applied microbiology

MDPI

Article

Breast Milk Microbiota at One Month Postpartum in Women from Arequipa, Peru, Assessed by Full-Length 16S rRNA Gene Nanopore Sequencing

Loisa R. Hancco-Vera [1], Ana M. Ahuate-Toribio [1], Kevin W. Lévano-Olín [1], Rodrigo A. Suarez-Silva [1], Yndalú K. Huillca-Pari [1], Luis C. Lizarraga-Vargas [2], Sandra Condori-Catachura [3] and Cinthya J. Lovon-Luque [2,*]

[1] Molecular Microbiology Research Laboratory, Universidad Nacional de San Agustín de Arequipa, Arequipa 04001, Peru; lhanccov@unsa.edu.pe (L.R.H.-V.); aahuate@unsa.edu.pe (A.M.A.-T.); klevano@unsa.edu.pe (K.W.L.-O.); rsuarezs@unsa.edu.pe (R.A.S.-S.); yhuillcap@unsa.edu.pe (Y.K.H.-P.)
[2] Department of Human Morphology, Faculty of Medicine, Universidad Nacional de San Agustín de Arequipa, Arequipa 04001, Peru; llizarragav@unsa.edu.pe
[3] Laboratory of Applied Microbiology and Biotechnology, Department of Bioscience Engineering, Universiteit Antwerpen, 2020 Antwerp, Belgium; sandra.condori@uantwerpen.be
* Correspondence: clovonl@unsa.edu.pe

Abstract

Human milk is a low-biomass fluid whose bacterial community may contribute to infant gut colonization; however, the milk microbiota of women from southern Peru remains uncharacterized. This study aimed to describe the bacterial composition and diversity of human milk at one month postpartum (days 30–33 of lactation) in women from Arequipa, Peru, using full-length 16S rRNA gene nanopore sequencing. In this cross-sectional study, milk was collected from 29 mothers, and 21 samples were analyzed after excluding eight with a relative abundance of Enterobacterales ≥30%. The microbiota was dominated by the phylum Bacillota (formerly Firmicutes; mean 92%), with *Streptococcus* (62.9%) and *Staphylococcus* (21.8%) forming the microbial core. Presumptive species-level assignments corresponded mainly to oral and skin commensals (*Streptococcus mitis*, *Streptococcus salivarius*, and *Staphylococcus epidermidis*). Alpha diversity was low (medians: Shannon index 0.66; Simpson index 0.30; observed richness 11 genera). In the principal coordinates analysis (Bray–Curtis), samples clustered according to their dominant genus. Observed richness was nominally higher in mothers with normal body mass index ($p = 0.0071$), whereas overall community composition was not associated with maternal, obstetric, or dietary variables (PERMANOVA, all $p \geq 0.073$). This study provides the first characterization of the human milk microbiota in women from Arequipa, establishing a baseline for future studies.

Keywords: human milk; microbiota; lactation; 16S rRNA; nanopore sequencing; *Streptococcus*; *Staphylococcus*

Check for updates

Academic Editor:
Costas Papagiannitsis

Received: 21 August 2026
Revised: 8 October 2026
Accepted: 12 October 2026
Published: 27 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Human milk is recognized as the optimal source of nutrition for infants during the first six months of life [1]. Beyond its nutritional and bioactive components, it harbors a bacterial community of commensal and potentially probiotic bacteria that may contribute to intestinal colonization in infants [2]. This microbial transfer is particularly important during the first months of life, a critical period in which the infant gut microbiota is established and exposure to maternal microorganisms may contribute to the maturation of the immune system and provide protection against potentially pathogenic microorganisms [3].

Research on the microbiota of human milk has shifted from culture-dependent methods to molecular techniques, predominantly 16S rRNA gene sequencing. These advances have significantly expanded the capacity to characterize bacterial diversity within this ecosystem [4]. However, differences in DNA extraction methods, the 16S rRNA gene regions analyzed, and the sequencing platforms used may contribute to the variability in composition observed across studies [4,5]. Most human milk studies have sequenced short hypervariable regions of the 16S rRNA gene, whose taxonomic resolution depends on the region analyzed, and is generally limited to the genus level [6]. Nanopore sequencing of the full-length 16S rRNA gene (V1–V9), combined with classifiers designed for long reads, improves this resolution and enables presumptive species-level assignments [7,8]. This approach has only recently been applied to human milk [9,10].

Studies conducted across different populations have described a microbiota dominated mainly by the phyla Bacillota (Firmicutes) and Proteobacteria, with recurrent genera including *Staphylococcus*, *Streptococcus*, *Acinetobacter*, *Pseudomonas*, and *Lactobacillus* [11,12]. Although these bacterial groups are frequently reported, their abundance and diversity vary across populations and may be related to maternal, perinatal, environmental, and geographical factors [13,14].

Despite a growing body of research on human milk microbiota composition, much of the evidence derives from North American, European, and Asian populations, while studies conducted in South America remain limited [15]. Available Latin American data come mainly from Brazil [16], Colombia [17], and Ecuador [9], whereas data from Peru remain scarce. Latin American populations have diverse environmental, geographical, and dietary characteristics [18], and Peru is marked by significant geographical diversity and variation in dietary patterns across populations [19]. Consequently, the bacterial profiles described in other populations do not necessarily represent the composition of human milk microbiota in Peruvian women.

In particular, the human milk microbiota of women from Arequipa, in southern Peru, has not yet been characterized. Therefore, this study aimed to describe the bacterial composition and diversity of human milk at one month postpartum (days 30–33 of lactation) in women residing in Arequipa, using full-length 16S rRNA gene nanopore sequencing. Additionally, we explored the association between bacterial profiles and selected maternal, obstetric, and dietary variables.

2. Materials and Methods

2.1. Study Design and Eligibility Criteria

A cross-sectional study was conducted in healthy lactating women aged 18–42 years with uncomplicated pregnancies and term deliveries (≥37 weeks), recruited at Hospital Regional Honorio Delgado Espinoza and primary health centers in the province of Arequipa. Women with diabetes, hypertension, complicated pregnancies, preterm delivery, mastitis, or other breast infections in the preceding month were excluded, as were those who had consumed antibiotics in the 30 days before sampling, or who regularly consumed antidepressants, alcohol in excess, or illicit substances. The Institutional Ethics Review Committee (CIEI) of Universidad Peruana Cayetano Heredia approved the study (SIDISI code: 25010), and all participants provided written informed consent.

2.2. Collection of Clinical and Dietary Data

Clinical data were collected through a questionnaire administered at the time of sampling. The questionnaire included maternal body mass index (BMI), classified according to WHO criteria (normal 18.50–24.99; overweight/obesity ≥ 25.00) [20], mode of delivery, number of previous pregnancies, and the presence of obstetric risk factors (history of

preterm birth, history of spontaneous abortion), as well as maternal dietary habits and supplement use. Infant age, recent health status, and feeding type (exclusive breastfeeding or formula feeding) were recorded.

2.3. Sample Collection

A single breast milk sample was collected from each participant at one month postpartum (days 30–33 of lactation). Participants collected breast milk by self-sampling, using the breast that had not been suckled for at least three hours. Prior to milk extraction, participants were instructed to disinfect the nipple and areola using 70% ethanol (Biogenics Lab, Lima, Peru) to minimize potential contamination during sampling [21]. After discarding the first drops, participants manually collected between 20 and 40 mL of milk. Samples were transported in sterile glass bottles on ice to the laboratory, where they were aliquoted and stored at −80 °C for subsequent processing.

2.4. DNA Extraction

An aliquot of six milliliters of breast milk was centrifuged at 10,000× *g* for 10 min at 4 °C, after which the supernatant and fat layer were discarded. DNA was extracted from the cellular pellet using the TIANamp Bacteria DNA Kit (Tiangen Biotech, Beijing, China) according to the manufacturer's instructions [22], with the following modification: 0.1 mm glass beads (USA Scientific, Ocala, FL, USA) were added and the mixture was vortexed (Labnet International, Edison, NJ, USA) at 3400 rpm for 10 min to promote bacterial cell disruption [5]. Blank extraction controls, consisting solely of kit reagents without a milk sample, were processed in parallel through all extraction steps. DNA concentration and purity were determined by spectrophotometry using a BioSpectrometer® (Eppendorf, Hamburg, Germany), evaluating the A260/A280 and A260/A230 absorbance ratios. Genomic DNA integrity was additionally verified by 1.5% agarose (Cleaver Scientific, Rugby, UK) gel electrophoresis and visualized using a gel imaging system (UVITEC, Cambridge, UK).

2.5. 16S rRNA Gene Amplification

Full-length 16S rRNA gene amplification (regions V1–V9) was performed using the primers included in the Rapid Sequencing DNA 16S Barcoding Kit V14 (SQK-16S114.24; Oxford Nanopore Technologies, Oxford, UK), following the manufacturer's protocol. For each PCR reaction, 20 ng of DNA was added together with 10 µL of primer and 25 µL of LongAmp Taq 2X Master Mix (M0287L, New England Biolabs, Ipswich, MA, USA). Amplification conditions comprised an initial denaturation at 95 °C for 1 min, followed by 35 cycles of denaturation at 95 °C for 20 s, annealing at 55 °C for 30 s, and extension at 65 °C for 2 min, with a final extension at 65 °C for 5 min; reactions were held at 4 °C until retrieval. The number of cycles was increased from 25 (manufacturer default) to 35 to improve amplicon yield from low-biomass samples. Previous full-length 16S rRNA gene studies of human milk have used 35–38 cycles [9,10].

2.6. Library Preparation and Third-Generation Sequencing

Libraries were prepared using the Rapid Sequencing DNA 16S Barcoding Kit V14 (SQK-16S114.24; Oxford Nanopore Technologies, Oxford, UK) according to the manufacturer's instructions. Samples were multiplexed in 24-plex runs (consistent with the kit's barcode capacity), including the blank extraction controls and technical extraction replicates. Sample amounts (in fmol) were equilibrated and the library was purified with AMPure XP magnetic beads (supplied with the kit). Fifty fmol of the library were loaded onto an FLO-MIN114 (R10.4.1) flow cell and sequenced on a MinION Mk1D device (Oxford Nanopore Technologies) using MinKNOW software (v25.09.16) for 14–16 h per run. Base-

calling, demultiplexing, and adapter and barcode trimming were performed with Dorado (v7.11.2; integrated in MinKNOW v25.09.16) in 'sup' configuration, with a minimum read length of 1300 bp and a minimum Q-score of 12.

2.7. Bioinformatic Analysis

Reads were filtered by length (1300–1800 bp) and quality (Q ≥ 12). Each sample was randomly subsampled to a maximum of 50,000 reads (seqkit v2.13.0, fixed seed = 100) [5,23,24]; samples below this threshold were kept in full. Chimeras were identified with VSEARCH v2.32.0 [25]. Because dereplication of long reads with high error rates yields almost exclusively singleton sequences, which precludes abundance-based de novo detection, reads were clustered at 97% identity (–cluster_size) to recover the abundance structure, and –uchime_denovo was applied. Reads belonging to the corresponding cluster inherited the chimeric assignment, whereas sequences with "borderline" results were retained for further analysis. Taxonomic classification was performed with EMU v3.6.2 [8], which combines minimap2 alignment (as implemented in EMU v3.6.2) with an expectation-maximization algorithm optimized for full-length 16S rRNA gene reads, using EMU's default reference database, composed of sequences from rrnDB v5.6 and NCBI 16S RefSeq. Per-sample relative abundances and estimated counts were retained. The complete bioinformatic processing code is available at https://github.com/rsuarezs-lab/milk-microbiota-16S-ONT (accessed on 7 October 2026).

2.8. Contamination Control and Sample Filtering

Reagent contaminants were identified with decontam v1.22.0 (threshold 0.1) Davis [26] using the blank extraction controls, and the flagged genera were removed. Counts were converted to relative abundances within each taxonomic level. Sequencing depth adequacy was assessed using rarefaction curves generated with vegan v2.6-4 [27] and Good's coverage estimator [28].

2.9. Statistical Analysis

All community-level analyses were performed in R v4.3.3 with the vegan package v2.6-4 [29]. Beta diversity was assessed from genus-level Bray–Curtis dissimilarities and visualized by principal coordinates analysis (PCoA). Differences in composition between groups were tested with PERMANOVA (adonis2) [30] and differences in dispersion with PERMDISP (betadisper) [31], both with 999 permutations; the association of maternal, obstetric, and dietary variables with the ordination was assessed by vector fitting (envfit, 999 permutations). Alpha diversity (observed richness, Shannon and Simpson [1-D] indices) was compared between groups with Mann–Whitney tests. The core microbiome was defined as taxa exceeding 0.1% relative abundance in ≥80% of samples, values within the ranges used in previous microbiome studies, including those of human milk [32,33]. Given the sample size, the statistical power of PERMANOVA was estimated by simulation across a range of effect sizes. Continuous variables are expressed as mean ± SD and categorical variables as n (%); $p < 0.05$ was considered significant. Association analyses with maternal, obstetric, and dietary variables were considered exploratory; therefore, no correction for multiple comparisons was applied. The complete statistical analysis code is available at https://github.com/rsuarezs-lab/nanopore-16s-vegan-pipeline (accessed on 7 October 2026).

3. Results

3.1. Participant and Sample Selection

Breast milk samples were initially collected from 29 mothers living in Arequipa, Peru, who met the eligibility criteria. Eight samples were excluded because they showed a relative abundance of Enterobacterales ≥30% and a markedly atypical bacterial profile, leaving a final analytical cohort of 21 samples. A single sample was collected from each participant at one month postpartum (days 30–33 of lactation). Maternal clinical and dietary characteristics are summarized in Table 1.

Table 1. Maternal, obstetric, dietary, and infant/lactation characteristics of the study cohort.

Characteristics	Final Study Sample (n = 21)
Maternal variables	
Age, years—mean ± SD	29.0 ± 7.3
BMI, kg/m^2—mean ± SD	25.1 ± 3.0
Normal BMI (18.5–24.9)—*n* (%)	11 (52.4%)
Overweight/obesity (≥25.0)—*n* (%)	10 (47.6%)
Gestational age, weeks—mean ± SD	38.8 ± 1.1
Obstetric variables	
Mode of delivery: cesarean—*n* (%)	12 (57.1%)
Mode of delivery: vaginal—*n* (%)	9 (42.9%)
Primigravida—*n* (%)	9 (42.9%)
Multigravida (≥2 pregnancies)—*n* (%)	12 (57.1%)
History of spontaneous abortion—*n* (%)	6 (28.6%)
History of preterm birth—*n* (%)	1 (4.8%)
Dietary variables	
Processed foods or sodas intake—*n* (%)	15 (71.4%)
Fish/seafood intake—*n* (%)	14 (66.7%)
Frequent fermented foods intake—*n* (%)	7 (33.3%)
Meat intake—*n* (%)	21 (100.0%)
Poultry intake—*n* (%)	20 (95.2%)
Fruit and vegetable intake—*n* (%)	21 (100.0%)
Infant and lactation variables	
Infant age, days—median (range)	31 (30–33)
Recent infant illness—*n* (%)	3 (14.3%)
Exclusive breastfeeding—*n* (%)	19 (90.5%)
Breastfeeding with formula—*n* (%)	2 (9.5%)
Maternal supplement use—*n* (%)	14 (66.7%)

Continuous variables are presented as mean ± SD and categorical variables as *n* (%).

3.2. Sequence Processing and Quality

Each library was subsampled to a maximum of 50,000 reads (seqkit, seed 100) and subjected to chimera removal (VSEARCH), which accounted for between 0.12% and 4.72% of reads per sample. For samples sequenced in more than one run, the run with the greater depth was retained. Final depth ranged from 1676 to 48,957 reads (median 46,329). Rarefaction curves reached the asymptote in all samples (Figure S1) and Good's coverage was 100% in all of them, indicating adequate sequencing depth to characterize the community at the genus level (Table S1). Notably, the lowest-depth sample (1676 reads) occupied a central position in the subsequent PCoA ordination, indicating that its reduced sequencing depth did not bias its placement in the beta-diversity analysis, consistent with its complete Good's coverage.

3.3. Bacterial Taxonomic Composition of Breast Milk

Of the 21 samples analyzed, the bacterial microbiota was dominated by the phylum Bacillota (previously known as Firmicutes), which accounted for a mean relative abundance of 92% per sample (Figure S2). At the family level, Streptococcaceae predominated (64.7%), followed by Staphylococcaceae (22.5%) (Figure 1a). The remaining families were minor: Enterobacteriaceae (5.1%), Veillonellaceae (3.2%), and Moraxellaceae (1.7%), with the rest below 0.5%. At the genus level, the community resolved mainly into *Streptococcus* (62.9%) and *Staphylococcus* (21.8%), with minor contributions from *Escherichia* (4.3%), *Gemella* (3.0%), *Veillonella* (3.0%), and *Acinetobacter* (1.6%), and the rest below 0.5% (Figure 1b). At the species level, the most abundant assignments corresponded to *Streptococcus mitis* (26.7%), *Streptococcus salivarius* (15.7%), *Staphylococcus epidermidis* (13.0%), and *Streptococcus* sp. (reference identifier FDAARGOS_192; 7.7%), with the rest below 5% (Figure 1c).

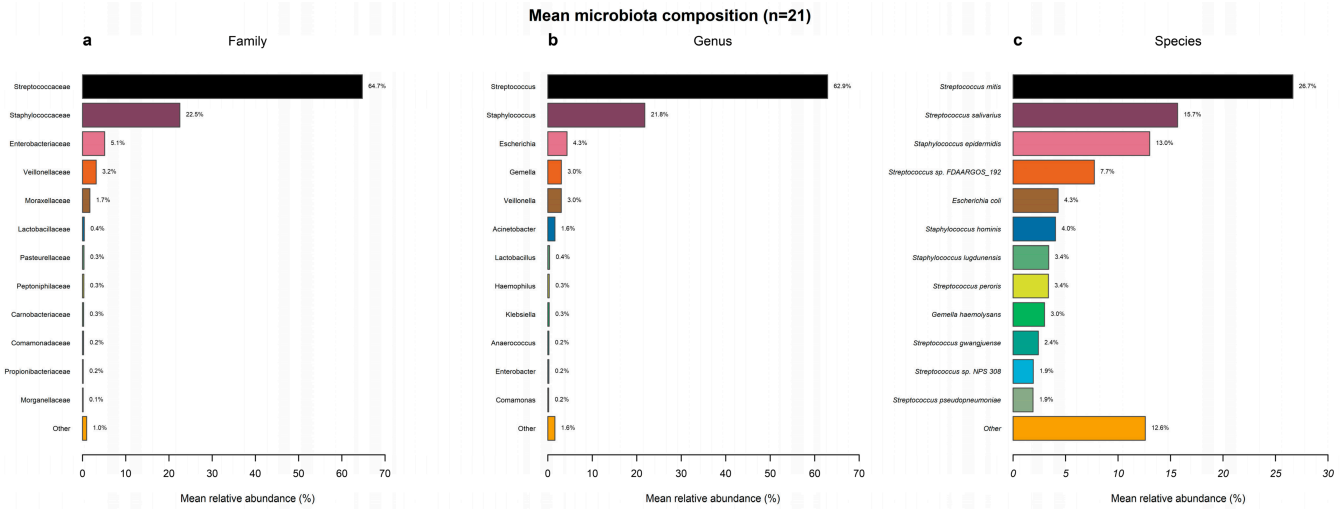


Figure 1. Mean microbiota composition in human milk. (**a**) Family level, (**b**) genus level, and (**c**) species level.

3.4. Diversity and Community Structure of Breast Milk Microbiota

Alpha diversity was low across all samples, a pattern characteristic of communities dominated by a small number of taxa. At the genus level, the median Shannon index was 0.66 (range 0.29–1.75) and observed richness had a median of 11 genera per sample (range 2–23). The Simpson index (1−D) was similarly low, with a median of 0.30 (range 0.11–0.78) (Figure 2).

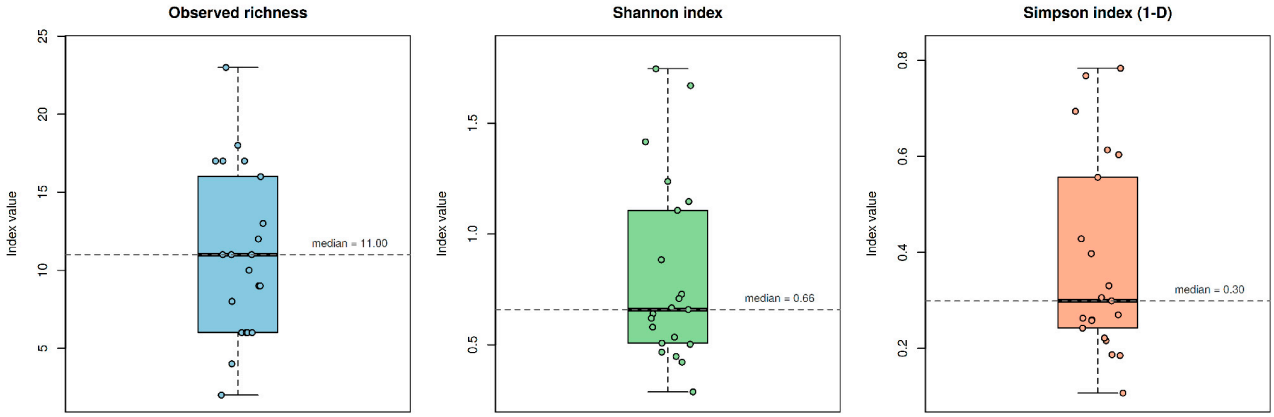


Figure 2. Alpha diversity indices at the genus level (n = 21). Vertical dashed lines indicate the whiskers, and the horizontal dashed line indicates the median value.

The low Simpson value indicates that two randomly drawn reads belong to the same genus in about 70% of cases, reflecting a structure in which abundance is concentrated mainly in *Streptococcus* and, to a lesser extent, *Staphylococcus*. Meanwhile, the remaining richness corresponds to genera of very low abundance. The wide range of the indices additionally reflects inter-individual heterogeneity across mothers.

To explore community structure, a heat map of the 20 most abundant genera was constructed with UPGMA hierarchical clustering based on Bray–Curtis distances for both samples and taxa (Figure 3). Samples were organized into blocks defined by their dominant genus (upper color bar), mainly *Streptococcus* and *Staphylococcus*, which co-occurred in nearly all samples, although one generally predominated. Below the dominant taxa, a set of low-abundance genera, such as *Gemella* and *Veillonella*, appeared recurrently in most samples, but did not reach the core prevalence threshold. Core microbiome analysis confirmed this predominance: at the genus level, it included only *Streptococcus* and *Staphylococcus*, present in 100% and 95% of samples, respectively. At the family level, the shared core included Streptococcaceae and Staphylococcaceae, both within the phylum Bacillota. The limited extent of the core, of just two genera, is consistent with the low alpha diversity described above.

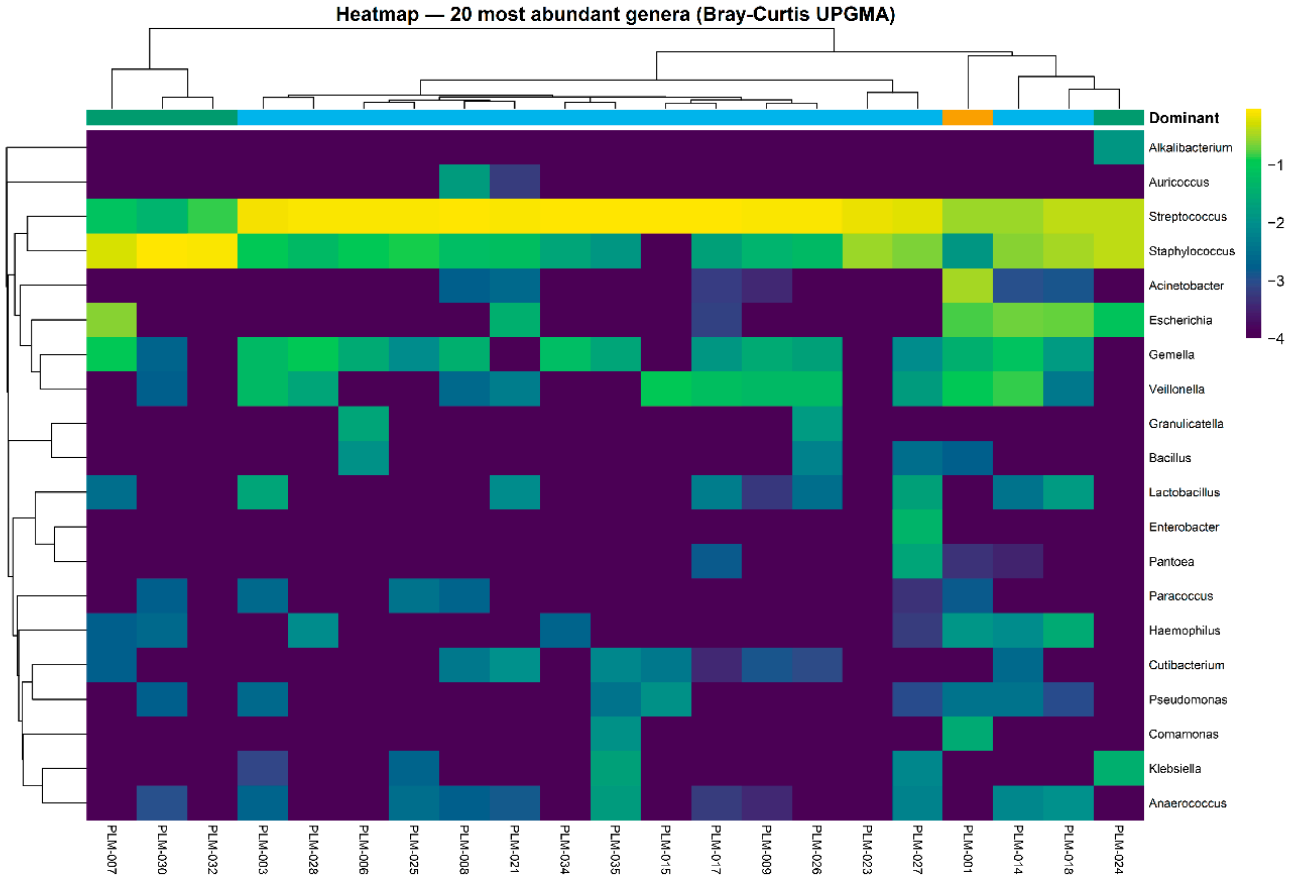


Figure 3. Heat map and hierarchical clustering of the 20 most abundant bacterial genera in milk samples.

This structure was accompanied by considerable interindividual heterogeneity in the dominant taxon (Figure 4). Most samples (16/21; 76%) were dominated by *Streptococcus*, four (19%) by *Staphylococcus*, and one (5%) by *Acinetobacter*. Thus, despite sharing the same core, each community centered on a single dominant genus that varied between individuals, reflecting the natural variability of the human milk microbiota.

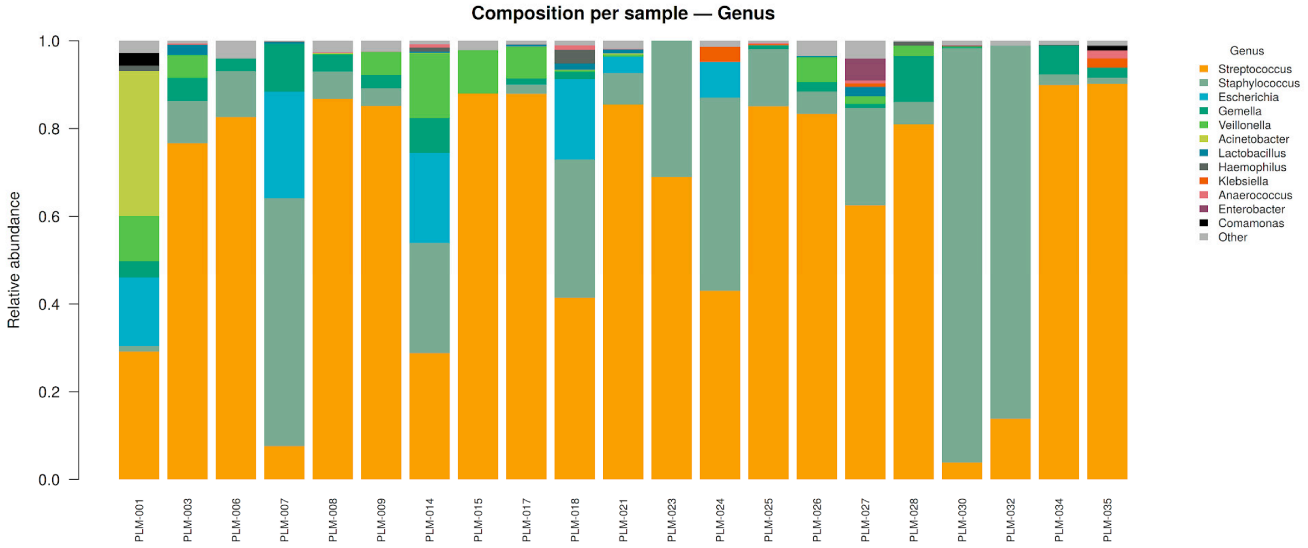


Figure 4. Microbial composition of human milk at one month postpartum at genus level.

3.5. Association of Microbial Diversity and Composition with Maternal, Obstetric, and Dietary Variables

Alpha-diversity indices were compared between groups using Mann–Whitney tests. Observed richness was nominally higher in the normal-BMI group than in the overweight/obesity group (median: 16 vs. 7 genera; $p = 0.0071$), whereas no significant differences were detected for Shannon or Simpson indices or for any alpha-diversity metric across the other variables evaluated. Given the small sample size, the multiple comparisons performed, and the exploratory nature of these analyses, this nominal difference should be interpreted with caution (Table S2). The association between microbiota composition and maternal, obstetric, and dietary variables was explored by projecting these variables onto the principal coordinates analysis (PCoA) ordination based on genus-level Bray–Curtis dissimilarity (PCo1: 68.9%; PCo2: 18.8%) using vector fitting (envfit) (Figure 5). The ordination was dominated by the primary genus of each sample (*Streptococcus*, *Staphylococcus*, or *Acinetobacter)* and no variable showed a significant association with community composition (PERMANOVA, all $p \geq 0.073$), with consistently low proportions of explained variance ($R^2 \leq 0.13$; Table 2). Concordantly, no envfit vector was significant; BMI had the largest R^2 (0.25), without reaching significance ($p = 0.077$).

In the dispersion analysis (PERMDISP), only BMI showed a nominally significant difference ($p = 0.011$), whereas the remaining variables did not reach significance (processed-food diet $p = 0.086$; fish intake $p = 0.154$; mode of delivery $p = 0.334$; maternal supplements $p = 0.488$; fermented foods $p = 0.831$). This signal reflects differing internal variability between groups rather than a difference in mean composition, given that the corresponding PERMANOVA was not significant ($p = 0.073$). The few samples with atypical profiles, dominated by *Staphylococcus* or *Acinetobacter* rather than *Streptococcus*, were unevenly distributed across BMI categories and concentrated in one group, increasing its apparent dispersion. This result should be interpreted with caution given the small sample size ($n = 21$) and the imbalance in some comparisons, conditions to which PERMDISP is sensitive. Therefore, this dispersion difference should not be interpreted as evidence of a BMI effect on community composition. Overall, the composition of the milk microbiota at one month postpartum showed no detectable association with the maternal, obstetric, and dietary variables evaluated and remained organized around the dominant genus of each sample.

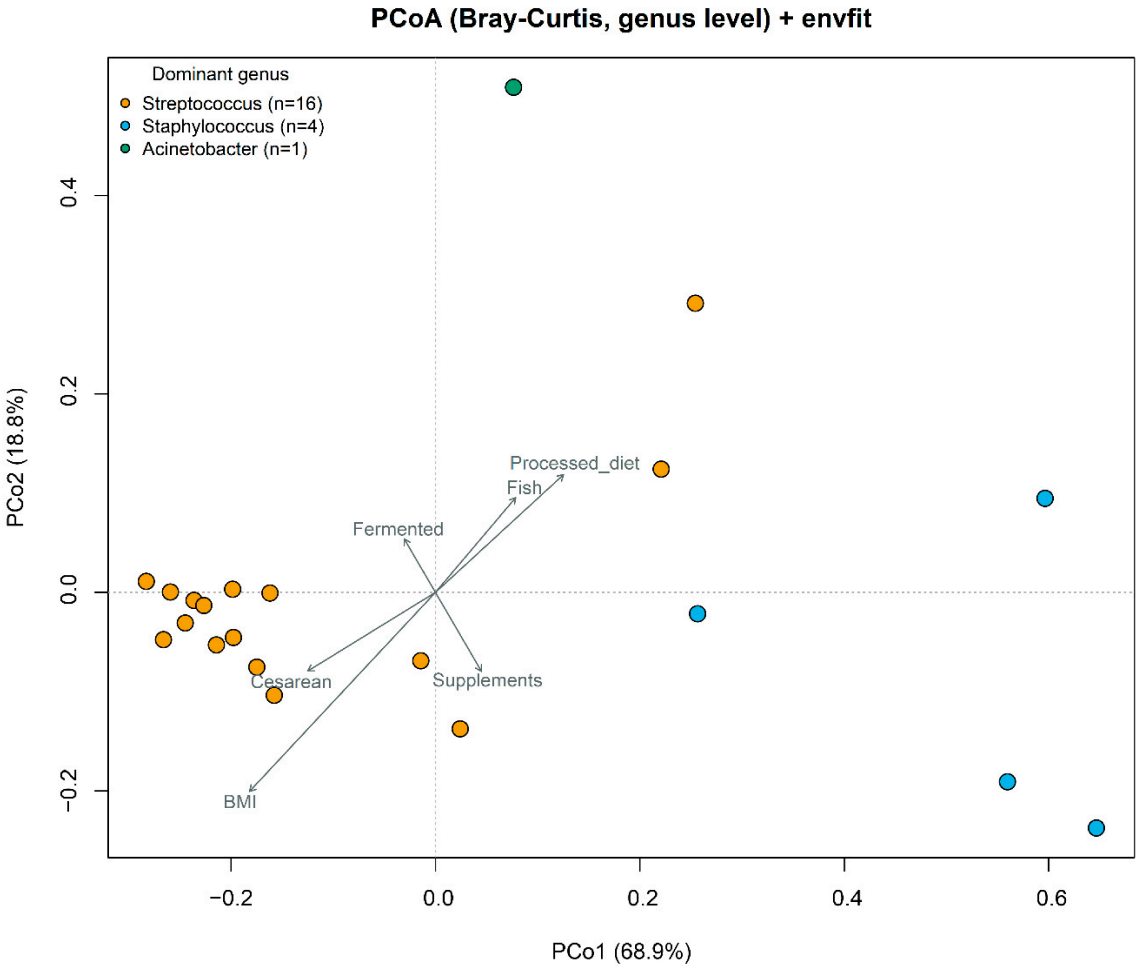


Figure 5. Principal coordinates analysis (PCoA) based on genus-level Bray–Curtis dissimilarities. Arrows represent the maternal, obstetric and dietary variables fitted with envfit.

Table 2. Association between maternal, obstetric, and dietary variables and microbiota composition ($n = 21$).

Variable	R^2 (PERMANOVA)	PERMANOVA (p)	PERMDISP (p)
BMI (Normal vs. Overweight/Obesity)	0.13	0.073	0.011
Mode of delivery (Cesarean vs. Vaginal)	0.052	0.314	0.334
Maternal supplement use (Yes vs. No)	0.02	0.713	0.488
Processed-food diet (Consumes vs. Avoids)	0.069	0.231	0.086
Fermented foods (High vs. Low)	0.021	0.724	0.831
Fish intake (Consumes vs. Never)	0.042	0.443	0.154

R^2 and p values from PERMANOVA (composition) and PERMDISP (dispersion) on genus-level Bray–Curtis dissimilarities (999 permutations).

4. Discussion

The present study analyzed the breast milk microbiota at one month postpartum in 21 women from Arequipa, characterized by a balanced distribution of maternal nutritional status, a high frequency of cesarean deliveries, mixed diets and a high rate of exclusive breastfeeding; these clinical and demographic characteristics are consistent with public health records for the Peruvian population [34], providing the sociodemographic context needed to interpret the structure and variability of the observed bacterial community. In this

context, full-length 16S rRNA gene sequencing showed that the breast milk microbiota was dominated by the genera *Streptococcus* (62.9%) and *Staphylococcus* (21.8%), which formed the microbial core shared among participants. This finding is consistent with reports in healthy Brazilian women by [16] at 30 ± 4 days postpartum, with abundances of 42% for *Streptococcus* and 22% for *Staphylococcus*, and by 24. Londoño-Sierra [17], who reported relative abundances of 32% for *Streptococcus* and 17.3% for *Staphylococcus* in Colombian women during the first three months of lactation. Although our population showed a higher abundance of *Streptococcus*, longitudinal studies [10,35] support both as core genera and attribute their variation to lactation stage, population characteristics and methodological differences. Nevertheless, these comparisons should be interpreted qualitatively, as the cited studies used Illumina sequencing of partial 16S regions (V4 in Padilha et al.; V3–V4 in Londoño-Sierra et al.) and also differed in DNA extraction method, taxonomic classification and reference database. All of these factors influence relative abundances [4,5]. At the species level, *Streptococcus mitis* (26.7%), *Streptococcus salivarius* (15.7%), *Staphylococcus epidermidis* (13.0%), and *Streptococcus* sp. (reference identifier FDAARGOS_192; 7.7%) were the most abundant. Previous studies have shown that the dominant species within *Staphylococcus* and *Streptococcus* in human milk correspond to commensal taxa typical of the skin and oral cavity, mainly *Staphylococcus epidermidis* and *Streptococcus mitis*, *S. salivarius* and *S. parasanguinis* [36,37]. Their origin is attributed to two routes: retrograde flow from the infant's oral cavity during suckling, and colonization of the nipple–areola complex by the maternal skin microbiota [21,38]. In contrast, the detection of *Escherichia coli* at lower abundance (4.3%) could reflect bacterial traces introduced during milk expression [39], an effect amplified in low-biomass samples, or translocation via the entero-mammary pathway [38]. Our design does not allow us to discriminate between these origins or to assess their viability.

Regarding diversity, the low values observed in our population (median Shannon index: 0.66; median Simpson index: 0.30) are consistent with the temporal dynamics described during early lactation. Xu [10], in an Australian longitudinal cohort, reported that milk microbiota richness and evenness were higher in the first week than at one month postpartum and then stabilized for the remainder of the first year. Similarly, Wan [12], in Chinese mothers, found that microbial diversity and richness were highest in colostrum and progressively decreased towards six weeks postpartum. These findings agree with those described in other cohorts Cabrera-Rubio and Liu [35,40], indicating that the low diversity observed in our population reflects an expected pattern at this stage of lactation.

Regarding community structure, samples were distributed mainly according to their dominant genus, predominantly *Streptococcus* and, to a lesser extent, *Staphylococcus*, a clustering pattern also described in Brazilian women [16]. Among the maternal, obstetric, and dietary variables evaluated, only BMI showed a nominal association with observed richness, which was higher in mothers with normal weight, in agreement with the lower diversity described in the milk of obese mothers [40]. In contrast, no variable was significantly associated with community composition, although BMI explained the largest proportion of variation (PERMANOVA $R^2 = 0.13$). Higher *Staphylococcus* abundance has been reported in the milk of mothers with overweight or excessive gestational weight gain [41], whereas in Mexican mothers BMI had a limited effect on milk microbiota overall but was associated with changes in specific genera [42]. Together, these observations suggest that the influence of BMI may be subtle and not necessarily reflected in overall community composition. Given the sample size, these analyses are exploratory, and the absence of associations should not be interpreted as evidence of no effect.

Although these findings offer a valuable perspective, methodological limitations constrain the interpretation. The sample size ($n = 21$) and the imbalance between some

comparison groups limit statistical power; therefore, the findings constitute a reference baseline that requires confirmation in larger populations, and differences in dispersion should be interpreted with caution [31]. At the molecular level, full-length 16S rRNA sequencing provides high taxonomic resolution at the phylum and genus levels and can also support species-level identification, particularly when long-read data are analyzed with classifiers optimized for full-length 16S sequences. However, species-level resolution is not uniform across all bacterial taxa, and closely related species such as members of the *Streptococcus mitis* group (*S. mitis*, *S. oralis*, *S. pneumoniae*) may not be unequivocally distinguished based on the 16S rRNA gene alone [6]. In addition, species-level assignments can be influenced by the classifier and reference database used [43]. Therefore, species-level assignments in the present study are considered presumptive, and orthogonal confirmation may be warranted for specific taxa using approaches such as culture followed by whole-genome sequencing or other higher-resolution genomic methods. Future research should combine longitudinal designs with more frequent sampling, to evaluate the temporal dynamics of the microbiota and its relationship with maternal factors [44], and metagenomic sequencing, to characterize the functional potential of these bacteria and their impact on the infant [45].

5. Conclusions

In conclusion, breast milk microbiota in women from Arequipa at one-month postpartum was dominated by *Streptococcus* and *Staphylococcus*, and overall community composition showed no detectable association with the maternal, obstetric and dietary factors evaluated. Species-level assignments corresponded mainly to oral and skin commensals (*Streptococcus mitis*, *Streptococcus salivarius*, and *Staphylococcus epidermidis*) and should be considered presumptive pending confirmation by complementary higher-resolution approaches. This study provides the first characterization of the breast milk microbiota in women from Arequipa using full-length 16S rRNA gene sequencing, establishing a baseline for future longitudinal and functional studies.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/applmicrobiol6110128/s1, Figure S1: Genus-level rarefaction curves ($n = 21$); Table S1: Sequencing depth, chimera removal and coverage per sample ($n = 21$); Figure S2: Mean bacterial composition at phylum level ($n = 21$); Table S2: Comparison of alpha-diversity indices across maternal, obstetric, and dietary groups ($n = 21$).

Author Contributions: Conceptualisation, L.R.H.-V., A.M.A.-T., R.A.S.-S., Y.K.H.-P., L.C.L.-V. and S.C.-C.; Methodology, L.R.H.-V., A.M.A.-T., R.A.S.-S., Y.K.H.-P. and L.C.L.-V.; Formal Analysis, R.A.S.-S.; Investigation, L.R.H.-V., A.M.A.-T., K.W.L.-O., R.A.S.-S., Y.K.H.-P., L.C.L.-V., S.C.-C. and C.J.L.-L.; Resources, R.A.S.-S. and Y.K.H.-P.; Data Curation, R.A.S.-S.; Writing—Original Draft Preparation, L.R.H.-V., A.M.A.-T., R.A.S.-S. and Y.K.H.-P.; Writing—Review and Editing, R.A.S.-S., Y.K.H.-P., L.C.L.-V., S.C.-C. and C.J.L.-L.; Visualisation, L.R.H.-V., A.M.A.-T. and R.A.S.-S.; Supervision, C.J.L.-L.; Project Administration, C.J.L.-L.; Funding Acquisition, C.J.L.-L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Universidad Nacional de San Agustín de Arequipa through Contract No. PI-02-2025-UNSA.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Ethics Review Committee (CIEI) of the Universidad Peruana Cayetano Heredia (SIDISI code: 25010, approved 25 July 2025).

Informed Consent Statement: Informed consent was obtained from all participants enrolled in the study.

Data Availability Statement: The raw full-length 16S rRNA gene sequencing reads generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession

number PRJNA1540999. The bioinformatic processing and statistical analysis code is publicly available at https://github.com/rsuarezs-lab/milk-microbiota-16S-ONT (accessed on 7 October 2026) and https://github.com/rsuarezs-lab/nanopore-16s-vegan-pipeline (accessed on 7 October 2026).

Acknowledgments: S.C.-C. was supported by the Isala sisterhood funding (University Research Fund, BOF, University of Antwerp). We sincerely thank all volunteer mothers for their invaluable participation. We also acknowledge Roberto Huamani from the Hospital Docente de la UNSA for his administrative support; Hospital Regional Honorio Delgado, Centro de Salud Maritza Campos Díaz, and MicroRed Buenos Aires de Cayma for the institutional facilities provided; and Smith Ollachica together with the team of the Laboratorio de Investigación en Microbiología Molecular of the UNSA for the technical support offered. During the preparation of this manuscript, the authors used Claude (Anthropic, Opus 4.7) to support text translation. The authors reviewed and edited the content and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shoji, H. Effect and concern of breastfeeding in infants. *Juntendo Med. J.* **2024**, *70*, 300–306. [CrossRef] [PubMed]
2. Yi, D.; Kim, S. Human breast milk composition and function in human health: From nutritional components to microbiome and microRNAs. *Nutrients* **2021**, *13*, 3094. [CrossRef] [PubMed]
3. Boudry, G.; Charton, E.; Le Huërou-Luron, I.; Ferret-Bernard, S.; Le Gall, S.; Even, S.; Blat, S. The relationship between breast milk components and the infant gut microbiota. *Front. Nutr.* **2021**, *8*, 629740. [CrossRef] [PubMed]
4. Cheema, S.K.; Li, R.; Cameron, S.J.S. Culturomics as a tool to better understand the human milk microbiota and host–microbiota interactions. *Microbiota Host* **2023**, *1*, e230001. [CrossRef]
5. Douglas, C.A.; Ivey, K.L.; Papanicolas, L.E.; Best, K.P.; Muhlhausler, B.S.; Rogers, G.B. DNA extraction approaches substantially influence the assessment of the human breast milk microbiome. *Sci. Rep.* **2020**, *10*, 123. [CrossRef] [PubMed]
6. Bose, N.; Moore, S.D. Variable region sequences influence 16S rRNA performance. *Microbiol. Spectr.* **2023**, *11*, e0125223. [CrossRef] [PubMed]
7. Aja-Macaya, P.; Conde-Pérez, K.; Trigo-Tasende, N.; Buetas, E.; Nasser-Ali, M.; Nión, P.; Rumbo-Feal, S.; Ladra, S.; Bou, G.; Mira, Á.; et al. Nanopore full length 16S rRNA gene sequencing increases species resolution in bacterial biomarker discovery. *Sci. Rep.* **2025**, *15*, 26486. [CrossRef] [PubMed]
8. Curry, K.D.; Wang, Q.; Nute, M.G.; Tyshaieva, A.; Reeves, E.; Soriano, S.; Wu, Q.; Graeber, E.; Finzer, P.; Mendling, W.; et al. Emu: Species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat. Methods* **2022**, *19*, 845–853. [CrossRef] [PubMed]
9. Baldeon, M.E.; Cardenas, P.; Arevalo, V.; Prado-Vivar, B.; Uchimiya, M.; Peña, L.; Benitez, A.D.; Suárez-Jaramillo, A.; Edison, A.S.; Herrera, A.; et al. Human milk Microbiota across lactation stages and free glutamate concentrations in healthy Ecuadorian women. *Nutrients* **2025**, *17*, 805. [CrossRef] [PubMed]
10. Xu, R.; Gridneva, Z.; Payne, M.S.; Nicol, M.P.; Cheema, A.S.; Geddes, D.T.; Stinson, L.F. Longitudinal profiling of the human milk microbiome from birth to 12 months reveals overall stability and selective taxa-level variation. *Microorganisms* **2025**, *13*, 1830. [CrossRef] [PubMed]
11. Li, Y.; Ren, L.; Wang, Y.; Li, J.; Zhou, Q.; Peng, C.; Li, Y.; Cheng, R.; He, F.; Shen, X. The effect of breast milk microbiota on the composition of infant gut microbiota: A cohort study. *Nutrients* **2022**, *14*, 5397. [CrossRef] [PubMed]
12. Wan, Y.; Jiang, J.; Lu, M.; Tong, W.; Zhou, R.; Li, J.; Yuan, J.; Wang, F.; Li, D. Human milk microbiota development during lactation and its relation to maternal geographic location and gestational hypertensive status. *Gut Microbes* **2020**, *11*, 1438–1449. [CrossRef] [PubMed]
13. Demmelmair, H.; Jiménez, E.; Collado, M.C.; Salminen, S.; McGuire, M.K. Maternal and perinatal factors associated with the human milk microbiome. *Curr. Dev. Nutr.* **2020**, *4*, nzaa027. [CrossRef] [PubMed]
14. Zimmermann, P.; Curtis, N. Breast milk microbiota: A review of the factors that influence composition. *J. Infect.* **2020**, *81*, 17–47. [CrossRef] [PubMed]
15. Kumar, H.; Du Toit, E.; Kulkarni, A.; Aakko, J.; Linderborg, K.M.; Zhang, Y.; Nicol, M.P.; Isolauri, E.; Yang, B.; Collado, M.C.; et al. Distinct patterns in human milk microbiota and fatty acid profiles across specific geographic locations. *Front. Microbiol.* **2016**, *7*, 1619. [CrossRef] [PubMed]
16. Padilha, M.; Danneskiold-Samsøe, N.B.; Brejnrod, A.; Hoffmann, C.; Cabral, V.P.; Iaucci, J.d.M.; Sales, C.H.; Fisberg, R.M.; Cortez, R.V.; Brix, S.; et al. The human milk microbiota is modulated by maternal diet. *Microorganisms* **2019**, *7*, E502. [CrossRef] [PubMed]

17. Londoño-Sierra, D.C.; Mesa, V.; Guzmán, N.C.; Bolívar Parra, L.; Montoya-Campuzano, O.I.; Restrepo-Mesa, S.L. Maternal diet may modulate breast milk microbiota—A case study in a group of Colombian women. *Microorganisms* **2023**, *11*, 1812. [CrossRef] [PubMed]
18. Parra-Soto, S.; Carpio-Arias, T.V.; Rios-Castillo, I.; Pérez-Armijo, P.; Landaeta-Díaz, L.; Murillo, A.G.; Araneda-Flores, J.; Cavagnari, B.M.; Gómez, G.; Morales, G.; et al. Dietary patterns and sustainable lifestyles: A multicenter study from Latin America and Spain. *Foods* **2025**, *14*, 2065. [CrossRef] [PubMed]
19. Quevedo Olaya, J.L.; Rodríguez León, A. Interacciones nutricionales en la dieta cotidiana peruana y su impacto en la biodisponibilidad. *Rev. Científica Pakamuros* **2025**, *13*, 69–85. [CrossRef]
20. World Health Organization. *Obesity: Preventing and Managing the Global Epidemic—Report of a WHO Consultation*; WHO Technical Report Series, No. 894; World Health Organization: Geneva, Switzerland, 2000.
21. Stinson, L.F.; Geddes, D.T. Typical skin and oral bacterial species present in human milk are not the result of contamination during the sampling process. *Lett. Appl. Microbiol.* **2025**, *78*, ovaf084. [CrossRef] [PubMed]
22. Yao, C.; Zhang, Y.; You, L.; E, J.; Wang, J. Comparative analysis of three experimental methods for revealing human fecal microbial diversity. *BMC Microbiol.* **2025**, *25*, 258. [CrossRef] [PubMed]
23. Schloss, P.D. Rarefaction is currently the best approach to control for uneven sequencing effort in amplicon sequence analyses. *mSphere* **2024**, *9*, e0035423. [CrossRef] [PubMed]
24. Shen, W.; Le, S.; Li, Y.; Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **2016**, *11*, e0163962. [CrossRef] [PubMed]
25. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **2016**, *4*, e2584. [CrossRef] [PubMed]
26. Davis, N.M.; Proctor, D.M.; Holmes, S.P.; Relman, D.A.; Callahan, B.J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **2018**, *6*, 226. [CrossRef] [PubMed]
27. Gotelli, N.J.; Colwell, R.K. Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **2001**, *4*, 379–391. [CrossRef]
28. Good, I.J. The population frequencies of species and the estimation of population parameters. *Biometrika* **1953**, *40*, 237–264. [CrossRef]
29. Oksanen, J.; Simpson, G.L.; Blanchet, F.G.; Kindt, R.; Legendre, P.; Minchin, P.R.; O'Hara, R.B.; Solymos, P.; Stevens, M.H.H.; Szoecs, E.; et al. *Vegan: Community Ecology Package*, version 2.6-4; R package; CRAN: Vienna, Austria, 2022. Available online: https://CRAN.R-project.org/package=vegan (accessed on 11 October 2026).
30. Anderson, M.J. A new method for non-parametric multivariate analysis of variance: Non-parametric manova for ecology. *Austral Ecol.* **2001**, *26*, 32–46. [CrossRef]
31. Anderson, M.J. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **2006**, *62*, 245–253. [CrossRef] [PubMed]
32. Lackey, K.A.; Williams, J.E.; Meehan, C.L.; Zachek, J.A.; Benda, E.D.; Price, W.J.; Foster, J.A.; Sellen, D.W.; Kamau-Mbuthia, E.W.; Kamundia, E.W.; et al. What's normal? Microbiomes in human milk and infant feces are related to each other but vary geographically: The INSPIRE Study. *Front. Nutr.* **2019**, *6*, 45. [CrossRef] [PubMed]
33. Neu, A.T.; Allen, E.E.; Roy, K. Defining and quantifying the core microbiome: Challenges and prospects. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2104429118. [CrossRef] [PubMed]
34. Instituto Nacional de Estadística e Informática. Perú: Encuesta Demográfica y de Salud Familiar 2025—Nacional y Departamental 2025. Available online: https://www.gob.pe/institucion/inei/informes-publicaciones/8284085-peru-encuesta-demografica-y-de-salud-familiar-endes-2025 (accessed on 7 October 2026).
35. Liu, B.; Zhao, J.; Liu, Y.; Qiao, W.; Jiang, T.; Chen, L. Diversity and temporal dynamics of breast milk microbiome and its influencing factors in Chinese women during the first 6 months postpartum. *Front. Microbiol.* **2022**, *13*, 1016759. [CrossRef] [PubMed]
36. Boix-Amorós, A.; Hernández-Aguilar, M.T.; Artacho, A.; Collado, M.C.; Mira, A. Human milk microbiota in sub-acute lactational mastitis induces inflammation and undergoes changes in composition, diversity and load. *Sci. Rep.* **2020**, *10*, 18521. [CrossRef] [PubMed]
37. Sindi, A.S.; Cheema, A.S.; Trevenen, M.L.; Geddes, D.T.; Payne, M.S.; Stinson, L.F. Characterisation of human milk bacterial DNA profiles in a small cohort of Australian women in relation to infant and maternal factors. *PLoS ONE* **2023**, *18*, e0280960. [CrossRef] [PubMed]
38. Ma, G.; Peng, L.; Qin, L.; Cai, R.; Tan, X.; Gao, R. Human milk microbiota: Origins, determinants, and roles in maternal-infant microbial transmission and infant microbiome assembly. *Front. Microbiol.* **2026**, *17*, 1790225. [CrossRef] [PubMed]
39. Serra, V.V.; Teves, S.; López de Volder, A.; Osorio, F.; Aguila, N.; Armadans, M. Comparison of the risk of microbiological contamination between samples of breast milk obtained at home and at a healthcare facility. *Arch. Argent. Pediatría* **2013**, *111*, 115–119. [CrossRef] [PubMed]

40. Cabrera-Rubio, R.; Collado, M.C.; Laitinen, K.; Salminen, S.; Isolauri, E.; Mira, A. The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. *Am. J. Clin. Nutr.* **2012**, *96*, 544–551. [CrossRef] [PubMed]
41. Collado, M.C.; Laitinen, K.; Salminen, S.; Isolauri, E. Maternal weight and excessive weight gain during pregnancy modify the immunomodulatory potential of breast milk. *Pediatr. Res.* **2012**, *72*, 77–85. [CrossRef] [PubMed]
42. Cervantes-Monroy, E.; Zarzoza-Mendoza, I.C.; Canizales-Quinteros, S.; Morán-Ramos, S.; Villa-Morales, J.; López-Contreras, B.E.; Carmona-Sierra, F.V.; Rodríguez-Cruz, M. Influence of early life factors on the breast milk and fecal microbiota of mother–newborn dyads. *Microorganisms* **2024**, *12*, 2142. [CrossRef] [PubMed]
43. Commichaux, S.; Luan, T.; Muralidharan, H.S.; Pop, M. Database size positively correlates with the loss of species-level taxonomic resolution for the 16S rRNA and other prokaryotic marker genes. *PLoS Comput. Biol.* **2024**, *20*, e1012343. [CrossRef] [PubMed]
44. Xie, Y.; Lu, Y.; Jiang, Z.; Dong, W.; Liu, Y.; Wang, Y.; Shao, J.; Yan, W.; Chai, Z. Dynamics and drivers of the human breast milk microbiome across lactation stages. *Front. Nutr.* **2026**, *13*, 1861294. [CrossRef] [PubMed]
45. Matchado, M.S.; Rühlemann, M.; Reitmeier, S.; Kacprowski, T.; Frost, F.; Haller, D.; Baumbach, J.; List, M. On the limits of 16S rRNA gene-based metagenome prediction and functional profiling. *Microb. Genom.* **2024**, *10*, 001203. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.