in run 1 compared to run 2 (Figure 2).

### 3.2 | Microbiota similarities between runs

The similarity between sequence runs was explored by PCoA, ANOSIM, and direct comparison of similarity between technical replicates from a similarity matrix. The PCoA revealed that the sequencing run had a bigger effect on the microbiota composition than the individual cow (Figure 3). The difference between sequencing runs was more pronounced when the Dice similarity index was applied to the data compared to the Bray-Curtis similarity index (Figure 3). Regardless of the similarity index used, the ANOSIM confirmed that the difference between the runs was significant (Bonferroni-corrected  $p < 0.01$ ). For three out of four individuals, there were significant differences between runs, when analyzed in ANOSIM, for both similarity indices used (Appendix Figure A1).

When direct comparisons of the similarity of type 1 replicates of milk samples from the two sequencing runs were analyzed the average similarity varied with the similarity index used. For the Dice similarity index, the average similarity was  $0.42 \pm 0.16$ , and for the Bray-Curtis similarity index, the average similarity was  $0.32 \pm 0.18$ , Figure 4A. In comparison, the average similarity of type 1 replicates of the bacterial mock community was  $0.95 \pm 0.05$  for the Dice similarity index and  $0.82 \pm 0.03$  for the Bray Curtis similarity (Figure 4C).

### 3.3 | Microbiota similarity within runs

The similarity between type 2 replicates, samples that were taken 1–2 h apart from the same individual (udder quarter), were analyzed

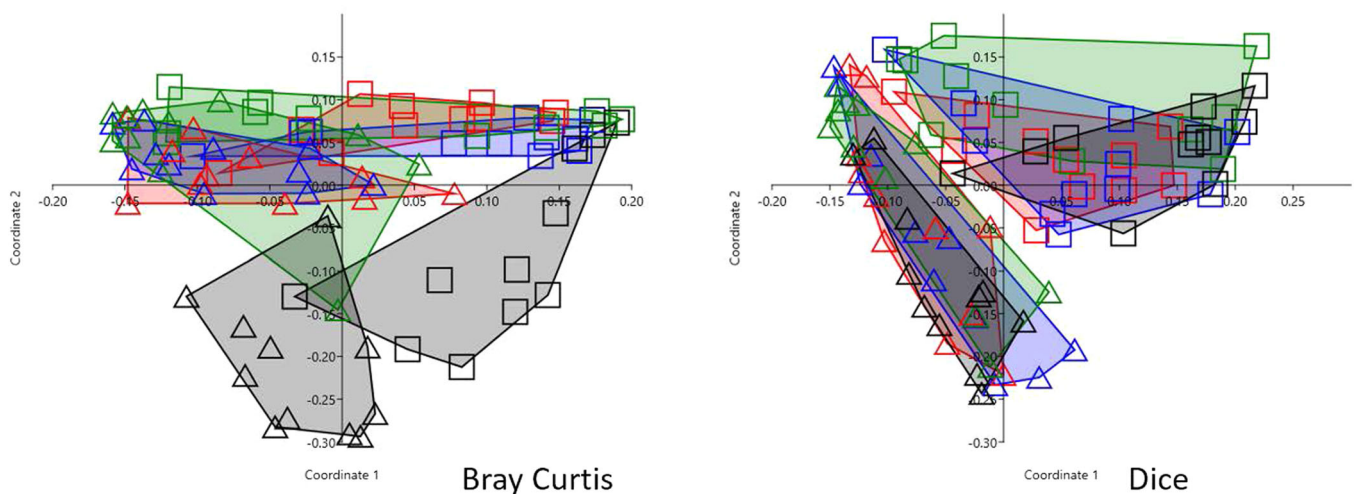
in the data set. The average Dice similarity between samples taken 1–2 h apart was  $0.47 \pm 0.12$  and the average Bray Curtis similarity was  $0.39 \pm 0.19$  (Figure 4B).

As we noted that the similarity between milk samples were comparable regardless if the comparison was between type 1 or type 2 replicates (as seen in Figure 4A,B), this phenomenon was further explored. For both similarity indices used, there was a significantly higher similarity between type 2 replicates in sequencing run 2 compared to sequencing run 1 (Figure 5) ( $p < 0.05$ ,  $t$ -test).

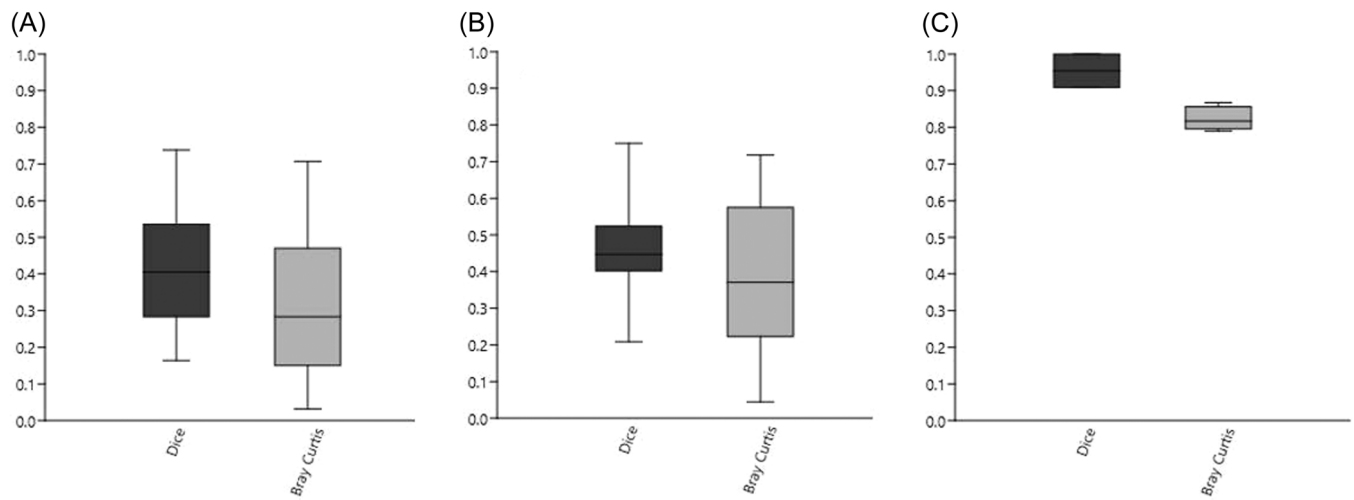
## 4 | DISCUSSION

In this text, the term type 1 replicates is used when comparing milk samples between different sequencing runs as the microbiota in identical samples were analyzed twice. However, some variations in the DNA extraction and PCR amplification existed between the two sequencing runs. Even though we used the same DNA extraction kit (MoBio PowerFood changed its name to DNeasy PowerFood after Qiagen purchased MoBio Laboratories), the batch number has been shown to influence results in microbiota studies (de Goffau et al. 2018, 2019; Salter et al., 2014). Another difference between the sequencing runs was the inclusion of the milk fat layer in sequencing run 1. It has previously been shown that milk fraction (whole milk, cell pellet, fat layer, or cell pellet and fat) included in the DNA extraction can affect the results (Lima et al., 2018; Sun et al., 2019). As such, some variation can be attributed to the different batches used and milk fractions included in DNA preparation.

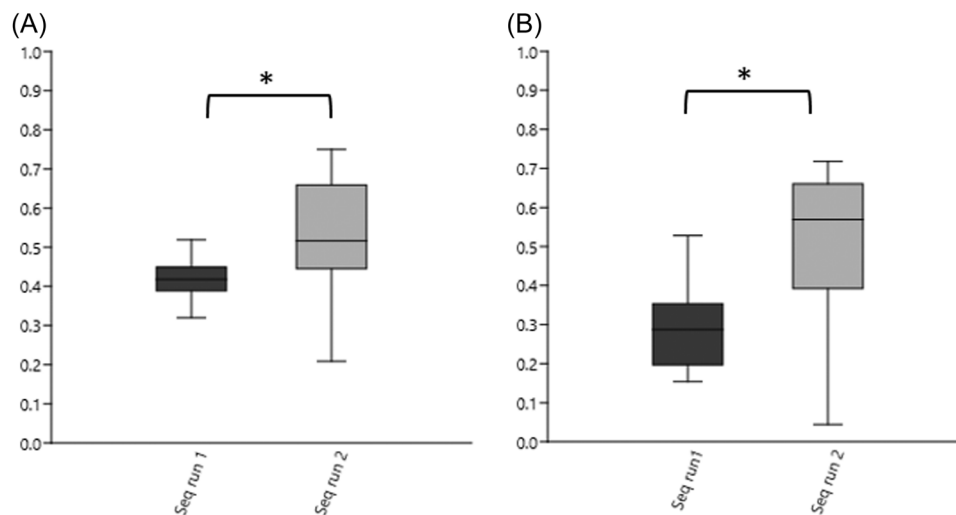
It has previously been shown that different sequencing runs can contribute largely to microbiota composition in sequence data (de Goffau et al., 2019; Salter et al., 2014; Taponen et al., 2019). We, and others, have shown that high bacterial biomass in samples decreases



**FIGURE 3** PCoA with 44 milk samples processed and sequenced twice, type 1 replicates. Samples are grouped by individual (color) and sequencing run (symbol). The similarity index used is indicated in the figure. For three out of four individuals, there were significant differences between runs, when analyzed in ANOSIM, for both similarity indexes used.



**FIGURE 4** Boxplots of direct comparison of similarity in type 1 and type 2 replicates using different similarity indexes. (A) Type 1 replicates of milk samples sequenced in two different runs ( $n = 44$ ). (B) Type 2 replicates of milk samples (taken 1–2 h apart) sequenced within the same run ( $n = 32$ ). (C) Type 1 replicates of a bacterial mock community sequenced in two different runs ( $n = 4$ ). For boxes, the 25%–75% quartiles are in the box, and max and min are shown as whiskers.



**FIGURE 5** Boxplots of direct comparison of similarity for type 2 replicates separated by sequencing run. (A) Dice similarity index. (B) Bray-Curtis similarity index. Number of comparisons within each run = 16. Brackets with stars above boxes indicate that they differ significantly in the *t*-test. For boxes, the 25%–75% quartiles are in the box, and max and min are shown as whiskers.

the risk that sequencing results are affected by contamination (Dahlberg et al., 2019; Salter et al., 2014). In the original experiment, data on bacterial biomass in milk samples were not collected, before the infusion all quarters were considered healthy and had a low somatic cell count that is typically seen in the absence of infection. Here, the direct comparison of similarity in a created mock community between sequence runs showed a high similarity for both indexes used. These results imply that samples with high bacterial biomass are less affected by introduced biases and indicate that high similarity between sequencing runs can be accomplished.

In this study, the similarity between sequencing runs was 42% when the binary Dice index was used. These numbers are in the same

range as Wen et al. (2017) presented; that similarity between technical replicates was in the range of 33%–44% when abundance not was taken into account. Schwenker et al. (2022), on the other hand, looked at variability introduced by re-sequencing and compared that to the total variability introduced by re-DNA-extraction and re-sequencing. They concluded that resequencing contributed to a major part (41%–178%) of the total variability seen.

When direct comparisons of similarity in milk samples were analyzed, within and between sequencing runs, two different similarity indexes were used. The Dice similarity index is a binary index, using the presence–absence of taxa, while the Bray-Curtis similarity index, on the other hand, accounts for the abundance of



taxa in the calculation. These indices were chosen to increase the robustness of the analysis.

For the comparison within the sequencing run, that is, the type 2 replicates, milk samples collected 1–2 h apart were chosen as replicates. During the first 4 h after LPS infusion, local and systemic clinical signs of inflammation were observed. At 4 h post LPS infusion, there was also an increase in somatic cell count (mainly neutrophils and macrophages) in the milk (see Johnzon et al., 2018, for details). The inflammatory response could affect the microbiota, although not likely within such a short timeframe (1–2 h). In support of this, we have recently shown that the inflammatory response to the LPS infusion did not affect the microbial composition in milk (Dahlberg et al., 2023). Further, Ganda et al. (2017) demonstrated only a minor bacterial clearance in 6 h after experimental infusion of live *E. coli* in lactating dairy cows, and also, in the wild, the estimated doubling time for *E. coli* is 15 h (Gibson et al., 2018). Both from a biological perspective and previous experiences we expected that the milk samples used in this study should be highly similar, especially in regard to their bacterial composition.

In the comparisons of the results from sequencing run 1 and 2, we observed more unique bacterial taxa per sample from run 1. However, the accumulated amount of data (i.e., reads) from the unique taxa in run 1 was much smaller than the accumulated data from the unique taxa in run 2, showing that the unique taxa in run 1 occurred in a very low abundance. This presence of low abundant taxa probably affects the similarity of samples within the sequencing run and thus explains why samples collected 1–2 h apart were less similar in run 1. Concurrently, it cannot be ruled out that the low abundant unique taxa found in sequencing run 1 could be attributed to contamination.

Other research groups have presented results that show a large variation in microbiota studies due to different batches of DNA-extraction kits (Salter et al., 2014) or resequencing (Schwenker et al., 2022). Methods that reduce the impact of contamination (Davis et al., 2018) and working methods that “substantially reduce the contamination-induced variability” (Moossavi et al., 2021; Wen et al., 2017) have been presented. We believe that these types of work packages can increase the repeatability of microbiota studies and that measurements of similarity between biological or technical replicates within the same sequencing run can be used as an indicator of repeatability.

We hope that the results presented here can contribute to an increased understanding of the factors that affect the results of microbiome studies and why results from different studies are not easily comparable.

## 5 | CONCLUSION

We conclude that a large variability can be introduced during sample processing and sequencing in microbiota studies. This is shown through a low degree of similarity found between identical milk samples processed and sequenced (with similar but not identical lab protocols) in two separate runs. We also conclude that for milk

samples collected from the same quarter within a short timeframe and analyzed in the same sequencing run, processing and sequencing can affect the level of similarity between these samples.

## AUTHOR CONTRIBUTIONS

**Josef Dahlberg:** Conceptualization (equal); Methodology (equal); Writing—original draft (lead); Writing—review and editing (equal). **Erik Pelve:** Data curation (lead); Software (lead); Writing—review and editing (equal). **Johan Dicksved:** Conceptualization (equal); Methodology (equal); Writing—review and editing (equal).

## ACKNOWLEDGMENTS

The authors would like to acknowledge Dr Li Sun at the Department of Molecular Sciences at the Swedish University of Agricultural Sciences for valuable support with the bioinformatics pipeline. This study was made possible through a grant from the Swedish Research Council Formas (20121365-2272942; 221-2012-1098). The National Genomics Infrastructure in Stockholm and Uppsala is funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure.

## CONFLICT OF INTEREST STATEMENT

None declared.

## DATA AVAILABILITY STATEMENT

Raw sequences data are available in the National Center for Biotechnology Information under BioProject PRJNA1013402: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1013402>.

## ETHICS STATEMENT

None required.

## ORCID

Josef Dahlberg  <http://orcid.org/0000-0001-8131-6725>

## REFERENCES

- Alipour, M. J., Jalanka, J., Pessa-Morikawa, T., Kokkonen, T., Satokari, R., Hynönen, U., Iivanainen, A., & Niku, M. (2018). The composition of the perinatal intestinal microbiota in cattle. *Scientific Reports*, 8(1), 10437.
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from illumina amplicon data. *Nature Methods*, 13(7), 581–583.
- Clooney, A. G., Fouhy, F., Sleator, R. D., O’ Driscoll, A., Cotter, P. D., Claesson, M. J., & Stanton, C. (2016). Comparing apples and oranges?: Next generation sequencing and its impact on microbiome analysis. *PLoS One*, 11(2), e0148028.
- Dahlberg, J., Johnzon, C. F., Sun, L., Pejler, G., Östensson, K., & Dicksved, J. (2023). Absence of changes in the milk microbiota during *Escherichia coli* endotoxin induced experimental bovine mastitis. *Veterinary Research*, 54(1), 46.
- Dahlberg, J., Sun, L., Persson Waller, K., Östensson, K., McGuire, M., Agenäs, S., & Dicksved, J. (2019). Microbiota data from low biomass

- milk samples is markedly affected by laboratory and reagent contamination. *PLoS One*, 14(6), e0218257.
- Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, 6(1), 226.
- Ganda, E. K., Gaeta, N., Sipka, A., Pomeroy, B., Oikonomou, G., Schukken, Y. H., & Bicalho, R. C. (2017). Normal milk microbiome is reestablished following experimental infection with *Escherichia coli* independent of intramammary antibiotic treatment with a third-generation cephalosporin in bovines. *Microbiome*, 5(1), 74.
- Gibson, B., Wilson, D. J., Feil, E., & Eyre-Walker, A. (2018). The distribution of bacterial doubling times in the wild. *Proceedings of the Royal Society B: Biological Sciences*, 285(1880), 20180789.
- de Goffau, M. C., Lager, S., Salter, S. J., Wagner, J., Kronbichler, A., Charnock-Jones, D. S., Peacock, S. J., Smith, G. C. S., & Parkhill, J. (2018). Recognizing the reagent microbiome. *Nature Microbiology*, 3(8), 851–853.
- de Goffau, M. C., Lager, S., Sovio, U., Gaccioli, F., Cook, E., Peacock, S. J., Parkhill, J., Charnock-Jones, D. S., & Smith, G. C. S. (2019). Human placenta has no microbiome but can contain potential pathogens. *Nature*, 572(7769), 329–334.
- Hammer, O., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4(1), 1–9.
- Hugerth, L. W., Wefer, H. A., Lundin, S., Jakobsson, H. E., Lindberg, M., Rodin, S., Engstrand, L., & Andersson, A. F. (2014). DegePrime, a program for degenerate primer design for broad-taxonomic-range PCR in microbial ecology studies. *Applied and Environmental Microbiology*, 80(16), 5116–5123.
- Iversen, K. N., Dicksved, J., Zoki, C., Fristedt, R., Pelve, E. A., Langton, M., & Landberg, R. (2022). The effects of high fiber rye, compared to refined wheat, on gut microbiota composition, plasma short chain fatty acids, and implications for weight loss and metabolic risk factors (the RyeWeight Study). *Nutrients*, 14(8), 1669.
- Johnzon, C. F., Dahlberg, J., Gustafson, A. M., Waern, I., Moazzami, A. A., Östensson, K., & Pejler, G. (2018). The effect of lipopolysaccharide-induced experimental bovine mastitis on clinical parameters, inflammatory markers, and the metabolome: A kinetic approach. *Frontiers in Immunology*, 9, 1487.
- Karstens, L., Asquith, M., Davin, S., Fair, D., Gregory, W. T., Wolfe, A. J., Braun, J., & McWeeney, S. (2019). Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. *mSystems*, 4(4), e00290–19.
- Kennedy, K., Hall, M. W., Lynch, M. D. J., Moreno-Hagelsieb, G., & Neufeld, J. D. (2014). Evaluating bias of illumina-based bacterial 16S rRNA gene profiles. *Applied and Environmental Microbiology*, 80(18), 5717–5722.
- Lauder, A. P., Roche, A. M., Sherrill-Mix, S., Bailey, A., Laughlin, A. L., Bittinger, K., Leite, R., Elovitz, M. A., Parry, S., & Bushman, F. D. (2016). Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome*, 4(1), 29.
- Lima, S. F., Bicalho, M. L. S., & Bicalho, R. C. (2018). Evaluation of milk sample fractions for characterization of milk microbiota from healthy and clinical mastitis cows. *PLoS One*, 13(3), e0193671.
- Marotz, C., Sharma, A., Humphrey, G., Gottel, N., Daum, C., Gilbert, J. A., Eloë-Fadrosh, E., & Knight, R. (2019). Triplicate PCR reactions for 16S rRNA gene amplicon sequencing are unnecessary. *Biotechniques*, 67(1), 29–32.
- McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8(4), e61217.
- Moossavi, S., Fehr, K., Khafipour, E., & Azad, M. B. (2021). Repeatability and reproducibility assessment in a large-scale population-based microbiota study: Case study on human milk microbiota. *Microbiome*, 9(1), 41.
- Nearing, J. T., Comeau, A. M., & Langille, M. G. I. (2021). Identifying biases and their potential solutions in human microbiome studies. *Microbiome*, 9(1), 113.
- O'Sullivan, D. M., Doyle, R. M., Temisak, S., Redshaw, N., Whale, A. S., Logan, G., Huang, J., Fischer, N., Amos, G. C. A., Preston, M. D., Marchesi, J. R., Wagner, J., Parkhill, J., Motro, Y., Denise, H., Finn, R. D., Harris, K. A., Kay, G. L., O'Grady, J., ... Huggett, J. F. (2021). An inter-laboratory study to investigate the impact of the bioinformatics component on microbiome analysis using mock communities. *Scientific Reports*, 11(1), 10590.
- Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The madness of microbiome: Attempting to find consensus “best practice” for 16S microbiome studies. *Applied and Environmental Microbiology*, 84(7), e02627–17.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(Database issue), 590–596.
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J., & Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, 12, 87.
- Schwenker, J. A., Friedrichsen, M., Waschina, S., Bang, C., Franke, A., Mayer, R., & Hölzel, C. S. (2022). Bovine milk microbiota: Evaluation of different DNA extraction protocols for challenging samples. *MicrobiologyOpen*, 11(2), e1275.
- Sun, L., Dicksved, J., Priyashantha, H., Lundh, Å., & Johansson, M. (2019). Distribution of bacteria between different milk fractions, investigated using culture-dependent methods and molecular-based and fluorescent microscopy approaches. *Journal of Applied Microbiology*, 127(4), 1028–1037.
- Taponen, S., McGuinness, D., Hiitö, H., Simojoki, H., Zadoks, R., & Pyörälä, S. (2019). Bovine milk microbiome: A more complex issue than expected. *Veterinary Research*, 50(1), 44.
- Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73(16), 5261–5267.
- Wen, C., Wu, L., Qin, Y., Van Nostrand, J. D., Ning, D., Sun, B., Xue, K., Liu, F., Deng, Y., Liang, Y., & Zhou, J. (2017). Evaluation of the reproducibility of amplicon sequencing with illumina MiSeq platform. *PLoS One*, 12(4), e0176716.
- Łukasiak, P., Newton, J. A., Sanders, J. G., Hu, Y., Moreau, C. S., Kronauer, D. J. C., O'Donnell, S., Koga, R., & Russell, J. A. (2017). The structured diversity of specialized gut symbionts of the New World army ants. *Molecular Ecology*, 26(14), 3808–3825.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Dahlberg, J., Pelve, E., & Dicksved, J. (2023). Similarity in milk microbiota in replicates. *MicrobiologyOpen*, 12, e1383. <https://doi.org/10.1002/mbo3.1383>



APPENDIX A

See Figure A1.

(a)

	run2 cow 1	run2 cow2	run2 cow3	run2 cow4	run1 cow1	run1 cow2	run1 cow3	run1 cow4
run2 cow 1		1	1	0.0028	0.0084	0.0028	0.0336	0.0028
run2 cow2	0.0278		1	0.0224	0.014	0.0028	0.0028	0.0028
run2 cow3	0.04222	0.1211		0.2268	0.2352	0.1848	0.322	0.0028
run2 cow4	0.357	0.2939	0.2197		0.0056	0.0028	0.0028	0.0056
run1 cow1	0.3678	0.4969	0.2406	0.5647		1	1	0.0028
run1 cow2	0.4515	0.5812	0.269	0.7262	0.06551		1	0.0028
run1 cow3	0.4072	0.574	0.2364	0.7029	0.01518	0.01127		0.0028
run1 cow4	0.8299	0.8023	0.7219	0.4712	0.4798	0.5812	0.597	

(b)

	run2 cow 1	run2 cow2	run2 cow3	run2 cow4	run1 cow1	run1 cow2	run1 cow3	run1 cow4
run2 cow 1		1	1	0.0868	0.0196	0.0112	0.0364	0.0028
run2 cow2	0.01878		1	0.196	0.0252	0.028	0.028	0.0084
run2 cow3	0.03689	0.04162		1	0.1232	0.2492	0.406	0.0056
run2 cow4	0.2651	0.2359	0.1453		0.0028	0.0028	0.0028	0.0028
run1 cow1	0.4131	0.414	0.2693	0.8011		1	1	0.1848
run1 cow2	0.3805	0.3597	0.2462	0.7393	0.09557		1	0.5376
run1 cow3	0.3847	0.4038	0.205	0.7782	0.01645	0.06296		0.3164
run1 cow4	0.4977	0.4693	0.4071	0.7198	0.194	0.1608	0.2005	

**FIGURE A1** Analysis of similarities (ANOSIM) with (A) Bray Curtis and (B) Dice similarity index for 44 milk samples processed and sequenced twice, type 1 replicates. A PCoA of the same data is presented in Figure 3 in the main text. For both tables, Bonferroni corrected *p*-values are presented above the diagonal and the *R* values below the diagonal.