Characterization of bovine vaginal microbiota using 16S rRNA sequencing: associations with host fertility, longevity, health, and production

L Brulin, S. Ducrocq, G. Even, M P Sanchez, S. Martel, S. Merlin, C. Audebert, P. Croiseau, J. Estellé

scientific reports

Check for updates

OPEN

Characterization of bovine vaginal microbiota using 16S rRNA sequencing: associations with host fertility, longevity, health, and production

L. Brulin[1,2]✉, S. Ducrocq[1,3], G. Even[1,3], M. P. Sanchez[2], S. Martel[1,3], S. Merlin[1,3], C. Audebert[1,3], P. Croiseau[2] & J. Estellé[2]

Due to their potential impact on the host's phenotype, organ-specific microbiotas are receiving increasing attention in several animal species, including cattle. Specifically, the vaginal microbiota of ruminants is attracting growing interest, due to its predicted critical role on cows' reproductive functions in livestock contexts. Notably, fertility disorders represent a leading cause for culling, and additional research would help to fill relevant knowledge gaps. In the present study, we aimed to characterize the vaginal microbiota of a large cohort of 1171 female dairy cattle from 19 commercial herds in Northern France. Vaginal samples were collected using a swab and the composition of the microbiota was determined through 16S rRNA sequencing targeting the V3–V4 hypervariable regions. Initial analyses allowed us to define the core bacterial vaginal microbiota, comprising all the taxa observed in more than 90% of the animals. Consequently, four phyla, 16 families, 14 genera and a single amplicon sequence variant (ASV) met the criteria, suggesting a high diversity of bacterial vaginal microbiota within the studied population. This variability was partially attributed to various environmental factors such as the herd, sampling season, parity, and lactation stage. Next, we identified numerous significant associations between the diversity and composition of the vaginal microbiota and several traits related to host's production and reproduction performance, as well as reproductive tract health. Specifically, 169 genera were associated with at least one trait, with 69% of them significantly associated with multiple traits. Among these, the abundances of *Negativibacillus* and *Ruminobacter* were positively correlated with the cows' performances (i.e., longevity, production performances). Other genera showed mixed relationships with the phenotypes, such as *Leptotrichia* being overabundant in cows with improved fertility records and reproductive tract health, but also in cows with lower production levels. Overall, the numerous associations underscored the complex interactions between the vaginal microbiota and its host. Given the large number of samples collected from commercial farms and the diversity of the phenotypes considered, this study marks an initial step towards a better understanding of the intimate relationship between the vaginal microbiota and the dairy cow's phenotypes.

Keywords Vaginal microbiota, Holstein breed, Fertility, Health, Production

Abbreviations

AI	Artificial insemination
ANCOM-BC	Analysis of compositions of microbiomes with bias correction
ASV	Amplicon sequence variant
$C\text{-}AI_f$	Time between calving and fertilizing artificial insemination
CI	Calving interval

[1]GD Biotech-Gènes Diffusion, 59000 Lille, France. [2]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France. [3]PEGASE-Biosciences, Institut Pasteur de Lille, 1 Rue du Professeur Calmette, 59019 Lille, France. ✉email: l.brulin@genesdiffusion.com

DADA2	Divisive amplicon denoising algorithm 2
FIS	First artificial insemination success
OTU	Operation taxonomic variant
PERMANOVA	Permutational multivariate analysis of variance

Numerous studies have analyzed the bovine microbiota, with a particular focus on the rumen microbial community. Ruminant microbiota are of major interest due to for instance the role ruminal microorganisms play in digestion, directly impacting production, feed efficiency, and methane emissions[1–3]. However, over the last decades selection for milk production has led to a decline in dairy cows' fertility[4,5], a phenomenon due to the negative correlations between fertility and production phenotypes[6]. These reproductive issues have resulted in increased calving intervals and culling rates. Pinedo et al.[7] reported that 17.7% of the animals from 38 states in the USA were culled due to reproductive issues, a figure mirrored in Canada, where a similar proportion of 14.5% has been reported in 2023[8]. Some of these infertility cases have also been associated with infections of the reproductive tract, such as metritis, endometritis, and pyometra. These infections not only negatively affect the welfare of cows[9], but also have detrimental effects on fertility[10] and production[11]. Therefore, given the importance of the reproductive tract in the dairy industry, the microorganisms it hosts have garnered interest in recent years, with consideration given to both the uterine and vaginal microbiota.

The composition and diversity of the uterine microbiota have been linked with uterine diseases, such as metritis[12,13] or endometritis[14]. Notably, dysbiosis associated with reproductive tract infections, such as metritis or endometritis, is generally characterized by a lower alpha-diversity assessed by various index (Chao1, Shannon, and Evenness)[12–14] along with an increased abundance of specific taxa, including Bacteroidetes, *Bacteroides*, *Porphyromonas* and *Fusobacterium necrophorum*[12,14]. However, significant differences between the uterine and vaginal microbiota[15,16] have been highlighted, the microorganisms inhabiting the bovine vagina, which are located at the interface between the uterus and the host's environment, can also provide valuable information[17]. First, and similar to the uterine microbiota, the vaginal microbiota in cows with metritis or endometritis exhibits lower diversity, as evidenced by the Observed Richness, Shannon index and Species Evenness[18,19]. The overall composition (β-diversity) and the abundances of certain taxa showed distinct patterns between healthy and cows with reproductive infections, such as an increase abundance in Bacteroidetes in cows with endometritis[18,19]. Secondly, the composition of the vaginal microbiota has identified some operational taxonomic units (OTU) associated with the outcome of artificial insemination (AI)[16,20,21], a finding also observed in ovine species[22]. However, only a limited number of host phenotypes have been studied alongside the features of the vaginal microbiota in dairy cows in large population.

The observed associations between reproductive tract microbiotas and host's traits hold promises for the use of microbiota data in dairy sector. We hypothesized that using a large population of commercial animals could significantly enhance and consolidate the current knowledge. Notably, a large-scale study would help in defining patterns associated with a healthy vaginal microbiota and identifying specific patterns or taxa of the vaginal microbiota associated with improved performance in fertility, health, longevity, or production. Thus, in this study, we used the 16S rRNA sequencing methodology to characterize the vaginal microbiota on a large cohort of Holstein female cattle raised under commercial conditions and subjected to diverse environmental conditions. Comprehensive animal records were available for our study, including milk yield, fat yield, protein yield, fat content, protein content, reproductive tract infection status, longevity, culling decision, success at first insemination, calving interval, and the duration between calving and fertilizing AI. This exploratory study investigates the relationships between the diversity and composition of the vaginal microbiota and phenotypes related to productive and reproductive performance, longevity and health in Holstein cows.

Results

Diversity of vaginal microbiotas among French Holstein female cattle

16S rRNA sequencing was applied on 1171 samples of Holstein female cattle: 890 samples were collected on non-gestating lactating cows between the first and the fifth parity and 281 samples were collected from heifers. These animals were raised in 19 commercial herds in Northern France. The 16S rRNA sequences were grouped into 37,840 amplicons sequence variants (ASVs) using the DADA2 method[23], averaging 19,917 reads per sample. Samples from lactating cows represented 21,654 reads, while those from heifers comprised 14,988 reads. However, since 6% cows in the population presented an infection of the reproductive tract at the date of sampling based on the veterinary diagnostics (i.e., metritis, pyometra, etc.), a subset of the population consisting of female cattle declared free from reproductive infections was created to better capture the microbiota of healthy animals. In animals without signs of reproductive infection at sampling, a total of 33 phyla were detected (Fig. 1A) with Firmicutes (42%), Proteobacteria (36%) and Bacteroides (12%) representing 90% of the reads. At the genus level, 17 genera had a relative abundance above 1%: *Escherichia-Shigella* (16.9%), *Photobacterium* (7.5%), *UCG-005* (6.9%), unknown genus from *UCG-010* family (6.9%), *Histophilus* (4.3%), *Ureaplasma* (3.8%), *Rikenellaceae RC9 gut group* (3.5%), an unknown genus from Oscillospirales family (3.1%), *Bacteroides* (2.8%), an unknown genus from *Pasteurellaceae* family (2.1%), *Alistipes* (1.8%), an unknown genus from *Lachnospiraceae* family (1.7%), an unknown genus from *Bacteroidales RF16 group* family (1.5%), *Prevotellaceae UCG-003* (1.3%), *Phyllobacterium* (1.2%), *Romboutsia* (1.2%) and *Monoglobus* (1.2%).

We analyzed the taxonomic prevalence (Fig. 1B) in this set where all animals with any clinical reproductive infection were removed in order to define a core vaginal microbiota using the taxa present in at least 90% of the animals. Despite a lenient threshold, justified by the number of samples and the diversity of environmental conditions, one ASV, 14 genera, 16 families and four phyla were consistently identified in the vaginas of Holstein female cattle that did not present signs of reproductive infection at the sampling. However, our prevalence

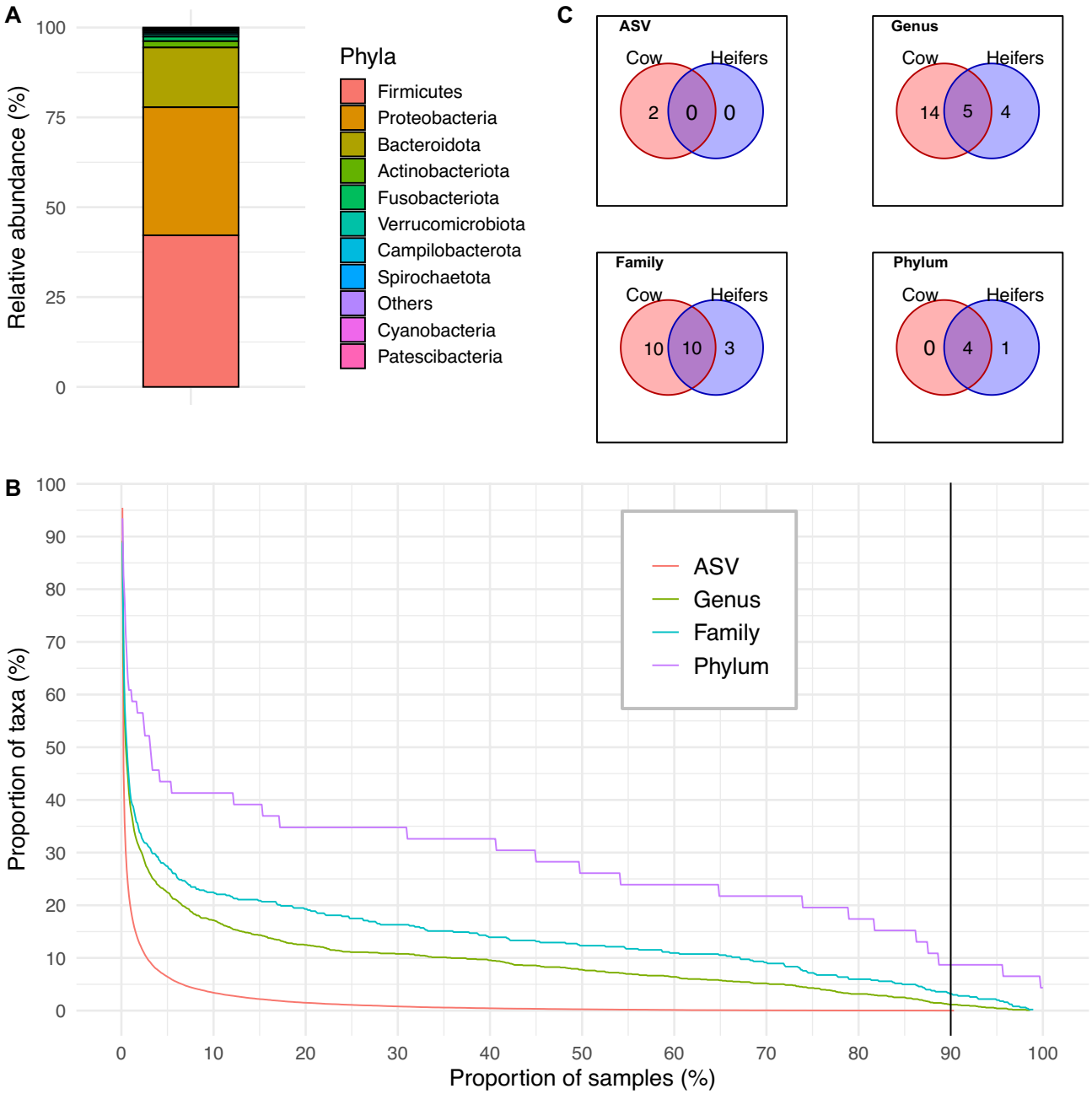

Figure 1. Characterization of the bovine vaginal microbiota (N = 1171). (**A**) Relative abundance of the 10 most abundant phyla in the vagina. The other observed phyla are included in the "Others" category; (**B**) Proportions of taxa shared by the samples for each taxonomic rank. The vertical line represents the minimum threshold (90%) for the taxa to belong to the microbiota core; (**C**) Number of taxa shared by the core microbiotas of the cows (N = 890) and heifers (N = 281) at different taxonomic ranks.

analysis (Fig. 1B) also revealed a sharp decrease, indicating that half of the taxa were shared by only 3% of the animals. This observation is further supported by Fig. 1C, which shows distinct vaginal core microbiotas in heifers and lactating cows.

This diversity of vaginal microbiotas, characterized by few common taxa and numerous rare taxa could be associated with the host's physiology and environment. After excluding all animals with any clinical reproductive tract infection, diversity indices were estimated on the samples of all lactating cows, subsampled at 7000 reads. Observed Richness and the Shannon index were used to illustrate α-diversity. These α-diversity indices were used as the descriptive variables in an analysis of variance (ANOVA), with the lab batch, herd, parity (i.e., lactation rank), sampling season, and day in milk (DIM) as explanatory variables. All the exploratory variables showed significant associations with both Observed Richness and the Shannon index (Table 1). Interestingly, α-diversity indices increased with two variables linked to the host's physiology: age and the interval between the previous calving and sampling. In other words, older animals, and animals more advanced in their lactation had a higher

	Samples	Observed Richness		Shannon		Beta-diversity	
		ASV	Genus	ASV	Genus	ASV	Genus
Factors							
Lab batch	823	< 2.2e−16	8.042e−09	2.258e−13	2.708e−13	0.0001	0.0001
Herd	823	3.204e−12	4.003e−12	1.654e−11	1.404e−10	0.0001	0.0001
Season	823	0.0001	0.009	0.0005	0.007	0.0001	0.0001
Parity	823	0.002	0.029	0.0001	0.018	0.0001	0.0001
Days in milk (DIM)	823	0.004	0.005	1.565e−05	0.0001	0.0002	0.0001

Table 1. Associations (*P*) between cofactors and the α and β-diversities for ASV and genus taxonomic ranks of the vaginal microbiota of adult cows. Bold p-values indicate significant p-values ($p \leq 0.05$).

α-diversity. Concurrently, the sequencing run, herd, parity, DIM, and sampling season were also explanatory variables in a permutational analysis of variance (PERMANOVA) to identify environmental variables associated β-diversity, expressed with a Bray–Curtis dissimilarity distance matrix on a subsampled dataset (7000 reads) (Table 1). This analysis revealed that 38.1% of the total variance in the vaginal microbiota was associated with these physiological or environmental factors.

Associations between the vaginal microbiota and the host's longevity

The longevity of lactating cows was assessed through two distinct phenotypes: culling, a binary trait associated with the decision to end the production life, and longevity, defined as the number of days from the first calving to the end of the last lactation. Using the models described in the "Methods" section, both α and β-diversities were significantly associated ($p < 0.05$) with culling across both taxonomic ranks, culled animals exhibiting higher α-diversity (Fig. 2). In contrast, longevity was only significantly associated with β-diversity (Fig. 2).

ANCOM-BC differential abundance analyses also highlighted associations between both longevity traits and the microbiota composition (Fig. 3). Specifically, two genera, *Negativibacillus* (0.09%) and *Methylobacterium-Methylorubrum* (0.06%), had abundances significantly associated with both traits. *Negativibacillus* was more abundant in animals that entered a subsequent lactation and in long-career animals, while *Methylobacterium-Methylorubrum*, the sole genus negatively associated with longevity, was overabundant in culled animals. Culling was significantly linked with 48 genera, 47 of which were overabundant in culled animals (Fig. 3A), notably, *Pseudomonas* (0.22%). Culled animals also presented increased abundance of one ASV associated *Escherichia-Shigella* (log-fold change = 0.19) and lower abundance of one ASV related to *UCG-002* genus (log-fold change = − 0.21). Conversely, longevity was associated with only seven genera, six of which had a beneficial relationship: *Ruminobacter* (0.17%), *Negativibacillus* (0.09%) (also less abundant in culled animals), *Parasutterella* (0.16%), *Anaerovibrio* (0.03%) and two unknown genera from *Paludibacteraceae* (0.57%) and *Peptococcaceae* (0.18%) families (Fig. 3B). At the ASV level (Supplementary Table S1), only one ASV from the *Peptococcaceae* family showed a significant association with longer career, with a log-fold change increase of 0.00084 for each additional unit of longevity.

Associations between the vaginal microbiota and the host's vaginal health

As mentioned previously, a portion of the lactating cows were sampled while enduring a reproductive infection. We hypothesized that this may impact the composition and diversity of their vaginal microbiota compared to cows that were not having clear signs of infection at the sampling. Even though the infection status of the bovine vagina did not correlate with the α and β-diversities index at the ASV level or the genus taxonomic rank (Fig. 2), the differential abundance analysis (ANCOM-BC) revealed significant associations between the reproductive tract infection status and 52 genera (Fig. 4). Pathogenic genera such as *Peptoniphilus* (0.08%), *Porphyromonas* (0.51%), and *Fusobacterium* (0.28%) were found to be overabundant in infected cows. In contrast, commensal genera, such as *Streptobacillus* (0.59%) and *Leptotrichia* (0.16%), were more abundant in the vaginas of cows that were declared free of infection at the sampling. The *Campylobacter* genus (0.51%), represented by a unique ASV assigned to the *C. lanienae* species (0.03%), was also strongly and positively associated with animal that did not present reproductive infection at the sampling. The entire list of the ASVs whose abundances were associated with the host's vaginal health could be found in Supplementary Table S1.

Associations between the vaginal microbiota and the host's fertility traits

Associations between the vaginal microbiota of lactating cows and reproduction performance were evaluated through three host traits. First, the time between calving and fertilizing AI (C-AI_f), which is the time required from the last calving to the AI that resulted in the next pregnancy. Second, the success at first artificial insemination (FIS), a binary trait used to associate a cow with the outcome (success/failure) of the first artificial insemination (AI). Lastly, the calving interval (CI), which is the time between the last calving before sampling and the following calving. No significant associations were highlighted between the α or β-diversities and the quantitative fertility traits CI and C-AI_f (Fig. 2). However, FIS presented significant associations with both the Observed Richness and the Shannon index: animals that did not become pregnant after the first AI generally had a higher α-diversity score. Additionally, the β-diversity was also significantly associated with the FIS trait. In parallel, in differential abundance analysis, *Streptobacillus* (0.59%) and *Methanosphaera* (0.004%) were both

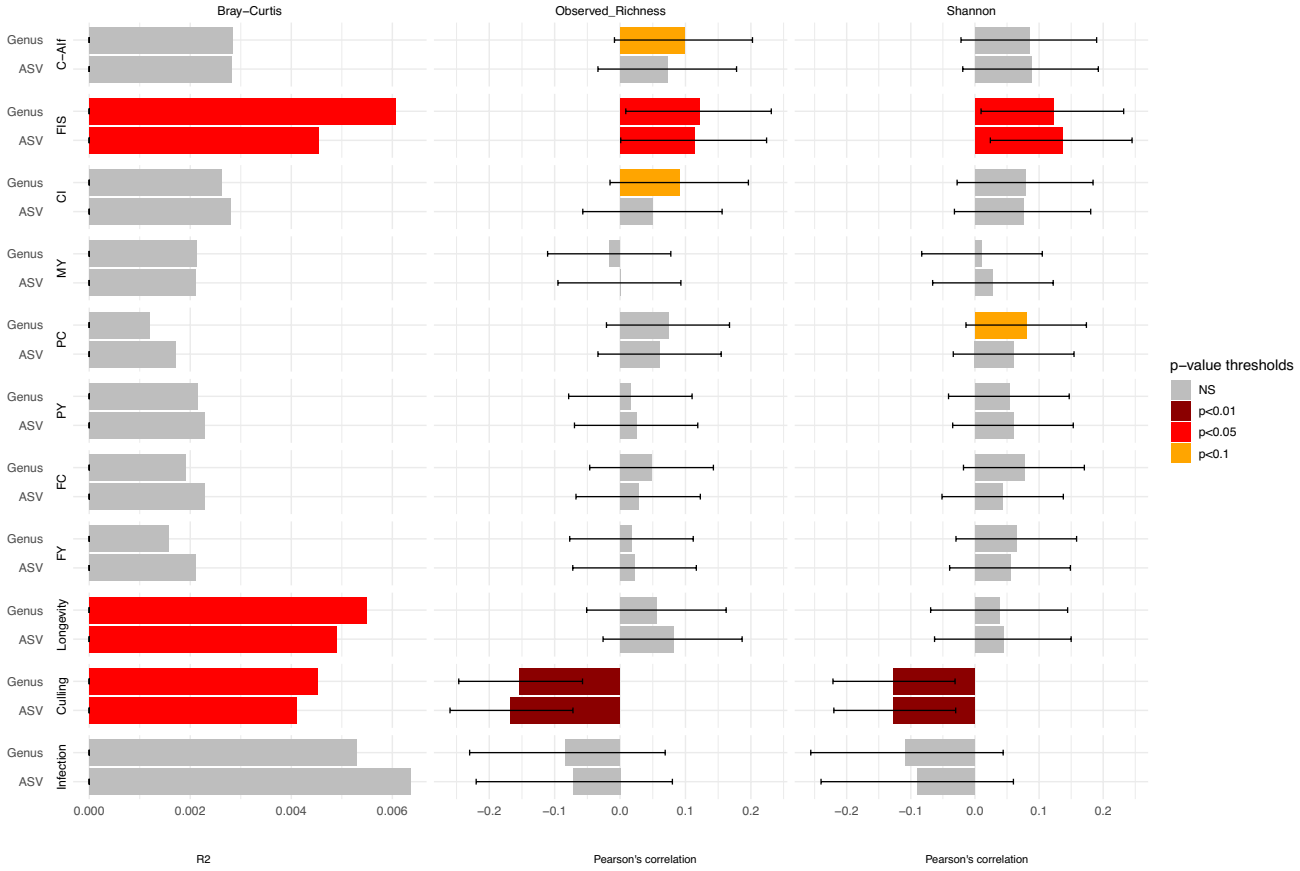

Figure 2. Associations between phenotypes and α- and β-diversities of the fecal microbiota of Holstein cows. R2 between Bray–Curtis (β-diversity) and phenotypes were estimated with PERMANOVA. Pearson's correlations were performed between α-diversity indices (Observed Richness and Shannon index) and phenotypes. α- and β-diversities indices were corrected for lab batch, herd, parity, sampling season and DIM. All phenotypes were corrected for the parity and herd effects except Longevity which was only corrected for herd effect. CI = calving interval (N = 338), *C-AI$_f$* = time between calving and fertilizing AI interval (N = 339), MY = milk yield (N = 433), PY = protein yield (N = 432), PC = protein content (N = 430), FY = fat yield (N = 431), FC = fat content (N = 428), Longevity = length of dairy career from first calving to the end of the last lactation (N = 336), Culling = animal cull or not at the end of the lactation (0/1) (N = 410), FIS = first insemination outcome by cow (0/1) (N = 302), Infection = absence/presence of a reproductive tract infection at the sampling date (0/1) (N = 167).

significantly more abundant in cows that became pregnant after a single AI (Fig. 5A). Interestingly, CI and C-AI$_f$ interval traits were also negatively associated with several genera (Fig. 5B,C). Among these, both *Leptotrichia* (0.16%) and *Fournierella* (0.007%) were more abundant in the vagina of cows with shorter C-AI$_f$ and CI. Overall, nine genera were found to vary in abundance with the reproductive performance of the cows. Among ASVs whose abundances varied along with the host's reproductive performance (see Supplementary Table S1), one ASV from the *Paludibacteraceae* family was associated with shorter CI and C-AI$_f$ intervals. Additionally, two other ASVs from *Clostridia UCG-014*, one ASV from *Leptotrichia* genus and one ASV from *Bacteroides* genus were associated with a shorter C-AI$_f$ length, with log-fold changes ranging between −0.0021 and −0.0034 for each unit increase in C-AI$_f$. Conversely, one ASV from the *Lachnospiraceae* family was associated with a longer C-AI$_f$ interval, with a log-fold change of 0.0032 for each unit increase in C-AI$_f$.

Associations between the vaginal microbiota and the host's production

Lactating cows were also recorded for dairy performance, including milk yield (MY), fat yield (FY) and protein yield (PY) during the first 305 days of lactation. None of these production-related traits showed significant associations with α or β-diversities in the vaginal microbiota (Fig. 2). In contrast, differential abundance analyses in ANCOM-BC detected 41 genera and 730 ASVs whose abundance fluctuated with the dairy performance of the animals (Fig. 6). Twenty-one genera and 434 ASVs were significantly more abundant in the vagina of animals with higher milk, protein and fat yields (Fig. 6A; Supplementary Table S1). Of these, *Lachnospiraceae AC2044 group* genus (0.09%) and an ASV from *Alloprevotella* genus had among strongest associations with the three phenotypes. Additionally, *Ruminobacter* (0.17%) and one ASV from *Bacteroidales RF16 group* genus were also among the ten most significant genera and ASV, respectively, for the three traits. In contrast, 20 other genera were significantly more abundant in animals that had reduced milk, fat, and protein yields (Fig. 5A). *Streptobacillus*

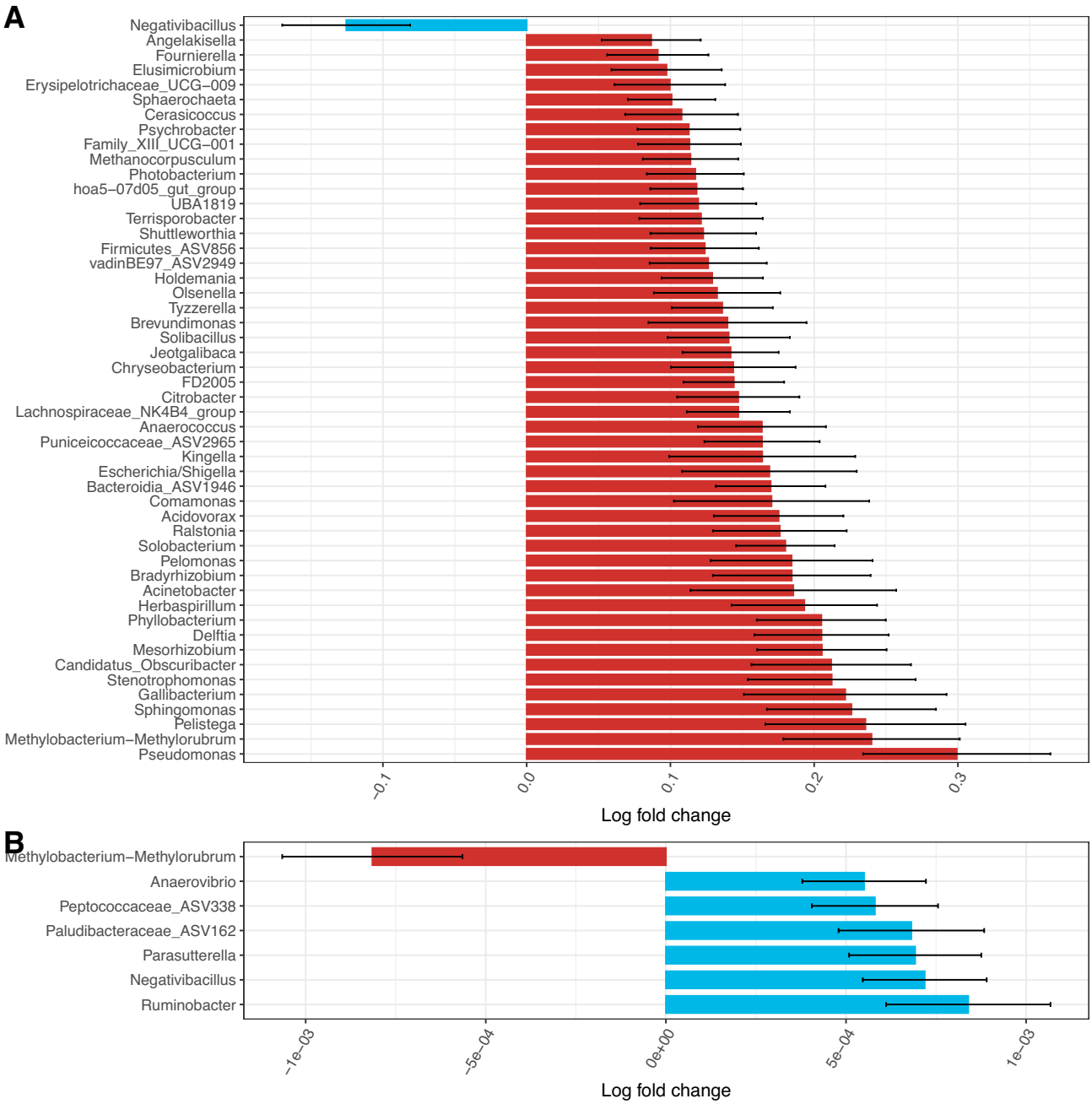

Figure 3. Log-fold change values of the genera of the fecal microbiota associated with the Culling (N = 522) and Longevity (N = 402) in Holstein cows. (**A**) Culling = animal cull or not at the end of the lactation (0/1), (**B**) Longevity = length of dairy career from first calving to the end of the last lactation. The red bar genera were associated with the cull or short-career animals whereas the blue bar genera more abundant in long-career animals or in those that were not culled at the end of the lactation.

(0.59%), *Histophilus* (4.3%), *Ureaplasma* (3.8%) and *Facklamia* (0.29%) were among the 10 genera with the strongest associations with reduced performances for the three traits (Fig. 6B–D). Furthermore, two ASVs from the *Turicibacter* genus and *Pasteurellaceae* family, along with two ASVs closely related to the 16S rRNA gene close to *Ureaplasma diversum* and *Histophilus somni*, were among the top 10 taxa at the ASV level that showed higher abundance in the vagina of animals with reduced performances in MY, FY, and PY. While significant associations between the milk fat and protein contents and specific taxa were also investigated, only tendencies were found: *Bifidobacterium* ($p = 0.096$) (0.35%), *Atopostipes* ($p = 0.096$) (0.044%), and *Clostridium sensu stricto 1* ($p = 0.096$) (0.58%) tended to be more abundant in animals with higher fat content. At a lower taxonomic rank, no ASV was found to be associated with the fat and protein contents in the milk.

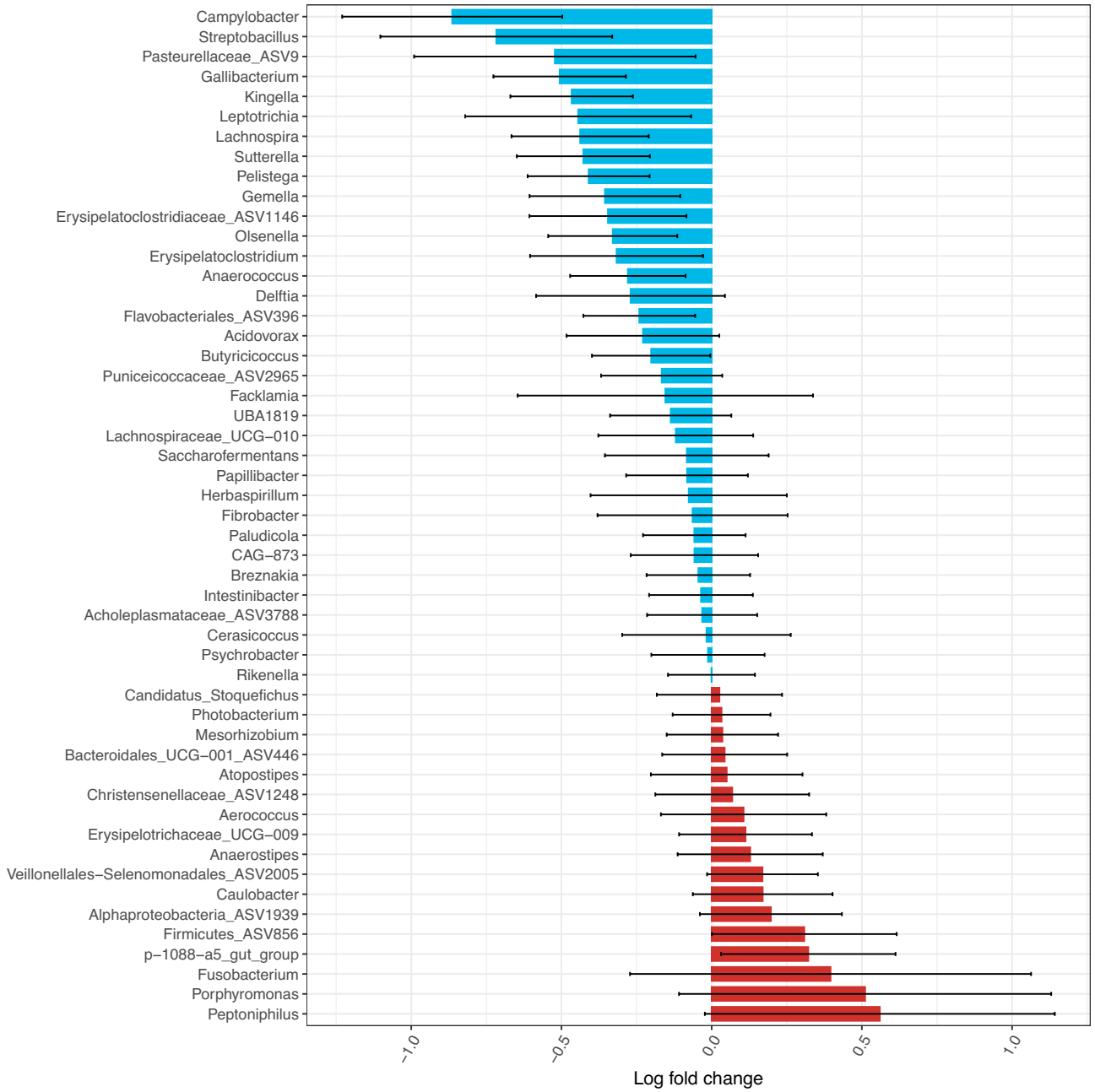

Figure 4. Log-fold change values of the genera of the fecal microbiota associated with the health status of the Holstein cows (N = 249). The red bar genera were associated with infected vaginas whereas the blue bar genera were associated with vaginas which were declared free of infection.

Identification of genera significantly associated with multiple host traits

The ANCOM-BC differential abundance analyses highlighted 186 genera that were negatively or positively associated with at least one longevity-, health-, or production-related trait of lactating cows. Interestingly, 69% of these genera were linked to at least two traits (Fig. 7), with beneficial and/or detrimental associations. Figure 6 summarizes the number of taxa shared among various phenotypes with matching associations (i.e., beneficial or detrimental). Forty-one genera associated with MY, FY or PY were also similarly associated with at least one other production trait.

The genus *Negativibacillus* was beneficially associated with the greatest number of phenotypes, including the three milk production traits, culling, and longevity. Similarly, *Ruminobacter* was more abundant in high-yield animals with a longer productive life. Conversely, some taxa, such as *Methylobacterium-Methylorubrum*, were detrimentally associated with multiple traits. Interestingly, 32 genera were significantly linked with various phenotypes in an antagonistic manner. For instance, *Leptotrichia* was more abundant in animals with shorter CI and C-AI_f, but it was also more abundant in animals with reduced performances in milk, protein and fat yields. Similarly, *Streptobacillus* appeared favorable for FIS and the health status but was detrimental regarding production performances.

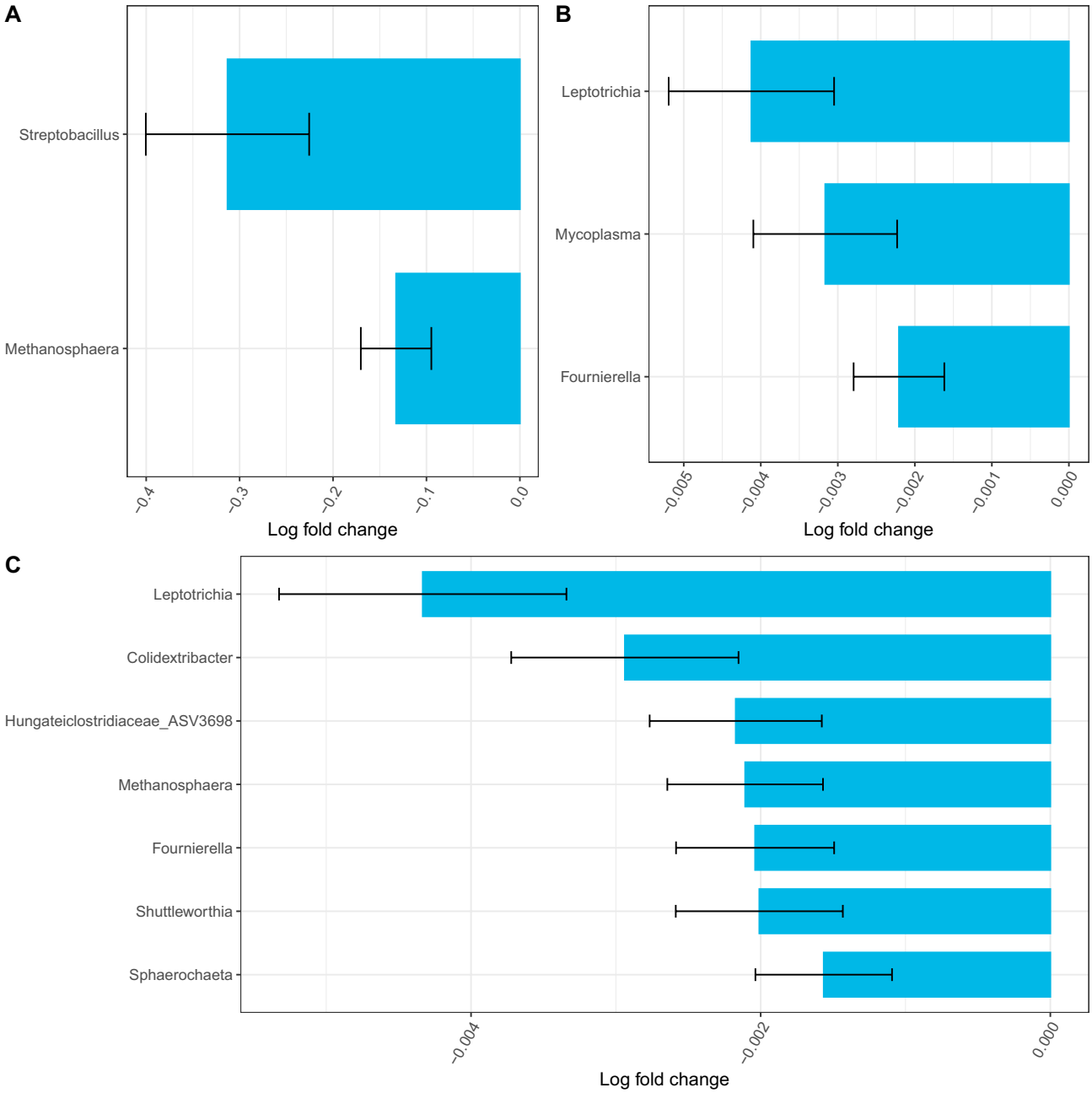

Figure 5. Log-fold change values of the genera of the fecal microbiota associated with fertility traits in Holstein cows. (**A**) FIS = first insemination outcome (0/1) (N = 386), (**B**) CI = calving interval (N = 422), (**C**) C-AI$_f$ = time between calving and fertilizing AI interval (i.e., start of pregnancy) (N = 430). The red bar genera were more abundant in animals with poorer fertility records whereas the blue bar genera were more abundant in animals with interesting fertility results.

Discussion

In this study, we present an extensive analysis of the vaginal microbial communities in dairy cows, utilizing what is, to our knowledge, the largest cohort of animals to date. Overall, these results highlight the complexity of this microbiota and unveil relevant links with several traits of major breeding and economic interest in the dairy industry. Furthermore, as our study was conducted on commercial herds, it is directly representative of current production systems. Due to its cost-effectiveness on large datasets, microbiota data were obtained through 16S rRNA sequencing. This choice restricted our analyses to bacteria and archaea and has inherent biases such as inequal copy number of 16S gene along different taxa. Whole metagenome sequencing could offer a deeper understanding of the complex relationships between microbial functions and host traits[24]. In contrast, the 16S approach is more resilient to host genome contamination, although new applications with the adaptive sequencing strategy seem able to partially overcome this limit[25]. Overall, in our opinion the choice of 16S rRNA sequencing still provides results that are more immediately applicable to large scale situation in commercial

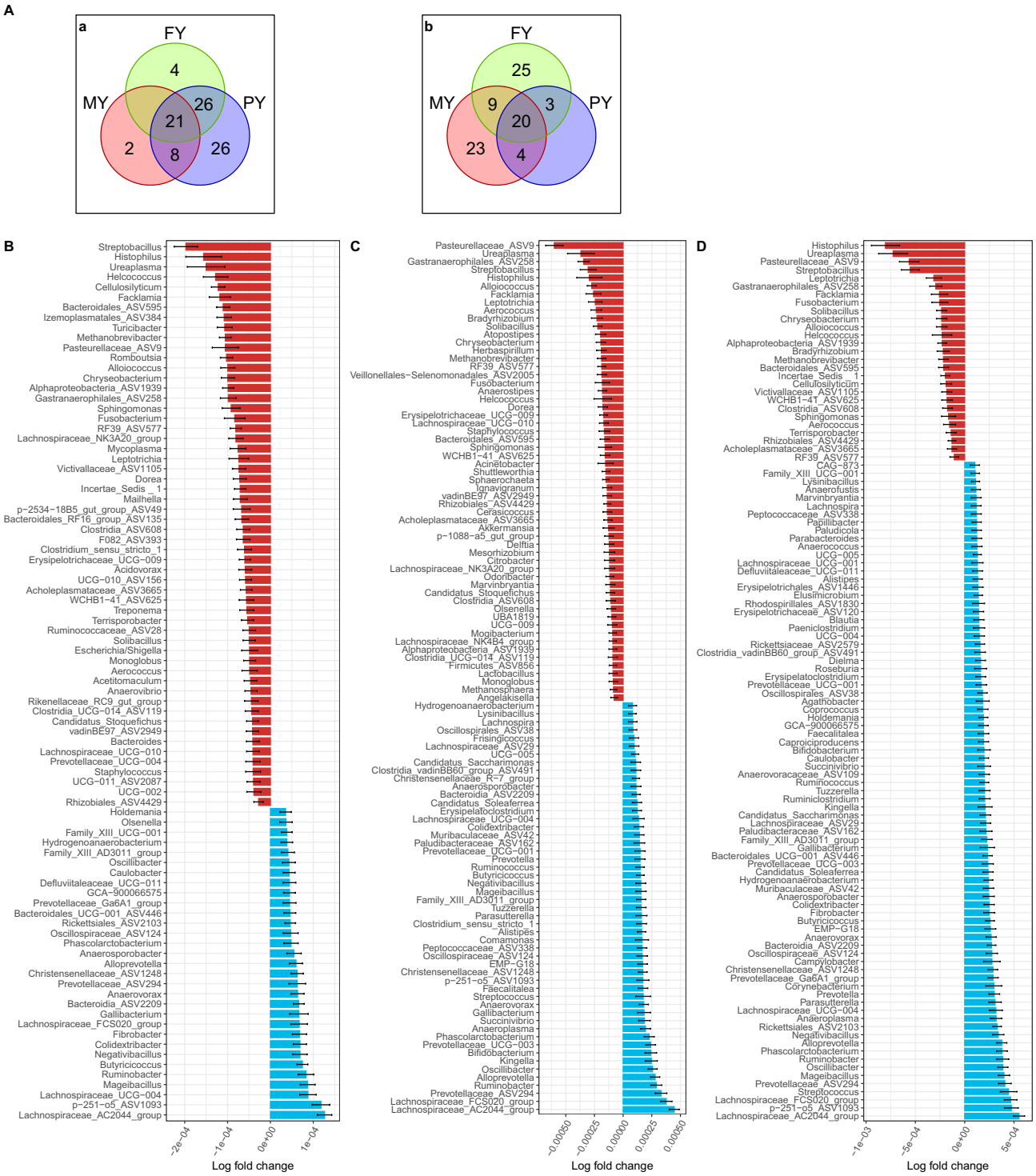

Figure 6. Results of the differential abundance analyses on production traits of Holstein cows. (**A**) Number of shared genera of the fecal microbiota between Milk Yield (MY), Protein Yield (PY) and Fat Yield (FY) which were associated with good performances (a) or bad performances (b); (**B**) Log-fold change results of the genera of the fecal microbiota associated with the MY (N = 545); (**C**) Log-fold change results of the genera of the fecal microbiota associated with the FY (N = 543); (**D**) Log-fold change results of the genera of the fecal microbiota associated with the PY (N = 543). For histograms, red bar genera were more abundant in animals with the poorer records whereas the blue bar genera were more abundant in animals with the highest production records.

farms. Our findings underscore the significant impact of the host's physiology and environment on the diversity (α and β-diversities) and composition of the vaginal microbiota. Within our sampled population, several factors

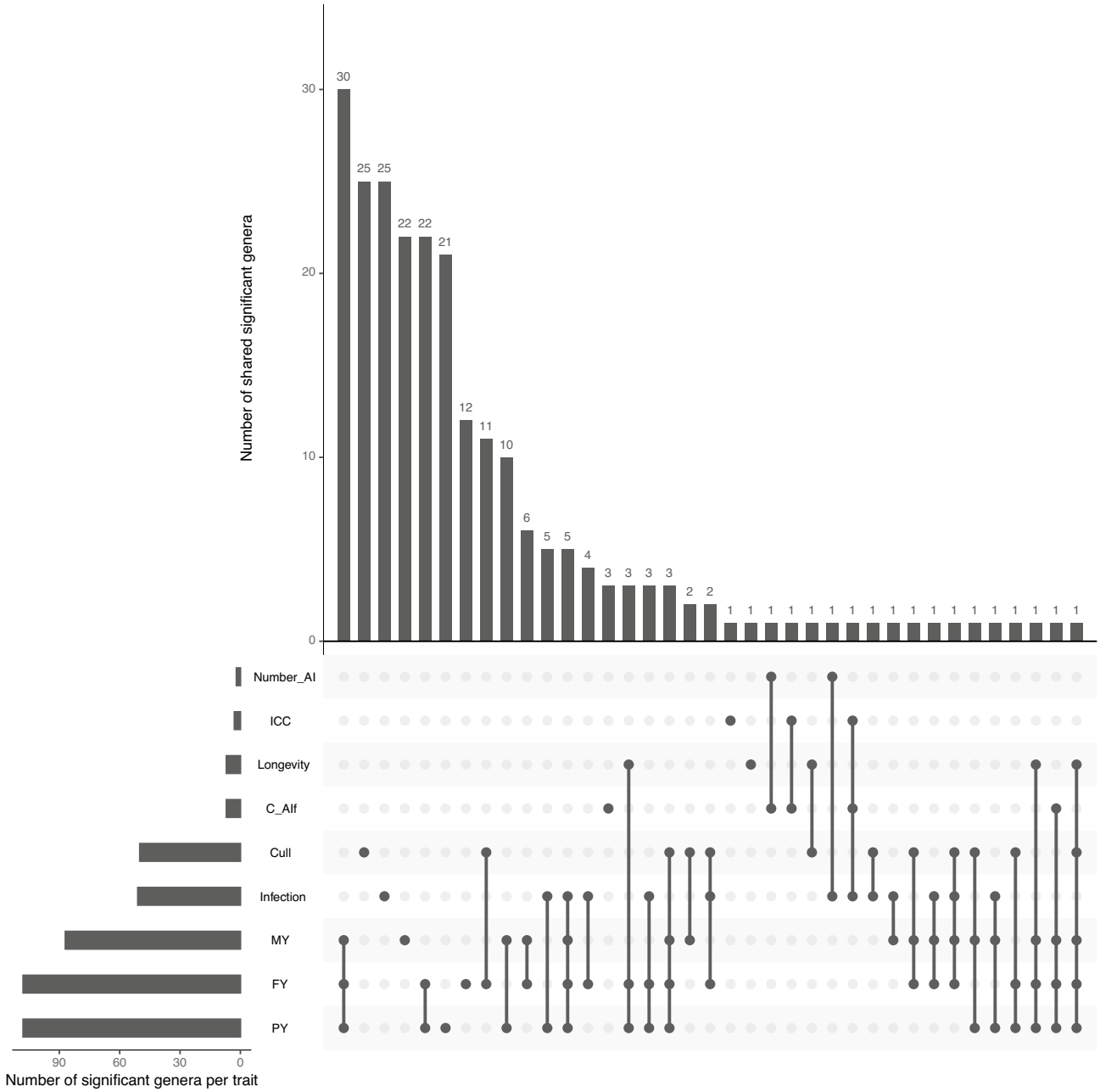

Figure 7. Number of genera of the fecal microbiota whose abundances are significantly associated with Holstein cow's trait and that are shared among traits with same direction relationships. CI = calving interval (N = 422), C-AI_f = time between calving and fertilizing AI interval (i.e., start of pregnancy) (N = 430), MY = milk yield (N = 545), PY = protein yield (N = 543), FY = fat yield (N = 543), Longevity = length of dairy career from first calving to the end of the last lactation (N = 402), Culling = animal cull or not at the end of the lactation (0/1) (N = 522), FIS = first insemination outcome by cow (0/1) (N = 386), Infection = absence/presence of a reproductive tract infection at the sampling date (0/1) (N = 249).

were identified to influence the vaginal microbiota, including farm and lab managements, sampling season, parity, and lactation stage. The impact of herd management was unsurprising, given its integration of various variables such as housing conditions, grazing access, and diet, all known to influence cattle microbiota[26,27]. Host physiology, particularly parity and lactation stage, has been also previously associated with vaginal microbiota diversity and composition[19,36]. The importance of parity and lactation stage may reflect the impact of pregnancy and calving on vaginal microbiota, due to physiological and hormonal changes. These findings are consistent with differences in β-diversity observed between the luteal and follicular phases, suggesting a potential influence of progesterone levels on vaginal microbiota composition[28].

Large-scale characterization of the vaginal microbiota of French Holstein cows

We confirmed that the vaginal microbiota is primarily composed of three phyla: *Firmicutes*, *Proteobacteria*, and *Bacteroidetes*. This finding aligns with previous studies conducted on smaller cohorts of dairy cattle[18,20,29,30], beef[15,21,31], and Nellore cattle[32,33]. In these studies, *Firmicutes* emerged as a dominant phylum in the vaginal microbiota with a relative abundance ranging from 32.4 to 65.9%. This range is consistent with the 42% relative abundance observed in our study. Interestingly, these phyla are also prevalent in other species such as in sows[34,35] and mares[36]. At the genus level, our results were more contrasted, with 17 genera accounting for 75% of the taxa. The *Escherichia-Shigella* genus had the highest relative abundance at 17%. This pattern mirrors the findings of Quereda et al.[28] and Clemmons et al.[15], who identified 17 and 10 genera, respectively, with a relative abundance exceeding 1% dairy heifers and cows. The high abundance of *Escherichia-Shigella* (17%) was surprising given its potential implication with metritis occurrence in cows[37] and in endometritis in sows[38]. However, this genus has also been observed in the vaginal microbiota of beef heifers[21]. Other studies have identified *Ureaplasma*[20,29] as a dominant taxon, despite its known pathogenicity. These observations confirm that dairy cows may naturally harbor potential/opportunistic pathogens in their reproductive tract, even in the absence of any apparent illness. Despite its overall abundance, *Escherichia-Shigella* was not classified as part of the core microbiota because it did not reach the minimum prevalence of 90%.

In this study, the core microbiota of the vaginal samples consisted of one ASV, 14 genera, 16 families, and four phyla. The definition of the core microbiota varies across studies. We adopted an approach (prevalence > 90%) aimed at providing a balanced estimate given our sample size and the diversity of the farms of origin. Quereda et al.[28], who defined the core microbiota as present in all samples, also observed the *UCG-005*, the *Bacteroides,* and an unknown genus from the *Ruminococcaceae* in the vaginal core microbiota of dairy heifers. Consequently, only a minor fraction of the ASVs, genera, families, and phyla were common across animals. Over half of the ASVs, genera, and families were present in less than 3% of the samples, indicating that a major portion of the taxa could be considered as rare. This finding highlights the diversity of vaginal microbiotas we observed in dairy cows, in agreement with a previous observation by Miranda-CasoLuengo et al.[19].

Association between the vaginal microbiota and multiple host's phenotypes

This study underlined interesting associations between the taxa present in the vagina and the host's traits, including health, production, and reproduction performance, and highlighted the potential of using certain vaginal microorganisms as biomarkers to predict host traits.

To our knowledge, no established relationships exist between the composition of the vaginal microbiota and dairy performance metrics (i.e., milk yield, protein yield, fat yield, protein content, fat content). In this study, most genera were associated with at least one production phenotype, and 41 genera with milk yield, protein yield, and fat yield. We also observed a large number of genera commonly found in the gastrointestinal tract, such as *Negativibacillus*. The latter appeared to be part of the genera that were positively associated with better performance: high production levels and longer productive life. The presence of this genus in high-performing animals is in agreement with its role in digestive mechanisms and good digestive performances[39]. This finding raises questions about potential interactions between the vagina and the gastrointestinal tract and supports the hypothesis that the composition of the vaginal microbiota may mirror at least partially that of the gastrointestinal tract. Conversely, other genera were associated with undesirable phenotypes, such as low milk production levels or low fertility. Interestingly, many of these genera were identified as pathogens. Among them, *Histophilus*, *Pseudomonas*, and *Porphyromonas* were found to be associated with metritis[13,40–42].

In contrast, a group of genera exhibited antagonistic effects. These interactions involved traits known to be negatively correlated, such as reproductive and production traits[6]. This was the case for *Leptotrichia* and *Streptobacillus,* which were associated with improved fertility performances but lower milk, protein, and fat yields. Their roles as lactic-acid bacteria could explain their associations with positive fertility records. In humans, Lactobacillus-dominating vaginal microbiota had been associated with reduction of pH, and reduced risks of infections[43–45]. However, the biological associations between their abundances and the production of milk, protein and fat were less obvious to explain and have never been reported, even in the digestive tract. Therefore, these bacteria could be biologically associated with only one phenotype and indirectly linked to other traits due to phenotypic correlations. Interestingly, some genera, such as *Fournierella*, which were significantly associated with fertility traits, were not negatively associated with production traits.

In summary, the numerous associations between the vaginal microbiota and cows' performance underscore the strong interactions that exist between the vaginal microorganisms and the host, and offer promising avenues for solutions aiming at improving cow performances.

Associations between the vaginal microbiota and the productive career length

The animals' productive lifespan is a significant concern for breeders, both in terms of cost-efficiency and environmental impact[46]. While we found no correlation between the α-diversity metrics of the vagina and the career length of the animals, we identified specific associations between the composition of the vaginal microbiota and the longevity of the animals by looking at the global β-diversity and the specific relative abundance of certain taxa. We observed increased abundances of certain genera in the vagina of long productive career cows including *Ruminobacter*, *Anaerovibrio*, *Negativibacillus*, *Parasutterella,* and two unknown genera from *Paludibacteriaceae* and *Peptococcaceae* families. These genera have been previously reported in the bovine vagina[28,47,48], but their specific roles in this reproductive organ remains largely unknown. Besides, *Ruminobacter, Anaerovibrio*, *Parasutterella* and *Negativibacillus* have been mostly described in the gut microbiota of ruminant species, where they are generally associated with diverse metabolisms[39,49–53]. Conversely, *Methylobacterium-Methylorubrum* was generally more abundant in animals with short careers. This genus, which has not been previously reported in any

vaginal microbiota or bovine microbiota, is a strictly aerobic neutrophile bacterium, while we expect the vagina to be an acidic and closed environment[31]. Its presence could indicate specific physiological parameters of the vagina. In addition, *Methylobacterium-Methylorubrum* was the second most differentially abundant genus in culled cows out of a total of 49 genera, with *Pseudomonas* showing the strongest association with the culling. *Pseudomonas* is known for its frequent antibiotic resistance[54] and has also been involved in bovine fertility disorders, such as endometritis[55] but also with other cattle infections, such as mastitis[56]. Other pathogenic bacteria, such as *Stenotrophomonas*[57] or *Gallibacterium*[58], were more abundant in culled animals. *Negativibacillus* was the sole genus observed with significantly lower abundance in culled animals. Overall, our study presents novel findings as, to our knowledge, it is the first one to explore associations between cow longevity and vaginal microbiota. Future research could benefit from evaluating these associations using survival analysis.

Associations between vaginal health status and microbiota

As discussed in the previous section, cows with shorter careers tended to carry higher abundances of infection-related bacteria, and were in consequence more prone to culling. Thus, the associations between vaginal health and its microbiota are of major interest in this cohort. However, compared to other studies in cows[14,19,40] or sows[35], we did not observe any difference in α or β-diversities between the infectious status of the reproductive tract (i.e., non-infected or infected reproductive tract). This could be due to the low prevalence of infection in our population or our broad definition of the infected status, which may have led to the aggregation of multiple infections, irrespective of the underlying pathogenic agent. Overall, various infectious origins could result in different types of dysbiosis, complicating their comparison with the vaginas of animals declared free of reproductive infections. However, the differential abundance analysis did point out significantly associated genera. Indeed, even though we did not observe some of the typical pathogenic bacteria as *Trueperella*[14], vagina of infected cows generally had increased abundances of pathogenic genera such as *Fusobacterium, Porphyromonas*, and *Peptoniphilus*, which are typically overabundant in the reproductive tracts with metritis[14,40–42,47]. Although this large-scale study did not identify specific taxa associated with particular infections, these findings hold promises in the definition of vaginal microbiota patterns associated with the occurrence of infections in the reproductive tract of commercial dairy cows. Such insights could be of main interest to prevent future infections. Concurrently, other genera were more abundant in animals declared free of infections, especially *Campylobacter*. Even through adult ruminants exhibit a high amount of *Campylobacter* in their digestive tracts[59,60], these bacteria are often considered as pathogens. In the present study, we primarily focused our analyses on the genus taxonomic rank to mitigate limitations of the 16S rRNA approach. Nevertheless, when relevant, we have also included results at the ASV level. Indeed, a deeper analysis at the ASV level (that is, close to the species taxonomic rank or even strain level), revealed *C. lanienae* as the most significantly overabundant *Campylobacter* species in vaginas without infection signs. This species is not considered as pathogenic in the literature[61] and its presence could prevent the occurrence of other similar pathogens by occupying their ecological niche. In this sense, *C. lanienae* has been detected in the large intestine of beef cattle and has been proposed to prevent the presence of *C. jejuni*[60]. *Streptobacillus* was another intriguing genus negatively associated with infections. Although it has been described as an abundant genus of the bovine vagina[31], this poorly known taxon was associated, through the *S. moniliformis* species, with rat bite fever[62]. However, this genus belongs to the *Leptotrichiaceae* family, along with *Leptotrichia*, which is known to produce acids, such as lactic acid. Hence, if cows are not sensitive to these genera, their presence could help to decrease the pH of the vagina and protect it from other pathogens, similar to the role of *Lactobacillus* in humans[43–45]. Other findings, such as the overabundance of *Lachnospiraceae* UCG-010, were consistent with the previous results of Moreno et al.[29] who observed a similar association between the *Lachnospiraceae* and healthy vaginas. Interestingly, most of the genera significantly overabundant in vaginas declared infection-free at sampling could thrive in anaerobic conditions. Therefore, they may not directly protect the vagina, but rather, they could be indicative of a physiological state that prevents the contamination by opportunistic aerobic pathogens. This hypothesis aligns with a proposed classification of vaginas based on the amount of oxygen available for the vaginal microbiota ecosystem[32]. In general, all genera enriched in animals without visible signs of infections are also of major interest. Indeed, the vagina being at the frontier between the uterus and the external environment, may harbor taxa that potentially prevent pathogenic infections. Given their detection in a large commercial population, these genera warrant further investigation to better understand the reasons underlying their enrichment in the vagina of healthy dairy cows.

The vaginal microbiota as a potential indicator of the fertility performances

The fertility of the cows is strongly dependent on the health of the reproductive tract[63]. In contrast to the work of Chen et al.[20], who did not find any correlation between the α and β-diversities in the vagina and the pregnancy status of dairy cows, we report here that both the α and β-diversities were significantly associated with the success of the first artificial insemination, with reduced diversity being beneficial for conception at the first service in Holstein cows. This strong association was further reinforced by our observation of certain genera, such as *Leptotrichia* and *Streptobacillus*, being overabundant both in animals without signs of reproductive infections at the sampling and those with enhanced reproductive performances. *Methanosphaera*, another lactic acid producer, was also associated with successful outcome of the first AI and negatively associated with the C-AI_f length. Although *Methanosphaera* is a well-known genus from the bovine gastrointestinal tract[64,65], it has also been observed in ewe vagina, with a specific enrichment in pregnant animals[22]. However, its role and niche in the cattle vagina remain uncertain. Other genera were also associated with the reproductive performances of the cows. For instance, *Fournierella* was associated with shorter calving interval. Although *Fournierella* (*Ruminococcaceae* family) is not commonly observed in the vagina, Chen et al.[20] have also highlighted an increased abundance of various unclassified genera from *Ruminococcaceae* family in the vaginas of bred animals. Thus, deeper analyses

need to be performed to better understand the potential roles or niches *Fournierella* occupies in cows' vaginas. The vaginal abundance of *Lachnospiraceae* also appeared of interest for improved fertility, as Chen et al.[20] noted an increase abundance of a genus from *Lachnospiraceae* in cows that successfully to became pregnant, and we observed the *Shuttleworthia* genus (*Lachnospiraceae* family) as being more abundant in animals with short calving intervals. This finding was not surprising as the presence of the *Lachnospiraceae* family was also considered beneficial by Moreno et al.[29]. Conversely, we noticed that an overabundance of one ASV from the *Lachnospiraceae* family was associated with increased length in C-AI_f, highlighting the diversity of taxa within this taxonomic group.

The vaginal microbiota is linked with the dairy production performances

Although the vaginal taxa of dairy cows were not anticipated to directly influence milk production, we explored potential associations between dairy traits of interest and the vaginal microbiota. We found no correlation with the vaginal α and β-diversities, but the abundance of 41 genera fluctuated with milk, fat, and protein yields. Among them, the *Lachnospiraceae AC2044 group* was one of the most overabundant in the best performing animals. This genus has not yet been linked to the vaginal microbiota, but Liu et al.[66] observed an increased amount of this genus in the rumen microbiota of Holstein cows with higher levels of total milk solid. Conversely, in cows without obvious infection in the reproductive tract, we found that lower production performances during the sampled lactation were generally associated with a higher abundance of pathogenic genera such as *Ureaplasma*[47,67], *Histophilus*[48] or *Fusobacterium*[14,47,48]. Interestingly, as previously discussed, *Fusobacterium* was also more abundant in animals enduring an infection, supporting the link between the presence of pathogenic bacteria and reduced milk production. The analyses performed at the ASV level also permitted to specifically identify ASVs with 16S rRNA sequences closely related to those of *Ureaplasma diversum* and *Histophilus somni*, both well-known for their pathogenicity[37,67]. To our knowledge, no previous studies have investigated the association between the vaginal microbiota and milk production traits. However, as for the productive longevity, many of these differentially abundant genera have been observed in the gut of dairy cows[66]. As production is tightly linked to the digestive tract[68,69], these genera may only reflect the gut microbiota composition without being responsible for the differences of performances. Thus, they could be considered as proxies of the gut microbiota an hypothesis that has already been brought forward by other studies[15,32] and that warrants further investigations.

Conclusions and perspectives for milk production industry

The associations between vaginal microbiota and traits of interest for the dairy industry suggest that the vaginal microbiota holds promises for breeding purposes. From one side, the taxa beneficially associated with host traits represent potential targets for the development of pro-biotic solutions for increasing the health and the performances of milking cows. On the other side, it will be relevant to explore how data on microbiota composition could be useful to improve the efficiency of breeding and selection schemes. Indeed, the host's genetics is also known to be a driver of the microbiota composition, for instance in cattle[70] and in the human vagina[71].

Through this large-scale exploratory analysis of the vaginal microbiota in French Holstein cows, we revealed interesting associations with various phenotypes related to fertility, health, and milk production performance. While α-diversity showed limited associations, the composition of the vaginal microbiota, particularly the abundance of several taxa, exhibited significant associations. Certain taxa, such as *Negativibacillus*, were more abundant in animals with favorable phenotypes (ex. high yield, long career, etc.), whereas others, like *Methylobacterium-Methylorubrum*, were more prevalent in animals with suboptimal performance. Moreover, numerous taxa showed associations with multiple traits, illustrating the intricate relationships between the host and its vaginal microbiota.

Therefore, our results confirm the potential value of studying the microbiota of this reproductive organ for the dairy industry. A better understanding of the interactions between the vaginal microbiota and its host, including the influence of the host genetics on the vaginal microbiota and the precise mechanisms underlying microorganisms–phenotype associations, could provide valuable insights for managing this ecosystem and enhancing the reproductive health of dairy cows.

Methods

Animal sampling and phenotyping

A total of 1171 samples were collected throughout 19 commercial farms located in Northern France, averaging 51 sampled cows per herd (min = 23, max = 108). Among these samples, 281 were samples from heifers (age: mean = 547 ± 107 days; min = 381 days; max = 1134 days) and 890 were collected on cows between parity one and five (mean = 88 ± 57 DIM; min = 20 DIM; max = 425 DIM). Sample collection was conducted between September 2017 and December 2018 by animal reproduction technicians from Gènes Diffusion company by performing vaginal swabs (Novolab, Geraardsbergen, Belgium) on non-gestating Holstein cows. The samples were kept in sterile tubes in an electric cooler at 4 °C ± 1 °C, and transported the same day to Gènes Diffusion research laboratory (Institut Pasteur de Lille, France). They were then stored in a freezer at − 20 °C until the DNA extraction was performed.

Phenotypic information related to health, production, and reproduction of adult cows was extracted from routinely collected data at the farms. Features of each trait are presented in Table 2. All animals being only submitted to AI, the reproductive performance was assessed through three different traits: calving to fertilizing AI interval (C-AI_f), defined as the duration between the last calving before sampling and the start of the pregnancy; success (1) or failure (0) at the first artificial insemination (FIS) for each cow; calving interval (CI), representing

Traits	Units	N samples	N herds	Min	Mean	Max	SD	Proportion (%)
CI	days	422 (338)	17 (17)	311 (311)	419.6 (422.3)	685 (685)	73.51 (75.70)	
C-AI_f	days	430 (339)	17 (17)	43 (43)	142.2 (141.1)	466 (396)	77.38 (72.98)	
MY	kg/305d	545 (433)	16 (16)	3 530 (3530)	9 072 (9026)	14,275 (14,275)	1792.37 (1793.60)	
PC	g/kg/305d	541 (430)	16 (16)	27.2 (27.2)	31.5 (31.4)	37.2 (37.0)	2.00 (2.06)	
PY	kg/305d	543 (432)	16 (16)	1 428 (1428)	2 847 (2 831)	4 262 (4 262)	555.68 (554.57)	
FC	g/kg/305d	529 (428)	16 (16)	29.2 (29.6)	38.9 (39.2)	49.9 (52.6)	4.07 (4.29)	
FY	kg/305d	543 (431)	16 (16)	1 751 (1 751)	3 531 (3 520)	5 896 (5 889)	755.12 (749.83)	
Longevity	days	402 (336)	16 (16)	157 (157)	1 258 (1 269)	2 562 (2 562)	543.35 (535.06)	
Culling	Binary	522 (410)	16 (14)					45.0%
FIS	Binary	386 (302)	13 (12)					49.0%
Infection	Binary	249 (167)	6 (5)					6.0%

Table 2. Descriptive statistics of the different phenotypes of interest of lactating Holstein cows. The numbers in brackets refer to the rarefied dataset. *CI* = calving interval, *C-AI$_f$* = calving to fertilizing AI interval, *MY* = milk yield, *PC* = protein content, *PY* = protein yield, *FC* = fat content, *FY* = fat yield, *Longevity* = length of dairy career from first calving to the end of the last lactation, *Culling* = animal not cull/cull at the end of the lactation (0/1), *FIS* = first insemination success (0/1), *Infection* = absence/presence of a reproductive tract infection at the sampling date (0/1).

the number of days between the last calving before sampling and the subsequent one. For all fertility traits, only the animals sampled before the first AI were evaluated. Production-related traits were milk, fat and protein yields, as well as fat and protein contents measured during the first 305 days of lactation (MY, FY, PY, FC, and PC, respectively). Only animals with a complete lactation of at least 300 days and no longer than 600 days were considered. The health status of the animals' reproductive tract was assessed by the AI technician and/or the veterinarian at the sampling date. A cow was declared as "infected" if the presence of a reproductive tract infection (i.e., affecting the uterus and/or vagina) was observed. Notably, most of the infections observed were metritis and pyometra. Only herds with at least two samples linked to cows with reproductive tract infection were included in the analysis. Finally, two longevity phenotypes were defined as the decision of culling at the end of the observed lactation (referred as Culling, 0/1) and length of the productive life (referred as Longevity, in days) which considered the time between the first calving to the end of the last lactation.

(Table 2).

Microbiota DNA extraction

The DNA extraction was performed using the Nucleospin® 96 Soil kit (Macherey Nagel, Düren, Germany) under aseptic conditions at room temperature and following manufacturer's recommendations. In brief, the cotton swabs used for sampling were first cut and transferred to 2 mL tubes where supplied ceramic beads were then transferred. Lysis buffer and Enhancer buffer were added to tubes and they were agitated at 30 Hz for 2 min with homogenizer (Tissue Lyzer, Qiagen, Hilden, Germany) to break and lyse the samples. The protocol followed the supplier's recommendations until the elution phase of the samples with 50 µL of TE 1× pH 8.0 preheated at 70 °C followed by a 1-h incubation at room temperature. A centrifugation at 6000×*g* for 2 min was performed. We conducted a qPCR 16S monitoring on DNA to assess the global bacterial load. This data enabled us to account for any potential contamination introduced during the process and to adjust for PCR conditions implemented during library preparations. The DNA was stored at − 20 °C.

Library preparation and 16S sequencing

The sequencing library is based on a dual-indexed paired-end sequencing strategy targeting the V3–V4 variable regions of the 16S rRNA gene, as previously described[26]. In brief, to achieve this, two PCRs were successively applied: from 2 µL of the extracted DNA diluted to 1/25, a first PCR was realized in a final volume of 50 µL, using 2.5 U of Precision Taq Polymerase (Applied Biological Materials Inc, Richmond, Canada), each primer had a final concentration of 500 nM. For this first PCR, forward and reverse primers had been designed with the 5′-Tag sequences 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′ and 5′-GTCTCGTGGGCT CGGAGATGTGTATAAGAGACAG-3′ for the forward and reverse primers, respectively, and 16S rRNA gene specific sequences 5′-CCTACGGGNGGCWGCAG-3′ and 5′-GACTACHVGGGTATCTAATCC-3′ for forward and reverse primers, respectively. According to *Escherichia coli* 16S rRNA sequence gene specific primers target a locus between position 341 and 785, resulting in the amplification of a locus of 445 bp. The amplification conditions were 3 min at 94 °C, 25 cycles of 15 s at 94 °C for denaturation, 15 s at 51 °C for primers annealing and 45 s at 68 °C for extension, followed by an incubation at 68 °C for 1 min. At the end of this first PCR, amplification products had been purified with NucleoFast® 96 PCR (Macherey Nagel, Düren, Germany) according to supplier recommendations except for the last step for which 30 µL of TE 1× pH 8.0 preheated at 70 °C had been used for elution. From 5 µL of the previously purified PCR products, a second PCR was performed in a final volume of 50 µL, 2.5 U of Precision Taq Polymerase (Applied Biological Materials Inc, Richmond, Canada). Each primer had a final concentration of 500 nM. The amplification conditions were the same as those of the

previous one except the number of cycles reduced to 8. In addition to the Tag sequences, these PCR2 primers contain a locus to index the samples (barcode sequence) and a locus sequence adapter suitable for the Illumina sequencing technology. A NucleoFast® purification step identical to the one presented above was performed followed by a Quant-iT PicoGreen ds DNA quantification (Life Technologies, Carlsbad, California, USA). An equimolar pool of the library was produced, 200 µL of this mixture was purified by NucleomagNGS® (Macherey Nagel, Düren, Germany). The purification was performed twice to 1.2× with 240 µL of beads suspension at each purification to conclude with a final elution in 50 µL of TE 1× pH 8.0. This purified pool library was then monitored by Bioanalyzer with the High sensitivity DNA Chips kit (Agilent, Santa Clara, California, USA) as a quality control and to estimate the average size of the pool. Then, a quantification of DNA was realized by Qubit assay (Invitrogen, Waltham, Massachusetts, USA). The concentration of DNA and the average size were used to assess molarity of the purified pool.

Sequencing library has been paired-end sequenced on Miseq platform (Illumina, San Diego, California, USA) with MiSeq Reagent Kit v3 allowing 600 sequencing cycles to be performed. At the end of the sequencing a quality control by FastQC v.0.11.9[72] was carried out.

Bioinformatic processing of 16S data and taxonomic assignment

The sequence analyses were conducted using the R software (v.4.2.1)[73] and the dada2 v.1.24-0 package[23] following the author's recommendations. Each sequencing run was analyzed separately for the quality filtering, denoising pair-end merging, and amplicon variant calling steps. Primers and indexes were trimmed and the forward and reverse reads were truncated using the Phred score Q20 as quality threshold[28]. Thus, forwards reads were trimmed at positions between 280 and 290 bp and reverse reads were trimmed at positions between 220 and 230 bp. The DADA2 method[23] was chosen to cluster the sequences with a pairwise identity threshold of 100% (Amplicon Sequence Variants—ASV) and to obtain a count matrix of samples by ASV. All lab batch-specific tables were merged into a unique count table with chimeras removed.

The SILVA v.138[74] was used for taxonomic assignment of the ASVs at all taxonomic ranks, from reign to the species level. To avoid sequence depth bias in diversity analyses, each sample was rarefied to a common depth of 7000 sequences by using the phyloseq (v.1.40.0)[75] R package.

Statistical analysis

For all statistical analyses, the level of significance was defined as $p \leq 0.05$, while tendencies were considered when $0.05 < p \leq 0.10$.

α-diversity was assessed on the rarefied dataset through the Observed Richness and the Shannon diversity index, calculated using the phyloseq and vegan (v.2.6-2)[76] R packages, respectively. To identify factors associated with α-diversity in lactating cows, we performed an ANOVA (Type II sum of squares) with heteroscedasticity correction (white.adjust = TRUE) with the car (v.3.1-0)[77] R package. The model was defined as follows:

$$\boldsymbol{y} = \mu + \boldsymbol{Xb} + \boldsymbol{e} \quad (1)$$

with $\boldsymbol{y}$ being a vector with either the Observed Richness or Shannon indices, μ being the mean for the α-diversity index $\boldsymbol{y}$, $\boldsymbol{X}$ the incidence matrix, $\boldsymbol{b}$ being the vectors of fixed effects, and $\boldsymbol{e}$, the residuals. Then, to assess the relationship between α-diversity indices and adult cow phenotypes, both variables were first corrected for various fixed effects with Model (1) to mitigate confounding effects. For qualitative traits (i.e., culling, infection and FIS), a binomial logistic regression was considered. Fixed effects for α-diversity were sequencing run, herd, parity, sampling season and DIM. For all cows' phenotypes, co-factors were herd and parity, except for longevity where only herd was considered. Then, Pearson's correlation was calculated between adjusted α-diversity indices and cow phenotypes.

β-diversity was computed using the Bray–Curtis dissimilarity matrix on the rarefied dataset with the vegan R package. Initially, associations between Bray–Curtis values and various co-factors were tested using PERMANOVA with the adonis2() function (vegan R package). We applied Model (1), with $\boldsymbol{y}$ being Bray–Curtis dissimilarity matrix. Subsequently, the association between β-diversity and lactating cows' phenotypes was estimated by adding the trait of interest to the fixed effects vector $\boldsymbol{b}$. In all PERMANOVA analyses, marginal effects of each factor were assessed using the by = "margin" option, and the number of permutations was set to 9999.

Finally, in order to identify genera associated with lactating cow's phenotypes, differential abundance analyses were performed using the analysis of compositions of microbiomes with bias correction (ANCOM-BC) method implemented in the ANCOMBC R package (v.1.4.0)[78]. Model (1) was applied with $\boldsymbol{y}$ being the vector of centered log-ratio (CLR) abundance for each genus, $\boldsymbol{X}$ the incidence matrix with the log-fold change values, and $\boldsymbol{b}$, the vector of fixed effects including the lab batch, herd, sampling season, parity, DIM, and the host's phenotype. Log-fold change values were obtained by comparing with a reference level for qualitative variables, and for one unit increase for quantitative variables. Taxa observed in less than 10% of samples were removed, as well as samples with less than 1000 sequences. To account for multiple testing, p-values were adjusted using the Benjamini–Hochberg procedure.

Ethics approval and consent to participate

This work was conducted on farm animals reared for commercial purposes in compliance with the French regulation (Code Rural et de la Pêche Maritime) and the European Council Directive 98/58/EC. In accordance with the legislation, no approval from the Institutional Animal Care and Use Committee or ethics committee was necessary as all the performed sampling operations were part of routine animal manipulations by duly authorized technicians. These technicians were employees of the Gènes Diffusion breeding company (Douai, France) who

took part in the routine reproduction monitoring. They were holders of the CAFTI diploma (Certificat d'Aptitude aux Fonctions de Technicien d'Insémination—Certificate of Fitness for Insemination Technician Functions), approved by the declaration to the EDE (Departmental Establishments of Breeding) that authorizes them for biological sampling in agreement with animal welfare regulation. The farmers involved in this study agreed to the use of their animals' samples for research purposes. Our study is reported in full compliance with the ARRIVE guidelines (https://arriveguidelines.org/).

Data availability

The data used in this study are available upon request from the corresponding author.

Received: 10 February 2023; Accepted: 7 August 2024
Published online: 20 August 2024

References

1. Wallace, R. J. *et al.* A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. *Sci. Adv.* **5**, eaav8391 (2019).
2. Lopes, D. R. G. *et al.* Assessing the relationship between the rumen microbiota and feed efficiency in Nellore steers. *J. Anim. Sci. Biotechnol.* **12**, 79 (2021).
3. Løvendahl, P. *et al.* Review: Selecting for improved feed efficiency and reduced methane emissions in dairy cattle. *Animal* **12**, s336–s349 (2018).
4. Royal, M. D. *et al.* Declining fertility in dairy cattle: Changes in traditional and endocrine parameters of fertility. *Anim. Sci.* **70**, 487–501 (2000).
5. López-Gatius, F. Is fertility declining in dairy cattle?: A retrospective study in northeastern Spain. *Theriogenology* **60**, 89–99 (2003).
6. Boichard, D. & Manfredi, E. Genetic analysis of conception rate in French Holstein cattle. *Acta Agric. Scand. Sect. A Anim. Sci.* **44**, 138–145 (1994).
7. Pinedo, P. J., De Vries, A. & Webb, D. W. Dynamics of culling risk with disposal codes reported by Dairy Herd Improvement dairy herds. *J. Dairy Sci.* **93**, 2250–2261 (2010).
8. CDIC, Canadian dairy informal center. Culling and replacement rates in dairy herds in Canada. https://agriculture.canada.ca/en/sector/animal-industry/canadian-dairy-information-centre/statistics-market-information/dairy-animal-genetics/culling-replacement (2024).
9. Stojkov, J., von Keyserlingk, M. A. G., Marchant, J. N. & Weary, D. M. Assessment of visceral pain associated with metritis in dairy cows. *J. Dairy Sci.* **98**, 5352–5361 (2015).
10. Ribeiro, E. S. *et al.* Carryover effect of postpartum inflammatory diseases on developmental biology and fertility in lactating dairy cows. *J. Dairy Sci.* **99**, 2201–2220 (2016).
11. Pérez-Báez, J. *et al.* The economic cost of metritis in dairy herds. *J. Dairy Sci.* **104**, 3158–3168 (2021).
12. Galvão, K. N. *et al.* Effect of chitosan microparticles on the uterine microbiome of dairy cows with metritis. *Appl. Environ. Microbiol.* **86**, e01066-20 (2020).
13. Jeon, S. J. *et al.* Uterine microbiota and immune parameters associated with fever in dairy cows with metritis. *PLoS One* **11**, e0165740 (2016).
14. Pascottini, O. B. *et al.* Dynamics of uterine microbiota in postpartum dairy cows with clinical or subclinical endometritis. *Sci. Rep.* **10**, 12353 (2020).
15. Clemmons, B. A. *et al.* Vaginal and uterine bacterial communities in postpartum lactating cows. *Front. Microbiol.* https://doi.org/10.3389/fmicb.2017.01047 (2017).
16. Webb, E. M. *et al.* Sequencing and culture-based characterization of the vaginal and uterine microbiota in beef cattle that became pregnant or remained open following artificial insemination. *Microbiol. Spectrum* **11**, e02732-23 (2023).
17. Luecke, S. M., Webb, E. M., Dahlen, C. R., Reynolds, L. P. & Amat, S. Seminal and vagino-uterine microbiome and their individual and interactive effects on cattle fertility. *Front. Microbiol.* **13**, 1029128 (2022).
18. Bicalho, M. L. S. *et al.* Dynamics of the microbiota found in the vaginas of dairy cows during the transition period: Associations with uterine diseases and reproductive outcome. *J. Dairy Sci.* **100**, 3043–3058 (2017).
19. Miranda-CasoLuengo, R. *et al.* Delayed differentiation of vaginal and uterine microbiomes in dairy cows developing postpartum endometritis. *PLoS One* **14**, e0200974 (2019).
20. Chen, S.-Y., Deng, F., Zhang, M., Jia, X. & Lai, S.-J. Characterization of vaginal microbiota associated with pregnancy outcomes of artificial insemination in dairy cows. *J. Microbiol. Biotechnol.* **30**, 804–810 (2020).
21. Deng, F. *et al.* The vaginal and fecal microbiomes are related to pregnancy status in beef heifers. *J. Anim. Sci. Biotechnol.* **10**, 92 (2019).
22. Koester, L. R., Petry, A. L., Youngs, C. R. & Schmitz-Esser, S. Ewe vaginal microbiota: Associations with pregnancy outcome and changes during gestation. *Front. Microbiol.* **12**, 745884 (2021).
23. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
24. Ong, C. T. *et al.* Adaptive sampling during sequencing reveals the origins of the bovine reproductive tract microbiome across reproductive stages and sexes. *Sci. Rep.* **12**, 15075 (2022).
25. Ong, C. T. *et al.* Technical note: Overcoming host contamination in bovine vaginal metagenomic samples with nanopore adaptive sequencing. *J. Anim. Sci.* **100**, skab344 (2022).
26. Liu, J. *et al.* Dairy farm soil presents distinct microbiota and varied prevalence of antibiotic resistance across housing areas. *Environ. Pollut.* **254**, 113058 (2019).
27. Weese, J. S. & Jelinski, M. Assessment of the fecal microbiota in beef calves. *J. Vet. Intern. Med.* **31**, 176–185 (2017).
28. Quereda, J. J. *et al.* Vaginal microbiota changes during estrous cycle in dairy heifers. *Front. Vet. Sci.* **7**, 371 (2020).
29. Moreno, C. G., Luque, A. T., Galvão, K. N. & Otero, M. C. Bacterial communities from vagina of dairy healthy heifers and cows with impaired reproductive performance. *Res. Vet. Sci.* **142**, 15–23 (2022).
30. Nesengani, L. T., Wang, J., Yang, Y., Yang, L. & Lu, W. Unravelling vaginal microbial genetic diversity and abundance between Holstein and Fleckvieh cattle. *RSC Adv.* **7**, 56137–56143 (2017).
31. Swartz, J. D. *et al.* Characterization of the vaginal microbiota of ewes and cows reveals a unique microbiota with low levels of lactobacilli and near-neutral pH. *Front. Vet. Sci.* **1**, 19 (2014).
32. Laguardia-Nascimento, M. *et al.* Vaginal microbiome characterization of Nellore cattle using metagenomic analysis. *PLoS One* **10**, e0143294 (2015).
33. Giannattasio-Ferraz, S. *et al.* A common vaginal microbiota composition among breeds of *Bos taurus indicus* (Gyr and Nellore). *Braz. J. Microbiol.* **50**, 1115–1124 (2019).
34. Wang, Y. *et al.* Lactoferrin for the treatment of COVID-19 (Review). *Exp. Ther. Med.* **20**, 272 (2020).
35. Poor, A. P. *et al.* Vaginal microbiota signatures in healthy and purulent vulvar discharge sows. *Sci. Rep.* **12**, 9106 (2022).

36. Barba, M. *et al.* Vaginal microbiota is stable throughout the estrous cycle in Arabian mares. *Animals* **10**, 2020 (2020).
37. Gonzalez Moreno, C., Fontana, C., Cocconcelli, P. S., Callegari, M. L. & Otero, M. C. Vaginal microbial communities from synchronized heifers and cows with reproductive disorders. *J. Appl. Microbiol.* **121**, 1232–1241 (2016).
38. Wang, J. *et al.* Characterization of vaginal microbiota of endometritis and healthy sows using high-throughput pyrosequencing of 16S rRNA gene. *Microb. Pathog.* **111**, 325–330 (2017).
39. Elolimy, A., Alharthi, A., Zeineldin, M., Parys, C. & Loor, J. J. Residual feed intake divergence during the preweaning period is associated with unique hindgut microbiome and metabolome profiles in neonatal Holstein heifer calves. *J. Anim. Sci. Biotechnol.* **11**, 13 (2020).
40. Galvão, K. N., Bicalho, R. C. & Jeon, S. J. Symposium review: The uterine microbiome associated with the development of uterine disease in dairy cows. *J. Dairy Sci.* **102**, 11786–11797 (2019).
41. Peng, Y., Wang, Y., Hang, S. & Zhu, W. Microbial diversity in uterus of healthy and metritic postpartum Holstein dairy cows. *Folia Microbiol.* **58**, 593–600 (2013).
42. Jeon, S. J. *et al.* Uterine microbiota progression from calving until establishment of metritis in dairy cows. *Appl. Environ. Microbiol.* **81**, 6324–6332 (2015).
43. O'Hanlon, D. E., Moench, T. R. & Cone, R. A. Vaginal pH and microbicidal lactic acid when lactobacilli dominate the microbiota. *PLoS One* **8**, e80074 (2013).
44. O'Hanlon, D. E., Moench, T. R. & Cone, R. A. In vaginal fluid, bacteria associated with bacterial vaginosis can be suppressed with lactic acid but not hydrogen peroxide. *BMC Infect. Dis.* **11**, 200 (2011).
45. Chee, W. J. Y., Chew, S. Y. & Than, L. T. L. Vaginal microbiota and the potential of *Lactobacillus* derivatives in maintaining vaginal health. *Microb. Cell Fact.* **19**, 203 (2020).
46. Grandl, F., Furger, M., Kreuzer, M. & Zehetmeier, M. Impact of longevity on greenhouse gas emissions and profitability of individual dairy cows analysed with different system boundaries. *Animal* **13**, 198–208 (2019).
47. Machado, V. S. *et al.* Investigation of postpartum dairy cows' uterine microbial diversity using metagenomic pyrosequencing of the 16S rRNA gene. *Vet. Microbiol.* **159**, 460–469 (2012).
48. Luna, E. Z. & Vega, R. C. *Characterization and diversity of the uterine microbiota in Simmental and Simbrah cows by sequencing the V4 region of the Gen rRNA-16s during an FTAI protocol.* https://www.researchsquare.com/article/rs-708181/v1. https://doi.org/10.21203/rs.3.rs-708181/v1 (2021).
49. Zhao, S., Min, L., Zheng, N. & Wang, J. Effect of heat stress on bacterial composition and metabolism in the rumen of lactating dairy cows. *Animals* **9**, 925 (2019).
50. Xing, L. *et al.* The anti-inflammatory effect of bovine bone-gelatin-derived peptides in LPS-induced RAW264.7 macrophages cells and dextran sulfate sodium-induced C57BL/6 mice. *Nutrients* **14**, 1479 (2022).
51. Zhang, J., Xu, C., Huo, D., Hu, Q. & Peng, Q. Comparative study of the gut microbiome potentially related to milk protein in Murrah buffaloes (*Bubalus bubalis*) and Chinese Holstein cattle. *Sci. Rep.* **7**, 42189 (2017).
52. Jewell, K. A., McCormick, C. A., Odt, C. L., Weimer, P. J. & Suen, G. Ruminal bacterial community composition in dairy cows is dynamic over the course of two lactations and correlates with feed efficiency. *Appl. Environ. Microbiol.* **81**, 4697–4710 (2015).
53. Zhang, Y. K. *et al.* Characterization of the rumen microbiota and its relationship with residual feed intake in sheep. *Animal* **15**, 100161 (2021).
54. Lupo, A., Haenni, M. & Madec, J.-Y. Antimicrobial resistance in *Acinetobacter* spp. and *Pseudomonas* spp. *Microbiol. Spectrum* **6**, 6301 (2018).
55. Mahmoud, S. F. *et al.* Genetic diversity, biofilm formation, and antibiotic resistance of pseudomonas aeruginosa isolated from cow, camel, and mare with clinical endometritis. *Vet. Sci.* **9**, 239 (2022).
56. Park, H. *et al.* Characterisation of *Pseudomonas aeruginosa* related to bovine mastitis. *Acta Vet. Hung.* **62**, 1–12 (2013).
57. Ohnishi, M. *et al.* Antimicrobial susceptibility and genetic relatedness of bovine *Stenotrophomonas maltophilia* isolates from a mastitis outbreak. *Lett. Appl. Microbiol.* **54**, 572–576 (2012).
58. Van Driessche, L. *et al.* Isolation of drug-resistant *Gallibacterium anatis* from calves with unresponsive bronchopneumonia, Belgium. *Emerg. Infect. Dis.* **26**, 721–730 (2020).
59. Thépault, A. *et al.* Prevalence of thermophilic *Campylobacter* in cattle production at slaughterhouse level in France and link between *C. jejuni* bovine strains and campylobacteriosis. *Front. Microbiol.* **9**, 471 (2018).
60. Inglis, G. D., Kalischuk, L. D., Busz, H. W. & Kastelic, J. P. Colonization of cattle intestines by *Campylobacter jejuni* and *Campylobacter lanienae*. *Appl. Environ. Microbiol.* **71**, 5145–5153 (2005).
61. Costa, D. & Iraola, G. Pathogenomics of emerging campylobacter species. *Clin. Microbiol. Rev.* **32**, e00072-18 (2019).
62. Graves, M. H. & Janda, J. M. Rat-bite fever (*Streptobacillus moniliformis*): A potential emerging disease. *Int. J. Infect. Dis.* **5**, 151–154 (2001).
63. Koeck, A., Egger-Danner, C., Fuerst, C., Obritzhauser, W. & Fuerst-Waltl, B. Genetic analysis of reproductive disorders and their relationship to fertility and milk yield in Austrian Fleckvieh dual-purpose cows. *J. Dairy Sci.* **93**, 2185–2194 (2010).
64. Smith, P. E., Kelly, A. K., Kenny, D. A. & Waters, S. M. Differences in the composition of the rumen microbiota of finishing beef cattle divergently ranked for residual methane emissions. *Front. Microbiol.* **13**, 855565 (2022).
65. Zhu, Z. *et al.* Changes in rumen bacterial and archaeal communities over the transition period in primiparous Holstein dairy cows. *J. Dairy Sci.* **101**, 9847–9862 (2018).
66. Liu, K. *et al.* Ruminal bacterial community is associated with the variations of total milk solid content in Holstein lactating cows. *Anim. Nutr.* **9**, 175–183 (2022).
67. Díaz, J. *et al.* Association of *Ureaplasma diversum* with reproductive disease in cattle. *N. Z. Vet. J.* **67**, 249–256 (2019).
68. Jami, E., White, B. A. & Mizrahi, I. Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. *PLoS One* **9**, e85423 (2014).
69. Xue, M., Sun, H., Wu, X., Guan, L. L. & Liu, J. Assessment of rumen microbiota from a large dairy cattle cohort reveals the pan and core bacteriomes contributing to varied phenotypes. *Appl. Environ. Microbiol.* **84**, e00970-18 (2018).
70. Li, F. *et al.* Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. *Microbiome* **7**, 92 (2019).
71. Si, J., You, H. J., Yu, J., Sung, J. & Ko, G. Prevotella as a hub for vaginal microbiota under the influence of host genetics and their association with obesity. *Cell Host Microbe* **21**, 97–105 (2017).
72. Andrews, S. Babraham Bioinformatics—FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
73. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022).
74. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
75. McMurdie, P. J. & Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).
76. Oksanen, J. *et al.* The vegan Package (2022).
77. Fox, J. & Weisberg, S. *An R Companion to Applied Regression* (SAGE Publications, 2019).
78. Lin, H. & Peddada, S. D. Analysis of compositions of microbiomes with bias correction. *Nat. Commun.* **11**, 3514 (2020).

Acknowledgements
The authors acknowledge the breeders and the technicians for the samples and data collection of the study as well as Gérard HEUMEZ (Gènes Diffusion) for its support in the data processing. In addition, the authors thank the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, https://doi.org/10.15454/1.5572390655343293E12) for providing the computational resources needed for our analyses.

Author contributions
CA, SD and GE conceptualized the study and conceived the experimental design. SMa and SMe conducted the whole lab work. LB performed the bioinformatic data processing and biostatistical analyses under the supervision of JE. LB, JE, SD, PC, CA, GE, and M-PS all participated in the interpretation of the results. LB drafted the manuscript. All authors conceptualized the manuscript, edited later drafts, read, and approved the final manuscript.

Funding
Research was funded by the Gènes Diffusion company (Douai, France). The PhD contract of L. Brulin was financially supported by the French "Association Nationale de la Recherche et de la Technologie" (ANRT; CIFRE n°2021/0713).

Competing interests
LB, SD, GE, SMa, SMe and CA were employed by GD Biotech/Gènes Diffusion company. LB, SD, GE, SMa, SMe and CA were employed by GD Biotech/Gènes Diffusion company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-69715-7.

Correspondence and requests for materials should be addressed to L.B.