Reproductive Tract Microbial Transitions from Late Gestation to Early Postpartum Using 16S rRNA Metagenetic Profiling in First-Pregnancy Heifers

Shaked Druker [1,2], Ron Sicsic [1], Shachar Ravid [1], Shani Scheinin [1,2] and Tal Raz [1,3,*]

[1] Koret School of Veterinary Medicine, Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, Israel.
[2] Hachaklait, Mutual Society for Veterinary Services, Caesarea Industrial Park, Caesarea, Israel
[3] Advanced Academic Programs, Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore, MD, USA.
* Correspondence: tal.raz@mail.huji.ac.il

Abstract: Studies in recent years indicate that reproductive tract microbial communities are crucial for shaping mammals' health and reproductive outcomes. Following parturition, uterine bacterial contamination often occurs due to the open cervix, which may lead to postpartum uterine inflammatory diseases, especially in primiparous individuals. However, investigations into spatio-temporal microbial transitions in the reproductive tract of primigravid females remain limited. Our objective was to describe and compare the microbial community compositions in the vagina at late gestation and in the vagina and uterus at early postpartum in first-pregnancy heifers. Three swab samples were collected from 33 first-pregnancy Holstein-Friesian heifers: one vaginal sample at gestation day 258±4, and vaginal and uterine samples at postpartum day 7±2. Each sample underwent 16S rRNA V4 region metagenetic analysis via Illumina MiSeq, with bioinformatics following Mothur MiSeq SOP. The reproductive tract bacterial communities were assigned to 1,255 genus-level OTUs across 30 phyla. Dominant phyla, accounting for approximately 90% of the communities, included Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes, and Fusobacteria. However, the results revealed distinct shifts in microbial composition between the prepartum vagina (Vag-pre), postpartum vagina (Vag-post), and postpartum uterus (Utr-post). The Vag-pre and Utr-post microbial profiles were the most distinct. The Utr-post group had lower relative abundances of Proteobacteria but higher abundances of Bacteroidetes, Fusobacteria, and Tenericutes compared to Vag-pre, while Vag-post displayed intermediate values for these phyla, suggesting a transitional profile. Additionally, the Utr-post group exhibited lower bacterial richness and diversity compared to both Vag-pre and Vag-post. The unsupervised probabilistic Dirichlet Multinomial Mixtures model identified two distinct community types: most Vag-pre samples clustered into one type and Utr-post samples into another, while Vag-post samples were distributed evenly between the two. LEfSe analysis revealed distinct microbial profiles at the genus level. Overall, specific microbial markers were associated with anatomical and temporal transitions, revealing a dynamic microbial landscape during the first pregnancy and parturition. These differences highlight the complexity of these ecosystems and open new avenues for research in reproductive biology and microbial ecology.

Keywords: 16S rRNA next-generation sequencing; uterus; vagina; pregnancy; primigravida; cow

1. Introduction

The intricate interplay between microbial communities and mammals' reproductive physiologic and pathologic dynamics has garnered significant attention in recent years [1-3]. Notably, the transition from late pregnancy to the early postpartum period, characterized by the opening of the cervix, presents a scenario where substantial uterine bacterial contamination commonly occurs, either from the vaginal microbiota or perhaps from the surrounding environment [4, 5]. This phenomenon is particularly pertinent for primigravid females, as this is their first-ever parturition. In dairy cows, postpartum uterine contamination has been documented in nearly 100% of animals [6-8], and the same was suggested in humans and other mammals [1, 9-11]. Moreover, in dairy cows and women, postpartum uterine inflammatory diseases are more prevalent among primiparous females than multiparous females [12-14]. In dairy cows, traditional culture-dependent microbiological analyses have identified several microbial species in the postpartum uteri, including *Escherichia coli*, *Trueperella pyogenes*, *Fusobacterium necrophorum*, *Prevotella melaninogenica*, *Bacteroides* spp., *Pseudomonas* spp., *Streptococcus* spp., and *Staphylococcus* spp., some of which have been associated with metritis [6, 7, 15-17]. However, recent years have witnessed advancements in culture-independent investigations of bacterial communities within postpartum bovine uteri, mainly next-generation sequencing (NGS) of 16S-rDNA amplicon libraries [2, 7, 8, 18]. These approaches have yielded findings that differ from culture-based studies and have enabled the identification of bacteria that are challenging to cultivate through traditional methods. Furthermore, these studies demonstrated that Koch's postulates are not fulfilled in postpartum uterine diseases [19].

Current research has primarily examined postpartum uterine microbial contamination in dairy cows, particularly concerning metritis. Nevertheless, there is a significant gap in our understanding regarding the transition of microbial communities from late pregnancy to early postpartum, especially in the context of first-pregnancy heifers. This is particularly noteworthy, not just because of the high susceptibility of primiparous cows to uterine inflammatory diseases, but also because the first pregnancy presents a distinctive scenario, offering an unparalleled opportunity to investigate the intricate dynamics of reproductive tract physiology and pathology, along with the environment, in what is probably the first most dramatic contamination of the female reproductive tract. Accordingly, the objective of this study was to describe and compare the community compositions in the vagina at late gestation and in the vagina and uterus at early postpartum, in first-pregnancy heifers using r16S amplicon NGS data.

2. Results

A total of 99 samples were obtained from 33 first-pregnancy heifers. Heifers included in the study were those with singleton pregnancies, normal spontaneous vaginal delivery, and no history of pregnancy-related illness during gestation or antibiotic treatments within six weeks before sampling. Moreover, heifers were excluded if they had retained fetal membranes or signs of postpartum septic metritis, as detailed in section 5.1. Three swab samples were collected from each animal: one from the vagina ('Vag-pre') at a

median gestation day of 258 (range 254-262 days), and two postpartum samples, from the vagina ('Vag-post') and the uterus ('Utr-post'), at 8 days (median) after calving (range 5-9 days). The DNA from each sample was used for metagenetic analysis of the V4 hypervariable region of the 16s rRNA gene via the Illumina MiSeq platform [20-22]. Bioinformatic analysis was performed predominantly according to Mothur MiSeq SOP [23]. In this analysis, a total of 3,887,476 sequences were initially identified, of which, 3,772,145 sequences (97%) remained after screening based on length criteria. Following deduplication, 2,214,729 unique sequences were identified and aligned to the customized SILVA reference database targeting the V4 region of the 16S rRNA gene. Then, an additional screening step resulted in 1,966,847 unique sequences (total number of sequences = 3,175,973), which were filtered to remove overhangs at both ends and underwent another deduplication step, resulting in 1,966,162 unique sequences. Following denoising by pre-clustering and chimeras removal, 1,463,757 unique sequences (total number of sequences = 3,090,757) were selected to be classified taxonomically. After identifying and removing undesirable non-bacterial sequences, 1,404,486 high-quality unique sequences remained (total number of sequences = 2,985,052), with an average read length of 291 bp.

The high-quality unique sequences were assigned to 1,255 genus-level OTUs. All OTUs were taxonomically assigned to the bacteria domain and were from 30 phyla, 88 classes, 193 orders, 413 families, and 1255 genera. At the per-sample level, the vast majority of samples had >10,000 sequences assigned to OTUs; however, 19 samples had a low sequence count (≤210). Examination of the rarefaction curves (Figure 1) demonstrated that these samples were unrepresentative and excluded from further analyses. Samples that remained in the downstream analyses had a high number of sequences (mean±SEM: 37,302±1760, Median: 41,244, Range: 10,240-78,602), and their rarefaction curves reached a plateau, indicating an adequate sequencing depth. The number of sequences was significantly higher among samples in the Vag-pre and Vag-post groups, compared to the Utr-post group (41,080±2,540, 41,080±3,397, 27,722±2,578 sequences, respectively; P=0.0007, Kruskal-Wallis One-Way AOV test). Wherever relevant, downstream analyses were performed with a subsampling of 10,000 sequences from each sample.

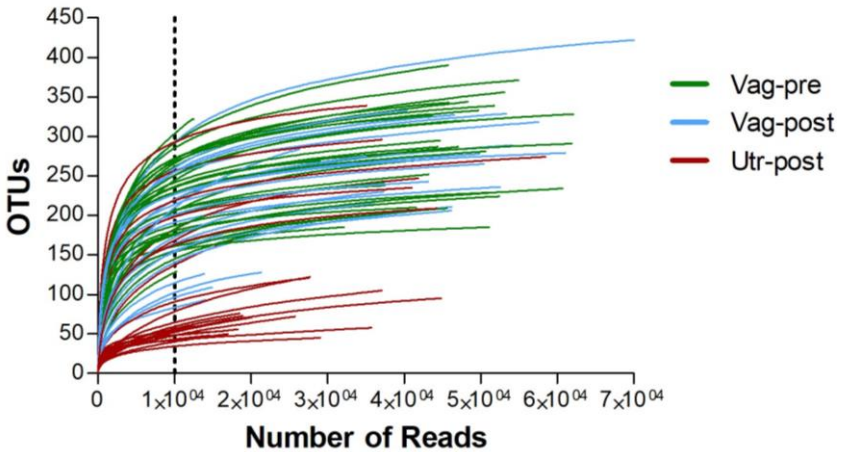

Figure 1. Rarefaction curves of 16S sequencing samples. Curves show the number of identified OTUs (Y-axis) as a function of the number of reads (X-axis). Samples from the Vag-pre, Vag-post, and Utr-post groups are represented in green, light blue, and red, respectively. Only samples with>10,000 reads (marked by the vertical dotted line) were included in further analyses.

The bacterial community compositions by phylum in individual samples are presented in Figure 2. The mean bacterial community compositions by phylum in the Vag-pre, Vag-post, and Utr-post groups are shown in Figure 3 and detailed in the corresponding Table 1. Regardless of the group, the five most dominant phyla by mean relative abundance were Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes, and Fusobacteria, accounting for approximately 90% of the taxonomically assigned community components. As listed in Table 1, a comparison of mean bacterial community composition according to phylum revealed significant differences, mainly between the Vag-pre and Utr-post groups, particularly in the phyla Proteobacteria, which was lower, and Bacteroidetes, Fusobacteria, and Tenericutes, which were higher in the Utr-post group. The Vag-post group had relative abundance values that were commonly in between the values of the Vag-pre and Utr-post groups.

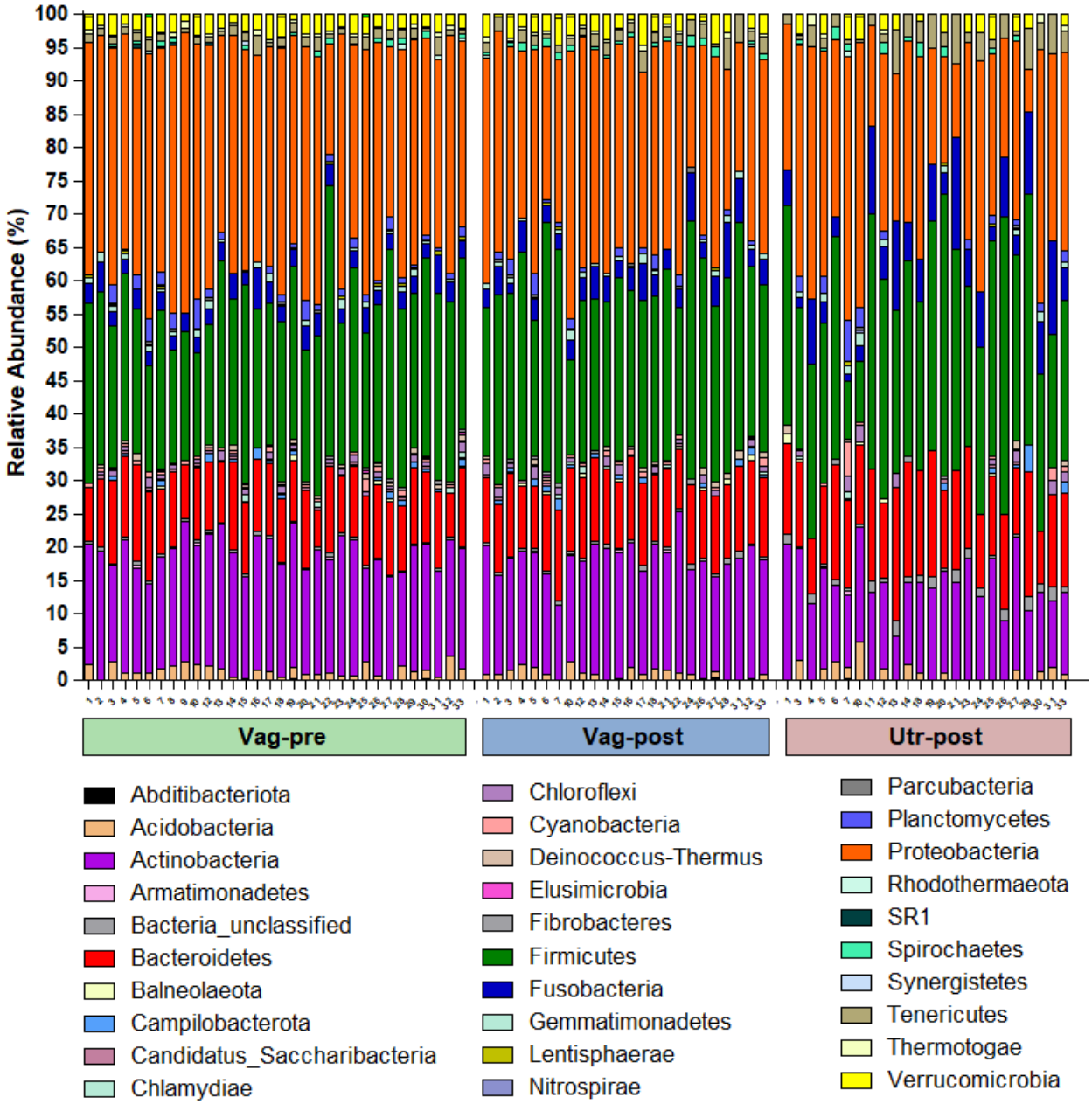

Figure 2. Individual stacked bar plots of bacterial community composition according to phylum. Each bar represents one sample, and the cow identity number is noted below the X-axis.

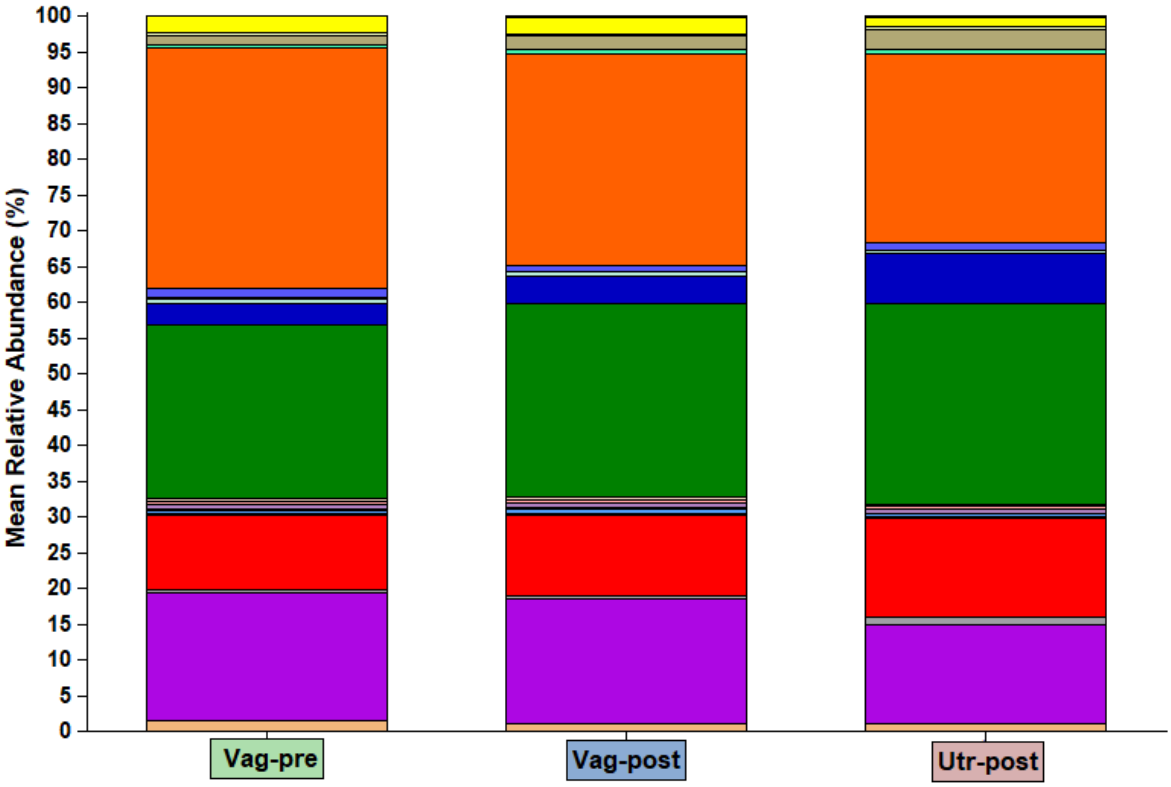

Figure 3. Group stacked bar plots of mean bacterial community composition according to phylum in the Vag-pre, Vag-post, and Utr-post groups. The color of each phylum is presented in Figure 2. The mean relative abundance values and the statistical comparisons among the groups are detailed in Table 1.

The sample coverage, number of observed OTUs, and Inverse Simpson diversity estimate parameters (average, 95% CI width, and Std) were evaluated for each sample. The analysis was standardized by subsampling 10,000 sequences from each sample 1,000 times, and the results were compared among the Vag-pre, Vag-post, and Utr-post groups. As illustrated in Figure 4, the sample coverages were all above 99%, indicating that all samples are highly representative. The mean sample coverage was slightly but significantly higher in the Utr-post group (99.57±0.03%, 99.56±0.03%, 99.73±0.03%; P=0.0002). The mean number of defined OTUs did not differ between the Vag-pre and Vag-post groups; however, it was significantly lower in the Utr-post group (221.7±8.4, 189.1±11.6, 103.7±15.8 OTUs, P<0.0001). Furthermore, the Inverse Simpson diversity estimate parameters (average, 95%CI width, and Std), also known as 1/D, did not differ between the Vag-pre and Vag-post groups but were significantly lower in the Utr-post group (1/D average: 11.8±0.8, 12.7±1.1, 6.9±1.3, P<0.0001; 1/D 95%CI width: 0.91±0.07, 0.84±0.08, 0.38±0.08, P<0.0001; 1/D Std: 0.19±0.02, 0.18±0.02, 0.08±0.02, P<0.0001). Overall, these results indicated that the postpartum uterus had lower bacterial richness and diversity compared to the vagina before and after parturition.

Table 1. Mean relative abundance of different phyla among the Vag-pre, Vag-post, and Utr-post groups.

Phylum	Vag-pre	Vag-post	Utr-post	*P* value
Proteobacteria	33.66 ± 0.88[a]	29.57 ± 1.10[b]	26.46 ± 1.87[b]	0.0007
Firmicutes	24.25 ± 0.93	26.90 ± 1.22	28.15 ± 1.80	0.0560
Actinobacteria	17.86 ± 0.41	17.46 ± 0.51	13.70 ± 0.71	0.9373
Bacteroidetes	10.57 ± 0.34[a]	11.36 ± 0.27[a]	13.85 ± 0.66[b]	<0.0001
Fusobacteria	2.92 ± 0.17[a]	3.83 ± 0.33[ab]	6.89 ± 0.91[b]	<0.0001
Verrucomicrobia	2.28 ± 0.12	2.33 ± 0.26	1.49 ± 0.28	0.0969
Acidobacteria	1.43 ± 0.16	0.99 ± 0.16	1.13 ± 0.29	0.1318
Tenericutes	1.21 ± 0.11[a]	1.80 ± 0.21[b]	2.96 ± 0.43[b]	0.0002
Planctomycetes	1.18 ± 0.19	0.75 ± 0.18	0.90 ± 0.30	0.1083
Chloroflexi	0.60 ± 0.08	0.69 ± 0.13	0.58 ± 0.19	0.3682
Gemmatimonadetes	0.59 ± 0.08	0.60 ± 0.08	0.43 ± 0.12	0.2451
Thermotogae	0.43 ± 0.04[a]	0.36 ± 0.05[a]	0.22 ± 0.09[b]	0.0006
Cyanobacteria/Chloroplast	0.43 ± 0.06[a]	0.38 ± 0.07[ab]	0.41 ± 0.22[b]	0.0145
Campilobacterota	0.41 ± 0.08	0.44 ± 0.10	0.34 ± 0.19	0.0363
Bacteria_unclassified	0.39 ± 0.02[a]	0.47 ± 0.04[a]	1.15 ± 0.13[b]	<0.0001
Deinococcus-Thermus	0.34 ± 0.06[a]	0.41 ± 0.08[a]	0.16 ± 0.08[b]	0.0021
Spirochaetes	0.33 ± 0.07	0.66 ± 0.09	0.49 ± 0.14	0.0401
Candidatus_Saccharibacteria	0.19 ± 0.03[a]	0.11 ± 0.04[ab]	0.07 ± 0.03[b]	0.0429
Lentisphaerae	0.16 ± 0.03	0.06 ± 0.03	0.08 ± 0.04	0.0297
Chlamydiae	0.12 ± 0.05	0.13 ± 0.05	0.05 ± 0.04	0.3195
Fibrobacteres	0.11 ± 0.03	0.12 ± 0.05	0.10 ± 0.05	0.5407
Balneolaeota	0.10 ± 0.04	0.20 ± 0.06	0.14 ± 0.07	0.2824
Parcubacteria	0.08 ± 0.03	0.14 ± 0.04	0.04 ± 0.03	0.1289
Rhodothermaeota	0.08 ± 0.04	0.04 ± 0.03	0.04 ± 0.04	0.5798
SR1	0.06 ± 0.03	0.00 ± 0.00	0.01 ± 0.01	0.0648
Armatimonadetes	0.06 ± 0.03	0.06 ± 0.03	0.04 ± 0.03	0.7827
Elusimicrobia	0.04 ± 0.02	0.02 ± 0.02	0.00 ± 0.00	0.1508
Synergistetes	0.03 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	0.2210
Abditibacteriota	0.02 ± 0.01	0.02 ± 0.02	0.01 ± 0.01	0.9373
Nitrospirae	0.01 ± 0.01	0.00 ± 0.00	0.02 ± 0.02	0.6310

[1] Data is related to Figure 3. Mean relative abundance (%) is presented as mean±SEM; Conditional formatting (i.e., blue horizontal column bars) was applied to mean values to visually represent the data and highlight key differences.

[a,b,c] Different letters above values within a line represent significant differences between groups.

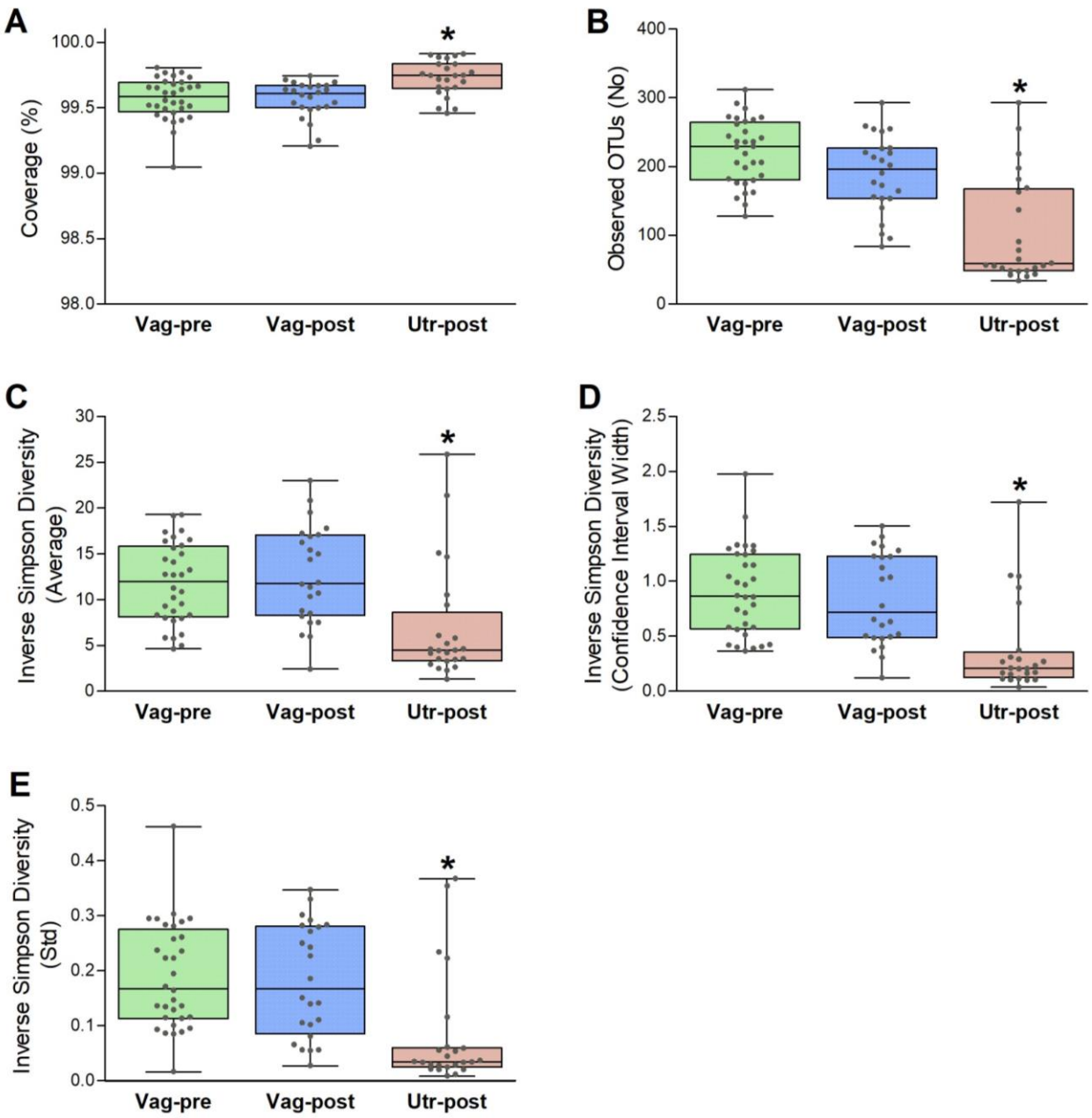

Figure 4. Alpha diversity metrics**.** The sample coverage, the number of observed OTUs, and the Inverse Simpson diversity estimate parameters (average, 95%CI width, and Std) were evaluated for each sample [23]; the analysis was standardized by subsampling 10,000 sequences from each sample, 1,000 times, and the results were compared among the Vag-pre, Vag-post, and Utr-post groups. Box and whisker plots are presented with the overlap of scatter plots. The median (central white line) is shown in the middle of each box, with the 25% and 75% percentile ranges (box depth) and the maximum and minimum (whiskers); each dot represents a value of a single sample. Kruskal–Wallis one-way AOV, followed by Dunn's all-pairwise comparisons. * $P \leq 0.0002$.

The non-metric multidimensional scaling (NMDS) ordination [24] of the bacterial communities from the Vag-pre, Vag-post, and Utr-post samples is shown in Figure 5 (3-dimensions; Stress=0.197; Coefficient of Determination, R^2=0.77). To explore the observed spatial separation of community structures, the AMOVA test was applied, indicating that the centers of the clouds representing each group significantly differ from each other (P<0.001). The HOMOVA test analysis revealed that the bacterial community

variations among samples in the Vag-pre and Vag-post groups did not differ (P=0.099), but it was significantly higher in the Utr-post group (SSwithin/(Ni-1) values = 0.170, 0.208, 0.325, respectively, P<0.001; Bonferroni pairwise error rate = 0.0167).

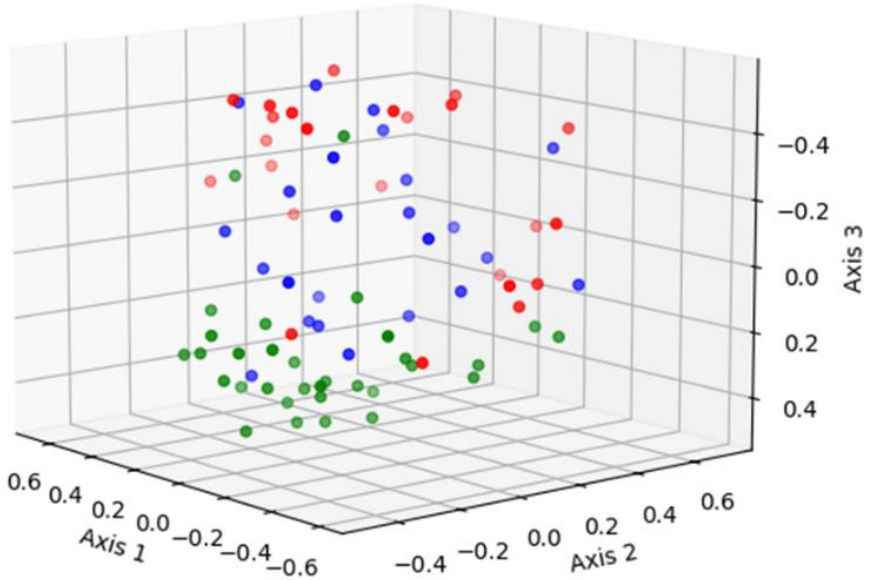

Figure 5. Non-metric Multidimensional Scaling (NMDS) ordination plot of the bacterial communities composition from the Vag-pre (green), Vag-post (blue), and Utr-post (red) samples (3-dimensions; Stress=0.197; R^2=0.77).

The comparative analysis among the Vag-pre, Vag-post, and Utr-post groups, utilizing non-parametric LEfSe analysis, uncovered distinct microbial profiles associated with each anatomical niche (Figures 6-8, Tables S1-3). In the comparison between Vag-pre and Vag-post, 65 discriminant features were identified (Figure 6, Table S1). Notable genera with the highest logMaxMean values (>4) and LDA scores (>3.5) in the Vag-pre group included *Trinickia, Micrococcaceae_unclassified, Intrasporangiaceae_unclassified, Staphylococcus,* and *Betaproteobacteria_unclassified*, among others. On the other hand, Vag-post exhibited higher logMaxMean values for *Fusobacterium*, *Leptotrichiaceae_unclassified*, *Histophilus*, *Mycoplasmopsis*, *Parvimonas*, *Bacteroidales_unclassified*, and *Fusobacteriaceae_unclassified*, among others. When comparing Vag-pre with Utr-post, 113 discriminant features were uncovered (Figure 7, Table S2). Significant top genera in the Vag-pre group included *Trinickia*, *Micrococcaceae_unclassified*, *Intrasporangiaceae_unclassified*, *Ruminococcaceae_unclassified*, *Glutamicibacter*, *Micrococcales_unclassified*, and *Streptococcus*, among others; while Utr-post showed higher logMaxMean values and LDA scores for *Gemella*, *Leptotrichiaceae_unclassified*, *Trueperella*, and *Streptobacillus*, among others. Lastly, in the comparison between the postpartum vagina and uterus, 71 discriminant features were identified (Figure 8, Table S3). Genera such as *Trinickia*, *Leptotrichiaceae_unclassified, Micrococcaceae_unclassified, Ruminococcaceae_unclassified, Fusobacteriaceae_unclassified* and *Clostridiales_unclassified* had the highest logMaxMean values and LDA scores in the Vag-post group, while *Trueperella*, *Streptococcus*, and *Peptoniphilus* were more prominent in the Utr-post group. The distinctive genera identified in each of these comparisons provide insights into the specific microbial dynamics associated with these anatomical niches.

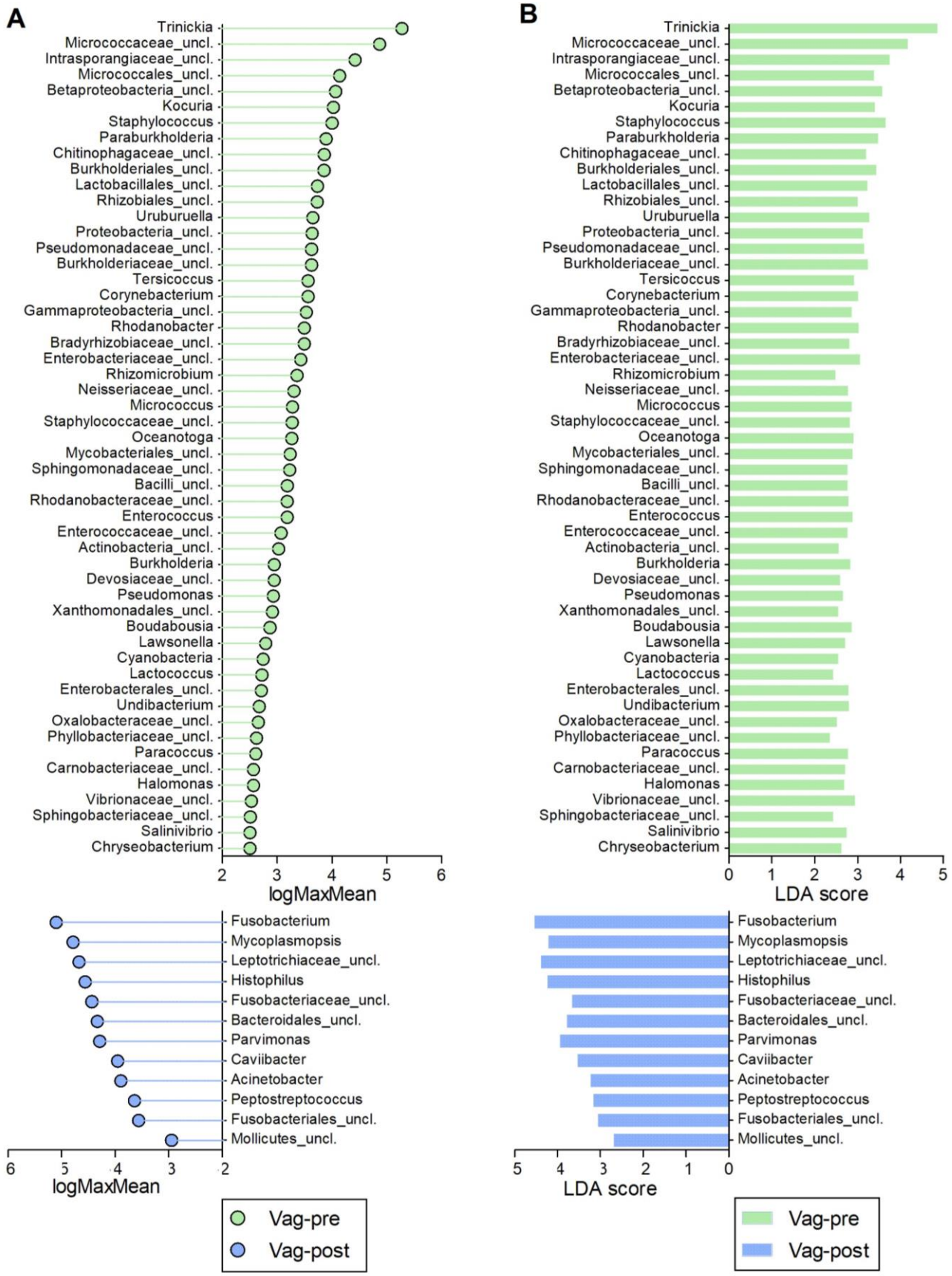

Figure 6. Discriminative genera obtained from LEfSe comparison between Vag-pre and Vag-post samples. (A) Lollipop chart of the logMaxMean values, indicating the logarithm value of the maximum mean abundance of the OTU across the specified groups. (B) Corresponding bar chart of the Linear Discriminant Analysis (LDA) scores. $P<0.05$ for all values. Vag-pre is indicated in green, and Vag-post is indicated in blue. Corresponding data are detailed in Table S1.

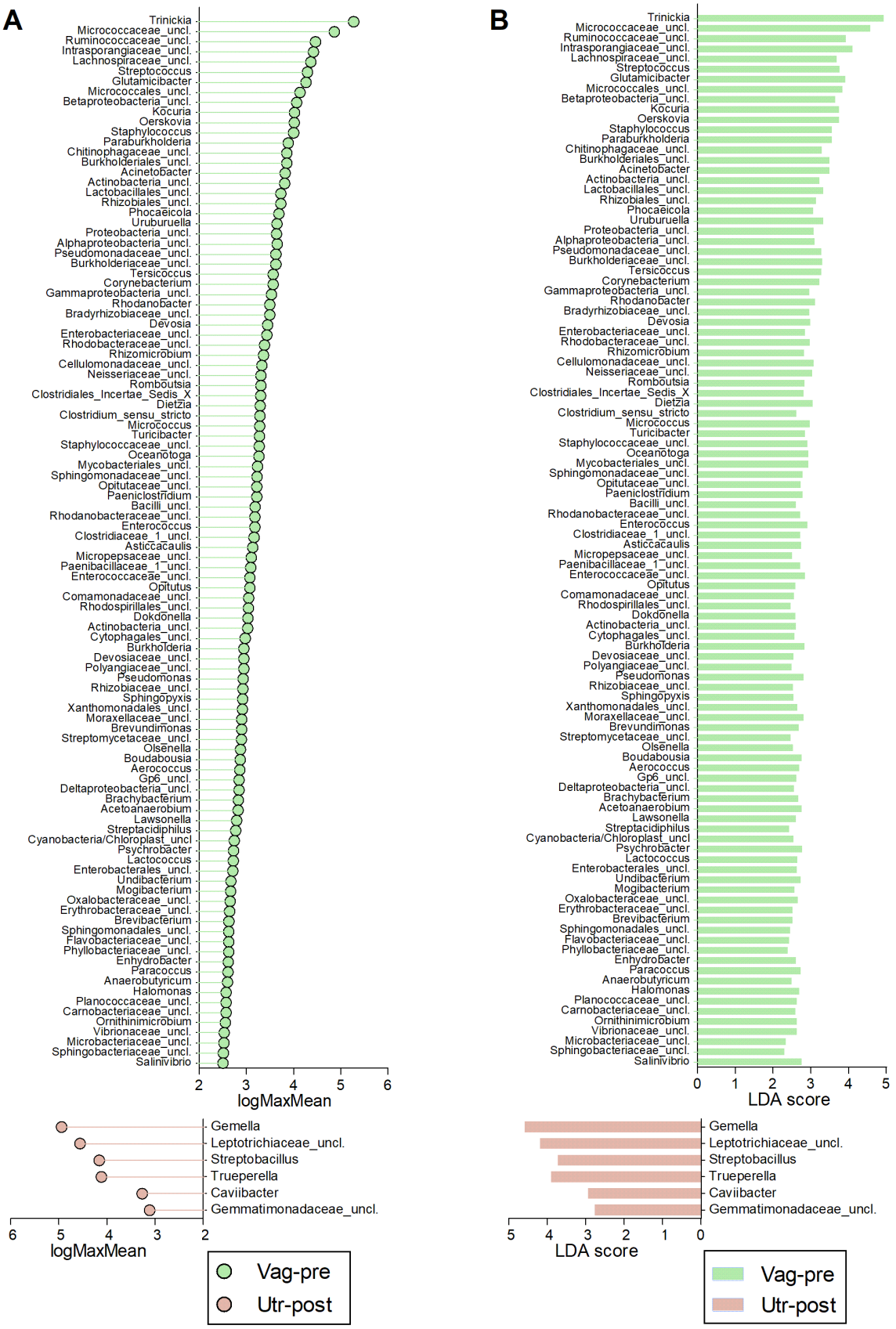

Figure 7. Discriminative genera obtained from LEfSe comparison between Vag-pre and Utr-post samples. (A) Lollipop chart of the logMaxMean values, indicating the logarithm value of the maximum mean abundance of the OTU across the specified groups. (B) Corresponding bar chart of the Linear Discriminant Analysis (LDA) scores. $P<0.05$ for all values. Vag-pre is indicated in green, and Utr-post is indicated in red. Corresponding data are detailed in Table S2.

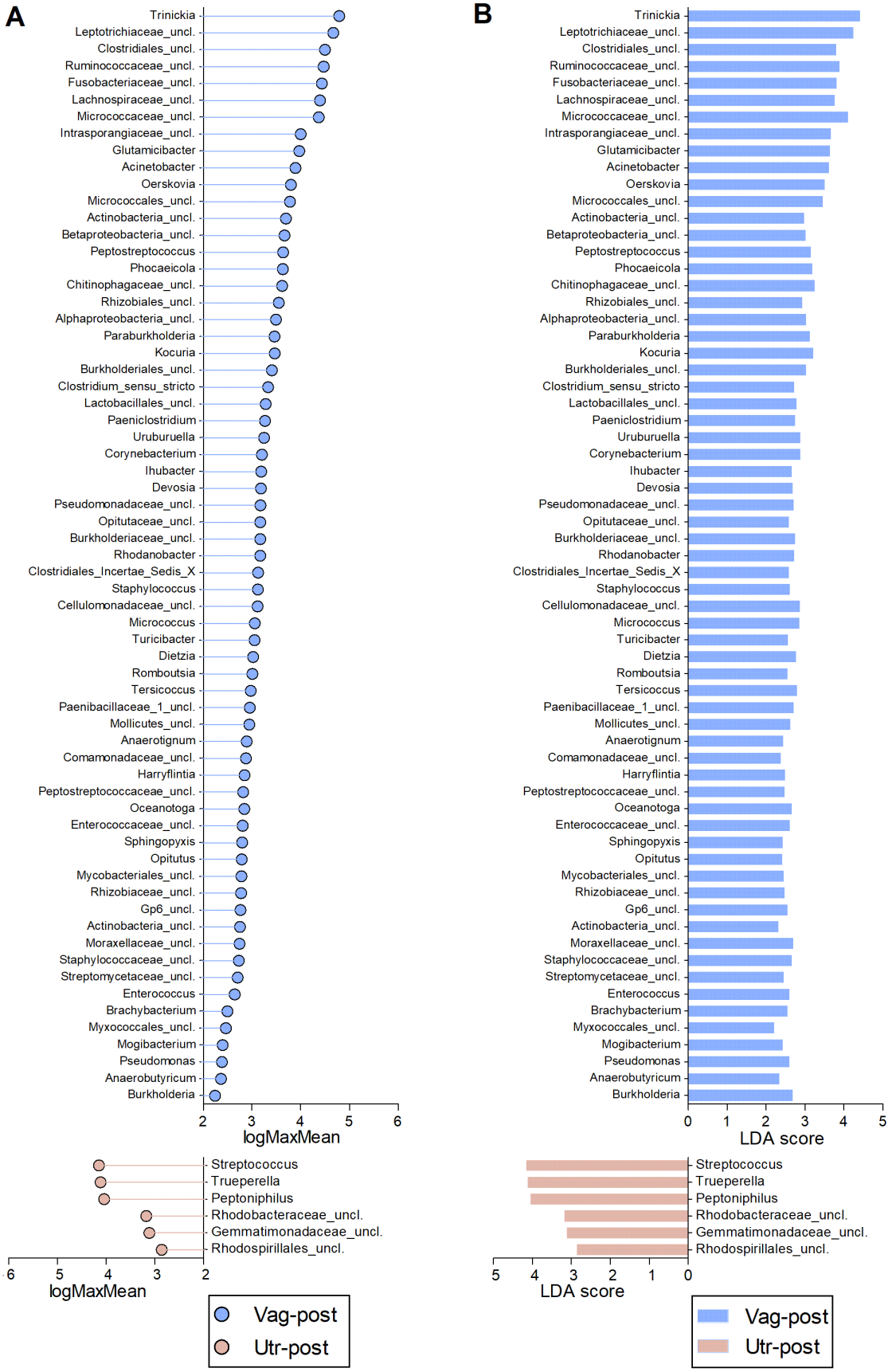

Figure 8. Discriminative genera obtained from LEfSe comparison between Vag-post and Utr-post samples. (A) Lollipop chart of the logMaxMean values, indicating the logarithm value of the maximum mean abundance of the OTU across the specified groups. (B) Corresponding bar chart of the Linear Discriminant Analysis (LDA) scores. $P<0.05$ for all values. Vag-post is indicated in blue, and Utr-post is indicated in red. Corresponding data are detailed in Table S3.

The probabilistic Dirichlet multinomial mixtures (DMM) model analysis revealed a minimum Laplace value at K=2 (Figure 9), indicating that this unsupervised analysis segregates the samples into two distinct community types based on their inherent community structures [25]. Of the 80 samples included in the analysis, 34 were allocated to one community type (community P1) and 46 to a second community (community P2). Interestingly, the vast majority of the samples from the Utr-post group (20/24, 83.3%) were segregated to community type P1, the vast majority of samples from the Vag-pre group (29/32, 90.6%) were segregated to community type P2, while the samples from Vag-post were divided quite evenly between the two (P1: 11/24, 45.8%; P2: 13/24, 54.2%).

The first 36 genera that were most responsible for separating the samples into the two community types in the DMM model analysis are presented in Table 2. This list of top genera represents the cumulative 50% proportion of the total abundance. Community type P1 predominantly includes the majority of Utr-post samples (83.3%), with notable genera such as *Histophilus* (2.73%), *Helcococcus* (2.77%), *Fusobacterium* (5.34%), *Firmicutes_unclassified* (1.39%), *Micrococcaceae_unclassified* (0.67%), *Fusobacteriaceae_unclassified* (1.98%), and *Bacteroides* (1.55%) showing higher relative abundances. In contrast, Community type P2 predominantly includes the majority of Vag-pre samples (90.6%), with significantly higher relative abundances of genera such as *Trinickia* (12.82%), *Micrococcaceae_unclassified* (9.11%), *Firmicutes_unclassified* (1.10%), *Ruminococcaceae_unclassified* (2.07%), *Lachnospiraceae_unclassified* (1.51%), *Bacteroides* (0.79%), *Betaproteobacteria_unclassified* (1.33%), and *Peptoniphilus* (1.02%). Key differences include the stark contrast in *Trinickia*, which has a mean abundance of 12.82% in P2 compared to 1.04% in P1, highlighting its association with the Vag-pre group, while *Histophilus* and *Helcococcus* are significantly more abundant in P1, indicating their association with the Utr-post group. Additionally, *Fusobacterium* shows a higher mean abundance in P1 (5.34%) compared to P2 (4.01%), and genera such as *Ruminococcaceae_unclassified*, *Lachnospiraceae_unclassified*, and *Peptoniphilus* are more abundant in P2, suggesting their association with the Vag-pre group. Overall, these findings suggest a distinct microbial profile associated with the prepartum vaginal environment (associated mainly with community type P2), a unique postpartum uterine microbial signature (associated mainly with community type P1), and two communities that imply a transitional phase or variability in the postpartum vaginal microbiome.

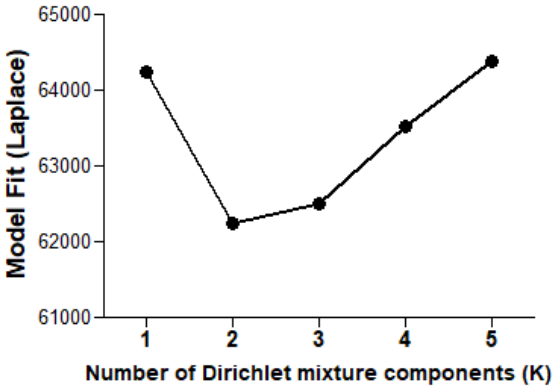

Figure 9. Model fit using the probabilistic Dirichlet Multinomial Mixtures (DMM) model analysis. The Y-axis represents the Laplace values as the number of Dirichlet mixture components K increases.

Table 2. Relative abundance of different genera in two distinct community types (P1 and P2) defined by unsupervised probabilistic Dirichlet Multinomial Mixtures (DMM) model analysis [1].

Genus	All samples mean	Community Type P1 mean (95%CI)	Community Type P2 mean (95%CI)	Difference (P1 vs.P2)	Cumulative Fraction
Trinickia	2.89	1.04 (0.71-1.51)	12.82 (11.8-13.9)	11.78	0.16
Histophilus	1.29	2.73 (2.01-3.71)	0.67 (0.52-0.86)	2.06	0.18
Micrococcaceae_uncl.	1.09	0.67 (0.43-1.03)	2.56 (2.19-3.01)	1.9	0.21
Helcococcus	1.53	2.77 (2.01-3.82)	0.90 (0.72-1.13)	1.87	0.24
Fusobacterium	4.55	5.34 (4.08-6.98)	4.01 (3.52-4.56)	1.33	0.25
Firmicutes_uncl.	1.39	2.23 (1.64-3.04)	0.95 (0.76-1.18)	1.29	0.27
Intrasporangiaceae_uncl.	0.82	0.54 (0.34-0.85)	1.61 (1.33-1.94)	1.07	0.28
Ureaplasma	1.01	1.57 (1.09-2.26)	0.64 (0.50-0.82)	0.93	0.30
Mycoplasmopsis	1.19	1.69 (1.16-2.46)	0.83 (0.66-1.05)	0.87	0.31
Ruminococcaceae_uncl.	1.40	1.21 (0.84-1.74)	2.07 (1.75-2.46)	0.86	0.32
Fusobacteriaceae_uncl.	1.46	1.98 (1.42-2.75)	1.16 (0.94-1.42)	0.82	0.33
Lachnospiraceae_uncl.	1.22	1.00 (0.67-1.47)	1.81 (1.52-2.17)	0.82	0.34
Fusobacteriales_uncl.	0.63	1.16 (0.81-1.67)	0.34 (0.26-0.45)	0.82	0.35
Bacteroidales_uncl.	1.08	1.55 (1.09-2.20)	0.79 (0.63-1.00)	0.76	0.36
Bacteria_uncl.	5.55	6.12 (4.78-7.83)	5.41 (4.83-6.06)	0.71	0.37
Betaproteobacteria_uncl.	0.82	0.65 (0.43-0.98)	1.33 (1.09-1.62)	0.68	0.38
Peptoniphilus	0.85	1.25 (0.88-1.78)	0.62 (0.48-0.80)	0.63	0.39
Parvimonas	0.49	0.89 (0.58-1.35)	0.27 (0.2-0.38)	0.61	0.40
Caviibacter	0.25	0.67 (0.44-1.01)	0.09 (0.06-0.15)	0.58	0.41
Bacteroidetes_uncl.	1.43	1.77 (1.27-2.48)	1.27 (1.04-1.55)	0.5	0.41
Bacteroides	0.51	0.82 (0.54-1.23)	0.32 (0.23-0.43)	0.5	0.42
Paraburkholderia	0.56	0.35 (0.20-0.59)	0.83 (0.66-1.05)	0.48	0.43
Erysipelotrichaceae_uncl.	0.57	0.86 (0.57-1.30)	0.38 (0.28-0.51)	0.48	0.43
Clostridia_uncl.	0.38	0.69 (0.47-1.02)	0.22 (0.15-0.30)	0.48	0.44
Mediterranea	0.30	0.62 (0.41-0.95)	0.15 (0.10-0.22)	0.47	0.44
Leptotrichiaceae_uncl.	1.00	1.29 (0.90-1.85)	0.82 (0.65-1.04)	0.47	0.45
Gemella	0.85	1.11 (0.74-1.66)	0.64 (0.50-0.83)	0.47	0.46
Peptoniphilaceae_uncl.	0.24	0.54 (0.35-0.85)	0.11 (0.07-0.16)	0.44	0.46
Micrococcales_uncl.	0.71	0.58 (0.37-0.90)	1.00 (0.80-1.25)	0.42	0.47
Pasteurellaceae_uncl.	0.34	0.57 (0.36-0.90)	0.19 (0.14-0.27)	0.38	0.47
Porphyromonas	1.09	1.33 (0.91-1.94)	0.96 (0.76-1.21)	0.37	0.48
Burkholderiales_uncl.	0.61	0.47 (0.29-0.74)	0.83 (0.66-1.04)	0.36	0.48
Mollicutes_uncl	0.18	0.44 (0.27-0.71)	0.08 (0.05-0.13)	0.36	0.49
Prevotella	0.27	0.50 (0.31-0.80)	0.14 (0.10-0.21)	0.35	0.49
Clostridiales_uncl.	2.04	2.37 (1.74-3.24)	2.02 (1.71-2.40)	0.35	0.50
Acinetobacter	0.54	0.40 (0.25-0.66)	0.74 (0.58-0.94)	0.33	0.50

[1] The DMM analysis segregated most Utr-post samples (83.3%) to community P1, most Vag-pre samples (90.6%) to community P2, while Vag-post samples were divided between P1 (45.8%) and P2 (54.2%). Relative abundances (%) of the first 36 genera most responsible for separating the samples into these two community types are presented (as mean and 95%CI). Data presented also include the relative abundances of all samples before segregation, the difference between P1's and P2's relative abundances, and the cumulative proportion of the total abundance explained by the listed genera. Conditional formatting (i.e., blue horizontal column bars) was applied to mean values to represent the data visually. *uncl.*, unclacified.

3. Discussion

The results of this study highlight the distinct transition of microbial communities in the reproductive tract of first-pregnancy heifers, from late pregnancy to early postpartum, focusing on the vagina and uterus. The study addressed a significant gap in understanding the dynamics of microbial colonization during this critical period, as the uterus is substantially contaminated. This is particularly interesting in primiparous cows, as the first parturition is a once-in-a-lifetime event, most likely characterized by the first and most massive microbial invasion into the relatively naïve uterus. Furthermore, our study may provide insights into the increased susceptibility of primiparous females to uterine inflammatory diseases. By employing next-generation sequencing of 16S-rDNA amplicon libraries, we aimed to comprehensively characterize the bacterial communities and shed light on their changes across anatomical sites and time points. Samples were collected from the vagina at gestation day 258±4 and from the vagina and uterus at postpartum day 7±2 to align with our study objectives and the gestation period of Holstein Friesian cows. With an average gestation length of 280 days (typical range of 270-290 days), sampling at day 258±4 was expected to capture the vaginal microbiota before major delivery-associated changes, establishing a baseline for late gestation before the cervix opens [26-28]. The postpartum sampling at day 7±2 allowed observation of early microbial colonization and shifts following parturition, a period marked by uterine involution and new microbial establishment. This timing is consistent with several previous studies focused on the postpartum uterine microbiome in dairy cows [8, 18, 29, 30]. Our results revealed notable shifts in microbial composition between the prepartum vagina (Vag-pre), postpartum vagina (Vag-post), and postpartum uterus (Utr-post), indicated by several analytic approaches. Notably, the postpartum uterus exhibited lower richness and diversity compared to the prepartum and postpartum vagina, indicating a dynamic microbial landscape associated with parturition. The lower diversity and richness in the Utr-post group suggest a selective environment in the postpartum uterus, possibly due to the luminal secretions, immune response, or other physiological changes occurring during and after parturition. These findings align with previous findings in dairy cows and women [1, 6, 8, 31, 32], suggesting an impact of delivery on uterine microbial colonization.

The distinct endocrine environment during late gestation and early postpartum may have significantly influenced the microbial community composition in the reproductive tract in our study [33]. During late pregnancy, elevated levels of progesterone and estrogen modulate the vaginal and uterine environments, impacting microbial colonization and community structure [33-35]. Progesterone, in particular, may promote the growth of certain bacteria while suppressing others, leading to a distinctive microbial profile [36-38]. After parturition, there is a dramatic shift in hormonal levels, with a decrease in progesterone and a possible increase in estrogen, which may contribute to changes in microbial communities [6, 33, 39, 40]. The drop in progesterone allows for the initiation of uterine involution, a process that may influence microbial dynamics in the postpartum uterus [41-43]. The physiological changes associated with involution, such as increased blood flow and changes in uterine secretions, can alter the microbial habitat, leading to shifts in community composition and diversity [34, 42]. Additionally, the postpartum period is marked by

a more open cervix and increased exposure to external environmental contaminants, which can introduce new microbial taxa into the reproductive tract [6, 43]. Therefore, endocrine regulation should be considered when interpreting our observed microbial diversity and composition shifts between prepartum and postpartum states, warranting further studies.

The meticulous data preparation and sequence quality control steps done via Mothur were crucial for ensuring the reliability of our results [23, 44]. By applying rigorous screening and denoising procedures, we reduced potential biases introduced by sequencing and PCR errors [23, 45]. Also, it enabled us to recognize and exclude samples whose data were likely to be unrepresentative, more commonly among postpartum samples. The presence of low-quality samples in both postpartum vagina and uterus, while scarce in the prepartum vagina, may be due to several factors. The increased difficulty in obtaining high-quality samples from the uterus, attributed to its internal location and potential for contamination during collection [46], may contribute to the higher incidence of low-quality uterine samples. Postpartum vaginal and uterine samples might also face challenges during collection, possibly due to residual fluids or tissues affecting DNA extraction [47]. The relatively low occurrence of low-quality prepartum vaginal samples could be associated with less invasive collection methods and a cleaner environment compared to the postpartum period. Additionally, hormonal and physiological changes during and after parturition might influence the microbial composition and DNA quality [48-50], further impacting the sample quality in these reproductive tract compartments. Overall, quality control measures enhanced the robustness of our findings and allowed for more accurate interpretations of the observed microbial dynamics.

The analysis of phylum-level relative abundance provided insights into the major players in the microbial communities. Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes, and Fusobacteria emerged as the dominant phyla, consistent with previous studies in the bovine reproductive tract [2, 8, 30]. Importantly, significant differences between the Vag-pre and Utr-post groups, particularly in Proteobacteria, Bacteroidetes, and Fusobacteria, suggest a distinct uterine microbial signature postpartum, as also observed in the genus-level analyses.

Alpha diversity metrics, including sample coverage, observed OTUs, and Inverse Simpson diversity estimates, provided valuable insights into community richness and diversity [20, 51]. The lower diversity in the Utr-post group indicated a shift in the uterine microbial landscape, potentially influenced by postpartum physiological changes and the initiation of uterine involution. These findings underscore the importance of considering both anatomical location and postpartum status when studying microbial diversity. The NMDS ordination plot illustrated spatial separation among Vag-pre, Vag-post, and Utr-post samples, emphasizing the distinct community structures associated with each anatomical niche [52]. AMOVA results further supported significant differences between the groups [53], highlighting the impact of anatomical location on bacterial community variation. The LEfSe analysis identified discriminative genera, providing a finer resolution of microbial differences between groups [54]. Key genera, such as

Trinickia, *Micrococcaceae_unclassified*, and *Streptococcus*, were more abundant in Vag-pre samples, while *Gemella* and *Trueperella* were enriched in Utr-post samples.

The DMM model analysis revealed two distinct community types, P1 and P2, segregating the samples. These distinct microbial community types identified by the DMM model analysis underscore the complex and dynamic nature of microbial transitions in the reproductive tract during late gestation and early postpartum. Community type P1, predominantly associated with the Utr-post samples, was characterized by higher relative abundances of genera such as *Histophilus*, *Helcococcus,* and *Fusobacterium*. In contrast, Community type P2, which includes most Vag-pre samples, shows significantly higher relative abundances of genera like *Trinickia*, *Micrococcaceae_unclassified*, and *Ruminococcaceae_unclassified*. These distinct microbial profiles suggest that the prepartum vaginal environment (Vag-pre) and the postpartum uterine environment (Utr-post) harbor unique microbial communities, while the postpartum vaginal samples (Vag-post) exhibit transitional characteristics. Identifying specific genera associated with each community type provides valuable insights into the microbial dynamics accompanying anatomical and temporal changes in the reproductive tract of first-pregnancy heifers.

As detailed above, our findings identify specific microbial markers associated with anatomical and temporal transitions. Interestingly, *Trinickia*, a genus of gram-negative, aerobic, and rod-shaped bacteria in the phylum Proteobacteria (recently renamed as Pseudomonadota) [55], was found in high abundance in the prepartum vagina of first-pregnancy heifers. To the best of our knowledge, this genus has not previously been reported in the bovine female reproductive tract; it is a relatively newly recognized genus with limited information available [56, 57]. *Histophilus* has been reported in the cow reproductive tract and was associated with high abundances in cows diagnosed with purulent vaginal discharge and metritis [30, 58]. The genera *Fusobacterium* and *Trueperella* have been reported extensively in the bovine uterus and have been associated with postpartum uterine diseases [8, 58]. A high prevalence of the genus *Gemella* has been reported in the vagina of pregnant women and was associated with alteration of the immune response, preterm delivery, and postpartum uterine complications [59, 60]. To the best of our knowledge, *Gemella* has not been reported in the cow reproductive tract but has recently been recognized in high abundance in the seminal microbiota of bull semen [61].

Based on the observed differences in bacterial composition among the Vag-pre, Vag-post, and Utr-post samples, it is plausible that the postpartum uterine bacterial composition results from both vaginal bacterial invasion and potential environmental sources. Significant shifts in microbial profiles from the late-pregnancy vaginal state to the postpartum uterine state suggest that vaginal microbiota may seed the uterine microbiome. The close anatomical proximity between the vagina and uterus supports bacterial migration from the lower to upper reproductive tract. However, some unique microbial signatures in postpartum uterine samples, differing from the vaginal microbiota, indicate that environmental exposures during and after delivery may also shape the uterine microbiota. Additionally, uterine contamination, bacterial community establishment, and involution processes influence uterine secretions, likely affecting

postpartum vaginal bacterial dynamics. These factors may explain changes in postpartum vaginal microbiota compared to late pregnancy, while some pre-calving vaginal characteristics remain, albeit modified. Thus, reproductive bacterial compositions are temporally and spatially affected by the unique characteristics of each niche, with different niches (vagina, uterus, and environment) likely influencing each other. The relative prominence of each scenario is challenging to determine definitively without detailed metagenomic analyses (such as Shotgun Metagenomic Sequencing, or Long Read Sequencing), additional functional data, and well-planned environmental sampling [62, 63]. Future studies utilizing advanced sequencing techniques and comprehensive environmental monitoring could further elucidate the dynamics of postpartum uterine bacterial colonization and clarify the relative contributions of vaginal and environmental sources.

The implementation of Next-Generation Sequencing technology and advanced analytical methods has significantly enhanced our understanding of microbial communities within the reproductive tract of first-pregnancy heifers. The high-resolution nature of our approach allowed for the identification of subtle changes in community composition, offering a detailed portrayal of the dynamics within these intricate ecosystems [20]. The assignment of OTUs to phylotypes, as conducted in our study, provides a taxonomic framework that aids in discerning the diversity and relationships among microbial populations [45, 64]. This method allows for a more nuanced analysis compared to calculating pairwise distances between aligned DNA sequences and clustering OTUs accordingly, as it may account for phylogenetic relatedness, potentially providing a more biologically meaningful representation of microbial diversity [65]. However, this approach has some disadvantages [45]. Assigning OTUs by phylotypes typically relies on sequence similarity thresholds for clustering, which may lead to overestimation or underestimation of species diversity based on the chosen threshold [66, 67]. This can introduce biases and affect the accuracy of taxonomic assignments. In contrast, calculating uncorrected pairwise distances provides a more continuous and quantitative measure of sequence similarity, allowing for a more focused assessment of genetic relationships between organisms [66, 68]. The use of phylotype-based OTU assignment may result in splitting single taxa into multiple OTUs, depending on the chosen threshold, which might have occurred in our study to some degree. This lack of resolution can limit the precision of microbial community analysis, especially in environments with high species diversity or closely related species [66-68]. On the other hand, clustering OTUs based on sequence distances allows for a more flexible and data-driven approach, capturing a finer-scale representation of microbial diversity. However, it is essential to consider the computational demands and potential biases associated with different distance metrics when employing this approach [44, 64]. The choice between phylotype-based OTU assignment and sequence distance-based clustering also depends on the research objectives, the nature of the microbial community under investigation, and the available computational resources [66]; all were part of our consideration when the data was analyzed.

There are other inherent limitations associated with our study approach. Potential biases introduced during sample collection and DNA extraction processes may impact the accuracy of our results. Variability in sampling techniques or the DNA extraction process could introduce biases in representing certain taxa, potentially influencing the observed microbial composition. Moreover, while our study contributes valuable insights into the taxonomic aspects of microbial dynamics, further investigations employing functional analyses would be instrumental in elucidating the ecological roles of these microbial communities [62, 63]. Future studies should also compare the reproductive tract microbiota before and after calving in both first-pregnancy heifers and multiparous cows and explore how their dynamics may be associated with postpartum uterine diseases. Integrating metagenomic or metatranscriptomic approaches could unveil functional attributes, shedding light on the potential contributions of these microbes to reproductive health and overall host physiology.

4. Conclusions

Our study provides data that support the conclusion that distinct microbial profiles exist within the reproductive tract of first-pregnancy heifers, influenced by anatomical and temporal transitions. Notably, the microbial profiles of the vagina at late gestation and the uterus at early postpartum are the most distinct, with the postpartum vagina serving as an intermediary link between them. At the phylum level, dominant phyla in the reproductive tract include Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes, and Fusobacteria; however, there is a significant reduction in Proteobacteria and an increase in Bacteroidetes, Fusobacteria, and Tenericutes in the postpartum uterus compared to the late gestation vagina. At the genus level, the late gestation vagina is characterized by higher abundances of *Trinickia* and *Micrococcaceae*, while there is a substantial increase in *Histophilus*, *Helcococcus*, and *Fusobacterium* in the postpartum uterus. Additionally, there is a substantial reduction in bacterial richness and diversity in the postpartum uterus compared to the vagina, indicating a selective niche possibly shaped by endocrine and other physiological changes associated with parturition and early postpartum uterine involution. Our findings support the hypothesis that vaginal microbiota may seed the postpartum uterine microbiome due to their close anatomical proximity. However, unique microbial signatures in the postpartum uterus, differing from vaginal microbiota, imply that environmental exposures during and after delivery may also play a role in shaping the uterine microbiota. Overall, our study highlights the distinct transition of microbial communities in the reproductive tract of first-pregnancy heifers from late pregnancy to early postpartum, emphasizing the need for further research using advanced sequencing techniques to fully understand the functionality of these microbial dynamics and their implications for reproductive physiology and health.

5. Materials and Methods

5.1. Animals and sampling procedures

All animal handling and sampling procedures were performed by a certified veterinarian in accordance with the ethical approval obtained from the Hebrew University Institute Animal Care and Use Committee (IACUC MD-13-13807-2). Three swab samples were collected from first-pregnancy Holstein-Friesian heifers, as detailed below; one from the vagina at gestation day 258±4 ('Vag-pre'), and two additional samples from the vagina ('Vag-post') and uterus ('Utr-post') at postpartum day 8±3. Heifers included in the study were from a commercial dairy farm in Israel. They were housed in loose housing systems in large, completely covered open sheds, and fed a total mixed ration, according to the Nutrient Requirement of Dairy Cattle recommendations. All pregnancies were achieved following artificial insemination with thawed frozen semen (fAI) while the mature heifer was on spontaneous estrus, and pregnancy diagnosis was performed by trans-rectal palpation of the uterus 40-50 days after the fAI date (i.e., indicated as the first gestation day) and reconfirmed before the first sampling at late gestation. Individual cow health and reproduction data were recorded using the herd management computer program NOA (Israeli Cattle Breeders Association, Caesarea, Israel). Heifers were excluded from the study if they had a history of pregnancy-related illness during gestation or if they were treated with antibiotics during the six weeks before the first sampling until after three swab samples were collected. Furthermore, heifers were excluded from the study if they had difficult assisted calving, twin pregnancy, postpartum retained fetal membranes, or systemic clinical signs indicating postpartum septic metritis. All heifers included in the study were calved by spontaneous vaginal delivery. Accordingly, the study included data from 99 swab samples collected from 33 first-pregnancy Holstein-Friesian heifers.

Swab samples were collected using acceptable methods [8, 69]. Before sampling, the vulva was cleaned with a brush and soap water, followed by wipes containing chlorhexidine gluconate and isopropyl alcohol. Sampling was performed using a sterile, disposable, double-guarded endometrial swab (76 cm long; Equivet, Denmark Cat # 290955) inserted into the vagina and managed by one hand while it was directed to the sampling location transrectally by the other hand. For vaginal sampling, the swab was inserted into the cranial vagina, where the inner protective sheath and the swab gradually extruded, and the protruded swab was used to sample the most cranial vagina and the external os of the cervix for approximately 10 seconds. For uterine sampling, the protected swab was guided transcervically into the uterine body, where the inner sheath and the swab gradually extruded, and the swab was pressed and rotated against the endometrium for approximately 10 seconds. Before retraction, the inner swab was inserted back into its protective sheaths to avoid contamination while withdrawing it from the reproductive tract. Outside the reproductive tract, the inner swab tip was kept in its designated sterile sheath and transported at 4 °C to the lab. In the lab, the tip of the swab was placed into a sterile vial containing 1ml sterile PBS; the vial was

vortexed (2 min, maximum speed, twice, 10 min apart) to release the biological material, the tip was discarded, and the fluid was kept frozen at -80 °C until DNA extraction.

5.2. DNA Extraction

DNA extraction was done using the DNeasy® Blood & Tissue kit (catalog number 69504, Qiagen, Hilden, Germany) according to the manufacturer's instructions. Extracted DNA was quantified in a Qubit 4 fluorometer using the dsDNA HS and the dsDNA BR assay kits (Thermo Scientific, MA, USA). Additional quality evaluation included NanoDrop 2000 spectrophotometer (Thermo Scientific, MA, USA) absorbance curve inspection and 260/280 and 260/230 ratio. DNA samples were kept at -20 °C until use.

5.3. 16S rRNA Gene Amplicon Libraries Preparation and NGS Sequencing

Genomic DNA was subjected to polymerase chain reaction (PCR) amplification using the 515F and 806R primers, targeting the V4 hypervariable region of the bacterial and archaeal 16S rRNA gene [20, 21]. The primers, CS1_515F (ACACTGACGACATGGTTCTACAGTGCCAGCMGCCGCGGTAA) and CS2_806R (TACGGTAGCAGAGACTTGGTCTGGACTACHVGGGTWTCTAAT), were designed with common sequences 1 (CS1) and 2 (CS2) at their 5' ends, as previously described [22]. Primers were purchased from Integrated DNA Technologies as standard primers and diluted to 10 μM concentration. DNA samples were adjusted to a 10-20 ng/μL concentration before the amplification process. Amplification reactions were conducted in 96-well plates with a 20 μL reaction volume. The master mix for the entire plate included 10 μL of OneTaq Quick-Load 2X Master Mix with standard buffer (New England Biolabs, MA, USA), 0.5 μL of each primer (10 μM), and 8 μL of nuclease-free water. One microliter of template DNA was added to each well. The cycling conditions consisted of an initial denaturation at 95 °C for 5 minutes, followed by 35 cycles of denaturation at 95 °C for 30 seconds, annealing at 56 °C for 45 seconds, and extension at 72 °C for 30 seconds. A final extension step of 3 minutes at 72 °C was performed. Each plate included a positive control containing E. coli DNA and a negative control (double-distilled water, DDW). Visible amplification was confirmed through 1.7% agarose gel electrophoresis. Subsequently, PCR products were purified and normalized using the SequalPrep normalization plate kit (Thermo Scientific, MA, USA) following the manufacturer's instructions. Size selection was achieved using Pippin Prep (Sage Science, MA, USA).

Following the normalization and size selection steps, a second round of PCR amplification was conducted in 10 μL reactions within 96-well plates. The master mix for the entire plate was prepared, with the composition of each reaction as follows: 5 μL of 2X AccuPrime SuperMix II (Life Technologies, MD, USA), 2 μL of double-distilled water (DDW), 1 μL of each primer (4 μM), and 1 μL of the PCR product obtained from the previous stage. Each reaction employed a distinct primer pair sourced from the Access Array Barcode Library for Illumina Sequences. Each primer pair featured a unique 10-base barcode (Fluidigm, South San Francisco, CA; Cat # 100-4876). Reactions with unique barcodes were designated for

the positive control, no-template control (reaction 1), and a second no-template control reaction containing only Access Array Barcode library primers.

The cycling conditions were programmed as follows: an initial denaturation at 95 °C for 5 minutes, followed by 8 cycles of denaturation at 95 °C for 30 seconds, annealing at 60 °C for 30 seconds, and extension at 68 °C for 30 seconds. A final elongation step of 7 minutes at 68 °C concluded the PCR. Validation of the PCR yield for positive and negative controls, along with select samples, was performed using Qubit analysis and size quantification utilizing an Agilent TapeStation 2200 device with D1000 ScreenTape (Agilent Technologies, Santa Clara, CA). Subsequently, samples were pooled in equal volumes and subjected to purification using solid-phase reversible immobilization (SPRI) cleanup with AMPure XP beads at a ratio of 0.6X (v/v) SPRI solution to sample. Quality control was performed using TapeStation 2200 and Qubit analysis, and the pool was diluted to 6.5 pM for emulsion PCR. Pooled libraries were then sequenced on the Illumina MiSeq platform using V2 reagent kit (150 bp paired-end; UIC Research Resources Center, University of Illinois) [20].

5.4. Bioinformatic Analysis

The bioinformatic analysis was performed by using the Mothur program (v.1.48.0) in an interactive mode via Linux terminal, based on the Mothur MiSeq SOP (https://mothur.org/wiki/miseq_sop/) [23]. Paired-end fastq files (R1, R2) from 99 samples obtained from 33 heifers were included in the metagenetic analysis. Data preparation and initial sequence quality control steps to minimize sequencing and PCR errors included read concatenation, contig formation, and length screening. Throughout the process, sequence numbers and their lengths were assessed, and screening parameters were adjusted accordingly to remove sequences based on length criteria. This included the removal of sequences in the upper and lower 2.5th percentiles and adjusting the remaining sequence lengths to ≤ 293 bp, which is suitable for the V4 region of the 16S rRNA gene. Subsequently, a sequence deduplication step was performed to reduce redundancy (merging duplicates) and streamline downstream analyses. The unique sequences were aligned to the SILVA reference database (Version 132), which was customized to target the V4 region. Coordinates for the V4 region (start=11895, end=25318) were set based on a parallel analysis conducted via Mothur by aligning the E. coli rrsB 16S rRNA gene sequence (NCBI Gene ID 948466) against the reference alignment. An additional screening step was applied to remove sequences in the upper and lower 2.5th percentiles to maintain sequences within the desired region. Sequences were then filtered to remove overhangs at both ends, and an additional deduplication step was performed to remove redundancy. Denoising was done by pre-clustering using a pseudo-single linkage algorithm developed by Huse et al. (2010), allowing up to two differences between sequences, aiming to remove sequences likely due to pyrosequencing errors [67]. Chimeras were removed by using the VSEARCH algorithm [70]. The sequences were then classified taxonomically using a Bayesian classifier with the mother-formatted RDP training set (Version 18), and non-bacteria reads, such as 16S rRNA from chloroplasts, mitochondria, Archaea, 18S rRNA, and random

sequences, were excluded from further analyses. The remaining high-quality sequences were summarized taxonomically and were used for downstream analyses.

The generation of operational taxonomic units (OTUs) was performed from phylotypes, which were classified by genus. Relative abundances were calculated by dividing the read abundance value of an OTU in a specific sample by the total number of reads for that sample. The subsequent steps involved calculating alpha diversity metrics and constructing rarefaction curves, which were used to identify unrepresentative samples that should be excluded from further analyses. The number of sequences, sample coverage, number of observed OTUs, and Inverse Simpson diversity estimates were calculated and compared among the groups (Vag-pre, Vag-post, Utr-post).

Beta diversity was assessed using the Bray-Curtis dissimilarity index, and a phylip-formatted distance matrix was used to generate an output file for a non-metric multidimensional scaling (NMDS) plot (done in 3-dimensions to achieve a stress value below 0.2) [71]. To explore the spatial separation of community structures observed in the NMDS plot, the molecular variance analysis (AMOVA test) was applied, indicating whether the centers of the clouds representing a group are more separated than the variation among samples of the same group. Furthermore, a homogeneity of molecular variance (HOMOVA test) was conducted to statistically compare variations among the Vag-pre, Vag-post, and Utr-post groups.

The non-parametric LEfSe analysis was conducted to provide additional insights into community composition and identify key microbial taxa distinguishing the different groups [54]. The analysis was performed with a threshold value of 2 on the logarithmic score for the Linear Discriminant Analysis (LDA). Furthermore, the Dirichlet multinomial mixtures model was used to calculate Laplace values, which were utilized in an unsupervised evaluation of the partitioning of samples into distinct community types [25]. This analysis was done to reveal inherent community structures (sample segregation and significant OTUs driving community segregation) without specifying sample groupings.

5.5. Additional Statistical Analysis & Plotting

Most statistical analyses were performed using Mothur, as detailed above [23]. However, additional statistical analyses were conducted when applicable using Statistix 10 software (Analytical Software, Tallahassee, FL, USA) and Prism 5.01 (GraphPad Software, San Diego, CA, USA). The choice of statistical tests was guided by the nature of the parameters analyzed to ensure adherence to the assumptions of each test. A non-parametric Kruskal–Wallis one-way analysis of variance (ANOVA) test was employed for non-normally distributed variables, complemented by Dunn's all-pairwise comparisons test when relevant. Repeated Measures ANOVA or a mixed model was applied to account for dependencies when applicable. Proportional data were compared using Fisher's exact test. All statistical tests were based on two-tailed hypotheses, with significance established at $P < 0.05$ and adjustments made for multiple comparisons. Tabular presentations were created using Microsoft Excel, while graphical visualizations were generated using Prism 5.01 (GraphPad Software) or Python (Python Software Foundation, Python Language

Reference, version 3.10.12). When applicable, data analysis and visualization were performed using in-house custom Python scripts, leveraging the Pandas library for data manipulation, the Matplotlib library for 2D visualizations, and the mplot3d toolkit for generating 3D plots, such as the NMDS plot.

Supplementary Materials:

The following supporting information is provided at the end of this manuscript. **Table S1**: Discriminative genera obtained from LEfSe comparison between Vag-pre and Vag-post samples (data related Figure 6). **Table S2**: Discriminative genera obtained from LEfSe comparison between Vag-pre and Utr-post samples (data related to Figure 7); **Table S3**: Discriminative genera obtained from LEfSe comparison between Vag-post and Utr-post samples (data related to Figure 8).

Acknowledgments

We thank the staff at Denron Dairy Farm (Kfar Daniel and Mevo Horon partnership) for enabling our research on their farm. Special thanks to Prof. Nahum Shpigel (The Hebrew University) for his insights, Prof. Joshua Orvis (Johns Hopkins University) for his bioinformatic advice, and Ms. Nathalie Weizmann (The Hebrew University) for her technical assistance.

Author Contributions

Conceptualization, S.A.D., S.S and T.R.; methodology, S.A.D., R.S., S.S, and T.R.; software, T.R.; validation, T.R.; formal analysis, T.R.; investigation, S.A.D., R.S., S.R., and S.S; resources, T.R.; data curation, T.R.; writing—original draft preparation, T.R.; writing—review and editing, all authors.; visualization, T.R.; supervision, T.R.; project administration, S.S., T.R.; funding acquisition, T.R. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

The study was conducted according to the ethical approval obtained from the Hebrew University Institute Animal Care and Use Committee (IACUC MD-13-13807-2).

Data Availability Statement

Sequencing data generated during this study will be deposited in the MG-RAST database and will be publicly available upon acceptance of this manuscript to publication.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Zhu, B.; Tao, Z.; Edupuganti, L.; Serrano, M. G.; Buck, G. A., Roles of the Microbiota of the Female Reproductive Tract in Gynecological and Reproductive Health. *Microbiol Mol Biol Rev* **2022,** 86, (4), e0018121.
2. Ong, C. T.; Turni, C.; Blackall, P. J.; Boe-Hansen, G.; Hayes, B. J.; Tabor, A. E., Interrogating the bovine reproductive tract metagenomes using culture-independent approaches: a systematic review. *Anim Microbiome* **2021,** 3, (1), 41.
3. Zong, Y.; Wang, X.; Wang, J., Research progress on the correlation between gut microbiota and preeclampsia: microbiome changes, mechanisms and treatments. *Front Cell Infect Microbiol* **2023,** 13, 1256940.
4. Timmons, B.; Akins, M.; Mahendroo, M., Cervical remodeling during pregnancy and parturition. *Trends Endocrinol Metab* **2010,** 21, (6), 353-61.
5. Yellon, S. M., Immunobiology of Cervix Ripening. *Front Immunol* **2019,** 10, 3156.
6. Sheldon, I. M.; Williams, E. J.; Miller, A. N.; Nash, D. M.; Herath, S., Uterine diseases in cattle after parturition. *Vet J* **2008,** 176, (1), 115-21.
7. LeBlanc, S. J., Review: Postpartum reproductive disease and fertility in dairy cows. *Animal* **2023,** 17 Suppl 1, 100781.
8. Sicsic, R.; Goshen, T.; Dutta, R.; Kedem-Vaanunu, N.; Kaplan-Shabtai, V.; Pasternak, Z.; Gottlieb, Y.; Shpigel, N. Y.; Raz, T., Microbial communities and inflammatory response in the endometrium differ between normal and metritic dairy cows at 5-10 days post-partum. *Vet Res* **2018,** 49, (1), 77.
9. Pascottini, O. B.; Aurich, C.; England, G.; Grahofer, A., General and comparative aspects of endometritis in domestic species: A review. *Reprod Domest Anim* **2023,** 58 Suppl 2, 49-71.
10. Pu, S.; Wang, M.; Wang, J.; Zhang, Q.; Ma, X.; Wang, R.; Yu, S.; Wang, L.; Pan, Y., Metagenomic analysis reveals a dynamic microbiome with diversified adaptive functions that respond to ovulation regulation in the mouse endometrium. *BMC Genomics* **2023,** 24, (1), 615.
11. Ababneh, M. M.; Degefa, T., Bacteriological findings and hormonal profiles in the postpartum Balady goats. *Reprod Domest Anim* **2006,** 41, (1), 12-6.
12. Chauhan, S. P.; Rice, M. M.; Grobman, W. A.; Bailit, J.; Reddy, U. M.; Wapner, R. J.; Varner, M. W.; Thorp, J. M., Jr.; Caritis, S. N.; Prasad, M.; Tita, A. T. N.; Saade, G. R.; Sorokin, Y.; Rouse, D. J.; Tolosa, J. E.; Eunice Kennedy Shriver National Institute of Child, H.; Human Development Maternal-Fetal Medicine Units, N., Neonatal and Maternal Composite Adverse Outcomes Among Low-Risk Nulliparous Women Compared With Multiparous Women at 39-41 Weeks of Gestation. *Obstet Gynecol* **2020,** 136, (3), 450-457.
13. Ngonzi, J.; Bebell, L. M.; Fajardo, Y.; Boatin, A. A.; Siedner, M. J.; Bassett, I. V.; Jacquemyn, Y.; Van Geertruyden, J. P.; Kabakyenga, J.; Wylie, B. J.; Bangsberg, D. R.; Riley, L. E., Incidence of postpartum infection, outcomes and associated risk factors at Mbarara regional referral hospital in Uganda. *BMC Pregnancy Childbirth* **2018,** 18, (1), 270.
14. Stephansson, O.; Sandstrom, A.; Petersson, G.; Wikstrom, A. K.; Cnattingius, S., Prolonged second stage of labour, maternal infectious disease, urinary retention and other complications in the early postpartum period. *BJOG* **2016,** 123, (4), 608-16.
15. Williams, E. J.; Fischer, D. P.; Pfeiffer, D. U.; England, G. C.; Noakes, D. E.; Dobson, H.; Sheldon, I. M., Clinical evaluation of postpartum vaginal mucus reflects uterine bacterial infection and the immune response in cattle. *Theriogenology* **2005,** 63, (1), 102-17.
16. Gonzalez Moreno, C.; Torres Luque, A.; Oliszewski, R.; Rosa, R. J.; Otero, M. C., Characterization of native Escherichia coli populations from bovine vagina of healthy heifers and cows with postpartum uterine disease. *PLoS One* **2020,** 15, (6), e0228294.

17. Sheldon, I. M.; Rycroft, A. N.; Dogan, B.; Craven, M.; Bromfield, J. J.; Chandler, A.; Roberts, M. H.; Price, S. B.; Gilbert, R. O.; Simpson, K. W., Specific strains of Escherichia coli are pathogenic for the endometrium of cattle and cause pelvic inflammatory disease in cattle and mice. *PLoS One* **2010,** 5, (2), e9192.
18. Chen, H.; Fu, K.; Pang, B.; Wang, J.; Li, H.; Jiang, Z.; Feng, Y.; Tian, W.; Cao, R., Determination of uterine bacterial community in postpartum dairy cows with metritis based on 16S rDNA sequencing. *Vet Anim Sci* **2020,** 10, 100102.
19. Segre, J. A., What does it take to satisfy Koch's postulates two centuries later? Microbial genomics and Propionibacteria acnes. *J Invest Dermatol* **2013,** 133, (9), 2141-2.
20. Caporaso, J. G.; Lauber, C. L.; Walters, W. A.; Berg-Lyons, D.; Huntley, J.; Fierer, N.; Owens, S. M.; Betley, J.; Fraser, L.; Bauer, M.; Gormley, N.; Gilbert, J. A.; Smith, G.; Knight, R., Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **2012,** 6, (8), 1621-4.
21. Wasimuddin; Schlaeppi, K.; Ronchi, F.; Leib, S. L.; Erb, M.; Ramette, A., Evaluation of primer pairs for microbiome profiling from soils to humans within the One Health framework. *Mol Ecol Resour* **2020,** 20, (6), 1558-1571.
22. Moonsamy, P. V.; Williams, T.; Bonella, P.; Holcomb, C. L.; Hoglund, B. N.; Hillman, G.; Goodridge, D.; Turenchalk, G. S.; Blake, L. A.; Daigle, D. A.; Simen, B. B.; Hamilton, A.; May, A. P.; Erlich, H. A., High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array System for simplified amplicon library preparation. *Tissue antigens* **2013,** 81, (3), 141-9.
23. Kozich, J. J.; Westcott, S. L.; Baxter, N. T.; Highlander, S. K.; Schloss, P. D., Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* **2013,** 79, (17), 5112-20.
24. Mather, P. M., *Computational methods of multivariate analysis in physical geography*. John Wiley and Sons: London, 1976; p 532.
25. Holmes, I.; Harris, K.; Quince, C., Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* **2012,** 7, (2), e30126.
26. Knott, J. C., A Study of the Gestation Period of Holstein-Friesian Cows1. *Journal of Dairy Science* **1932,** 15, (2), 87-98.
27. Eaglen, S. A.; Coffey, M. P.; Woolliams, J. A.; Wall, E., Direct and maternal genetic relationships between calving ease, gestation length, milk production, fertility, type, and lifespan of Holstein-Friesian primiparous cows. *J Dairy Sci* **2013,** 96, (6), 4015-25.
28. Taverne, M. A.; Breeveld-Dwarkasing, V. N.; van Dissel-Emiliani, F. M.; Bevers, M. M.; de Jong, R.; van der Weijden, G. C., Between prepartum luteolysis and onset of expulsion. *Domest Anim Endocrinol* **2002,** 23, (1-2), 329-37.
29. Jeon, S. J.; Vieira-Neto, A.; Gobikrushanth, M.; Daetz, R.; Mingoti, R. D.; Parize, A. C.; de Freitas, S. L.; da Costa, A. N.; Bicalho, R. C.; Lima, S.; Jeong, K. C.; Galvao, K. N., Uterine Microbiota Progression from Calving until Establishment of Metritis in Dairy Cows. *Appl Environ Microbiol* **2015,** 81, (18), 6324-32.
30. Tasara, T.; Meier, A. B.; Wambui, J.; Whiston, R.; Stevens, M.; Chapwanya, A.; Bleul, U., Interrogating the Diversity of Vaginal, Endometrial, and Fecal Microbiomes in Healthy and Metritis Dairy Cattle. *Animals (Basel)* **2023,** 13, (7).
31. DiGiulio, D. B.; Callahan, B. J.; McMurdie, P. J.; Costello, E. K.; Lyell, D. J.; Robaczewska, A.; Sun, C. L.; Goltsman, D. S.; Wong, R. J.; Shaw, G.; Stevenson, D. K.; Holmes, S. P.; Relman, D. A., Temporal and spatial variation of the human microbiota during pregnancy. *Proc Natl Acad Sci U S A* **2015,** 112, (35), 11060-5.
32. Fox, C.; Eichelberger, K., Maternal microbiome and pregnancy outcomes. *Fertil Steril* **2015,** 104, (6), 1358-63.

33. Mallott, E. K.; Borries, C.; Koenig, A.; Amato, K. R.; Lu, A., Reproductive hormones mediate changes in the gut microbiome during pregnancy and lactation in Phayre's leaf monkeys. *Sci Rep* **2020,** 10, (1), 9961.
34. Motomura, K.; Miller, D.; Galaz, J.; Liu, T. N.; Romero, R.; Gomez-Lopez, N., The effects of progesterone on immune cellular function at the maternal-fetal interface and in maternal circulation. *J Steroid Biochem Mol Biol* **2023,** 229, 106254.
35. Garcia-Gomez, E.; Vazquez-Martinez, E. R.; Reyes-Mayoral, C.; Cruz-Orozco, O. P.; Camacho-Arroyo, I.; Cerbon, M., Regulation of Inflammation Pathways and Inflammasome by Sex Steroid Hormones in Endometriosis. *Front Endocrinol (Lausanne)* **2019,** 10, 935.
36. Ozcan, G.; Tanyolac Talay, Z. G.; Paerhati, E.; Eren, O. C.; Coskun, N.; Sahin, D.; Alnajjar, I.; Albayrak, O.; Gursoy, A.; Keskin, O.; Celik, E.; Can, F., Dysbiosis in pregnant mice induced by transfer of human vaginal microbiota followed by reversal of pathological changes in the uterus and placenta via progesterone treatment. *BMC Pregnancy Childbirth* **2024,** 24, (1), 427.
37. Bradley, F.; Franzen Boger, M.; Kaldhusdal, V.; Ahlberg, A.; Edfeldt, G.; Lajoie, J.; Bergstrom, S.; Omollo, K.; Damdimopoulos, A.; Czarnewski, P.; Manberg, A.; Oyugi, J.; Kimani, J.; Nilsson, P.; Fowke, K.; Tjernlund, A.; Broliden, K., Multi-omics analysis of the cervical epithelial integrity of women using depot medroxyprogesterone acetate. *PLoS Pathog* **2022,** 18, (5), e1010494.
38. Carosso, A.; Revelli, A.; Gennarelli, G.; Canosa, S.; Cosma, S.; Borella, F.; Tancredi, A.; Paschero, C.; Boatti, L.; Zanotto, E.; Sidoti, F.; Bottino, P.; Costa, C.; Cavallo, R.; Benedetto, C., Controlled ovarian stimulation and progesterone supplementation affect vaginal and endometrial microbiota in IVF cycles: a pilot study. *J Assist Reprod Genet* **2020,** 37, (9), 2315-2326.
39. Zhang, K.; Liu, G.; Wu, Y.; Zhang, T.; Guo, M.; Lei, Y.; Cao, X.; Suo, L.; Brugger, D.; Wang, X.; Yang, Y.; Chen, Y., Gut Microbial Succession Patterns and Metabolic Profiling during Pregnancy and Lactation in a Goat Model. *Microbiol Spectr* **2023,** 11, (1), e0295522.
40. Kaur, S.; Sarma, S. J.; Marshall, B. L.; Liu, Y.; Kinkade, J. A.; Bellamy, M. M.; Mao, J.; Helferich, W. G.; Schenk, A. K.; Bivens, N. J.; Lei, Z.; Sumner, L. W.; Bowden, J. A.; Koelmel, J. P.; Joshi, T.; Rosenfeld, C. S., Developmental exposure of California mice to endocrine disrupting chemicals and potential effects on the microbiome-gut-brain axis at adulthood. *Sci Rep* **2020,** 10, (1), 10902.
41. Wan, L.; Huang, R. J.; Luo, Z. H.; Gong, J. E.; Pan, A.; Manavis, J.; Yan, X. X.; Xiao, B., Reproduction-Associated Hormones and Adult Hippocampal Neurogenesis. *Neural Plast* **2021,** 2021, 3651735.
42. Dockery, P.; Rogers, A. W., The effects of steroids on the fine structure of the endometrium. *Baillieres Clin Obstet Gynaecol* **1989,** 3, (2), 227-48.
43. Miettinen, P. V., Uterine involution in Finnish dairy cows. *Acta Vet Scand* **1990,** 31, (2), 181-5.
44. Schloss, P. D., The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput Biol* **2010,** 6, (7), e1000844.
45. Schloss, P. D.; Westcott, S. L.; Ryabin, T.; Hall, J. R.; Hartmann, M.; Hollister, E. B.; Lesniewski, R. A.; Oakley, B. B.; Parks, D. H.; Robinson, C. J.; Sahl, J. W.; Stres, B.; Thallinger, G. G.; Van Horn, D. J.; Weber, C. F., Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **2009,** 75, (23), 7537-41.
46. Barlund, C. S.; Carruthers, T. D.; Waldner, C. L.; Palmer, C. W., A comparison of diagnostic techniques for postpartum endometritis in dairy cattle. *Theriogenology* **2008,** 69, (6), 714-23.
47. Gilbert, J. A.; Jansson, J. K.; Knight, R., The Earth Microbiome project: successes and aspirations. *BMC Biol* **2014,** 12, 69.
48. Walther-Antonio, M. R.; Jeraldo, P.; Berg Miller, M. E.; Yeoman, C. J.; Nelson, K. E.; Wilson, B. A.; White, B. A.; Chia, N.; Creedon, D. J., Pregnancy's stronghold on the vaginal microbiome. *PLoS One* **2014,** 9, (6), e98514.

49. MacIntyre, D. A.; Chandiramani, M.; Lee, Y. S.; Kindinger, L.; Smith, A.; Angelopoulos, N.; Lehne, B.; Arulkumaran, S.; Brown, R.; Teoh, T. G.; Holmes, E.; Nicoholson, J. K.; Marchesi, J. R.; Bennett, P. R., The vaginal microbiome during pregnancy and the postpartum period in a European population. *Sci Rep* **2015,** 5, 8988.
50. Bezirtzoglou, E.; Voidarou, C.; Papadaki, A.; Tsiotsias, A.; Kotsovolou, O.; Konstandi, M., Hormone therapy alters the composition of the vaginal microflora in ovariectomized rats. *Microb Ecol* **2008,** 55, (4), 751-9.
51. Nikodemova, M.; Holzhausen, E. A.; Deblois, C. L.; Barnet, J. H.; Peppard, P. E.; Suen, G.; Malecki, K. M., The effect of low-abundance OTU filtering methods on the reliability and variability of microbial composition assessed by 16S rRNA amplicon sequencing. *Front Cell Infect Microbiol* **2023,** 13, 1165295.
52. Ramette, A., Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol* **2007,** 62, (2), 142-60.
53. Excoffier, L.; Smouse, P. E.; Quattro, J. M., Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **1992,** 131, (2), 479-91.
54. Segata, N.; Izard, J.; Waldron, L.; Gevers, D.; Miropolsky, L.; Garrett, W. S.; Huttenhower, C., Metagenomic biomarker discovery and explanation. *Genome Biol* **2011,** 12, (6), R60.
55. Oren, A.; Garrity, G. M., Valid publication of the names of forty-two phyla of prokaryotes. *Int J Syst Evol Microbiol* **2021,** 71, (10).
56. Estrada-de Los Santos, P.; Palmer, M.; Chavez-Ramirez, B.; Beukes, C.; Steenkamp, E. T.; Briscoe, L.; Khan, N.; Maluk, M.; Lafos, M.; Humm, E.; Arrabit, M.; Crook, M.; Gross, E.; Simon, M. F.; Dos Reis Junior, F. B.; Whitman, W. B.; Shapiro, N.; Poole, P. S.; Hirsch, A. M.; Venter, S. N.; James, E. K., Whole Genome Analyses Suggests that Burkholderia sensu lato Contains Two Additional Novel Genera (Mycetohabitans gen. nov., and Trinickia gen. nov.): Implications for the Evolution of Diazotrophy and Nodulation in the Burkholderiaceae. *Genes (Basel)* **2018,** 9, (8).
57. Wang, Y. Q.; Zhao, B. N.; Guo, X. Y.; Wu, K. C.; Qiu, L. H., Trinickia mobilis sp. nov. and Trinickia acidisoli sp. nov., isolated from soil. *Int J Syst Evol Microbiol* **2023,** 73, (6).
58. Adnane, M.; Chapwanya, A., A Review of the Diversity of the Genital Tract Microbiome and Implications for Fertility of Cattle. *Animals (Basel)* **2022,** 12, (4).
59. Movassagh, M.; Bebell, L. M.; Burgoine, K.; Hehnly, C.; Zhang, L.; Moran, K.; Sheldon, K.; Sinnar, S. A.; Mbabazi-Kabachelor, E.; Kumbakumba, E.; Bazira, J.; Ochora, M.; Mulondo, R.; Nsubuga, B. K.; Weeks, A. D.; Gladstone, M.; Olupot-Olupot, P.; Ngonzi, J.; Roberts, D. J.; Meier, F. A.; Irizarry, R. A.; Broach, J. R.; Schiff, S. J.; Paulson, J. N., Vaginal microbiome topic modeling of laboring Ugandan women with and without fever. *NPJ Biofilms Microbiomes* **2021,** 7, (1), 75.
60. Florova, V.; Romero, R.; Tarca, A. L.; Galaz, J.; Motomura, K.; Ahmad, M. M.; Hsu, C. D.; Hsu, R.; Tong, A.; Ravel, J.; Theis, K. R.; Gomez-Lopez, N., Vaginal host immune-microbiome interactions in a cohort of primarily African-American women who ultimately underwent spontaneous preterm birth or delivered at term. *Cytokine* **2021,** 137, 155316.
61. Koziol, J. H.; Sheets, T.; Wickware, C. L.; Johnson, T. A., Composition and diversity of the seminal microbiota in bulls and its association with semen parameters. *Theriogenology* **2022,** 182, 17-25.
62. Kumar, B.; Lorusso, E.; Fosso, B.; Pesole, G., A comprehensive overview of microbiome data in the light of machine learning applications: categorization, accessibility, and future directions. *Front Microbiol* **2024,** 15, 1343572.
63. Kim, C.; Pongpanich, M.; Porntaveetus, T., Unraveling metagenomics through long-read sequencing: a comprehensive review. *J Transl Med* **2024,** 22, (1), 111.

64. Schloss, P. D.; Handelsman, J., Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **2005,** 71, (3), 1501-6.
65. Massello, F. L.; Donati, E., Effect of heavy metal-induced stress on two extremophilic microbial communities from Caviahue-Copahue, Argentina. *Environ Pollut* **2021,** 268, (Pt B), 115709.
66. Schloss, P. D.; Westcott, S. L., Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl Environ Microbiol* **2011,** 77, (10), 3219-26.
67. Huse, S. M.; Welch, D. M.; Morrison, H. G.; Sogin, M. L., Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* **2010,** 12, (7), 1889-98.
68. Nebel, M.; Pfabel, C.; Stock, A.; Dunthorn, M.; Stoeck, T., Delimiting operational taxonomic units for assessing ciliate environmental diversity using small-subunit rRNA gene sequences. *Environ Microbiol Rep* **2011,** 3, (2), 154-8.
69. Druker, S. A.; Sicsic, R.; van Straten, M.; Goshen, T.; Kedmi, M.; Raz, T., Cytological endometritis diagnosis in primiparous versus multiparous dairy cows. *J Dairy Sci* **2022,** 105, (1), 665-683.
70. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahe, F., VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **2016,** 4, e2584.
71. Taguchi, Y. H.; Oono, Y., Relational patterns of gene expression via non-metric multidimensional scaling analysis. *Bioinformatics* **2005,** 21, (6), 730-40.

Supplementary material

Correspondence: tal.raz@mail.huji.ac.il

Table S1. Discriminative genera obtained from LEfSe comparison between Vag-pre and Vag-post samples. The logMaxMean represents the logarithm value of the maximum mean abundance of the OTU across the specified group (indicated as a class). LDA, Linear Discriminant Analysis score. Conditional formatting was applied (i.e., blue and green horizontal column bars) to logMaxMean and LDA values to visually represent the data.

Genus	logMaxMean	Class	LDA	P value
Trinickia	5.28	Vag-pre	4.86	<0.0001
Micrococcaceae_unclassified	4.87	Vag-pre	4.18	0.0071
Intrasporangiaceae_unclassified	4.42	Vag-pre	3.75	0.0019
Staphylococcus	4.00	Vag-pre	3.65	0.0009
Betaproteobacteria_unclassified	4.06	Vag-pre	3.58	<0.0001
Paraburkholderia	3.89	Vag-pre	3.48	0.0002
Burkholderiales_unclassified	3.85	Vag-pre	3.44	0.0001
Kocuria	4.02	Vag-pre	3.40	0.0040
Micrococcales_unclassified	4.14	Vag-pre	3.38	0.0030
Uruburuella	3.65	Vag-pre	3.28	0.0006
Burkholderiaceae_unclassified	3.62	Vag-pre	3.24	0.0001
Lactobacillales_unclassified	3.73	Vag-pre	3.23	0.0001
Chitinophagaceae_unclassified	3.86	Vag-pre	3.20	0.0347
Pseudomonadaceae_unclassified	3.63	Vag-pre	3.16	0.0015
Proteobacteria_unclassified	3.64	Vag-pre	3.12	0.0003
Enterobacteriaceae_unclassified	3.43	Vag-pre	3.06	0.0002
Rhodanobacter	3.49	Vag-pre	3.03	0.0005
Corynebacterium	3.56	Vag-pre	3.02	0.0012
Rhizobiales_unclassified	3.73	Vag-pre	3.00	0.0432
Vibrionaceae_unclassified	2.53	Vag-pre	2.94	0.0168
Tersicoccus	3.56	Vag-pre	2.92	0.0079
Oceanotoga	3.27	Vag-pre	2.91	0.0039
Enterococcus	3.18	Vag-pre	2.89	0.0035
Mycobacteriales_unclassified	3.24	Vag-pre	2.89	0.0002
Gammaproteobacteria_unclassified	3.53	Vag-pre	2.87	0.0072
Micrococcus	3.28	Vag-pre	2.87	0.0403
Boudabousia	2.87	Vag-pre	2.87	0.0165
Burkholderia	2.95	Vag-pre	2.83	0.0001
Staphylococcaceae_unclassified	3.27	Vag-pre	2.82	0.0110
Bradyrhizobiaceae_unclassified	3.49	Vag-pre	2.81	0.0283
Undibacterium	2.67	Vag-pre	2.80	0.0004
Rhodanobacteraceae_unclassified	3.18	Vag-pre	2.79	0.0009
Enterobacterales_unclassified	2.71	Vag-pre	2.79	0.0051
Paracoccus	2.61	Vag-pre	2.78	0.0082
Neisseriaceae_unclassified	3.31	Vag-pre	2.78	0.0156
Enterococcaceae_unclassified	3.07	Vag-pre	2.77	0.0090
Sphingomonadaceae_unclassified	3.23	Vag-pre	2.77	0.0047
Bacilli_unclassified	3.18	Vag-pre	2.77	0.0089
Salinivibrio	2.50	Vag-pre	2.75	0.0047
Carnobacteriaceae_unclassified	2.57	Vag-pre	2.71	0.0015
Lawsonella	2.79	Vag-pre	2.71	0.0034

Halomonas	2.57	Vag-pre	2.69	0.0480
Pseudomonas	2.93	Vag-pre	2.66	0.0150
Chryseobacterium	2.50	Vag-pre	2.63	0.0245
Devosiaceae_unclassified	2.95	Vag-pre	2.60	0.0314
Actinobacteria_unclassified	3.03	Vag-pre	2.56	0.0307
Cyanobacteria	2.74	Vag-pre	2.55	0.0292
Xanthomonadales_unclassified	2.91	Vag-pre	2.55	0.0020
Oxalobacteraceae_unclassified	2.65	Vag-pre	2.52	0.0354
Rhizomicrobium	3.36	Vag-pre	2.49	0.0372
Lactococcus	2.72	Vag-pre	2.43	0.0343
Sphingobacteriaceae_unclassified	2.51	Vag-pre	2.43	0.0488
Phyllobacteriaceae_unclassified	2.62	Vag-pre	2.36	0.0293
Fusobacterium	5.10	Vag-post	4.53	0.0003
Leptotrichiaceae_unclassified	4.67	Vag-post	4.38	0.0123
Histophilus	4.56	Vag-post	4.23	0.0223
Mycoplasmopsis	4.78	Vag-post	4.20	0.0425
Parvimonas	4.29	Vag-post	3.93	0.0056
Bacteroidales_unclassified	4.33	Vag-post	3.77	0.0083
Fusobacteriaceae_unclassified	4.43	Vag-post	3.65	0.0179
Caviibacter	3.95	Vag-post	3.52	<0.0001
Acinetobacter	3.89	Vag-post	3.22	0.0120
Peptostreptococcus	3.64	Vag-post	3.15	0.0149
Fusobacteriales_unclassified	3.56	Vag-post	3.05	0.0001
Mollicutes_unclassified	2.95	Vag-post	2.68	0.0017

Table S2. Discriminative genera obtained from LEfSe comparison between Vag-pre and Utr-post samples. The logMaxMean represents the logarithm value of the maximum mean abundance of the OTU across the specified group (indicated as a class). LDA, Linear Discriminant Analysis score. Conditional formatting was applied (i.e., blue and green horizontal column bars) to logMaxMean and LDA values to visually represent the data. *uncl.*, unclacified.

Genus	logMaxMean	Class	LDA	P value
Trinickia	5.28	Vag-pre	4.93	<0.0001
Micrococcaceae_unclassified	4.87	Vag-pre	4.57	<0.0001
Intrasporangiaceae_unclassified	4.42	Vag-pre	4.10	<0.0001
Ruminococcaceae_unclassified	4.47	Vag-pre	3.92	0.0003
Glutamicibacter	4.26	Vag-pre	3.91	<0.0001
Micrococcales_unclassified	4.14	Vag-pre	3.84	<0.0001
Streptococcus	4.30	Vag-pre	3.76	0.0003
Kocuria	4.02	Vag-pre	3.76	<0.0001
Oerskovia	4.02	Vag-pre	3.75	<0.0001
Lachnospiraceae_unclassified	4.37	Vag-pre	3.68	0.0011
Betaproteobacteria_unclassified	4.06	Vag-pre	3.65	<0.0001
Paraburkholderia	3.89	Vag-pre	3.56	<0.0001
Staphylococcus	4.00	Vag-pre	3.56	<0.0001
Burkholderiales_unclassified	3.85	Vag-pre	3.50	<0.0001
Acinetobacter	3.82	Vag-pre	3.49	<0.0001

Uruburuella	3.65	Vag-pre	3.33	<0.0001
Lactobacillales_unclassified	3.73	Vag-pre	3.33	<0.0001
Burkholderiaceae_unclassified	3.62	Vag-pre	3.30	<0.0001
Chitinophagaceae_unclassified	3.86	Vag-pre	3.29	<0.0001
Tersicoccus	3.56	Vag-pre	3.28	<0.0001
Pseudomonadaceae_unclassified	3.63	Vag-pre	3.28	<0.0001
Corynebacterium	3.56	Vag-pre	3.23	<0.0001
Actinobacteria_unclassified	3.81	Vag-pre	3.22	0.0001
Rhizobiales_unclassified	3.73	Vag-pre	3.14	<0.0001
Rhodanobacter	3.49	Vag-pre	3.11	<0.0001
Alphaproteobacteria_unclassified	3.65	Vag-pre	3.09	0.0001
Proteobacteria_unclassified	3.64	Vag-pre	3.08	<0.0001
Cellulomonadaceae_unclassified	3.33	Vag-pre	3.08	<0.0001
Phocaeicola	3.69	Vag-pre	3.07	0.0068
Dietzia	3.30	Vag-pre	3.05	<0.0001
Neisseriaceae_unclassified	3.31	Vag-pre	3.05	0.0003
Devosia	3.45	Vag-pre	2.99	0.0001
Micrococcus	3.28	Vag-pre	2.98	<0.0001
Rhodobacteraceae_unclassified	3.39	Vag-pre	2.97	<0.0001
Gammaproteobacteria_unclassified	3.53	Vag-pre	2.96	0.0004
Bradyrhizobiaceae_unclassified	3.49	Vag-pre	2.96	0.0004
Mycobacteriales_unclassified	3.24	Vag-pre	2.94	<0.0001
Oceanotoga	3.27	Vag-pre	2.93	<0.0001
Enterococcus	3.18	Vag-pre	2.92	<0.0001
Staphylococcaceae_unclassified	3.27	Vag-pre	2.91	<0.0001
Enterobacteriaceae_unclassified	3.43	Vag-pre	2.85	<0.0001
Turicibacter	3.28	Vag-pre	2.85	0.0002
Enterococcaceae_unclassified	3.07	Vag-pre	2.84	<0.0001
Romboutsia	3.30	Vag-pre	2.83	0.0001
Burkholderia	2.95	Vag-pre	2.83	<0.0001
Rhizomicrobium	3.36	Vag-pre	2.82	0.0006
Clostridiales_Incertae_Sedis_XIII_unclas	3.30	Vag-pre	2.82	0.0001
Moraxellaceae_unclassified	2.90	Vag-pre	2.80	<0.0001
Pseudomonas	2.93	Vag-pre	2.80	<0.0001
Paeniclostridium	3.22	Vag-pre	2.79	0.0007
Sphingomonadaceae_unclassified	3.23	Vag-pre	2.78	<0.0001
Psychrobacter	2.73	Vag-pre	2.77	0.0002
Salinivibrio	2.50	Vag-pre	2.76	0.0001
Boudabousia	2.87	Vag-pre	2.76	<0.0001
Acetoanaerobium	2.82	Vag-pre	2.75	<0.0001
Asticcacaulis	3.13	Vag-pre	2.75	0.0126
Nocardioidaceae_unclassified	2.33	Vag-pre	2.74	0.0082
Paracoccus	2.61	Vag-pre	2.74	0.0002
Opitutaceae_unclassified	3.22	Vag-pre	2.74	0.0003
Undibacterium	2.67	Vag-pre	2.73	<0.0001
Halomonadaceae_unclassified	2.39	Vag-pre	2.73	0.0002

Clostridiaceae_1_unclassified	3.16	Vag-pre	2.72	0.0081
Rhodanobacteraceae_unclassified	3.18	Vag-pre	2.72	<0.0001
Paenibacillaceae_1_unclassified	3.09	Vag-pre	2.71	0.0001
Halomonas	2.57	Vag-pre	2.70	0.0002
Aerococcus	2.86	Vag-pre	2.70	0.0005
Brevundimonas	2.90	Vag-pre	2.69	0.0022
Brachybacterium	2.83	Vag-pre	2.67	<0.0001
Oxalobacteraceae_unclassified	2.65	Vag-pre	2.66	<0.0001
Lactococcus	2.72	Vag-pre	2.64	<0.0001
Xanthomonadales_unclassified	2.91	Vag-pre	2.64	0.0012
Ornithinimicrobium	2.55	Vag-pre	2.64	0.0002
Enterobacterales_unclassified	2.71	Vag-pre	2.63	0.0001
Planococcaceae_unclassified	2.57	Vag-pre	2.63	0.0029
Vibrionaceae_unclassified	2.53	Vag-pre	2.62	0.0003
Clostridium_sensu_stricto	3.29	Vag-pre	2.62	0.0091
Gp6_unclassified	2.85	Vag-pre	2.62	0.0008
Lawsonella	2.79	Vag-pre	2.61	<0.0001
Bacilli_unclassified	3.18	Vag-pre	2.61	0.0414
Chryseobacterium	2.50	Vag-pre	2.60	0.0010
Enhydrobacter	2.62	Vag-pre	2.60	0.0123
Actinobacteria_unclassified	3.03	Vag-pre	2.60	<0.0001
Opitutus	3.07	Vag-pre	2.60	0.0007
Dokdonella	3.03	Vag-pre	2.60	0.0133
Carnobacteriaceae_unclassified	2.57	Vag-pre	2.60	<0.0001
Mogibacterium	2.67	Vag-pre	2.57	0.0004
Cytophagales_unclassified	2.97	Vag-pre	2.57	0.0035
Comamonadaceae_unclassified	3.05	Vag-pre	2.56	0.0006
Deltaproteobacteria_unclassified	2.85	Vag-pre	2.55	0.0225
Devosiaceae_unclassified	2.95	Vag-pre	2.54	0.0053
Sphingopyxis	2.92	Vag-pre	2.54	0.0011
Cyanobacteria /Chloroplast _ unclassified	2.74	Vag-pre	2.54	0.0019
Rhizobiaceae_unclassified	2.92	Vag-pre	2.53	0.0048
Olsenella	2.87	Vag-pre	2.53	0.0025
Erythrobacteraceae_unclassified	2.64	Vag-pre	2.51	0.0080
Brevibacterium	2.63	Vag-pre	2.51	0.0009
Micropepsaceae_unclassified	3.10	Vag-pre	2.51	0.0187
Anaerobutyricum	2.60	Vag-pre	2.50	0.0037
Polyangiaceae_unclassified	2.95	Vag-pre	2.49	0.0182
Streptomycetaceae_unclassified	2.90	Vag-pre	2.47	0.0153
Rhodospirillales_unclassified	3.04	Vag-pre	2.46	0.0016
Sphingomonadales_unclassified	2.62	Vag-pre	2.46	0.0144
Streptacidiphilus	2.77	Vag-pre	2.43	0.0334
Flavobacteriaceae_unclassified	2.62	Vag-pre	2.43	0.0083
Phyllobacteriaceae_unclassified	2.62	Vag-pre	2.39	0.0076
Microbacteriaceae_unclassified	2.52	Vag-pre	2.34	0.0412

Sphingobacteriaceae_unclassified	2.51	Vag-pre	2.30	0.0039
Gemella	4.94	Utr-post	4.59	0.0241
Leptotrichiaceae_unclassified	4.56	Utr-post	4.18	0.0095
Trueperella	4.12	Utr-post	3.90	0.0063
Streptobacillus	4.16	Utr-post	3.73	0.0039
Caviibacter	3.26	Utr-post	2.94	0.0009
Gemmatimonadaceae_unclassified	3.11	Utr-post	2.76	0.0336

Table S3: Discriminative genera obtained from LEfSe comparison between Vag-post and Utr-post samples. The logMaxMean represents the logarithm value of the maximum mean abundance of the OTU across the specified group (indicated as a class). LDA, Linear Discriminant Analysis score. Conditional formatting was applied (i.e., blue and green horizontal column bars) to logMaxMean and LDA values to visually represent the data. *uncl.*, unclacified.

Genus	logMaxMean	Class	LDA	P value
Trinickia	4.79	Vag-post	4.41	<0.0001
Leptotrichiaceae_unclassified	4.67	Vag-post	4.24	0.0021
Micrococcaceae_unclassified	4.38	Vag-post	4.11	<0.0001
Ruminococcaceae_unclassified	4.47	Vag-post	3.90	0.0032
Fusobacteriaceae_unclassified	4.43	Vag-post	3.81	0.0057
Clostridiales_unclassified	4.50	Vag-post	3.81	0.0312
Lachnospiraceae_unclassified	4.40	Vag-post	3.78	0.0012
Intrasporangiaceae_unclassified	4.00	Vag-post	3.67	<0.0001
Glutamicibacter	3.97	Vag-post	3.65	0.0001
Acinetobacter	3.89	Vag-post	3.62	0.0016
Oerskovia	3.80	Vag-post	3.51	<0.0001
Micrococcales_unclassified	3.78	Vag-post	3.46	<0.0001
Chitinophagaceae_unclassified	3.62	Vag-post	3.26	0.0026
Kocuria	3.47	Vag-post	3.21	0.0002
Phocaeicola	3.63	Vag-post	3.19	0.0197
Peptostreptococcus	3.64	Vag-post	3.16	0.0467
Paraburkholderia	3.46	Vag-post	3.12	0.0001
Alphaproteobacteria_unclassified	3.49	Vag-post	3.04	0.0062
Burkholderiales_unclassified	3.41	Vag-post	3.03	0.0001
Betaproteobacteria_unclassified	3.67	Vag-post	3.02	0.0031
Actinobacteria_unclassified	3.70	Vag-post	2.98	0.0113
Rhizobiales_unclassified	3.55	Vag-post	2.94	0.0085
Uruburuella	3.25	Vag-post	2.88	0.0001
Corynebacterium	3.20	Vag-post	2.88	0.0002
Cellulomonadaceae_unclassified	3.12	Vag-post	2.87	<0.0001
Micrococcus	3.06	Vag-post	2.86	0.0001
Tersicoccus	2.98	Vag-post	2.80	0.0004
Lactobacillales_unclassified	3.28	Vag-post	2.79	0.0004
Dietzia	3.02	Vag-post	2.77	0.0013
Burkholderiaceae_unclassified	3.17	Vag-post	2.75	0.0020
Paeniclostridium	3.27	Vag-post	2.75	0.0010
Rhodanobacter	3.17	Vag-post	2.72	0.0007

Clostridium_sensu_stricto	3.33	Vag-post	2.72	0.0366
Paenibacillaceae_1_unclassified	2.95	Vag-post	2.72	0.0006
Pseudomonadaceae_unclassified	3.18	Vag-post	2.72	0.0002
Moraxellaceae_unclassified	2.75	Vag-post	2.69	0.0023
Burkholderia	2.24	Vag-post	2.69	0.0032
Devosia	3.18	Vag-post	2.68	0.0014
Staphylococcaceae_unclassified	2.73	Vag-post	2.67	0.0115
Oceanotoga	2.85	Vag-post	2.66	0.0009
Ihubacter	3.19	Vag-post	2.66	0.0497
Mollicutes_unclassified	2.95	Vag-post	2.62	0.0077
Staphylococcus	3.12	Vag-post	2.62	0.0084
Enterococcaceae_unclassified	2.81	Vag-post	2.62	0.0018
Enterococcus	2.65	Vag-post	2.60	0.0006
Pseudomonas	2.38	Vag-post	2.60	0.0005
Clostridiales_Incertae_Sedis_XIII_unclas	3.13	Vag-post	2.59	0.0189
Opitutaceae_unclassified	3.17	Vag-post	2.59	0.0072
Turicibacter	3.05	Vag-post	2.57	0.0008
Gp6_unclassified	2.77	Vag-post	2.56	0.0132
Brachybacterium	2.50	Vag-post	2.56	0.0005
Romboutsia	3.01	Vag-post	2.55	0.0038
Harryflintia	2.85	Vag-post	2.50	0.0489
Peptostreptococcaceae_unclassified	2.82	Vag-post	2.48	0.0112
Rhizobiaceae_unclassified	2.78	Vag-post	2.48	0.0281
Mycobacteriales_unclassified	2.78	Vag-post	2.46	0.0008
Streptomycetaceae_unclassified	2.71	Vag-post	2.46	0.0079
Anaerotignum	2.89	Vag-post	2.44	0.0179
Sphingopyxis	2.80	Vag-post	2.44	0.0298
Mogibacterium	2.40	Vag-post	2.43	0.0066
Opitutus	2.79	Vag-post	2.42	0.0299
Comamonadaceae_unclassified	2.88	Vag-post	2.38	0.0221
Anaerobutyricum	2.37	Vag-post	2.34	0.0214
Actinobacteria_unclassified	2.75	Vag-post	2.32	0.0060
Myxococcales_unclassified	2.46	Vag-post	2.21	0.0245
Trueperella	4.12	Utr-post	3.79	0.0442
Streptococcus	4.15	Utr-post	3.70	0.0080
Peptoniphilus	4.04	Utr-post	3.55	0.0378
Rhodobacteraceae_unclassified	3.18	Utr-post	2.85	0.0035
Gemmatimonadaceae_unclassified	3.11	Utr-post	2.68	0.0103
Rhodospirillales_unclassified	2.86	Utr-post	2.61	0.0088