

<https://doi.org/10.1038/s41522-025-00860-1>

Longitudinal study of the udder microbiome using genome-centric metagenomics uncovers pathogen-driven adaptation and succession

Vinicius da Silva Duarte^{1,2}✉, Fiona Valerie Franklin¹, Alicja Krysmann¹ & Davide Porcellato¹

Bovine mastitis remains a major disease affecting dairy herds globally due to its complex and multi-etiological nature. To address gaps in microbial and immunological understanding, this longitudinal study examined the udder microbiome across lactation in 24 Norwegian Red cows. Somatic cell count (SCC) and microbiota composition varied by lactation stage, with low SCC (< 100,000 cells/mL) more frequent in early (80%) and middle (78.9%) than late lactation (53%) and dry-off (53.1%). Microbial diversity was shaped by SCC, lactation stage, and individual variability. Temporal profiling identified persistent infections involving *Staphylococcus aureus* and *Staphylococcus chromogenes*, while samples with low SCC were enriched in beneficial genera including *Corynebacterium*, *Bradyrhizobium*, and *Lactococcus*. Shotgun metagenomics revealed pathogen-specific metabolic traits, and genome-centric analysis recovered 142 MAGs characterized via sequence typing, virulence, and resistance profiling. These findings offer valuable insights into microbial adaptation and succession, informing strategies to better manage and prevent mastitis.

Bovine mastitis remains the leading disease affecting dairy herds worldwide¹. Due to its multi-etiological nature, it is a challenging condition to eradicate and is primarily managed through husbandry practices and the use of antibiotics to treat bacterial infections², which are the primary cause of intramammary infections (IMIs).

Mastitis can be classified into clinical and subclinical forms based on clinical features. Clinical mastitis (CM) is marked by distinct symptoms, such as the presence of flakes, clots, or watery secretions in milk. Affected quarters typically show swelling, warmth, and pain. In acute cases, systemic signs like hyperthermia, anorexia, and depression may also occur. The consequences of CM can be severe, often leading to cow mortality, agalactia, or premature culling³. Subclinical mastitis (SCM), on the other hand, is primarily identified by an increased somatic cell count (SCC) and reduced milk production, with SCC being considered the diagnostic gold standard for detecting SCM⁴.

Microbiological culture of milk remains the gold standard for pathogen identification, despite some limitations⁵. Few studies have investigated the microbial composition of milk samples classified as “mixed growth,” which are often considered contaminated during sampling⁶, culture-negative samples from clinical mastitis⁷, and samples from cows or quarters with high

SCC (subclinically affected quarters)⁸. Overall, both mixed growth and culture-negative samples constitute a substantial proportion (30% or more) of the milk samples submitted for mastitis diagnostics⁹.

Recent advancements in high-throughput next-generation sequencing (NGS) technology and bioinformatics tools over the past decade have facilitated a shift from traditional clinical microbiology to ecological and genomic characterization of the microbiome associated with infections in different research fields¹⁰, including udder health¹¹. This transition does not aim to replace traditional clinical microbiology but rather emerges as a complementary diagnostic tool and a means to study the interactions between the host and its microbiome^{12,13}. For example, studies employing metataxonomic analysis have shown that culture-based diagnoses generally align with the most prevalent organisms identified through metagenomic sequencing, although additional potential pathogens have been detected in culture-negative samples¹⁴.

The dynamics of udder immunity and infection are complex, influenced by factors such as the cow's health status, environmental conditions, and the resident microbiota¹⁵. Research has identified distinct changes in microbiome composition between healthy and mastitic milk, with key phyla such as *Proteobacteria*, *Bacteroidetes*, *Firmicutes*, and *Actinobacteria* playing

¹Faculty of Chemistry, Biotechnology and Food Science Norwegian University of Life Sciences (NMBU), Ås, Norway. ²Research Institute of Internal Medicine, Oslo University Hospital, Oslo, Norway. ✉e-mail: v.d.s.duarte@ous-research.no

significant roles¹⁶. Metagenomic analysis has further revealed previously unreported opportunistic strains in clinical mastitis samples, along with functional pathways related to bacterial colonization and antibiotic resistance¹⁷. However, the etiology of udder dysbiosis remains unclear, particularly whether it serves as a cause or consequence of disturbances in host-specific factors, health status, environmental conditions, xenobiotic exposure, or inadequate hygiene. While comprehensive studies directly addressing this question in bovine udders are limited, research on lactating mothers can provide valuable insights¹⁸. Additionally, as reported by Urrutia-Angulo et al. (2024)¹⁹, the clinical definition of udder health status does not consistently align with the microbial profile.

Evidence suggests that SCM may have a distinct pathological origin compared to CM, with unique microbial signatures and sensory protein profiles²⁰. These findings highlight the significant potential of metagenomic approaches to improve mastitis diagnosis and advance our understanding of its pathophysiology. Metagenomic studies provide valuable insights into the temporal dynamics of microbial communities and²¹, when paired with metagenome reconstruction, reveal possible interactions within the microbiota (ecological studies) and between the microbiota and the host^{22,23}. Furthermore, when integrated with community proteogenomics, these approaches become instrumental in identifying potential biomarkers for the early diagnosis of subclinical mastitis and in elucidating the pathophysiology of both subclinical and clinical mastitis, which remains incompletely understood^{24,25}.

Optimizing the udder microbiome has been proposed as a promising strategy for preventing mastitis in dairy cattle, as a healthy microbiome

contributes to protection against pathogen colonization and the overgrowth of opportunistic pathobionts^{26,27}. Therefore, this longitudinal study aimed to investigate the udder microbiome throughout lactation using metatransomic and shotgun sequencing approaches. Studies of this nature contribute to establishing pathogen-based therapies and provide insights into the composition of the milk microbiome, which remains far from fully characterized. This knowledge gap motivated the use of shotgun metagenomics and analysis at the metagenome-assembled genome (MAG) level in this study since such comprehensive investigations remain relatively scarce.

Results

Changes in SCC and pathogen distribution across lactation periods reveal significant variability

In this study, we collected 342 quarter-level milk samples from 24 Norwegian Red cows (8 primiparous, 16 multiparous). Based on SCC, 233 samples were classified as low (< 100,000 cells/mL) and 109 as high (> 100,000 cells/mL). The proportion of low-SCC quarters was higher during EL-2023 (80%) and ML-2023 (78.9%) compared to DO-2022 (53.1%) and LL-2023 (53%), with corresponding high-SCC rates of 20%, 21.1%, 46.9%, and 47%. As shown in Fig. 1A, SCC varied significantly across lactation periods (DO-2022/EL-2023, $p = 2.92 \times 10^{-6}$; EL-2023/LL-2023, $p = 5.92 \times 10^{-5}$; DO-2022/ML-2023, $p = 1.42 \times 10^{-5}$; LL-2023/ML-2023, $p = 2.72 \times 10^{-4}$), except for DO-2022/LL-2023 and EL-2023/ML-2023. IBC followed a similar trend, with significantly higher values in high-SCC samples (mean = 3.74 ± 0.33) compared to low-SCC samples (mean = 3.31 ± 0.48).

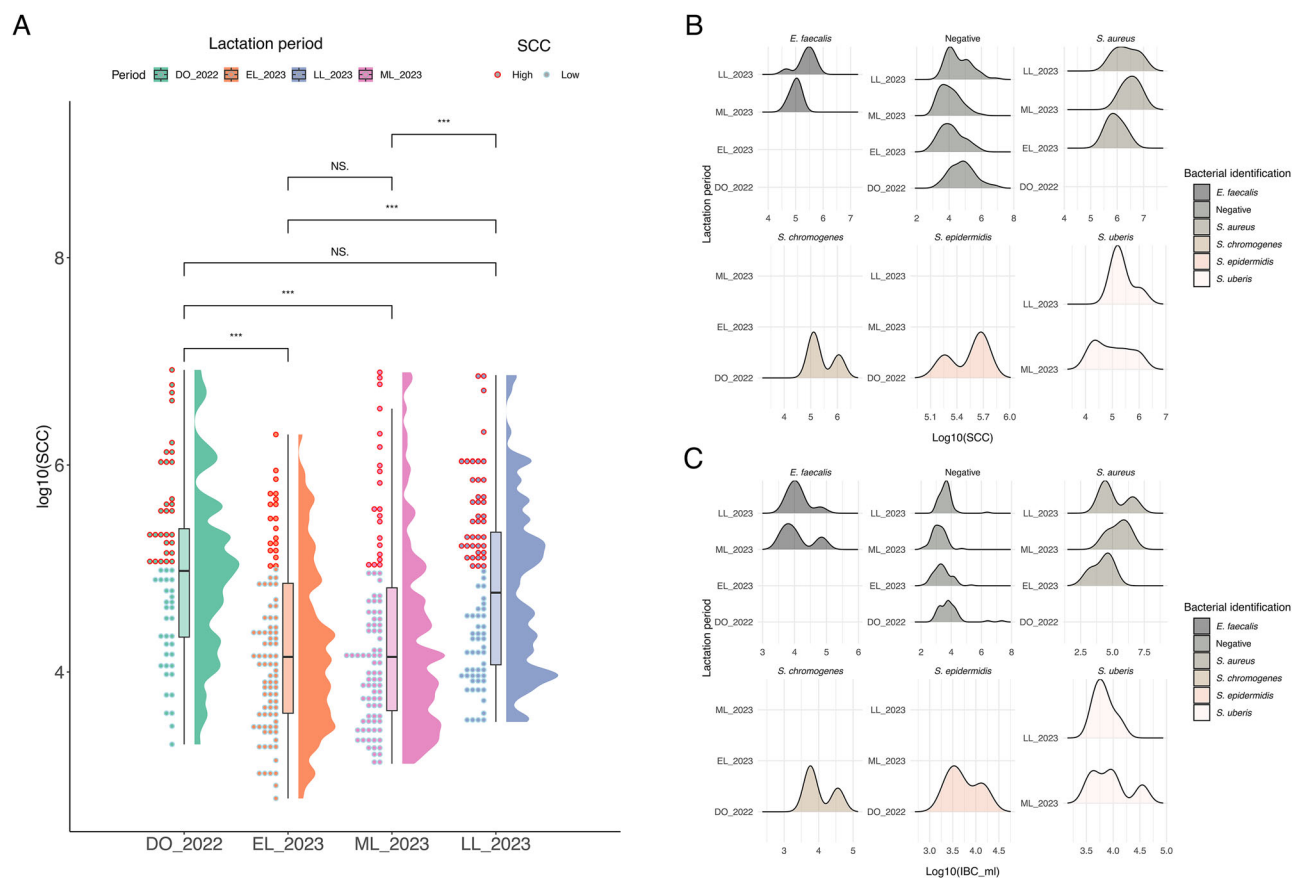


Fig. 1 | Somatic cell count dynamics and bacterial load across lactation periods. Raincloud plot (A) illustrating somatic cell count ($\log_{10}(\text{SCC})$) across different lactation periods (DO-2022, EL-2023, ML-2023, and LL-2023). Individual samples are represented by dots, color-coded as high SCC (red) or low SCC (light blue). Side histograms depict data distribution, while central bars indicate the median and interquartile range. Statistical comparisons between lactation periods are denoted by

significance levels: *** $p < 0.001$, ** $p < 0.01$, and NS (not significant). Density plot showing the relationship between SCC (B) and bacterial count ($\log_{10}(\text{IBC})$) (C) across lactation periods, stratified by bacterial species identified using MALDI-ToF. Density curves represent distinct pathogens (*E. faecalis*, *S. aureus*, *S. chromogenes*, *S. epidermidis*, and *S. uberis*) and samples classified as “negative”, highlighting variations in bacterial prevalence and load among lactation periods.

The percentage of quarters testing positive or negative for mastitis pathogens, identified via MALDI-Tof, also reflected the trends observed for SCC and IBC (Fig. 1B, C). The lowest detection of mastitis-causing pathogens was recorded during EL-2023, with only 10.5% of samples testing positive. Among the mastitis-causing pathogens identified in this study, *E. faecalis* emerged as the predominant etiological agent, accounting for 31% of the cases. *S. aureus*, *S. uberis*, and *S. chromogenes* were detected in 22%, 21%, and 10% of the samples, respectively. The bacterial species *S. epidermidis* (7%), *C. bovis* (3%), *S. dysgalactiae* (3%), and *S. haemolyticus* (2%) were identified to a lesser extent.

We also assessed the relationship between SCC and IBC in relation to specific pathogens. SCC was significantly higher in samples with *S. epidermidis* compared to pathogen-negative samples (Supplementary Fig. 1). Additional analysis of bacterial growth (48 h, 37 °C), combined with IBC, SCC, and pathogen data, revealed that many pathogen-negative samples still showed substantial colony counts (Fig. 2A). Of 129 samples with 30–300 CFUs, 57 showed SCC > 100,000 (median: 2.9×10^5) and a median IBC of 7.6×10^3 , suggesting that undetected or non-classical microbes may elevate SCC despite negative pathogen screening.

Alpha and beta diversity analyses show significant effects of the lactation period on diversity indices

We investigated the microbial composition of 306 milk samples encompassing a whole lactation cycle. By conducting amplicon sequencing, we obtained a total of 1.3×10^7 high-quality filtered sequences (low SCC, median = 37,589 sequences; high SCC, median = 47,282 sequences).

We investigated alpha diversity related to SCC and the lactation period. Low SCC samples showed greater microbial diversity and reduced species dominance (Fig. 2B). Five of seven diversity indices were higher in DO-2022 compared to LL-2023, and all indices were significantly different between EL-2023 and LL-2023 (Fig. 2C). Volatility plots (Fig. 2D, E) indicated a decline in microbial diversity at the end of lactation for samples with high SCC. The generalized additive model revealed significant effects of the lactation period and SCC on the Shannon and Simpson indices. The lactation period influenced diversity modestly (Shannon: edf = 2.3, $F = 3.7$, $p = 1.4 \times 10^{-2}$; Simpson: edf = 2.3, $F = 3.5$, $p = 2.1 \times 10^{-3}$), and SCC had a strong impact (Shannon: edf = 0.9, $F = 17.5$, $p = 3.0 \times 10^{-3}$; Simpson: edf = 0.9, $F = 24.9$, $p = 3.4 \times 10^{-4}$). Individual animal variability was also significant (Shannon: edf = 13.2, $F = 1.8$, $p = 6.1 \times 10^{-4}$; Simpson: edf = 12.0, $F = 1.4$, $p = 2.9 \times 10^{-3}$), while the parity did not influence the two metrics (Shannon: edf = 1.0, $F = 1.4$, $p = 0.24$; Simpson: edf = 1.0, $F = 1.4$, $p = 0.23$). Overall, the model explained 22.8% and 21.9% of the deviance for the Shannon and Simpson indices, respectively, with adjusted R-squared values of 17.8% and 17.2%, underscoring the relevance of the assessed factors.

PERMANOVA analyses using unweighted and weighted UniFrac distances identified significant factors influencing microbial community structure (Supplementary Fig. 2). In the unweighted UniFrac analysis, lactation period ($F = 3.2$, $p = 0.01$) and individual animal variability ($F = 1.3$, $p = 0.01$) together explained approximately 18% of the variance. In contrast, the weighted UniFrac analysis revealed that somatic cell count (SCC) ($F = 3.2$, $p = 0.01$), lactation period ($F = 7.6$, $p = 0.01$), individual animal variability ($F = 2.7$, $p = 0.01$), and the presence of mastitis pathogens ($F = 3.3$, $p = 0.01$) accounted for around 30% of the variability.

Differential abundance analysis and multivariable associations reveal shifts in Actinobacteriota across lactation stages

After clustering at a 99% similarity threshold and removing *g-Mitochondria*, *o-Chloroplast*, and *d-Archaea*, we identified a total of 48,622 ASVs. Four phyla comprised approximately 99% of the milk bacterial population: *Firmicutes* (median: 63%), *Actinobacteriota* (median: 20%), *Proteobacteria* (median: 7%), and *Bacteroidota* (median: 1%). At the genus level, ten taxa accounted for about 65% of the microorganisms: *Corynebacterium* (median: 12.9%), *Staphylococcus* (median: 4.0%), *Bradyrhizobium* (median: 1.8%), *Streptococcus* (median: 0.2%), *Aerococcus* (median: 2.6%), *Romboutsia* (median:

3.4%), *Enterococcus* (median: 0.2%), *Kocuria* (median: 0.9%), *Lactococcus* (median: 0.2%), and *Weissella* (median: 0.6%).

Multivariable association analysis with MaAsLin2 showed significant differences for 15 phyla (Supplementary Table 1). Low SCC samples showed a higher abundance of eight phyla, notably *Proteobacteria*, *Actinobacteriota*, and *Bacteroidota*. Across lactation periods, *Firmicutes* abundance remained stable ($q > 0.25$), while *Proteobacteria* was enriched in EL-2023 and declined thereafter. Conversely, *Actinobacteriota* increased during ML-2023 and LL-2023. Considering sampling day as a continuous variable, *Actinobacteriota* fluctuated, peaking at baseline (DO-2023), dropping in early lactation, and then rising again in middle and late lactation.

The genus-level analysis identified 66 differentially abundant genera (Supplementary Table 2). Samples with low SCC were enriched in several genera, markedly *Corynebacterium*, *Bradyrhizobium*, *Romboutsia*, and *Lactococcus*. In contrast, *Staphylococcus* was significantly enriched in samples classified as high SCC. By lactation period, *Bradyrhizobium*, unidentified *Lactobacillaceae*, *Pediococcus*, unidentified *Aerococcaceae*, and *Weissella* were enriched in EL-2023 when compared to DO-2022, with two additional genera from *Aerococcaceae* and *Lactobacillaceae* enriched in ML-2023. Over time, *Bifidobacterium* and *Bacillus* increased in abundance, while unidentified *Lactobacillaceae* decreased.

The integration between t-SNE and culturomics improves milk microbiota classification

Due to the high variability in udder microbiota, we applied t-SNE for dimensionality reduction on 295 samples, grouping them into 18 k-means clusters based on similar microbial profiles, determined by the Elbow method (Supplementary Fig. 3A–C). These clusters were analyzed for taxonomic composition and correlated with IBC, SCC, and alpha diversity indices (Fig. 3). Statistical analysis showed significant differences in IBC and SCC levels among several clusters compared to the global mean. *Corynebacterium*, *Staphylococcus*, and *Aerococcus* were prevalent and widely distributed across clusters, while *Enterococcus* and *Streptococcus* were restricted mainly to clusters 13 and 14, respectively, present sporadically in other clusters.

Considering a minimum mean relative abundance of 10% per cluster (Fig. 3, Supplementary Fig. 3), we observed a notably high relative abundance of *Staphylococcus* in five clusters ($p < 0.05$) (cluster 4: 80.50%, cluster 6: 77.88%, cluster 7: 58.85%, cluster 10: 56.96%, cluster 15: 52.69%). Combined with the culturomics, cluster 4 was enriched for *S. chromogenes* (4 positive samples out of 12), whereas cluster 10 was enriched for *S. aureus* (7 positive samples out of 9). Only one species of *S. haemolyticus* was identified by MALDI-Tof (cluster 6).

We conducted a MegaBLAST analysis on the top 20 ASVs to identify species within clusters dominated by *Staphylococcus*, *Corynebacterium*, and *Streptococcus* (Supplementary Fig. 5). In cluster 4, samples showed a homogeneous distribution of the four most abundant ASVs, all annotated as *S. chromogenes* (99.53–100% identity). Cluster 6 was dominated by an ASV with 100% identity to *S. borealis*, *S. pragensis*, *S. petrasii*, and *S. croceilyticus*. Cluster 10 was primarily dominated by an ASV matching different species of *Staphylococcus*, though sample H193 displayed a distinct profile lacking *S. aureus* dominance. The dominant ASV in cluster 7 matched *S. epidermidis*, *S. caprae*, and *S. capitis* with 100% identity, while four samples (H298, H300, H305, H307) also exhibited high levels of ASVs from the *S. borealis* group. Cluster 15 contained ASVs identified as *S. epidermidis*, *S. caprae*, *S. capitis*, and *S. hominis*.

For *Corynebacterium*, eleven of eighteen clusters had >10% relative abundance, notably clusters 2 (37.7%), 8 (64.9%), and 18 (44.6%) (Fig. 3, $p < 0.05$). Culturomics detected *Corynebacterium bovis* only in two samples (H087, H109). ASV-level analysis revealed many best hits were uncultured bacteria (Supplementary Fig. 5). Cluster 2 showed diverse *Corynebacterium* species with enrichment of *C. bovis*, while clusters 8 and 18 had more homogeneous ASV distributions, including uncultured bacterium/*C. bovis* variants (99.76–100% identity). Finally, clusters 3 and 14 were enriched in *Streptococcus*. Cluster 3 contained three ASVs matching *S. uberis/porcinus*

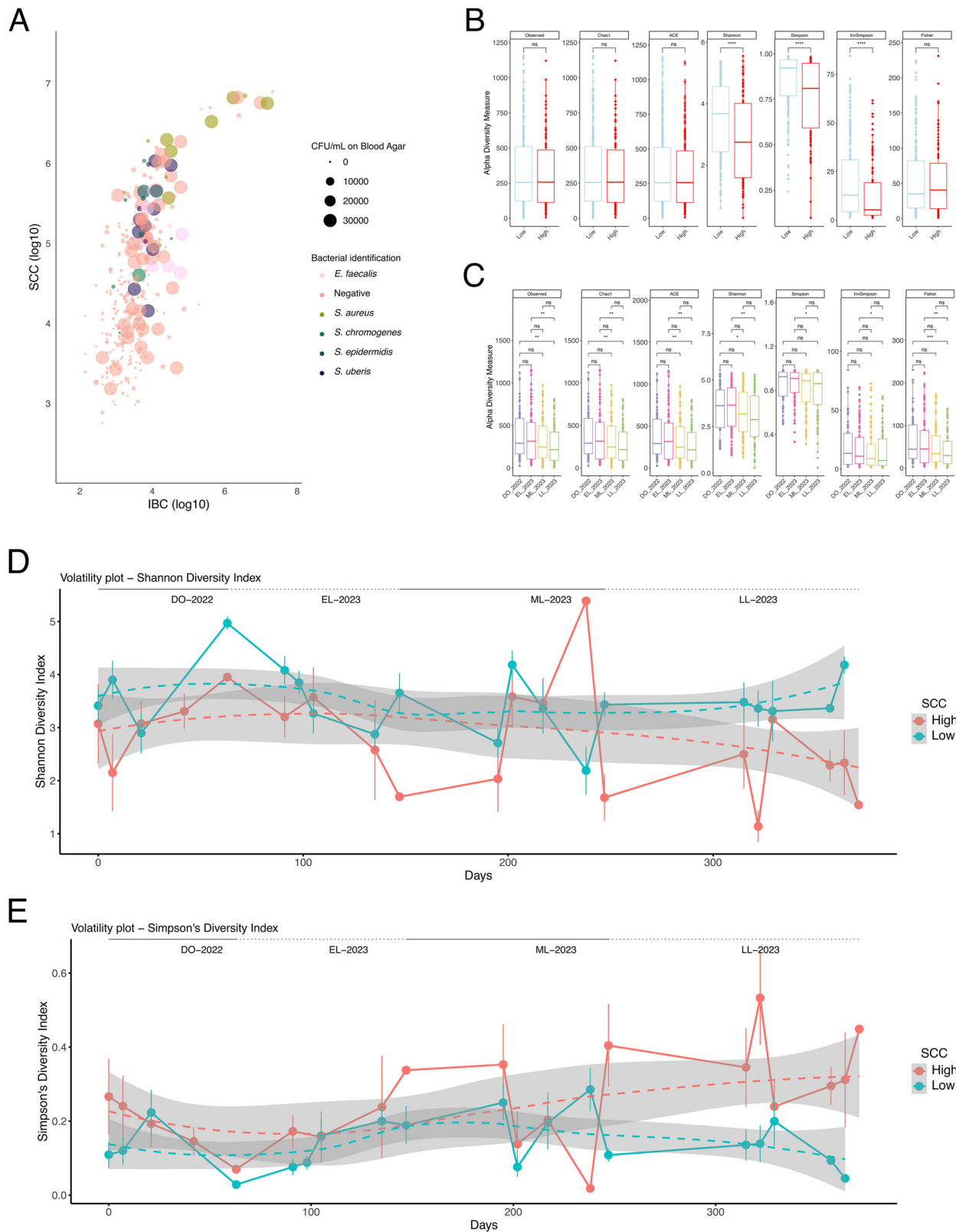


Fig. 2 | Bacterial load, diversity indices, and temporal dynamics in relation to somatic cell count. Bubble chart (A) displaying the relationship between somatic cell count (SCC) and individual bacterial count (IBC) across samples, categorized by bacterial species. Bubble sizes represent CFU/mL on blood agar, while colors indicate bacterial species (*E. faecalis*, *S. aureus*, *S. chromogenes*, *S. epidermidis*, *S. uberis*), and milk samples classified as “negative” for specific mastitis-causing pathogens. Box plots comparing Shannon (B) and Simpson (C) diversity indices between high

and low SCC groups across different time points and lactation periods. Statistical significance is denoted as *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. Volatility plot of Shannon (D) and Simpson (E) diversity indices over time, showing trends across lactation periods (DO-2023, EL-2023, ML-2023, and LL-2023). Solid and dashed lines represent high and low SCC groups, respectively, with shaded areas indicating 95% confidence intervals with group-wise comparisons highlighted.

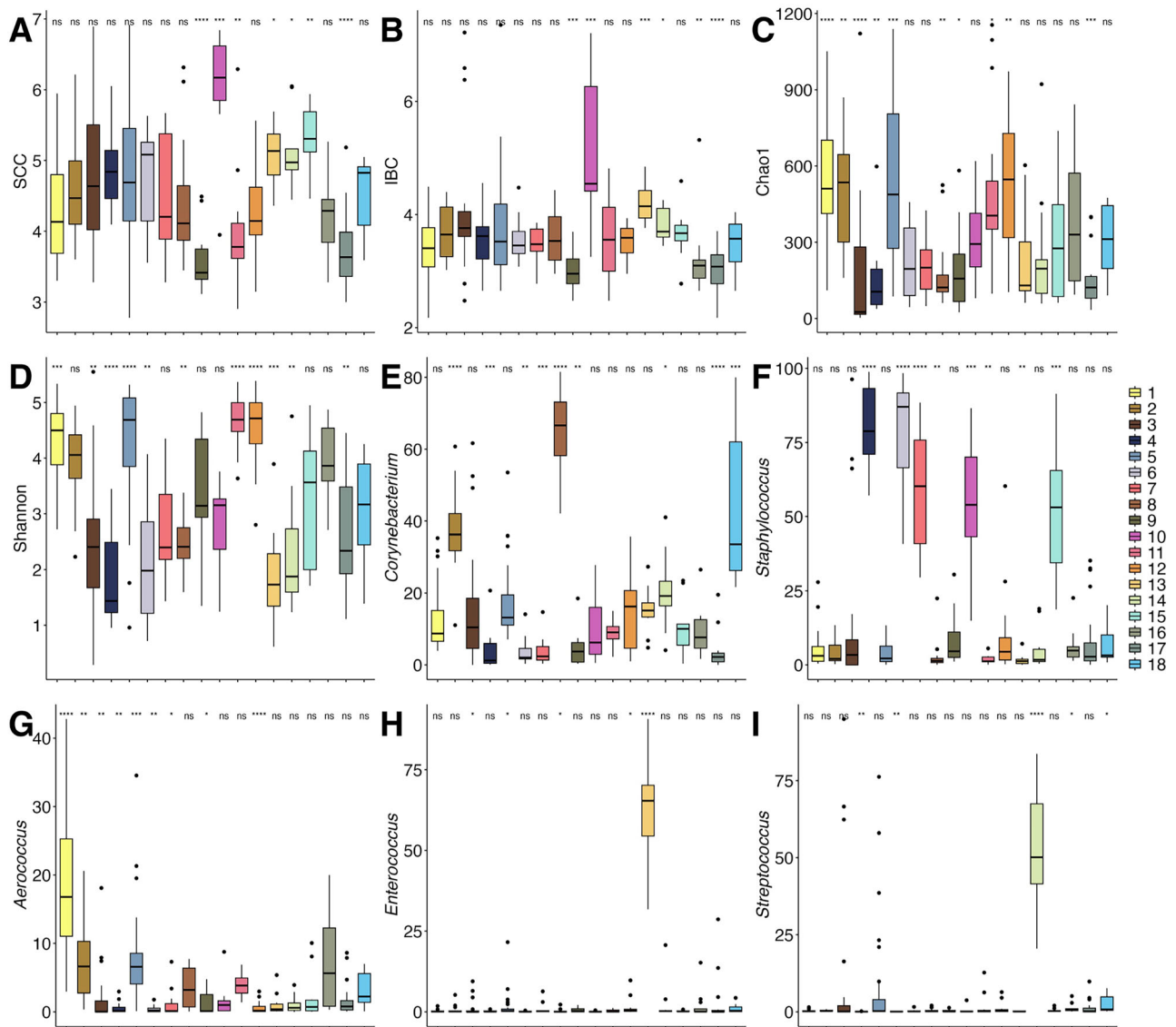


Fig. 3 | Microbiota cluster structure in relation to somatic cell count, bacterial load, diversity indices, and dominant genera. Distribution of bacterial count (A), somatic cell count (B), and alpha diversity indices (C, D) across the 18 microbiota clusters. Each box represents the interquartile range (IQR), with the median indicated by a horizontal line and whiskers extending to 1.5 times the IQR. Outliers are shown as individual points. Statistical significance of differences in relative abundance between clusters and the global mean is denoted above each box plot ($p < 0.05$; $p < 0.01$; $p < 0.001$; $p < 0.0001$; ns not significant). Box plots (E–I) showing the

distribution of the relative abundance of the five dominant genera (*Corynebacterium*, *Staphylococcus*, *Aerococcus*, *Enterococcus*, and *Streptococcus*) across the 18 microbiota clusters. Each box represents the interquartile range (IQR), with the median indicated by a horizontal line and whiskers extending to 1.5 times the IQR. Outliers are shown as individual points. Statistical significance of differences in relative abundance between clusters and the global mean is denoted above each box plot ($p < 0.05$; $p < 0.01$; $p < 0.001$; $p < 0.0001$; ns not significant).

and *S. alactolyticus/macedonicus/pasteurianus* (100% identity), while cluster 14 was dominated by a single ASV identified as *S. uberis* (100% identity).

Udder microbiota dynamics reveal persistent infection by *S. aureus* and *S. chromogenes*, but not *S. uberis*

The result from the t-SNE analysis enhanced sample classification and elucidated underlying patterns in the dataset. We therefore conducted a quarter-level analysis to track temporal fluctuations in genus-level taxa before and after detecting the pathogen identified by the gold standard method. Focus was given to samples positive for *Staphylococcus*, *Enterococcus*, and *Streptococcus*, primarily grouped in clusters 4, 10, 13, and 14. Only quarters with at least three sampling points were included (Fig. 4). To assess microbiota resilience post-dysbiosis, we combined a microbial dysbiosis index (MDI) with the Shannon diversity index. Quarters with a

Shannon index below 3.5 (mean for quarters with SCC < 100,000 cells/mL) and MDI > 0 were classified as dysbiotic.

In cluster 10 (enriched for *S. aureus*), analysis of 5 quarters showed persistent infection with *S. aureus* across lactation stages. Despite fluctuations in bacterial counts, SCC remained consistently elevated. Significant depletion of *Corynebacterium*, *Aerococcus*, *Romboutsia*, and other low-abundance taxa was observed. Co-infections with *Streptococcus* species occurred at two time points (samples H195 and H349). Overall, the MDI confirmed that post-infection samples did not revert to eubiosis. Cluster 4, enriched for *S. chromogenes*, also showed persistent infection, but SCC and bacterial counts were not significantly different from the global mean (Fig. 3, 4). In quarter 7032_VF, co-infection with *S. aureus* prevented microbiota recovery, a pattern also evident in quarters 6691_HB and 6691_VB. Our results suggest that these bacteria may

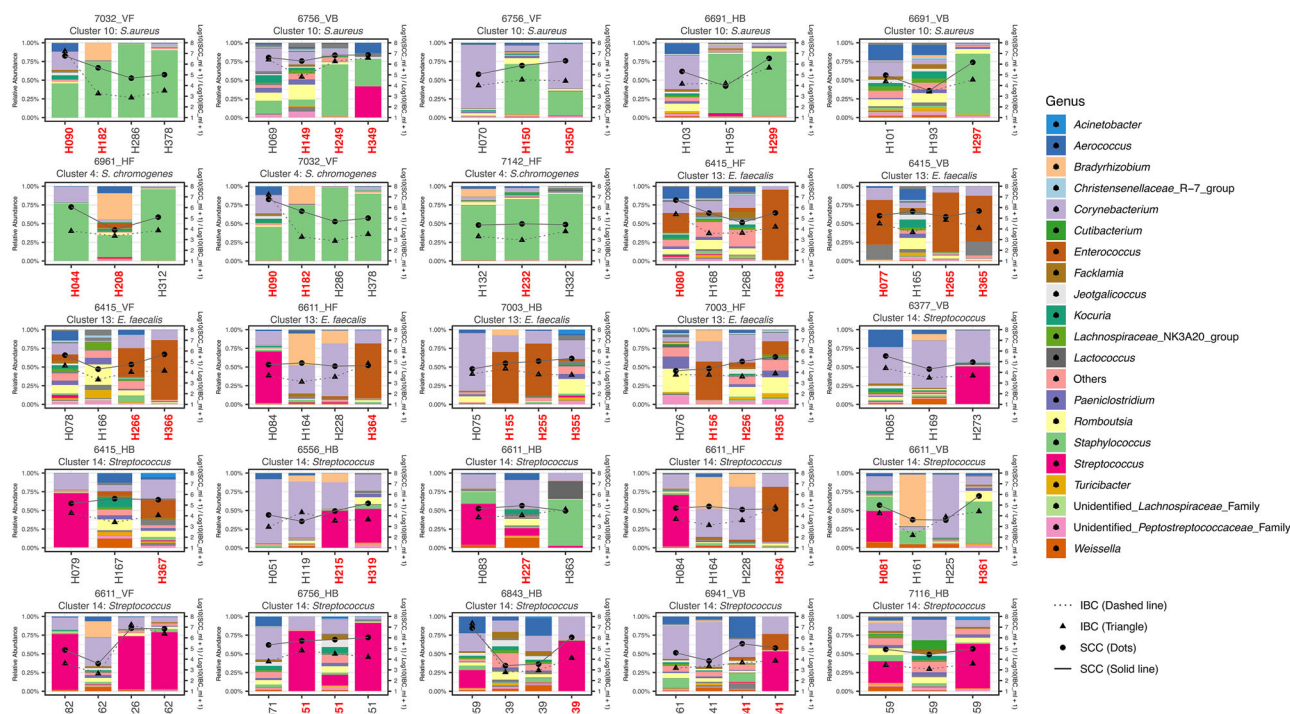


Fig. 4 | Udder microbiota dynamics. Stacked bar plots depicting the relative abundance of the 20 most abundant genera across different sampling quarters. Each bar represents the proportional composition of genera, with colors corresponding to specific genera as indicated in the legend. Genera with lower abundances are grouped under “Others.” Bacterial and somatic cell counts are overlaid: triangles connected by dashed lines represent IBC levels, while dots connected by solid lines

indicate SCC levels, illustrating changes in microbiota composition in relation to these metrics over sampling time. The analyzed animal and sampling quarter, along with microbiota clusters, are labeled above each panel, with sample identification numbers displayed below each plot. Samples highlighted in red indicate the detection of a mastitis-causing pathogen identified via MALDI-ToF analysis.

engage in competitive interactions that prolong dysbiosis by preventing recolonization of the native microbiota.

Cluster 13, enriched for *E. faecalis*, showed milder dysbiosis with reductions in *Corynebacterium*, *Aerococcus*, and *Romboutsia*, but the native microbiota often recovered, as seen in quarters 6415_VB, 7003_HB, and 7003_HF. Lastly, in cluster 14, dominated by *S. uberis*, depletion of *Corynebacterium* was less severe than in *Staphylococcus* or *Enterococcus* clusters. Although *S. uberis* caused dysbiosis, five quarters returned to eubiosis, and one was borderline (samples H084, H164; quarters 6611_HF, 6415_HB, 6611_HB, 6611_HF, 6941_VB).

Metagenomic shotgun sequencing uncovers pathogen-specific metabolic signatures in the bovine mammary gland

After trimming, quality filtering, and separating bacterial reads from bovine sequences, we obtained ~269.5 GB of high-quality paired-end reads from 73 samples for functional profiling with HUMAnN3. This identified 289 MetaCyc metabolic pathways, analyzed for differential abundance via MaAsLin2 using “SCC” and “pathogen” as fixed effects. The heatmap (Supplementary Fig. 6A) showed no clear clustering by SCC or lactation period, indicating high heterogeneity, though four pathogen-negative clusters lacking metabolic enrichment and clusters positive for *S. aureus* and *E. faecalis* were discernible. Among differentially abundant pathways ($q < 0.25$), 134 were enriched in high SCC samples and four in low SCC samples (Supplementary Fig. 6B, C). Within the high SCC group, numerous pathways associated with central metabolism were shared predominantly by *Staphylococci* and *Streptococci*.

We also assessed the enrichment of metabolic pathways by bacterial species identified via MALDI-ToF, compared to samples classified as negative. This analysis aimed to identify potential metabolic distinctions associated with pathogen presence and uncover infection-specific metabolic signatures unique to the mammary gland. In total, five pathogens were analyzed: *S. aureus*, *S. chromogenes*, *E. faecalis*,

S. epidermidis, and *S. uberis* (Supplementary Fig. 7). Based on the previously identified pathways with a q -value < 0.25 , a Venn diagram was constructed to identify unique and shared MetaCyc pathways (Fig. 5). For *S. aureus*, 14 unique MetaCyc pathways were identified showing a versatile metabolic profile, encompassing central carbon metabolism, amino acid and nucleotide biosynthesis, stress tolerance, and virulence factors. *S. chromogenes* exhibited unique metabolic pathways associated with amino acid biosynthesis, lipid metabolism, and energy production, reflecting its ability to adapt to resource-limited environments and its potential role in infection dynamics. The *S. uberis* group showed enrichment in pathways related to carbohydrate metabolism, including glycogen biosynthesis, lactose and galactose utilization, and nucleotide sugar metabolism, as well as processes involved in cell-wall biosynthesis, indicating its capacity to thrive in nutrient-rich environments, such as the mammary gland during infection. For *S. epidermidis*, eight unique metabolic pathways were annotated for this species and involved, for instance, in energy production, cell-wall structural maintenance, stress tolerance, and nutrient acquisition. *E. faecalis* displayed a metabolic profile enriched in pathways related to amino acid biosynthesis, nucleotide metabolism, and lipid-related processes, as well as pathways supporting secondary metabolite production and cofactor biosynthesis, underscoring its metabolic flexibility and potential resilience in diverse environmental conditions.

Genome-resolved metagenomic data analysis

The genome-centric approach applied to shotgun reads from milk microbiomes enabled the reconstruction of 142 MAGs, of which 26 were obtained using a co-assembly method and 116 using an individual assembly method. Among these, 38 genomes were classified as high-quality (completion $> 90\%$, contamination $< 5\%$), 50 as medium-quality (completion $> 50\%$, contamination $< 10\%$), and 39 as low-quality (completion $< 50\%$, contamination $> 10\%$) (Supplementary Table 3). MAGs with contamination

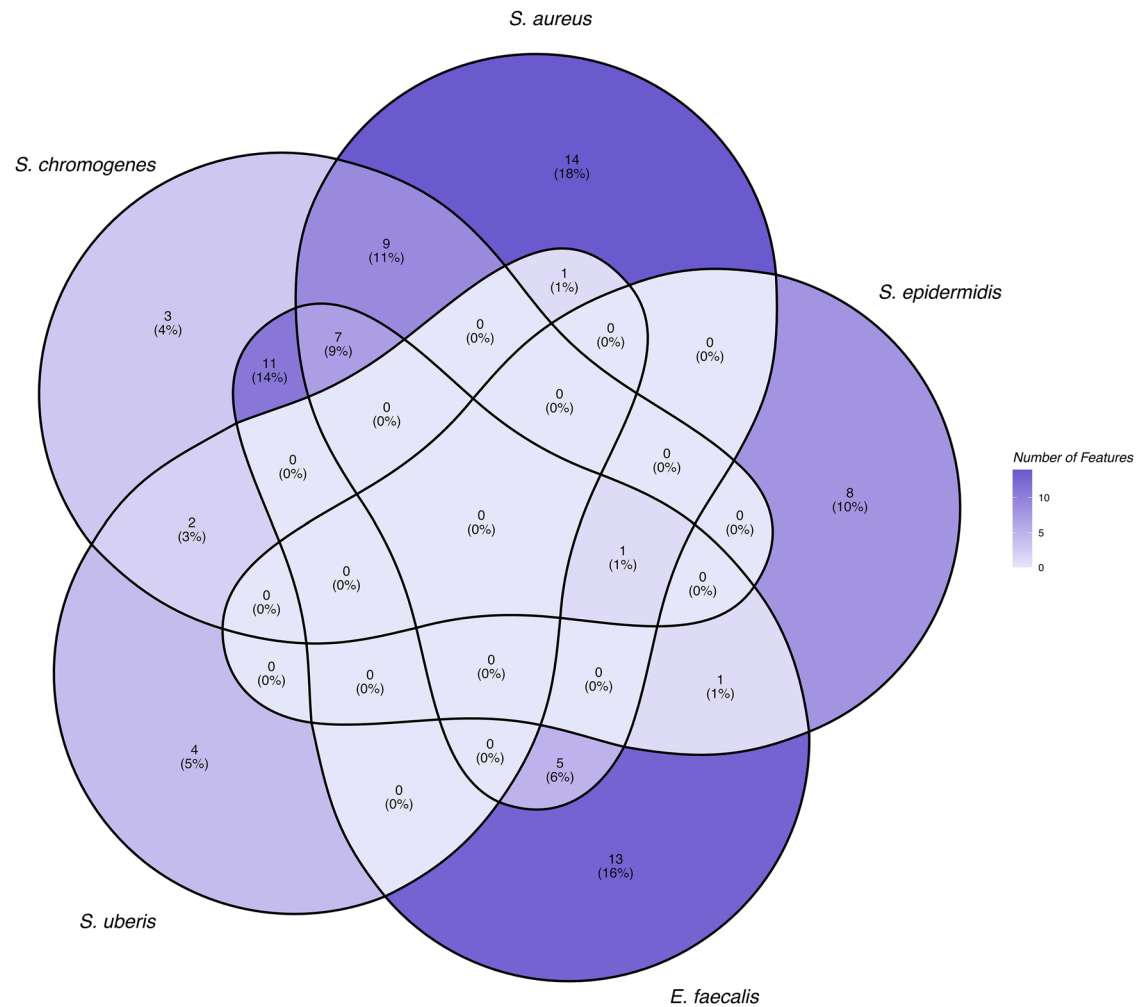


Fig. 5 | Shared and unique MetaCyc pathways across microorganisms. The Venn diagram shows the number and percentage of pathways that are unique or shared among different species. The shading gradient reflects pathway abundance, with darker shades indicating higher values.

higher than 10% were excluded from subsequent analyses. Taxonomic assignments were consolidated at the species level when always possible.

After genome dereplication and the exclusion of highly contaminated MAGs, 26 MAGs were selected as representative of the metagenomes. These MAGs were subsequently used to map the reads and estimate their relative abundances across the samples (Fig. 6). The MAG assigned as *Cutibacterium acnes* was identified as a putative contaminant in our dataset and removed from downstream analysis.

To explore microbiome functional potential, we predicted microbial genes from key MAGs and annotated protein sequences using EggNOG, normalizing COG categories by total proteins per MAG. We focused on *Streptococcaceae*, *Aerococcaceae*, *Staphylococcaceae*, and *Corynebacteriaceae* due to their prominence. The top 20 COG categories mainly involved metabolism, cellular processes, and information storage, with a large portion labeled “function unknown,” reflecting many poorly characterized proteins.

Our analysis identified four COG categories with significant differences across the bacterial families (Fig. 7). COG category G, related to carbohydrate transport and metabolism, was more abundant in *Aerococcaceae* and *Streptococcaceae* compared to *Corynebacteriaceae* and *Staphylococcaceae*. In contrast, COG category Q, associated with secondary metabolites biosynthesis, transport, and catabolism, was most abundant in *Corynebacteriaceae*. For COG category C, which involves energy production and conversion, *Staphylococcaceae* showed enrichment relative to *Corynebacteriaceae* and *Streptococcaceae* but was similar

to *Aerococcaceae*. Finally, COG category U, encompassing intracellular trafficking, secretion, and vesicular transport, was more prevalent in *Streptococcaceae* than in *Staphylococcaceae*. These findings reveal important differences in functional potential and ecological adaptation among the bacterial families.

MLST, Virulence Factors (VFs) and Antimicrobial Resistance Genes (AMRGs)

All metagenomes, along with high- and medium-quality MAGs, were subjected to MLST analysis. Results revealed that four bacterial species were associated with a single sequence type (*S. aureus*: ST133; *E. faecalis*: ST40; *S. dysgalactiae*: ST302; *S. uberis*: ST1409). In contrast, two sequence types were identified for *S. epidermidis* (ST100 and ST575), while *S. chromogenes* displayed greater diversity with three sequence types (ST1, ST59, and ST104).

The distribution of VFs across multiple bacterial taxa found in milk samples was analyzed to assess their biosynthetic and pathogenic potential (Fig. 8). For *Staphylococcus* spp., *S. aureus* showed dominant VFs categories associated with immune modulation (24.53%) and exoenzymes (14.34%), followed by nutritional/metabolic factors (13.58%), exotoxins (12.08%), adherence (11.70%), and effector delivery systems (11.70%). Bacterial species annotated as *E. faecalis* showed a broad distribution of VF categories, with adherence genes comprising the largest fraction (43.90%). Exoenzymes (15.70%), immune modulation (12.50%), and stress survival (12.50%) were also well-represented. Other categories included biofilm formation (11.34%), nutritional/metabolic factors (3.78%), and effector delivery

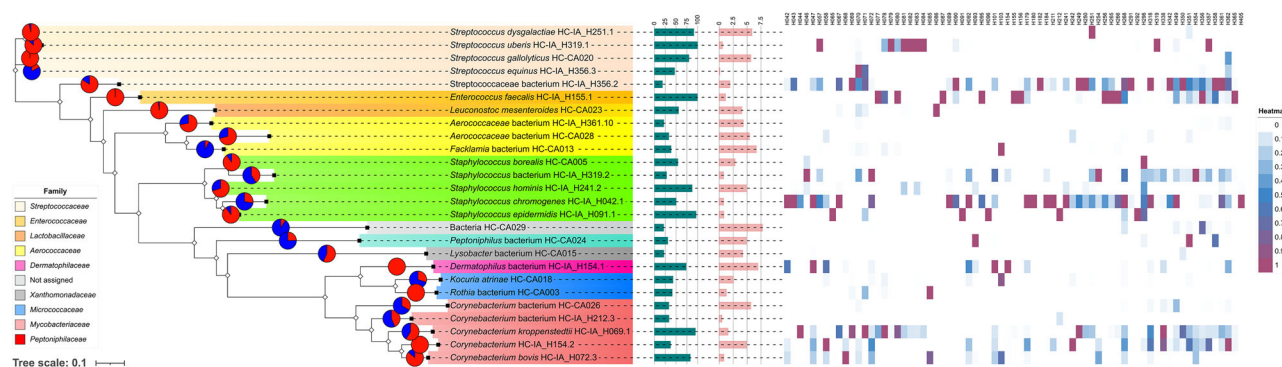


Fig. 6 | Phylogenetic tree and relative abundance of the species. From left to right tree shows pie charts with species abundance in samples with Low (blue) and High (red) somatic cell count, phylogenetic relationships colored at the family level,

barplots reporting genome completeness/contamination and heatmap representing relative abundance of species per sample.

systems (0.29%). In species of *Streptococcus*, adherence, immune modulation, and stress survival were equally represented (33.33% each) for *S. bovis*. *S. dysgalactiae* demonstrated a focus on adherence (50%), followed by exoenzymes and stress survival (14.29% each), with other categories such as immune modulation, invasion, and nutritional/metabolic factors at 7.14% each.

In the analysis of AMRGs (Supplementary Fig. 8), a total of 144 genes were annotated using three databases: ResFinder, CARD, and ARG-ANNOT. The top five antibiotic classes represented in the dataset were fluoroquinolones (15%), macrolides (10.2%), tetracyclines (9.6%), aminoglycosides (7.6%), and rifamycins (5.8%). When stratified by bacterial species, distinct patterns of AMR gene distribution were observed. For the genera *Corynebacterium* spp. and *Staphylococcus* spp., the highest abundance of AMR genes was associated with rifamycins (20.4%) and fluoroquinolones (18.2%), respectively. Similarly, *Streptococcus* spp. displayed a predominance of genes linked to fluoroquinolones (24.2%), while for *E. faecalis*, AMR genes were most frequently associated with macrolides (13.6%), fluoroquinolones (11.3%), lincosamides (8.1%), tetracyclines (8.1%), and aminoglycosides (8.0%). These findings underscore the diversity of AMR gene profiles across MAGs reconstructed from metagenomes obtained from milk samples, highlighting species-specific adaptations to antimicrobial pressures.

Prediction of bacteriocins and bioactive compounds

We used BAGEL4 to identify gene clusters in the MAGs DNA involved in the biosynthesis of Ribosomally synthesized and post-translational modified Peptides (RiPPs) and (unmodified) bacteriocins. In total, 27 gene clusters were predicted (Supplementary Fig. 9A). Bacterial species annotated as *S. aureus* the highest number of different classes, and four areas of interest were identified in MAGs assigned to *E. faecalis*.

Genome mining using AntiSMASH identified 322 biosynthetic gene clusters (BGCs) across all analyzed MAGs. These BGCs were classified into 26 distinct types of biosynthetic pathways associated with the production of secondary/specialized metabolites. The three most abundant pathways were non-ribosomal peptide synthetase (NRPS) (12.7%), terpene (11.1%), and non-ribosomal peptide siderophore (NI-siderophore) (10.8%). When the analysis was stratified by bacterial species and limited to hits with similarity associated with known clusters, several species-specific putative metabolites were identified (Supplementary Fig. 9B). The AntiSMASH analysis revealed a group of bioactive compounds shared among several *Staphylococcus* species, including staphyloferrin A, staphylopin, and kijanimicin. Unique metabolic signatures were observed in specific species, such as *S. aureus*, *C. kroppenstedtii*, *S. uberis*, and *C. bovis*. These findings suggest a potential common production of certain bioactive compounds, here predominantly identified among *Staphylococcus*, while other microorganisms, such as *C. kroppenstedtii*, appear to exhibit a more species-specific metabolite spectrum.

Discussion

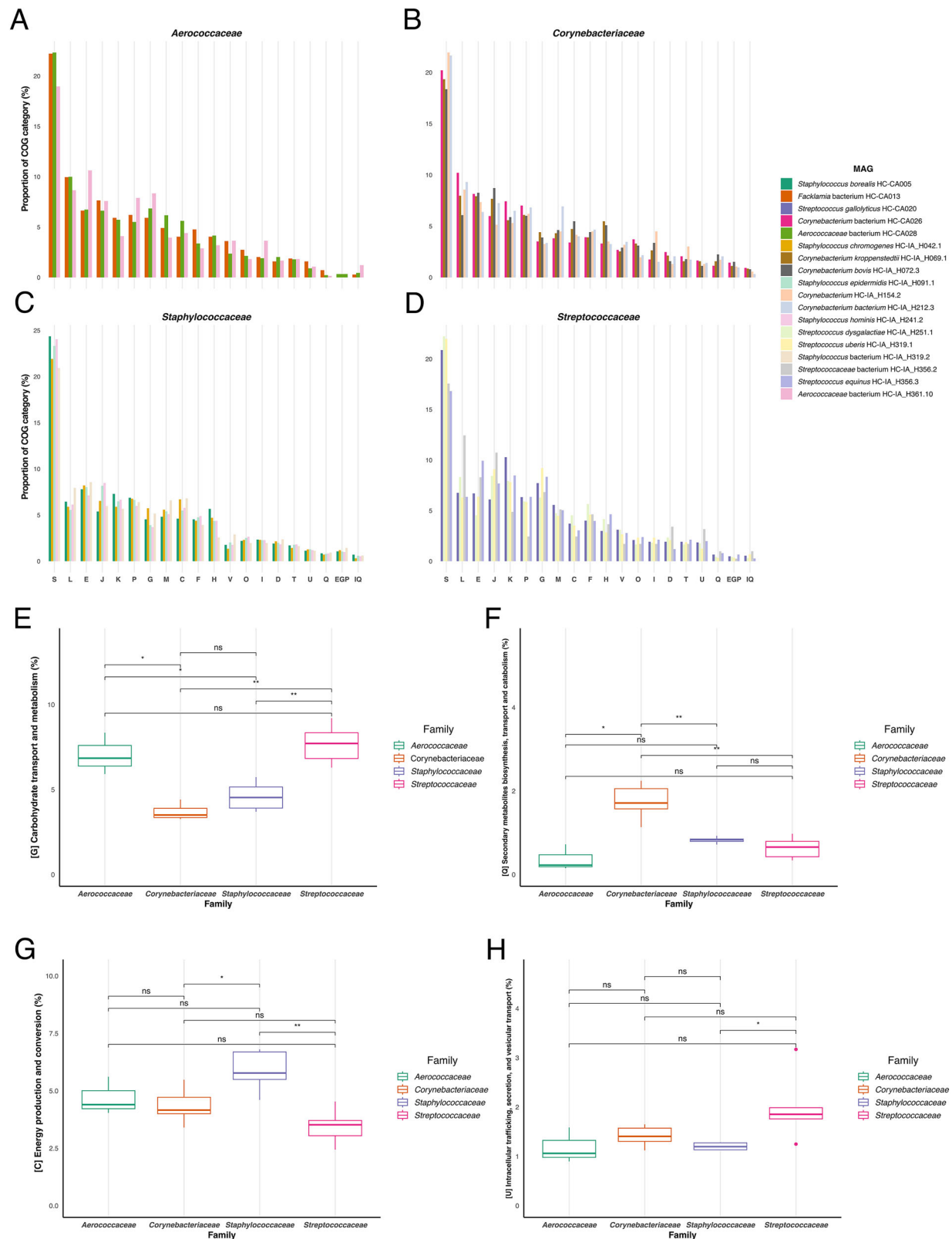
Bovine mastitis is the most economically significant disease in dairy cattle, and its complex, multifactorial etiology presents substantial challenges for effective management through both zootechnical and veterinary interventions^{28,29}.

Mastitis-causing pathogens are a primary driver of elevated SCC in the udder, with various bacterial species implicated. In our study, major pathogens such as *S. aureus*, *Enterococcus* spp., and environmental streptococci were predominant, while minor pathogens, including *S. chromogenes*, *S. haemolyticus*, *S. epidermidis*, and *C. bovis*, were less common. Apart from an elevated prevalence of *E. faecalis* during mid-lactation in 2023, which coincides with the transition from freestalls to pasture, our results align with previous surveys in Norwegian herds³⁰. As both a commensal and opportunistic pathogen, *Enterococcus* is highly resilient and commonly part of the animal gut microbiota, enabling its spread under certain environmental or management conditions³¹.

In this longitudinal study, we analyzed over 300 quarter-level milk samples to investigate microbial succession in bovine mastitis. High SCC samples exhibited reduced microbial diversity, indicating a shift toward dysbiosis driven by dominant taxa, as noted in several studies^{19,23,32}. Dimensionality reduction revealed the consistent presence of *Staphylococcus*, *Corynebacterium*, and *Aerococcus* across clusters, underscoring their ubiquity in both healthy and mastitic quarters. *Staphylococcus* species are notably contagious, and their prevalence is strain-dependent. While the role of *Corynebacterium* in udder health is still debated³³, it has been hypothesized that natural infections may confer protection against major mastitis pathogens. This is consistent with our findings and reported in previous studies^{34–36}. Meanwhile, *Aerococcus* has been primarily associated with healthy milk samples, and its occurrence with non-aureus staphylococci has been negatively correlated with *E. coli* in clinical mastitis^{26,37}.

Milk microbiota varied both between cows and between udder quarters. Notably, intramammary infections caused by *Staphylococcus* led to persistent dysbiosis, unlike infections by *Streptococcus*, which showed signs of microbial recovery. While cross-sectional studies have described the mammary microbiota^{38–40}, few have tracked its longitudinal dynamics at the quarter level. Our findings support previous research showing that *S. aureus* is difficult to eliminate once established²¹, likely due to its immune evasion strategies, such as biofilm formation, small colony variants, and its ability to invade both professional and nonprofessional cells. In contrast, *Streptococcus* infections, particularly by the environmental species *S. uberis*, were associated with transient dysbiosis and microbiota recovery, which is consistent with a study conducted by Urrutia-Angulo et al. (2024)¹⁹. These results underscore pathogen-specific dynamics and reinforce *S. aureus* as a key challenge in mastitis management.

Genome-centric metagenomics advanced our understanding of the milk microbiome by uncovering microbial interactions within the udder that are not captured by traditional amplicon sequencing. The reconstructed



MAGs reflected the dominant genera identified through 16S rRNA sequencing, including *Staphylococcus*, *Corynebacterium*, *Streptococcus*, *Enterococcus*, *Aerococcus*, *Kocuria*, and *Romboutsia*. Notably, a broad diversity of draft genomes was recovered, particularly within *Staphylococcus* and *Corynebacterium*. Among them, *Corynebacterium kroppenstedtii* HC-IA_H069.1 was highly abundant. This species is part of the

C. kroppenstedtii-like group, which includes *C. parakroppenstedtii* and *C. pseudokroppenstedtii*, organisms increasingly associated with breast abscesses and granulomatous mastitis in women⁴¹. However, their role in bovine udder health remains poorly understood and merits further investigation. *Corynebacterium* species are also recognized for their metabolic versatility and biotechnological potential, exemplified by *C. glutamicum*⁴².

Fig. 7 | Comparative analysis of clusters of orthologous groups categories among different bacterial families. Bar plots represent the top 20 COG categories across all samples for the bacterial families: *Aerococcaceae* (A), *Corynebacteriaceae* (B), *Staphylococcaceae* (C), and *Streptococcaceae* (D). Each color indicates a specific metagenome-assembled genome as shown in the legend. COG categories - (S Function unknown, L Replication, E Amino acid transport and metabolism, J Translation, ribosomal structure and biogenesis, K Transcription, P Inorganic ion transport and metabolism, G Carbohydrate transport and metabolism, M Cell wall/membrane/envelope biogenesis, C Energy production and conversion, F Nucleotide

transport and metabolism, H Coenzyme transport and metabolism, V Defense mechanisms, O Posttranslational modification, protein turnover, chaperones, I Lipid transport and metabolism, D Cell cycle control, cell division, chromosome partitioning, T Signal transduction mechanisms, U Intracellular trafficking, secretion, and vesicular transport, Q: Secondary metabolites biosynthesis, transport and catabolism—are represented on the x-axis, and their normalized proportion on the y-axis. Boxplots (E–H) represent COG categories that differed significantly among bacterial families based on the Kruskal–Wallis test. Significant pairwise differences are indicated with * ($p < 0.05$) and ns (not significant).

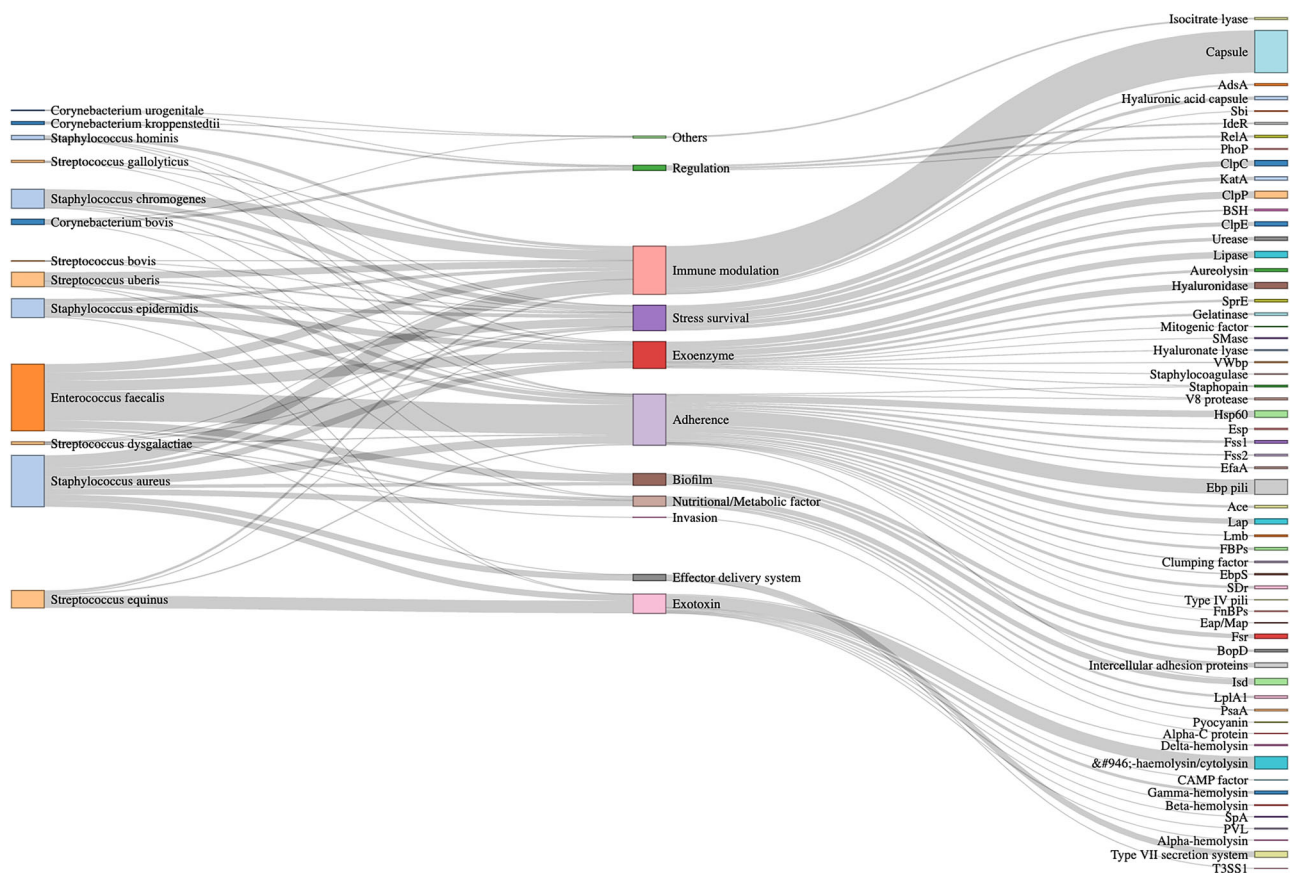


Fig. 8 | Sankey plot illustrating the relationship between bacterial taxa, virulence factor categories, and specific virulence factors. The thickness of the connecting lines represents the relationships between the taxa, virulence factor categories, and

specific virulence factors. Color coding differentiates bacterial taxa and highlights the distribution of virulence traits across the species.

Future studies should explore the adaptation of *C. kroppenstedtii*-like strains to the bovine mammary gland and conduct comparative genomics with human isolates to clarify their potential role in bovine health and disease.

Building on our genome-centric analysis, we investigated genes traditionally classified as virulence factors, many of which likely function as “niche factors” in commensal bacteria by supporting adaptation rather than pathogenicity⁴³. These elements often contribute to bacterial survival across diverse ecological contexts, including host-associated environments⁴⁴. Distinct patterns emerged across taxa. Regulatory proteins such as IdeR and PhoP were prominent in *Corynebacterium* spp., suggesting a role in niche adaptation^{45,46}. In *E. faecalis*, adherence-related genes accounted for a large proportion of the identified factors. Among *Staphylococcus* spp., immune evasion, particularly capsule biosynthesis, was the dominant strategy, consistent with their capacity for persistent and recurrent infections.

Shotgun metagenomic sequencing has recently emerged as a powerful tool for quantifying antimicrobial resistance dynamics at the human–livestock interface⁴⁷. Norway is notable for its low antibiotic usage and minimal reliance on antimicrobials in mastitis treatment, where

benzylpenicillin procaine remains the first-line therapy³⁰. Treatments are typically reserved for moderate to severe cases during lactation and for subclinical infections caused by *S. aureus*, *S. dysgalactiae*, *S. uberis*, or *S. agalactiae* at dry-off³⁰. From the health records of all cows, we confirmed that β -lactams (specifically Penicillin G) were the only antibiotics administered during the study period, and only on two occasions. All milk samples included in our analysis were collected at least two weeks before or after these treatments. Yet, our metagenomic analysis revealed antibiotic resistance genes (ARGs) extending beyond those anticipated from this limited and targeted β -lactam exposure. In particular, we detected ARGs conferring resistance to multiple antibiotic classes, inconsistent with the treatment history. These observations suggest that such resistance determinants may constitute part of the intrinsic milk resistome. Despite the restrictive use of antibiotics, *Corynebacterium* species, frequently detected in milk samples, exhibit broad resistance to macrolides, lincosamides, streptogramins, and often to β -lactams, clindamycin, erythromycin, azithromycin, ciprofloxacin, and gentamicin^{48–50}. However, most strains remain susceptible to vancomycin, minocycline, and linezolid⁵⁰. The high

prevalence and diversity of *Corynebacterium* in this study may reflect their intrinsic resistance to beta-lactams^{51,52} potentially contributing to their persistence within the herd.

In addition to antimicrobial resistance, bacterial competition through bacteriocin production also shapes microbial community structure in the mammary gland⁵³. In this study, diverse bacteriocin gene clusters were identified across multiple species, highlighting their role in microbial competition. *S. aureus*, for example, can produce lantibiotics such as BsaA2, conferring a competitive advantage in community-associated strains⁵⁴. Nukacins, also produced by *Staphylococcus* spp., show promise for mastitis control^{55,56}. These bacteria further rely on peptide-based quorum sensing systems like the accessory gene regulator (*agr*), which uses auto-inducing peptides to coordinate gene expression⁵⁷. *E. faecalis* contributes to competitive interactions through enterolysin A, a bacteriocin active against Gram-positive species including *S. aureus* and *Listeria monocytogenes*^{58,59}.

Beyond bacteriocins, *Staphylococcus* species also produce a range of secondary metabolites that contribute to microbiome interactions and nutrient acquisition, particularly iron, enhancing their competitive fitness in diverse environments. A recent genome-wide CRISPRi-seq screen by Märlä et al. (2024)⁶⁰ identified key fitness determinants for *S. aureus* growth in milk, underscoring the importance of metal acquisition pathways in this niche. Our analysis also predicted several bioactive compounds in *Corynebacterium* spp., including ϵ -poly-L-lysine—a potent antimicrobial widely used in food preservation and known to influence microbial dynamics in cheese and skin ecosystems⁶¹. In *C. kroppenstedtii*, AntiSMASH analysis revealed additional compounds such as a triscatecholamide siderophore, the catechol-peptide griseobactin, and phenazines. These molecules likely support iron acquisition under limiting conditions and may enhance ecological fitness and pathogenic potential in the mammary gland^{62,63}. Altogether, the potential for bacteriocin and secondary metabolite production highlights new avenues for microbiome modulation and alternative strategies to manage bovine mastitis.

In conclusion, bovine mastitis remains a major challenge for global dairy production, highlighting the need for innovative management strategies. This longitudinal study used metataxonomic and shotgun metagenomic approaches to track udder microbiome dynamics across lactation stages. We observed significant variations in somatic cell count and microbiota composition, with dominant species linked to high SCC samples. Dimensionality reduction identified health-associated bacterial profiles and microbial shifts at the quarter level, confirming *Staphylococcus* as a key mastitis pathogen. Metagenomics revealed pathogen-specific metabolic pathways and reconstructed 142 MAGs, offering insights into pathogen adaptation, virulence, and resistance. The potential protective roles of *Aerococcus*, *Kocuria*, and coryneform bacteria were also highlighted. Furthermore, the discovery of bacteriocin and biosynthetic gene clusters suggests new avenues for targeted antimicrobial development. Overall, this study enhances our understanding of the bovine udder microbiome and its role in the pathophysiology of mastitis, providing a foundation for microbiome-based strategies to enhance dairy herd health and productivity.

Methods

Experimental design and sample collection

Twenty-four Norwegian Red cows (8 primiparous; 16 multiparous – 8 in their second lactation and 8 beyond second lactation) were selected from a herd of 43 cows with expected calving dates between October 2022 and February 2023. Eight cows were excluded because they were enrolled in another study or were under antibiotic treatment. From the remaining 35 cows, selection was based on parity to capture herd diversity, resulting in 8 cows in their first lactation and 16 with one or more lactations. The selected cows were then divided into groups of five, ensuring that only 20 samples were collected and processed on each sampling day. The animals were housed at the Centre for Livestock Production, Norwegian University of Life Sciences, and managed in accordance with Norwegian Food Safety Authority regulations. From October to April, cows were housed in freestall barns with rubber mats and woodchip bedding, and from May to

September, they grazed on pasture. Feeding occurred twice daily, with a maize and silage mix provided from October to April (2022–2024). Concentrate was automatically dispensed during milking via robotic systems or automatic feeders. Animals were grouped by performance and lactation stage when feasible, and feeding was adjusted accordingly, using two different roughage mixes. Health records were reviewed for all cows included in the study to document treatments administered both prior to and during the study period, including antibiotic use. Only two treatments with antibiotics were recorded, both involving Penicillin G: one case for mastitis and one for interdigital phlegmon, each in a different cow. All milk samples analyzed in this study were collected at least two weeks before or after these treatments.

A total of 342 quarter-level milk samples were collected across four lactation periods spanning two lactation cycles: baseline pre-drying in 2022 (DO-2022), and early (EL-2023, median 22 days in milk (DIM)), middle (ML-2023, median 126 DIM), and late lactation (LL-2023, median 244 DIM) in 2023. Primiparous cows were sampled only during 2023, and one multiparous cow was culled after the 2022 sampling. Milking was conducted with an automatic robot, which was inaccessible from 10 PM on the evening before sampling.

On sampling days, approximately 350 mL of hindmilk was aseptically collected from each quarter following National Mastitis Council (NMC) guidelines (www.nmconline.org), including teat cleaning with iodine and 70% ethanol after milking. Milk was collected using non-invasive standard farm practices that are employed in the Norwegian dairy production system and did not alter the cows' normal routine, health, or welfare. Farm owners provided consent for sampling and data use. Samples were transported on ice, and fresh samples were used for microbiological analysis, while those for metagenomics were immediately frozen at -20 °C.

Culturing, identification of isolates, and BacSomatic

Milk samples from all quarters were sent to the TINE Mastitis Laboratory (Molde, Norway) for microbiological analysis and species identification using standard culturing protocols and MALDI-ToF MS (Microflex LT, Bruker Daltonics)^{2,3}. The aliquot of 10 μ L of milk was plated on cattle blood agar with esculin and incubated at 37 °C for 24 and 48 hours. Species identification was then carried out using MALDI-ToF MS (Microflex LT system, Bruker Daltonics). Lastly, individual bacterial count (IBC) and SCC were measured using the BacSomatic instrument (Foss Electric, Denmark). Samples were classified as positive or negative for mastitis-causing pathogens based on the official mastitis report.

Metagenomic DNA extraction and 16S rRNA gene amplicon sequencing

The metagenomic DNA (mgDNA) was extracted from bacterial pellets obtained from 40 mL of milk, following the protocol of Winther et al. (2022)²¹. A total of 342 samples were processed using the DNeasy PowerFood Microbial Kit (Qiagen, Germany). Pellets were bead-beaten three times (30 s each with 5 min cooling intervals) in a FastPrep-24 5 G (MP Biomedicals) using the *Lactococcus* program, and DNA extraction followed the manufacturer's protocol. DNA was eluted in 50 μ L and stored at -20 °C.

For library preparation, the V3–V4 regions of the 16S rRNA gene were amplified using primers Uni340F and Bac806R, with PCR conditions as described by Porcellato et al. (2020)⁶⁴. Negative controls (reagents and nuclease-free water) were included to monitor contamination. Sub-libraries were cleaned and normalized using the SequalPrep Normalization Plate Kit (Thermo Fisher Scientific), pooled, and quantified with Qubit 2 (dsDNA HS kit). Sequencing was performed across four batches on an Illumina NovaSeq 6000 platform (2 × 250 bp, V3 kit) at Novogene (Cambridge, UK).

Metagenomic DNA extraction and shotgun sequencing

For metagenomic shotgun sequencing, 73 milk samples were selected from 16 of the 24 cows included in the study, following a multi-step selection strategy. First, samples were chosen based on the presence of mastitis-causing pathogens reported in official mastitis records. Second, Bovine

Cytokine/Chemokine bead-based multiplex cytokine panel assay results (Merck KGaA, Darmstadt, Germany) (data not shown) were used to identify samples with immunological profiles of interest. Finally, additional samples were included to ensure coverage across lactation stages and to capture intra-cow variation by sampling adjacent healthy quarters. In total, 36 samples were collected before the dry period in 2022, 7 during early lactation in 2023, 15 during mid-lactation in 2023, and 15 during late lactation in 2023.

The bacterial pellet was prepared and processed as described by Duarte & Porcellato (2024)²³ using the manufacturer's protocol (Molzym GmbH & Co. KG, Bremen, Germany). MDA amplification was performed using the REPLI-g Single Cell kit (Qiagen) following the manufacturer's instructions, with a 16-hour isothermal reaction on a SimpliAmp thermal cycler (Applied Biosystems). DNA concentration was measured using the Qubit HS dsDNA assay (Thermo Fisher Scientific) and stored at -20°C . For Illumina library preparation, mgDNA was measured again with the Qubit HS kit. Samples below 0.5 ng/ μL were used undiluted (5 μL), while more concentrated samples were diluted to 0.2 ng/ μL . Library construction followed the Illumina Nextera XT protocol, and sequencing was performed on the Illumina NovaSeq 6000 platform (2 $\times$ 150 bp) at Novogene (Cambridge, UK).

Bioinformatics processing data

For 16S rRNA gene amplicon sequencing, demultiplexed paired-end reads were processed in QIIME2 (v2021.4) using the Casava 1.8 pipeline⁶⁵. DADA2 was employed to improve taxonomic resolution through error correction and accurate identification of sequence variants. This workflow included filtering, trimming, denoising, dereplication, merging of paired reads, and chimera removal⁶⁶. Amplicon sequence variants (ASVs) were then used to construct a phylogenetic tree via the align-to-tree-mafft-fasttree pipeline in the q2-phylogeny plugin⁶⁷. Taxonomy assignment for the 16S data was performed using a Naïve Bayes pre-trained Silva-138-99-nb-classifier⁶⁸. For downstream metataxonomic analysis, QIIME2 artifacts were imported into R (v3.6.2) using the qiime2R package (v0.99.20). Contaminant ASVs were identified and removed separately for each sequencing batch using the Decontam package (v1.12), applying the frequency method with a 0.5 threshold⁶⁹. After decontamination and before diversity analysis, ASVs classified as *g-Mitochondria*, *o-Chloroplast*, and *d-Archaea*, or unassigned at the phylum level were excluded. To refine taxonomic resolution and enable species-level identification where possible, ASVs were manually verified using MegaBLAST against the 16S ribosomal RNA database (Bacteria and Archaea).

For shotgun sequencing, short reads from the metagenomic datasets were processed using the metaWRAP v1.3⁷⁰ pipeline, following the bioinformatics workflow for genome-centric metagenomics of the bovine hindmilk microbiome as previously outlined by Duarte & Porcellato (2024)²³. Taxonomic and functional profiles were obtained using short reads with MetaPhlAn 3.0⁷¹ and HUMAnN 3.0⁷¹, respectively. All bioinformatics analyses were performed with the Orion Cluster at NMBU.

MAGs were classified into high, medium, and low quality according to the minimum information about metagenome-assembled genome guidelines (MIMAG)⁷². The Genome Taxonomy Database (GTDB) and associated taxonomic classification toolkit (GTDB-Tk) (v2.2.5, reference data version r207_v2)⁷³, was used for taxonomic classification. Identifiers were assigned to the MAGs based on taxonomic level. Dereplicated high-quality MAGs were used to generate a phylogenetic tree using FastTree (v2.1.11)⁷⁴, and their relationships were drawn with the Interactive Tree Of Life (iTOL)⁷⁵. CoverM (v0.6.1)⁷⁶ was used to retrieve MAGs's relative abundance and read counts. Draft genomes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP, release 2024-07-18.build7555)⁷⁷ and protein sequences were forwarded for functional annotation with EggNOG-mapper^{78,79}. MAGs were also processed through TORMES⁸⁰ version 1.3.0 (options --gene_min_id 30 --gene_min_cov 30) to detect antimicrobial resistance (AMR) and virulence genes. For AMR prediction, three databases were adopted by TORMES (ResFinder⁸¹, CARD⁸², and

ARG-ANNOT⁸³, whereas virulence genes were detected through the Virulence Factor Database (VFDB)⁸⁴. Multi-locus sequence typing (MLST) was performed with the mlst software (Seemann T, mlst Github <https://github.com/tseemann/mls>) and the PubMLST database⁸⁵. To investigate molecules potentially associated with microbial and host interactions in the mammary gland, BAGEL4⁸⁶ was employed to identify genes predicted to encode bacteriocins in the MAGs, while antiSMASH⁸⁷, with the strictness parameter set to "relaxed," was used to predict bioactive compounds as part of their secondary or specialized metabolism.

Statistical analysis

Unless otherwise noted, all statistical analyses were performed in R (v4.4.1) using RStudio (v1.2.5033). Two generalized additive models were fitted to assess the effects of the lactation period, SCC, parity, and individual animal variation on Shannon and Simpson diversity indices. The model formula included random effect smoothers for period, SCC, and animal (s(variable, bs = "re")), and a penalized spline smoother with six basis functions (k = 6) for parity to capture potential nonlinear effects. Model estimation was performed using restricted maximum likelihood (REML). Data handling and model fitting were conducted using the R packages phyloseq (v1.34.0)⁸⁸, mgcv (v1.9-3)⁸⁹, and dplyr (v1.1.4)⁹⁰.

Metataxonomic analyses and visualizations were conducted using R packages including MicrobiomeR (v1.31.2) (<https://github.com/microbiome/microbiome>), dplyr⁹⁰, ggplot2 (v 3.5.2) (<https://ggplot2.tidyverse.org>), phyloseq⁸⁸, tidyr (v1.3.1) (<https://github.com/tidyverse/tidyr>), vegan (v2.7-1) (<https://github.com/vegandevs/vegan>), and pairwise Adonis (v0.4) (<https://github.com/pmartinezarbizu/pairwiseAdonis>). Parametric data were analyzed with one-way ANOVA and Tukey's post hoc test; non-parametric data with Kruskal-Wallis and Dunn's test. Two-group comparisons used the Wilcoxon test. Alpha diversity differences by SCC and lactation period were assessed using the microbiome package (v2.1.24) with Wilcoxon's test. Beta diversity (weighted and unweighted UniFrac) was evaluated using PERMANOVA (999 permutations). NMDS plots were generated with microViz (v0.12.7)⁹¹. Results with $P < 0.05$ were considered significant.

Taxa associations with metadata (SCC, period, sampling day) were assessed using MaAsLin2 (v1.7.3)⁹² with a negative binomial model, Benjamini-Hochberg correction ($q < 0.25$), and animal ID as a random effect, based on absolute abundances at phylum and genus levels. The same approach analyzed microbial features (MetaCyc pathways) against metadata (e.g., SCC and lactation period) and random effects (animal), applying the same correction without normalization or transformation.

The resilience of resident microbiota following dysbiosis was evaluated based on a microbial dysbiosis index (MDI), calculated as follows: $\text{MDI}_{\text{sample}} = \log(\text{Relative Abundance of non-pathobionts} / \text{Relative Abundance of pathobionts})$. The genera *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Trueperella*, *Klebsiella*, *Serratia*, *Peptostreptococcus*, *Corynebacterium*, and *Escherichia-Shigella* were defined as pathobionts. A quarter was considered dysbiotic if it had a Shannon index below the mean of samples with $\text{SCC} < 100,000$ cells/mL and $\text{MDI} > 0$.

Lastly, samples were rarefied to 4000 sequences before dimensionality reduction. t-SNE was performed using the Rtsne package (v0.17)⁹³ (3D output), followed by hierarchical clustering (hclust) to group samples based on microbiota profiles.

Data availability

The sequencing data supporting this study have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA950968. Associated metadata are available upon reasonable request from the corresponding author.

Received: 31 January 2025; Accepted: 4 November 2025;

Published online: 09 December 2025

References

- Morales-Ubaldo, A. L. et al. Bovine mastitis, a worldwide impact disease: Prevalence, antimicrobial resistance, and viable alternative approaches. *Vet Anim Sci*, **21**, 100306 (2023).
- Diseases of the Mammary Gland. *Vet. Med.* 1904–2001(2017).
- Cobirka, M., Tancin, V. & Slama, P. Epidemiology and Classification of Mastitis. *Animals* **10**, 2212 (2020).
- Milk — Enumeration of somatic cells — Part 1: Microscopic method (Reference method). Preprint at <https://www.iso.org/standard/40446.html> (2008).
- Adkins, P. R. F. & Middleton, J. R. Methods for Diagnosing Mastitis. *Vet. Clin. North Am.: Food Anim. Pract.* **34**, 479–491 (2018).
- Astrup, L. B., Pedersen, K. & Farre, M. Microbiological Diagnoses on Clinical Mastitis—Comparison between Diagnoses Made in Veterinary Clinics versus in Laboratory Applying MALDI-TOF MS. *Antibiotics* **2022** **11**, 271 (2022).
- Taponen, S., Salmikivi, L., Simojoki, H., Koskinen, M. T. & Pyörälä, S. Real-time polymerase chain reaction-based identification of bacteria in milk samples from bovine clinical mastitis with no growth in conventional culturing. *J. Dairy Sci.* **92**, 2610–2617 (2009).
- Bradley, A. J., Leach, K. A., Breen, J. E., Green, L. E. & Green, M. J. Survey of the incidence and aetiology of mastitis on dairy farms in England and Wales. *Vet. Rec.* **160**, 253–258 (2007).
- Langhorne, C. et al. Bacterial culture and susceptibility test results for clinical mastitis samples from Australia's subtropical dairy region. *J. Dairy Sci.* **107**, 1151–1163 (2024).
- Yadav, D. et al. Next-Generation sequencing transforming clinical practice and precision medicine. *Clin. Chim. Acta* **551**, 117568 (2023).
- Oikonomou, G. et al. Microbiota of Cow's Milk; Distinguishing Healthy, Sub-Clinically and Clinically Diseased Quarters. *PLoS One* **9**, e85904 (2014).
- Sakwinska, O. & Bosco, N. Host microbe interactions in the lactating mammary gland. *Front Microbiol.* **10**, 448318 (2019).
- Cheema, S. K., Li, R. & Cameron, S. J. S. Culturomics as a tool to better understand the human milk microbiota and host-microbiota interactions. *Microbiota and Host* **1**, 1–11 (2023).
- Oikonomou, G., Machado, V. S., Santisteban, C., Schukken, Y. H. & Bicalho, R. C. Microbial Diversity of Bovine Mastitic Milk as Described by Pyrosequencing of Metagenomic 16s rDNA. *PLoS One* **7**, e47671 (2012).
- Srithanasuwan, A., Pangprasit, N., Mektrirat, R., Suriyasathaporn, W. & Chuammitri, P. Divergent Immune Responses to Minor Bovine Mastitis-Causing Pathogens. *Vet. Sci.* **11**, 262 (2024).
- Hoque, M. N. et al. Metagenomic deep sequencing reveals association of microbiome signature with functional biases in bovine mastitis. *Sci. Rep.* **9**, 13536 (2019).
- Hoque, M. N. et al. Microbiome dynamics and genomic determinants of bovine mastitis. *Genomics* **112**, 5188–5203 (2020).
- Oikonomou, G. et al. Milk Microbiota: What Are We Exactly Talking About? *Front Microbiol.* **11**, 1–15 (2020).
- Urrutia-Angulo, L. et al. Unravelling the complexity of bovine milk microbiome: insights into mastitis through enterotyping using full-length 16S-metabarcoding. *Anim. Microbiome* **6**, 1–14 (2024).
- Bhar, S. & Bose, T. Evidence from metagenomic study indicate that subclinical mastitis may have a different pathological origin than clinical mastitis. *bioRxiv*. <https://doi.org/10.1101/2024.05.24.595548> (2024).
- Winther, A. R. et al. Longitudinal dynamics of the bovine udder microbiota. *Anim. Microbiome* **4**, 26 (2022).
- Winther, A. R., da Silva Duarte, V. & Porcellato, D. Metataxonomic analysis and host proteome response in dairy cows with high and low somatic cell count: a quarter level investigation. *Vet. Res* **54**, 32 (2023).
- Duarte, V., Da, S. & Porcellato, D. Host DNA depletion methods and genome-centric metagenomics of bovine hindmilk microbiome. *mSphere* **9**, e0047023 (2024).
- Farkaš, V. et al. Biomarkers for subclinical bovine mastitis: a high throughput TMT-based proteomic investigation. *Vet. Res Commun.* **48**, 2069–2082 (2024).
- Satheesan, L. et al. Comparative Profiling of Milk Somatic Cells Proteomes Revealed Key Players in Mammary Immune Mechanisms During Mastitis in Tropical Sahiwal (*Bos indicus*) Cows. *Proteom. Clin. Appl* **18**, e202400054 (2024).
- Jung, D., Park, S., Kurban, D., Dufour, S. & Ronholm, J. The occurrence of *Aerococcus urinaeequi* and non-aureus staphylococci in raw milk negatively correlates with *Escherichia coli* clinical mastitis. *mSystems*. <https://doi.org/10.1128/MSYSTEMS.00362-24> (2024)
- Caballero-Flores, G., Pickard, J. M. & Núñez, G. Microbiota-mediated colonization resistance: mechanisms and regulation. *Nat. Rev. Microbiol.* **21**, 347–360 (2022).
- Krishnamoorthy, P., Goudar, A. L., Suresh, K. P. & Roy, P. Global and countrywide prevalence of subclinical and clinical mastitis in dairy cattle and buffaloes by systematic review and meta-analysis. *Res Vet. Sci.* **136**, 561–586 (2021).
- Ruegg, P. L. A 100-Year Review: Mastitis detection, management, and prevention. *J. Dairy Sci.* **100**, 10381–10397 (2017).
- Smistad, M., Bakka, H. C., Sølvørød, L., Jørgensen, H. J. & Wolff, C. Prevalence of udder pathogens in milk samples from Norwegian dairy cows recorded in a national database in 2019 and 2020. *Acta Vet. Scand.* **65**, 19 (2023).
- Paschoalini, B. R. et al. The Emergence of Antimicrobial Resistance and Virulence Characteristics in *Enterococcus* Species Isolated from Bovine Milk. *Antibiotics* **12**, 1243 (2023).
- Alessandri, G. et al. Metataxonomic analysis of milk microbiota in the bovine subclinical mastitis. *FEMS Microbiol Ecol.* **99**, 1–10 (2023).
- Lücken, A., Woudstra, S., Wente, N., Zhang, Y. & Krömker, V. Intramammary infections with *Corynebacterium* spp. in bovine lactating udder quarters. *PLoS One* **17**, e0270867 (2022).
- Silva, V. M. et al. Milk lymphocyte profile and macrophage functions: new insights into the immunity of the mammary gland in quarters infected with *Corynebacterium bovis*. *BMC Vet. Res* **17**, 1–8 (2021).
- Lam, T. J. G. M. et al. Effect of natural infection with minor pathogens on susceptibility to natural infection with major pathogens in the bovine mammary gland. *Am. J. Vet. Res* **58**, 17–22 (1997).
- Seshadri, R. et al. Expanding the genomic encyclopedia of Actinobacteria with 824 isolate reference genomes. *Cell Genomics* **2**, 100213 (2022).
- Park, S. et al. A longitudinal census of the bacterial community in raw milk correlated with *Staphylococcus aureus* clinical mastitis infections in dairy cattle. *Anim. Microbiome* **4**, 1–13 (2022).
- Derakhshani, H., Plaizier, J. C., De Buck, J., Barkema, H. W. & Khafipour, E. Composition and co-occurrence patterns of the microbiota of different niches of the bovine mammary gland: potential associations with mastitis susceptibility, udder inflammation, and teat-end hyperkeratosis. *Anim. Microbiome* **2**, 11 (2020).
- Metzger, S. A. et al. A cohort study of the milk microbiota of healthy and inflamed bovine mammary glands from dryoff through 150 days in milk. *Front Vet. Sci.* **5**, 412587 (2018).
- da Silva Duarte, V. et al. Milk microbial composition of Brazilian dairy cows entering the dry period and genomic comparison between *Staphylococcus aureus* strains susceptible to the bacteriophage vB_SauM-UFV_DC4. *Sci. Rep.* **10**, 5520 (2020).
- Stevenson, D. R. et al. *Corynebacterium kroppenstedtii* breast abscesses in context, a retrospective cohort study. *J. Med Microbiol.* **71**, 001616 (2022).
- Milke, L., Kallscheuer, N., Kappelmann, J. & Marienhagen, J. Tailoring *Corynebacterium glutamicum* towards increased malonyl-CoA availability for efficient synthesis of the plant pentaketide noreugenin. *Micro. Cell Fact.* **18**, 1–12 (2019).
- Hill, C. Virulence or Niche Factors: What's in a Name?. *J. Bacteriol.* **194**, 5725 (2012).

44. Herzberg, C., van Meegen, E. N. & van Hasselt, J. G. C. Interplay of virulence factors shapes ecology and treatment outcomes in polymicrobial infections. *Math. Biosci.* **377**, 109293 (2024).
45. Schmitt, M. P., Predich, M., Doukhan, L., Smith, I. & Holmes, R. K. Characterization of an iron-dependent regulatory protein (IdeR) of *Mycobacterium tuberculosis* as a functional homolog of the diphtheria toxin repressor (DtxR) from *Corynebacterium diphtheriae*. *Infect. Immun.* **63**, 4284–4289 (1995).
46. Marcos-Torres, F. J., Juniar, L. & Griese, J. J. The molecular mechanisms of the bacterial iron sensor IdeR. *Biochem. Soc. Trans.* **51**, 1319–1329 (2023).
47. Wee, B. A., Muloi, D. M. & van Bunnik, B. A. D. Quantifying the transmission of antimicrobial resistance at the human and livestock interface with genomics. *Clin. Microbiol. Infect.* **26**, 1612 (2020).
48. Ortiz-Pérez, A. et al. High frequency of macrolide resistance mechanisms in clinical isolates of *Corynebacterium* species. *Microb. Drug Resistance* **16**, 273–277 (2010).
49. Olender, A. Antibiotic resistance and detection of the most common mechanism of resistance (MLSB) of opportunistic *Corynebacterium*. *Chemotherapy* **59**, 294–306 (2014).
50. Barberis, C. M. et al. Comparison between disk diffusion and agar dilution methods to determine in vitro susceptibility of *Corynebacterium* spp. clinical isolates and update of their susceptibility. *J. Glob. Antimicrob. Resist.* **14**, 246–252 (2018).
51. Soriano, F., Zapardiel, J. & Nieto, E. Antimicrobial susceptibilities of *Corynebacterium* species and other non-spore-forming gram-positive bacilli to 18 antimicrobial agents. *Antimicrob. Agents Chemother.* **39**, 208–214 (1995).
52. Feßler, A. T. & Schwarz, S. Antimicrobial Resistance in *Corynebacterium* spp., *Arcanobacterium* spp., and *Trueperella pyogenes*. *Microbiol. Spectr.* **5**, 1–15 (2017).
53. Zhang, D. et al. A systematically biosynthetic investigation of lactic acid bacteria reveals diverse antagonistic bacteriocins that potentially shape the human microbiome. *Microbiome* **11**, 1–20 (2023).
54. Daly, K. M. et al. Production of the Bsa lantibiotic by community-acquired *Staphylococcus aureus* strains. *J. Bacteriol.* **192**, 1131–1142 (2010).
55. Sadaoka, N. et al. Opposing genetic polymorphisms of two ABC transporters contribute to the variation of nukacin resistance in *Streptococcus mutans*. *Appl Environ. Microbiol.* **90**, e0208423 (2024).
56. Ceotto, H. et al. Nukacin 3299, a lantibiotic produced by *Staphylococcus simulans* 3299 identical to nukacin ISK-1. *Vet. Microbiol.* **146**, 124–131 (2010).
57. Thoendel, M., Kavanaugh, J. S., Flack, C. E. & Horswill, A. R. Peptide signaling in the *Staphylococci*. *Chem. Rev.* **111**, 117–151 (2011).
58. Nilsen, T., Nes, I. F. & Holo, H. Enterolysin A, a cell wall-degrading bacteriocin from *Enterococcus faecalis* LMG 2333. *Appl Environ. Microbiol.* **69**, 2975–2984 (2003).
59. Zhang, T. et al. Molecular cloning and antimicrobial activity of enterolysin A and helveticin J of bacteriolysins from metagenome of Chinese traditional fermented foods. *Food Control* **31**, 499–507 (2013).
60. Märlä, M. T. et al. Genome-wide analysis of fitness determinants of *Staphylococcus aureus* during growth in milk. *PLoS Pathog.* **21**, e1013080 (2025).
61. Jiang, X. et al. Distribution of ϵ -Poly-L-Lysine Synthetases in Coryneform Bacteria Isolated from Cheese and Human Skin. *Appl Environ. Microbiol.* **87**, 1–8 (2021).
62. Hyvönen, P. et al. Concentrations of bovine lactoferrin and citrate in milk during experimental endotoxin mastitis in early-versus late-lactating dairy cows. *J. Dairy Res.* **77**, 474–480 (2010).
63. Patzer, S. I. & Braun, V. Gene cluster involved in the biosynthesis of griseobactin, a catechol-peptide siderophore of *Streptomyces* sp. ATCC 700974. *J. Bacteriol.* **192**, 426–435 (2010).
64. Porcellato, D., Meisal, R., Bombelli, A. & Narvhus, J. A. A core microbiota dominates a rich microbial diversity in the bovine udder and may indicate presence of dysbiosis. *Sci. Rep.* **10**, 21608 (2020).
65. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
66. Callahan, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
67. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkf436> (2002).
68. 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).
69. Davis, N. M., Proctor, D. I. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 1–14 (2018).
70. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
71. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with biobakery 3. *Elife* **10**, e65088 (2021).
72. Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
73. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022).
74. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One*. <https://doi.org/10.1371/journal.pone.0009490> (2010).
75. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
76. Aroney, S. T. N. et al. CoverM: Read coverage calculator for metagenomics. *Zenodo* <https://doi.org/10.5281/ZENODO.10531254> (2024).
77. Tatusova, T. et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624 (2016).
78. Huerta-Cepas, J. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
79. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
80. Quijada, N. M., Rodríguez-Lázaro, D., Eiros, J. M. & Hernández, M. TORMES: an automated pipeline for whole bacterial genome analysis. *Bioinformatics* **35**, 4207–4212 (2019).
81. Florensa, A. F., Kaas, R. S., Clausen, P. T. L. C., Aytan-Aktug, D. & Aarestrup, F. M. ResFinder – an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Micro. Genom.* **8**, 000748 (2022).
82. Alcock, B. P. et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res.* **51**, D690–D699 (2023).
83. Gupta, S. K. et al. ARG-annot, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* **58**, 212–220 (2014).
84. Liu, B., Zheng, D., Zhou, S., Chen, L. & Yang, J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* **50**, D912–D917 (2022).

85. Jolley, K. A. & Maiden, M. C. J. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinforma.* **11**, 1–11 (2010).
86. Van Heel, A. J. et al. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* **46**, W278 (2018).
87. Blin, K. et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **51**, W46–W50 (2023).
88. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* <https://doi.org/10.1371/journal.pone.0061217> (2013).
89. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Series B Stat. Methodol.* **73**, 3–36 (2011).
90. Wickham, H., François, R., Henry, L. & Müller, K. dplyr: A Grammar of Data Manipulation. R package version. *Media Preprint* at (2019).
91. Barnett, D., Arts, I. & Penders, J. microViz: an R package for microbiome data visualization and statistics. *J. Open Source Softw.* **6**, 1–4 (2021).
92. Mallick, H. et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol.* **17**, e1009442 (2021).
93. Van Der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

Acknowledgements

This work received economic support from the Norwegian Research Council (grant number 314733), the Faculty of Chemistry, Biotechnology, and Food Science at the Norwegian University of Life Sciences.

Author contributions

V.S.D. contributed to conceptualization, data curation, formal analysis, investigation, methodology, software development, validation, visualization, and manuscript writing (original draft, review and editing). D.P. provided conceptualization, formal analysis, investigation, project administration, supervision, and resources, as well as securing funding and contributing to

manuscript review and editing. F.V.F. and A.K. were involved in methodology development and manuscript review, and editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41522-025-00860-1>.

Correspondence and requests for materials should be addressed to Vinicius da Silva Duarte.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025