Longitudinal study of the udder microbiome of Norwegian Red dairy cows using metataxonomic and shotgun metagenomic approaches: Insights into pathogen-driven microbial adaptation and succession

Vinicius da Silva Duarte, Fiona Valerie Franklin, Alicja Krysmann & Davide Porcellato

Faculty of Chemistry. Biotechnology and Food Science. Norwegian University of Life Sciences. NMBU. P.O. Box 5003. N-1432 Ås. Norway

Corresponding author: vinicius.da.silva.duarte@nmbu.no

Abstract

Bovine mastitis remains the most significant disease affecting dairy herds globally, driven by its multi-etiological nature and the complex dynamics of udder immunity and infection. While research addressing the microbial and immunological aspects of the bovine udder is limited, optimizing the udder microbiome has emerged as a promising strategy for preventing mastitis. This longitudinal study aimed to investigate the udder microbiome throughout lactation using both metataxonomic and shotgun metagenomic approaches, including analysis at the metagenome-assembled genome (MAG) level. The use of such methodologies provides a deeper understanding of the microbial composition and dynamics within the udder, bridging critical gaps in knowledge and revealing potential interactions between the microbiota and host. Milk samples were collected from 342 individual quarters of 24 Norwegian Red dairy cows. Significant variations in somatic cell count and microbiota composition were observed across lactation stages. Quarters with low somatic cell count were notably higher during early lactation (80%) and mid-lactation (78.9%) compared to dry-off (53.1%) and late lactation (53%), with high somatic cell countobserved in 20–47% of samples. Diversity indices based on Shannon and Simpson metrics indicated significant effects of lactation stage, somatic cell count, and individual animal variability on microbial diversity. PERMANOVA analyses confirmed that individual animal variability (15.73%) and lactation period (5.52%) were the strongest factors influencing microbiota structure, with dysbiosis linked to mastitis-causing pathogens contributing 7.17% of the variance. Key pathogens identified included *Enterococcus faecalis*, *Staphylococcus aureus*, *Streptococcus uberis*, and *Staphylococcus chromogenes*, with persistent infections observed for *S. aureus* and *S. chromogenes*, but not *S. uberis*. Samples with low somatic cell count were enriched in beneficial genera such as *Corynebacterium*, *Bradyrhizobium*, and *Lactococcus*, while *Staphylococcus* predominated in milk samples with high somatic cell count. Dimensionality reduction integration with culturomics enhanced milk microbiota classification, providing novel insights into udder microbiota dynamics and their role in bovine mastitis. Metagenomic shotgun sequencing revealed pathogen-specific metabolic signatures in the bovine mammary gland, identifying 289 MetaCyc pathways. Genome-centric analysis reconstructed 142 metagenome-assembled genomes, including 26 from co-assembly and 116 from individual assembly. Multi-locus sequence typing, virulence factors, and antimicrobial resistance gene profiling provided insights into pathogen adaptation and

persistence in the bovine mammary gland. Notably, 27 bacteriocin gene clusters and 322 biosynthetic gene clusters were predicted using genome mining tools. Our findings contribute to the establishment of pathogen-based therapies and enhance our understanding of the milk microbiome, which remains far from fully characterized. Such insights are vital for developing effective strategies to combat mastitis and improve dairy cattle health.

Keywords: Bovine mastitis, metataxonomic analysis, shotgun sequencing, udder health, milk microbiome

1. Introduction

Bovine mastitis remains the leading disease affecting dairy herds worldwide [1]. 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 infections [2], 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 culling [3]. 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 SCM [4].

Microbiological culture of milk remains the gold standard for pathogen identification, despite some limitations [5]. Few studies have investigated the microbial composition of milk samples classified as "mixed growth," which are often considered contaminated during sampling [6], culture-negative samples from clinical mastitis [7], 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 diagnostics [9].

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 fields [10], including udder health [11]. 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 samples [14].

The dynamics of udder immunity and infection are complex, influenced by factors such as the cow's health status, environmental conditions, and the resident microbiota [15]. Research has identified distinct changes in microbiome composition

between healthy and mastitic milk, with key phyla such as *Proteobacteria*, *Bacteroidetes*, *Firmicutes*, and *Actinobacteria* playing significant roles[16]. Metagenomic analysis has further revealed previously unreported opportunistic strains in clinical mastitis samples, along with functional pathways related to bacterial colonization and antibiotic resistance [17]. 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 insights [18]. 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 profiles [20]. 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 and [21], 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 metataxonomic 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 gap in knowledge 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.

2. Material and methods

124 **2.1. Experimental design and sample collection**

125 Twenty-four Norwegian Red cows were chosen from the "Centre for Livestock
126 Production" at the Norwegian University of Life Sciences. The farm adheres to the
127 regulations set by the Norwegian Food Safety Authority for food production and animal
128 welfare. Sample collection and the use of related information were approved by the
129 farm owners, and no invasive procedures were involved in the study. The dairy cows
130 were kept in freestalls equipped with rubber mats and raw wood chips for bedding.
131 Their diet included silage available at all times and pelleted feed, adjusted according
132 to each cow's milk production.

133 The experimental period took place between the autumns of 2022 to 2023. Milk
134 samples were collected at four different time points, covering two lactation cycles: i)
135 before dry-off in 2022 (DO-2022), ii) early lactation (EL-2023), iii) mid-lactation (ML-
136 2023), and iv) late lactation (LL-2023). On each sampling day, hindmilk was aseptically
137 collected from each quarter at the end of the regular milking routine, as described by
138 Porcellato et al. (2020) [28]. In short, after the milking apparatus was removed, the teats
139 were cleaned with iodine and then 70% ethanol, and approximately 350 mL of milk
140 was manually collected. The procedure followed the guidelines of the National Mastitis
141 Council (NMC; www.nmconline.org). Once collected, the milk samples were
142 transported on ice to the laboratory and then divided into polystyrene tubes. Fresh milk
143 samples were used for microbiological analyses (item 2.2), while milk samples used
144 for metagenomic analyses were immediately frozen at -20°C until further processed
145 (items 2.3 and 2.4).

146

147 **2.2. BacSomatic, culturing and identification of isolates**

148 Milk samples from all quarters were sent to the TINE Mastitis Laboratory in
149 Molde, Norway, for microbiological analysis and species identification (gold standard
150 diagnostic). Bacteriological culturing followed standard protocols 2.3. In summary, 10
151 µL of milk was plated on cattle blood agar with esculin and incubated at 37°C. The
152 plates were examined after 24 and 48 hours. Species identification was then carried
153 out using MALDI-Tof MS (Microflex LT system. Bruker Daltonics).

154 Concomitantly, the same fresh milk samples were also plated in four different
155 media at SciFood (KBM. NMBU) and incubated at 37°C for 24, 48 and 72 hours as
156 follows: Tryptic Soy Agar (TSA, Sigma-Aldrich) plus 0.01% Tween 20, Blood agar
157 (Thermo Scientific), brilliance CRE agar (detection of *Enterobacteriaceae* Producing

158 KPC- and Metallo-Carbapenemase, ThermoFisher Scientific) and brilliance ESBL
159 agar (detection of extended-spectrum beta-lactamase-producing *Enterobacteriaceae*,
160 Thermo Scientific). Lastly, the samples were analyzed for somatic cell count (SCC)
161 and individual bacterial count (IBC) using the BacSomatic instrument (Foss Electric,
162 Hillerød, Denmark). In this study, samples were classified as either negative or positive
163 for mastitis-causing pathogens based on the official mastitis report.
164

165 **2.3. Metagenomic DNA extraction and 16S rRNA gene amplicon sequencing**

166 **2.3.1. Bacterial pellet preparation and metagenomic DNA extraction with**
167 **DNeasy PowerFood Microbial Kit**

168 For metagenomic DNA (mgDNA) extraction, a bacterial pellet was isolated from
169 40 mL of milk as outlined by Winther et al. (2022) [21]. A total of 345 milk samples were
170 processed. Initially, 40 mL of milk was thawed overnight at 4°C and centrifuged at
171 8,000 × g for 10 minutes at 4°C. The whey fraction was discarded, and the bacterial
172 pellet along with the fat layer was washed on an orbital shaker at 250 rpm for 15
173 minutes using a solution of 2% citrate water and 0.1% Tween 20. The mixture was
174 then centrifuged again at 8,000 × g for 10 minutes at 4°C, after which the supernatant
175 was removed, and the fat layer was gently collected and after discarded using a sterile
176 cotton swab soaked in 70% ethanol. The bacterial pellet was then harvested and
177 washed with 2% citrate water (16,200 × g for 3 minutes).

178 The metagenomic DNA (mgDNA) was extracted using the DNeasy PowerFood
179 Microbial Kit (Qiagen. Düsseldorf. Germany). The bacterial pellet was transferred
180 directly into a PowerBead tube and placed in a FastPrep-24 5G Instrument (MP
181 Biomedicals European HQ) for three rounds of bead-beating at (*Lactococcus* program,
182 30 seconds with a 5-minute cooling period between each round). The samples were
183 then processed following the manufacturer's instructions. Finally, the mgDNA was
184 eluted in 50 µL of elution buffer and stored at −20°C until further processed.
185

186 **2.3.2. Library preparation and 16S rRNA gene amplicon sequencing**

187 Library preparation for amplicon sequencing followed the method previously
188 described in section 2.3.1. Specifically, the V3 and V4 regions of the 16S rRNA genes
189 were amplified using the primers Uni340F (CCTACGGGRBGCASCAG) and Bac806R
190 (GGACTACYVGGGTATCTAAT). The PCR reagents and amplification conditions
191 matched those outlined by Porcellato et al. (2020)[28]. Negative controls consisting of

192 only reagents and nuclease free water were included to check for contamination during
193 DNA extraction and library preparation. Sub-libraries were cleaned and normalized
194 using the SequalPrep Normalization Plate (96) Kit (Thermo Fisher Scientific. USA)
195 and then pooled. The final library concentration was measured with Qubit 2 using the
196 dsDNA HS kit (Thermo Fisher Scientific. USA), and sequencing was performed on an
197 Illumina NovaSeq 6000 platform (Illumina) using the 2 × 250 bp V3 kit (Illumina) at
198 Novogene (Cambridge. UK). A total of four sequencing batches were completed.
199

200 **2.4. Metagenomic DNA extraction and shotgun metagenomic sequencing**

201 **2.4.1. Metagenomic DNA extraction with MolYsis complete5 kit**

202 For metagenomic shotgun sequencing, 73 samples were selected based on
203 their SCC levels and their representation of all the major mastitis-causing pathogens
204 identified among the samples included in the study. The bacterial pellet was obtained
205 as outlined in section 2.3.1 and processed using the manufacturer's protocol (Molzym
206 GmBH & Co. KG. Bremen. Germany). In brief, cells were resuspended in 1 mL of
207 buffer SU, and host cells were lysed by adding 250 µL of chaotropic buffer (buffer CM).
208 The released nucleic acids were degraded using an enzyme (MolDNase B). Microbial
209 cells were then pelleted and lysed with specific reagents and proteinase K. The
210 mgDNA was isolated using spin columns, and 50 µL of DNA was eluted and stored at
211 −20°C for further processing.
212

213 **2.4.2. Multiple displacement amplification by Phi29 DNA polymerase**

214 Following the manufacturer's instructions, whole metagenome amplification
215 was conducted on 73 samples using multiple displacement amplification (MDA) with
216 the REPLI-g Single Cell kit (Qiagen: 150345). Briefly, 5 µL of metagenomic DNA
217 (mgDNA) was added to a microcentrifuge tube along with 5 µL of buffer D1 and mixed
218 by vortexing. After a 3-minute incubation at room temperature (15–25°C), 10 µL of
219 buffer N1 (stop solution) was added. To the 20 µL of denatured mgDNA, 30 µL of
220 master mix (comprising 29 µL REPLI-g mini reaction buffer and 1 µL REPLI-g mini
221 DNA polymerase) was added. The final mixture (50 µL) was gently mixed, incubated
222 at 30°C for 16 hours, and then heat-inactivated at 65°C for 3 minutes using the
223 SimpliAmp (Applied Biosystems). The mgDNA concentration was measured using the
224 Qubit HS dsDNA assay Fluorometer (Thermo Fisher Scientific. USA) and stored at
225 −20°C for further analysis.

226

227 **2.4.3. Library preparation and sequencing**

228 For Illumina sequencing library preparation, the mgDNA concentration of the
229 MDA-treated samples was measured using the Qubit HS dsDNA kit. If the extracted
230 DNA concentration was below 0.5 ng/µL, 5 µL of the undiluted sample was used.
231 Otherwise, the DNA was diluted to 0.2 ng/µL. DNA preparation for sequencing followed
232 the guidelines of the Illumina Nextera XT Library Preparation Kit. The DNA
233 concentration was measured using the Qubit HS dsDNA assay, and the concentration
234 was calculated before being diluted and pooled at equimolar ratios. The DNA library
235 was sequenced on the Illumina NovaSeq 6000 system (2 × 150 bp read length) at
236 Novogene (Cambridge. UK).

237

238 **2.5. Bioinformatics processing data**

239 **2.5.1. 16S rRNA gene amplicon sequencing**

240 The demultiplexed raw paired-end reads obtained from sequencing were
241 uploaded and processed in QIIME2 (version 2021.4) using the Casava 1.8 paired-end
242 pipeline [31]. DADA2 was selected for its ability to enhance taxonomic resolution through
243 precise identification and error correction of sample sequences that differ by as little
244 as a single nucleotide. This process involved several sequential steps, including
245 filtering, trimming, denoising, dereplicating, merging paired reads, and removing
246 chimeric sequences [32]. Subsequently, amplicon sequence variants were used to
247 generate a phylogenetic tree through the align-to-tree-mafft-fasttree pipeline in the q2-
248 phylogeny plugin [33]. Taxonomy assignment for the 16S data was performed using a
249 Naïve Bayes pre-trained Silva-138–99-nb-classifier [34].

250 For downstream metataxonomic analysis, QIIME artifacts were imported into R
251 (R Core Team 3.6.2. 2019) using the qiime2R package v.099.20
252 (https://github.com/jbisanz/qiime2R; accessed on 5 November 2024). *In silico*
253 contaminant identification and removal of amplicon sequence variants (ASVs) were
254 conducted for each sequencing batch using the R package Decontam (version 1.12)
255 [35], employing the frequency method with a user-defined threshold of 0.5. After
256 decontaminating the samples with the help of negative controls and before calculating
257 any diversity indices, ASVs assigned to *g-Mitochondria*. *o-Chloroplast*, and *d-Archaea*,
258 assigned as Unknown at the phylum level, were excluded from the dataset. To improve
259 the taxonomical assignment and make inferences at the species level when

260 appropriate, ASVs were manually verified using the MegaBLAST search against the
261 16S ribosomal RNA sequences (Bacteria and Archaea) database.
262
263 **2.5.2. Shotgun sequencing**
264 Short reads from the metagenomic datasets were processed using the metaWRAP
265 v1.3[36] pipeline, following the bioinformatics workflow for genome-centric
266 metagenomics of the bovine hindmilk microbiome as previously outlined by Duarte &
267 Porcellato (2024)[23]. In short, the quality of the raw sequence data was verified using
268 the FastQC tool (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The
269 raw reads were trimmed with Trim-galore v0.4.3 [37], and then, the bovine-derived reads
270 were removed with bmtagger v3.101 (default settings). The high-quality reads were
271 assembled using both the individual or co-assembly methods with MegaHit v1.0[38].
272 Subsequently, the assemblies were binned with metaBAT v2.12.1 (-m 1500 and --
273 unbinned parameters) [39], Maxbin v2.2.4 (-markerset 40 option) [40] and CONCOCT
274 v0.4.0 (default settings) [41] using the metaWRAP-Binning module. The resulting three
275 bin sets were consolidated with metaWRAP-Bin-refinement, and the completion and
276 contamination of the resulting bins were evaluated with CheckM v1.0.7 (default
277 settings) [42]. Taxonomic and functional profiles were obtained using short reads with
278 MetaPhlAn 3.0 [43] and HUMAnN 3.0 [43], respectively.
279 All bioinformatics analyses were performed with the Orion Cluster at NMBU. The
280 raw reads were deposited into the Sequence Read Archive database
281 (http://www.ncbi.nlm.nih.gov/sra) under the BioProject PRJNA950968.
282
283 **2.5.3. MAG data analysis**
284 Metagenome-assembled genomes (MAGs) were classified into high, medium,
285 and low quality according to the minimum information about metagenome-assembled
286 genome guidelines (MIMAG) [44]. GTDB-Tk (v2.2.5, reference data version r207_v2) [45]
287 was used for taxonomic classification. Identifiers were assigned to the MAGs based
288 on taxonomic level. Dereplicated high-quality MAGs were used to generate a
289 phylogenetic tree using FastTree (v2.1.11) [46] and their relationships drawn with iTOL
290 [47]. CoverM (v0.6.1) [48] was used to retrieve MAGs's relative abundance and read
291 counts.
292 MAGs were annotated using the NCBI Prokaryotic Genome Annotation Pipeline
293 (PGAP, release 2024-07-18.build7555)[49] and protein sequences forwarded for

294 functional annotation with EggnNOG-mapper [50,51]. MAGs were also processed
295 through TORMES [52] version 1.3.0 (options --gene_min_id 30 --gene_min_cov 30) to
296 detect antimicrobial resistance (AMR) and virulence genes. For AMR prediction, three
297 databases were adopted by TORMES (ResFinder [53], CARD [54] and ARG-ANNOT [55],
298 whereas virulence genes were detected through the VFDB [56]. Multi-locus sequence
299 typing (MLST) was performed with the mlst software (Seemann T, mlst Github
300 https://github.com/tseemann/mls) and the PubMLST database [57]. To investigate
301 molecules potentially associated with microbial and host interactions in the mammary
302 gland, BAGEL4 [58] was employed to identify genes predicted to encode bacteriocins in
303 the MAGs, while antiSMASH [59], with the strictness parameter set to "relaxed," was
304 used to predict bioactive compounds as part of their secondary or specialised
305 metabolism.
306

307 **2.6. Statistical analysis**

308 **2.6.1. Generalized Additive Model (GAM)**

309 Two GA models were developed to investigate the influence of lactation period
310 (Period, DO-2022, EL-2023, ML-2023, LL-2023), somatic cell count classification
311 (SCC, “High” and “Low”), and individual animal variability (animal) on microbial
312 diversity metrics, specifically Shannon and Simpson indices. Both models were fitted,
313 treating each categorical variable as a random effect by including smooth terms (bs =
314 "re"). The REML method was selected for estimation to optimize the accuracy of
315 parameter estimates and random effect variances. Data handling, model fitting, and
316 visualization were conducted using R (version 4.4.1) with the R packages phyloseq [60],
317 mgcv [61], and dplyr [62].
318

319 **2.6.2. Metataxonomic analysis**

320 Statistical analysis and graph construction were performed with RStudio
321 (v.1.2.5033) using one or the combination of the following R packages: MicrobiomeR
322 (https://github.com/microbiome/microbiome), dplyr [62], ggplot2
323 (https://ggplot2.tidyverse.org), phyloseq [60], tidyr (https://github.com/tidyverse/tidyr),
324 vegan (https://github.com/vegandevs/vegan), and pairwise Adonis
325 (https://github.com/pmartinezarbizu/pairwiseAdonis).
326 Groups with parametric distribution were analyzed by one-way ANOVA followed
327 by Tukey’s *post hoc*. The Wilcoxon test was used to compare two unpaired groups.

328 Differences were considered significant at $P < 0.05$. Significant differences in alpha
329 diversity between the groups with high and low SCC, as well as between the different
330 lactation periods, were determined using the alpha function in microbiome R package
331 v.2.1.24, adopting Wilcoxon's test. For beta diversity, weighted and unweighted
332 UniFrac distances were subjected to permutational multivariate analysis of variance
333 (ANOVA) to assess significant differences (pseudo-F test) in bacterial community
334 composition and structure among the groups with a permutation number of 999. Non-
335 metric multidimensional scaling (NMDS) plots were constructed with microViz [63]
336 To investigate the association between microbial taxa and key variables, we
337 employed the MaAsLin 2 [64] package with a negative binomial regression model with
338 Benjamini-Hochberg correction (max_significance = 0.25), incorporating both fixed
339 and random effects. The input data matrix included the absolute abundances of
340 microbial taxa at the phylum and genus level, while the metadata comprised variables
341 of interest for the model, including "SCC", "Period", and "sampling day", with "animal"
342 as a random effect.
343 The resilience of resident microbiota following dysbiosis was evaluated based
344 on a microbial dysbiosis index (MDI), calculated as follows: (MDI_{sample}=log(Relative
345 Abundance of non-pathobionts/Relative Abundance of pathobionts). The genera
346 *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Trueperella*, *Klebsiella*, *Serratia*,
347 *Peptostreptococcus*, *Corynebacterium*, and *Escherichia-Shigella* were defined as
348 pathobionts. A quarter was defined as dysbiotic if it exhibited a Shannon diversity index
349 lower than the mean of quarters with somatic cell counts (SCC) below 100,000
350 cells/mL and an MDI greater than 0, indicating a predominance of dysbiotic
351 microorganisms within the specific sample.
352 To reduce the dimensionality of the metataxonomic data, t-distributed
353 stochastic neighbor embedding (t-SNE) was run using the R function Rtsne from the
354 Rtsne package [65] and selecting the three-dimensionality output option. A hierarchical
355 clustering method using the "hclust" function in R was then used on the output to
356 identify groups of samples with the same microbiota profile.
357
358 **2.6.3. Microbial differential abundance analyses based on microbial functional**
359 **profile**
360 To explore associations between microbial features and metadata variables,
361 microbial differential abundance analyses were conducted using the Maaslin2

362 package [64] with negative binomial regression models, incorporating fixed effects (e.g.,
363 "Pathogen", "SCC", and "Lactation period"), random effects (e.g., "animal"), and
364 Benjamini-Hochberg correction (max_significance = 0.25), without normalization or
365 transformation.
366

367 **3. RESULTS**

368 **3.1. Changes in SCC and pathogen distribution across lactation periods**
369 **reveal significant variability**

370 In this study, we collected 342 individual quarter-level milk samples from 24
371 Norwegian Red dairy cows (8 primiparous and 16 multiparous). These samples were
372 collected during four lactation periods and encompassed two lactation cycles. The pre-
373 drying period in 2022 (DO-2022) was used as the baseline. In 2023, milk samples
374 were collected at three distinct stages of the lactation cycle based on the days in milk
375 (DIM) of each animal. Samples were obtained during early lactation (EL-2023, median
376 DIM: 22 days), mid-lactation (ML-2023, median DIM: 126 days), and late lactation (LL-
377 2023, median DIM: 244 days), corresponding to the respective stages of the lactation
378 cycle.

379 Based on SCC, a total of 233 samples were classified as low SCC (< 100.000
380 cells/mL), while 109 samples were categorized as high SCC (> 100.000 cells/mL). The
381 percentage of quarters with low SCC was notably higher during EL-2023 (80%) and
382 ML-2023 (78.9%) compared to DO-2022 (53.1%) and LL-2023 (53%), with high SCC
383 percentages of 20%, 21.1%, 46.9%, and 47%, respectively. As depicted in **Figure 1A**,
384 there was a statistical difference between the different lactation periods, except
385 between DO-2022/LL-2023 and EL-2023/ML-2023. The distribution of individual
386 bacterial counts (IBC) exhibited a comparable pattern, revealing a significant
387 difference between samples classified as low SCC and those categorized as high
388 SCC. Specifically, low SCC samples had a mean IBC of 3.31 ± 0.475, while high SCC
389 samples showed a mean IBC of 3.74 ± 0.334.

390 The percentage of quarters testing positive or negative for mastitis pathogens,
391 identified via MALDI-Tof, also reflected the trends observed for SCC and IBC **(Figures**
392 **1B and 1C)**. The lowest detection of mastitis-causing pathogens was recorded during
393 EL-2023, with only 10.5% of samples testing positive. Among the mastitis-causing
394 pathogens identified in this study (from 60 samples), *E. faecalis* emerged as the
395 predominant etiological agent, accounting for 30% of the cases. *S. aureus*, *S. uberis*,

396 and *S. chromogenes* were detected in 25%, 20%, and 10% of the samples,
397 respectively. The bacterial species *S. epidermidis* (7%), *C. bovis* (3%), *S. dysgalactiae*
398 (3%), and *S. haemolyticus* (2%) were identified to a lesser extent.

399 The relationship between SCC and IBC ratio concerning specific mastitis-causing
400 pathogens was also investigated. Groups with fewer than three samples, specifically
401 *C. bovis*, *S. dysgalactiae*, and *S. haemolyticus*, were excluded from this analysis. Our
402 findings indicate that the level of somatic cells was significantly higher in samples
403 associated with *S. epidermidis* compared to those classified as negative, with an
404 adjusted p-value of 0.014. This suggests that a small number of *S. epidermidis* cells
405 may be sufficient to trigger a strong pro-inflammatory response in the mammary gland.

406 When we examined these same samples in relation to bacterial counts on blood
407 agar plates (after 48 hours of incubation at 37°C), alongside individual bacterial count
408 (IBC), somatic cell count (SCC), and the identified pathogen **(Figure 2A)**, we observed
409 that many samples classified as negative (i.e.. absence of mastitis-causing pathogens
410 according to the gold standard) still showed high plate counts. Specifically, 129
411 samples had colony counts ranging between 30 and 300 CFUs. From these samples,
412 57 showed SCC higher than 100,000 (median: 2.9 x 10^5 cells/mL) and a median IBC
413 of 7.6 x 10^3 cells/mL. Our findings indicate that these samples harbor microorganisms
414 distinct from the major mastitis-causing pathogens, which, although not routinely
415 identified, may still contribute to increased somatic cell counts in the mammary gland.
416

417 **3.2. Metataxonomic analysis**

418 **3.2.1. Alpha and beta diversity analyses show significant effects of lactation**
419 **period on diversity indices**

420 In this longitudinal study, we investigated the microbial composition of 306 milk
421 samples encompassing a whole lactation cycle. By conducting amplicon sequencing
422 of the hypervariable region V3-V4 of the 16S rRNA gene, we obtained a total of 1.3 ×
423 10^7 high-quality filtered sequences after the removal of chimeric, contaminants, and
424 spurious sequences (e.g.. mitochondrial and chloroplast sequences). The median
425 number of sequences per sample in the low SCC group was 37,589, while the high
426 SCC group had a median of 47,282 sequences. Across all groups, the overall median
427 number of sequences per sample was 42,970.

428 We first investigated alpha diversity analysis considering the variables SCC
429 (low and high) and lactation period (DO-2022, EL-2023, ML-2023, and LL-2023).

When we analyzed the variable SCC, statistical significance was observed for the indices Shannon, Simpson, and inverse Simpson **(Figure 2B)**. Overall, samples with low SCC exhibited greater microbial diversity and lower dominance of individual species. When the different periods were considered, five out of seven indices showed higher diversity measurements for the periods DO-2022 when compared to LL-2023. All seven indices were statistically significant for the pairwise comparison between EL-2023 and LL-2023 **(Figure 2C)**. As depicted in the volatility plots **(Figures 2D and 2E)** for the Shannon and Simpson indices, there is a clear trend indicating a reduction in both indices at the end of lactation in samples containing more than 100,000 cells/mL.

We also employed the generalized additive model (GAM) to analyze how the Shannon and Simpson diversity indices were influenced by different predictor variables. The GAM analysis revealed significant effects of the lactation period, somatic cell count, and individual animal variability on both the Shannon and Simpson indices, while bacterial count did not show a significant impact. The smooth term for the lactation period indicated a modest but statistically significant effect on microbial diversity (Shannon: edf = 2.10, F = 2.72, p = 0.039; Simpson: edf = 2.11, F = 2.66, p = 0.047), suggesting variation in diversity across lactation stages. Somatic cell count had a highly significant effect (Shannon: edf = 0.92, F = 28.95, $p < 0.001$; Simpson: edf = 0.94, F = 38.44, $p < 0.001$), indicating that the degree of mammary gland inflammation plays a critical role in shaping the microbiota. Additionally, the random effect of individual animals was significant (Shannon: edf = 16.37, F = 2.25, $p < 0.001$; Simpson: edf = 14.7, F = 1.72, p = 0.002), highlighting considerable inter-animal variability in microbial diversity. However, the bacterial count was not a significant predictor of diversity (Shannon: Estimate = 1.49e-08, p = 0.681; Simpson: Estimate = 4.353e-09, p = 0.387). The model explained 24.9% and 23.4% of the deviance, with adjusted R-squared values of 19.5% and 18.4% for the Shannon and Simpson indices, respectively, underscoring the relevance of the factors considered in the model.

The PERMANOVA analyses using unweighted and weighted UniFrac distances revealed significant effects of various factors on the microbial community structure **(Figure 3)**. For the unweighted UniFrac analysis, the lactation period explained 3.01% of the variance in microbiota composition (F = 3.20, p = 0.01), indicating shifts in microbial diversity across different stages of lactation. Individual animal variability also contributed 9.42% to the observed variance (F = 1.25, p = 0.01), highlighting notable inter-animal differences in microbial composition. However, somatic cell count and the

464 history of mastitis-causing pathogens identified in milk samples were not significant
465 predictors, accounting for only 0.30% and 2.85% of the variance, respectively. The
466 majority of the variation, 83.97%, remained unexplained by the measured factors,
467 suggesting other influences on the microbial community.

468 In contrast, the weighted UniFrac analysis indicated more pronounced effects
469 of the measured variables. Somatic cell count had a small but significant effect,
470 explaining 0.78% of the variance (F = 3.22, p = 0.01). The lactation period played a
471 more substantial role, explaining 5.52% of the variance (F = 7.57, p = 0.01), reinforcing
472 the influence of lactation stages on microbial composition. Individual animal variability
473 was the largest contributor to variance, explaining 15.73% (F = 2.69, p = 0.01), while
474 the identification of major mastitis-causing pathogens in milk samples contributed
475 7.17% (F = 3.28, p = 0.01), highlighting the impact of dysbiosis on microbial
476 communities in the udder. The remaining 65.18% of the variance was attributed to
477 residual (i.e., unmeasured factors), reflecting biological complexity and other unknown
478 influences shaping the udder microbiota.

479

480 **3.2.2. Differential abundance analysis and multivariable associations reveal**
481 **shifts in *Actinobacteriota* across lactation stages**

482 After clustering at a 99% similarity threshold and preprocessing steps (removal
483 of *g-Mitochondria*, *o-Chloroplast*, and *d-Archaea*), we identified a total of 48,622 ASVs.
484 Four phyla – *Firmicutes* (median: 63%; mean: 60.1%), *Actinobacteria* (median: 20%;
485 mean = 25.2%), *Proteobacteria* (median: 7%; mean: 11.4%), and *Bacteroidota* (1%;
486 mean: 2.3%) – comprised approximately 99% of the milk bacterial population. At the
487 genus level, ten taxa stand out and comprise approximately 65% of the
488 microorganisms at this taxonomic level: *Corynebacterium* (median: 12.9%; mean:
489 19.4%), *Staphylococcus* (median: 4.0%; mean: 17.0%), *Bradyrhizobium* (median:
490 1.8%; mean: 5.4%), *Streptococcus* (median: 0.2%; mean: 4.9%), *Aerococcus*
491 (median: 2.6%. mean: 4.8%), *Romboutsia* (median: 3.4%; mean: 4.5%), *Enterococcus*
492 (median: 0.2%; mean: 3.1%), *Kocuria* (median: 0.9%; mean: 2.0%), *Lactococcus*
493 (median: 0.2%; mean: 1.7%), and *Weissella* (median: 0.6%; mean: 1.6%).

494 We performed a differential abundance analysis using MaAsLin2 to determine
495 multivariable associations across samples grouped by somatic cell count (SCC),
496 lactation period, sampling day, and microbial composition at both the phylum and
497 genus levels. We found significant results for 15 phyla **(Supplementary Table 1)**.

498 Samples with low SCC showed a significantly greater abundance of eight phyla, with
499 *Proteobacteria*, *Actinobacteriota*, and *Bacteroidota* particularly prominent based on
500 their prevalence in our dataset (belonging to the top four phyla). Additionally, shifts in
501 bacterial phyla were observed across different lactation periods. While the abundance
502 of the *Firmicutes* phylum remained relatively stable throughout the lactation periods
503 (q-value > 0.25), we observed an enrichment of the *Proteobacteria* phylum in the early
504 lactation period (EL-2023), which gradually decreased in the subsequent periods. In
505 contrast, members of the *Actinobacteriota* phylum were enriched in the mid- (ML-
506 2023) and late lactation (LL-2023) periods. When considering the continuous variable
507 "day", we observed a distinct fluctuation in *Actinobacteriota* abundance across the
508 studied periods. Specifically, *Actinobacteriota* was highly abundant during the baseline
509 period DO-2023, but this abundance decreased in the early lactation period (EL-2023),
510 followed by two subsequent increases during the mid- (ML-2023) and late lactation
511 (LL-2023) periods.

512 Our genus-level analysis identified 66 differentially abundant genera
513 **(Supplementary Table 2)**. Samples with low SCC showed enrichment in several
514 genera, with *Corynebacterium*, *Bradyrhizobium*, *Romboutsia*, and *Lactococcus*
515 standing out among the top 10 genera. Conversely, *Staphylococcus* was significantly
516 enriched in samples with high SCC. Examining differentially abundant genera by
517 lactation period, *Bradyrhizobium*, Unidentified *Lactobacillaceae* family. *Pediococcus*,
518 Unidentified *Aerococcaceae* Family, and *Weissella* were enriched in samples collected
519 during EL-2023 compared to DO-2022. Additionally, two genera from the families
520 *Aerococcaceae* and *Lactobacillaceae* were enriched in ML-2023. When analyzed by
521 sampling day, *Bifidobacterium* and *Bacillus* displayed positive coefficients, indicating
522 a progressive increase in abundance over time, whereas the abundance of
523 Unidentified *Lactobacillaceae* family decreased over time.
524

525 **3.3. The integration between t-SNE and culturomics improves milk microbiota**
526 **classification**

527 The microbial composition of the 295 samples was subjected to dimensionality
528 reduction t-SNE analysis. An optimal number of k-mean clusters was identified by
529 using the “Elbow method,” and the microbiota was divided into eighteen clusters
530 **(Supplementary Figure 1A-C)**. The eighteen clusters were then used to investigate
531 the microbial composition, which significantly differed in main ASVs and genera. Each

532 cluster was then used as a factor to map Chao1 and Shannon diversity indices and
533 other variables such as SCC and IBC. As depicted in **Figure 4**, cluster analysis
534 revealed that IBC levels varied significantly from the global mean in clusters 9, 10, 13,
535 14, 16, and 17. For SCC, statistically significant differences were observed in clusters
536 9, 10, 11, 13, 14, 15, and 17. The diversity indices Chao1 and Shannon also exhibited
537 substantial variation across clusters. For the Chao1 index, clusters 6, 7, 10, 13, 14,
538 15, 16, and 18 did not differ significantly from the global mean. In contrast, for the
539 Shannon index, all clusters except 2, 7, 9, 10, 15, 16, and 18 showed significant
540 differences from the global mean.

541 We conducted an analysis to examine the distribution of the predominant
542 bacterial genera across different clusters **(Figure 5)**. The genera *Corynebacterium*,
543 *Staphylococcus*, and *Aerococcus* show high prevalence and widespread distribution
544 across clusters, indicating that these genera are commonly found across diverse milk
545 samples used in this study and are not restricted to any specific group. In contrast,
546 *Enterococcus* and *Streptococcus* were majority limited to a single cluster each,
547 respectively, clusters 13 and 14.

548 Considering a minimum mean relative abundance of 10% per cluster
549 **(Supplementary Figure 2)**, we observed a notably high relative abundance of
550 *Staphylococcus* in five clusters (cluster 4: 80.50%, cluster 6: 77.88%, cluster 7:
551 58.85%, cluster 10: 56.96%, cluster 15: 52.69%). Combined with the culturomics'
552 analysis, cluster 4 was enriched for *S. chromogenes* (4 positive samples out of 12),
553 whereas cluster 10 was enriched for *S. aureus* (7 positive samples out of 9). Only one
554 species of *S. haemolyticus* was identified by MALDI-Tof (cluster 6).

555 We also conducted a MegaBLAST analysis on the top 20 ASVs to investigate
556 species-level identities within clusters with high prevalence of *Staphylococcus* and
557 *Corynebacterium* and *Streptococcus* **(Figure 6)**. For cluster 4, samples showed a
558 homogenous distribution of the four most abundant ASVs, here annotated as *S.*
559 *chromogenes* (identity scores ranging from 99.53 to 100%). Cluster 6 showed a high
560 abundance of an ASV that had as top hits (identity of 100%) with *S. borealis*,
561 *S.taphylococcus pragensis*, *S.taphylococcus petrasii,* and *S.taphylo croceilyticus*.

562 Cluster 10 was primarily dominated by a single ASV with 100% identity,
563 annotated as *S. aureus*, *S. schweitzeri*, *S. argenteus*, *S. roterodami*, and *S. simiae*.
564 Notably, sample H193 displayed a microbial profile distinct from the other samples in
565 this cluster, diverging from the typical dominance of *S. aureus*. Cluster 7 was

566 dominated by an ASV with 100% identity matching *S. epidermidis*, *S. caprae*, and S.
567 *capitis*. Additionally, four samples (H298, H300, H305, and H307) showed a high
568 abundance of an ASV annotated with 100% identity as *S. borealis*, *S. pragensis*, *S.*
569 *petrasii*, and *S. croceilyticus*. In the final *Staphylococcus* cluster analyzed (cluster 15),
570 ASVs were annotated as *S. epidermidis, S. caprae, S. capitis,* and *S. hominis*.
571 Regarding *Corynebacterium*, eleven out of eighteen clusters showed a relative
572 abundance greater than 10%. Clusters 2 (37.68%), 8 (64.92%), and 18 (44.56%)
573 exhibited particularly high abundances. It is important to note that from culturomics'
574 data, the detection of *Corynebacterium* species was limited to only two samples (H087
575 and H109), both positive for *Corynebacterium bovis*. The ASV-level investigation
576 provided insights into the composition of these clusters. Our analysis revealed that
577 several ASVs had the best hits with uncultured bacteria. Cluster 2 stood out for its
578 heterogeneity of *Corynebacterium* species, a pattern not observed in clusters 8 and
579 18, and overall enrichment for *C. bovis*. Cluster 8 showed a more homogeneous
580 distribution of ASVs, and specifically two sequence variants annotated as uncultured
581 bacterium/*C. bovis* (identities of 99.76% and 100%).
582 Lastly, clusters 3 and 14 showed a high abundance of members of the genus
583 *Streptococcus*. Cluster 3 was enriched for three ASVs (100% identity) annotated as
584 *Streptococcus uberis* /*Streptococcus porcinus*, *Streptococcus*
585 *alactolyticus*/*Streptococcus macedonicus*/*Streptococcus pasteurianus*, whereas
586 cluster 14 was dominated by a single ASV identified as *Streptococcus uberis* (identity:
587 100%).
588
589 **3.3.1. Udder microbiota dynamics reveal persistent infection by *S. aureus* and**
590 ***S. chromogenes*, but not *S. uberis***
591 Following t-SNE analysis to enhance sample classification and elucidate
592 underlying patterns in the dataset, we conducted a quarter-level analysis to track
593 temporal fluctuations in genus-level taxa before and after detecting the pathogen
594 identified by the gold standard method. We focused on samples testing positive for
595 Staphylococci, Enterococci, and Streptococci. Notably, these milk samples
596 predominantly grouped into clusters 4, 10, 13, and 14. For robustness, only quarters
597 with a minimum of three sampling points were included in the analysis **(Figure 7)**. To
598 estimate the resilience of resident microbiota following a dysbiosis event, we
599 established a microbial dysbiosis index combined with the Shannon diversity index. A

quarter was classified as dysbiotic if it exhibited a Shannon index below 3.5 (the mean value for quarters with somatic cell counts below 100,000 cells/mL) and an MDI greater than 0, indicating a predominance of dysbiotic microorganisms within that sample.

For cluster 10, which was identified as enriched with *S. aureus*, 5 quarters were analyzed and the longitudinal analysis revealed a persistent state of infection by *S. aureus*. Once an infection was established, *S. aureus* remained consistently present in samples regardless of the lactation stage. Although Intramammary bacterial counts varied over time, somatic cell counts consistently remained elevated, typically exceeding, or in a few cases, hovering around 100,000 cells/mL. A marked depletion of specific taxa, such as *Corynebacterium*, *Aerococcus*, and *Romboutsia*, was observed, along with reductions in other low-abundance taxa grouped here as "Others." Additionally, a co-infection with species of the genus *Streptococcus* was detected at two specific time points (samples H195 and H349). When the MDI index was calculated for samples in this cluster, it was evident that samples following an episode of high *S. aureus* abundance did not return to a state of eubiosis.

Cluster 4 showed enrichment for *S. chromogenes* and persistent infection behavior was also observed. However, it's noteworthy that the infection-associated somatic cell counts and intramammary bacterial counts in this cluster are significantly lower than those found in cluster 10. In the case of quarter 7032_VF, both *S. aureus* and *S. chromogenes* were isolated and identified. Interestingly, once this quarter was infected by both species, the microbiota failed to revert to its original state. This phenomenon is also clearly demonstrated in quarters 6691_HB and 6691_VB, which suggests that *S. aureus* and *S. chromogenes* exhibit specific competitive mechanisms that are reflected in a pattern of microbiota depletion. This pattern impedes the recolonization of the native microbiota, leading to a more prolonged dysbiosis status, as evidenced by the dysbiosis indices calculated for those samples.

The analysis of cluster 13, enriched for *E. faecalis*, revealed a comparatively less severe dysbiosis than that observed in quarters infected with *S. aureus* and *S. chromogenes*. In this cluster, a reduction in the relative abundance of taxa such as *Corynebacterium*, *Aerococcus*, and *Romboutsia* was noted. However, the native microbiota appears capable of re-establishing itself, a resilience not observed in clusters 4 and 10. This phenomenon could be observed for the quarters 6415_VB (sample H077: Shannon = 1.73, MDI = 1.03; sample H165: Shannon = 4.77, MDI = -

634 2.35), 7003_HB (sample H255: Shannon = 2.53, MDI = 1.23; sample H355: Shannon
635 = 5.17, MDI = -0.72), and 7003_HF (sample H156: Shannon = 2.04, MDI = 1.05;
636 sample H256: Shannon = 4.43, MDI = -0.49).

637 For cluster 14 (characterized by enrichment in *Streptococcus uberis*), we
638 observed that the reduction in *Corynebacterium* was not as pronounced as in clusters
639 dominated by *Staphylococcus* and *Enterococcus*. Although *S. uberis* induced a
640 dysbiotic state in the mammary gland, we identified five instances in which a dysbiotic
641 quarter returned to a eubiotic state, and one which we classified as in the borderline
642 of a eubiotic state (samples H084 and H164), including quarters such as 6611_HF
643 (sample H082: Shannon = 1.65, MDI = 2.33; sample H162: Shannon = 3.92, MDI = -
644 0.65), 6415_HB (sample H079: Shannon = 1.60, MDI = 2.04; H167 = Shannon = 4.58,
645 MDI = -1.94), 6611_HB (sample H083: Shannon, 2.44, MDI: 1.98), 6611_HF (sample
646 H084: Shannon = 2.36, MDI = 1.55; sample H164: Shannon = 2.96, MDI = -0.21), and
647 6941_VB (sample H141: Shannon = 3.38, MDI = 0.34; sample H241: Shannon = 4.10,
648 MDI = -0.65).

649

650 **3.4. Metagenomic shotgun sequencing uncovers pathogen-specific**
651 **metabolic signatures in the bovine mammary gland**

652 After trimming, quality filtering, and in silico separation of bacterial reads from
653 bovine sequences, we obtained approximately 269.5 GB of high-quality PE reads
654 across 73 samples, which were then processed for functional profiling using
655 HUMAnN3. This profiling identified 289 MetaCyc metabolic pathways, which were
656 then assessed for differential abundance with MaAsLin2 considering "SCC" and
657 "pathogen" as fixed effects. At first glance, the heatmap **(Figure 8A)** constructed with
658 the 289 predicted MetaCyc pathways revealed that the samples did not form well-
659 defined clusters when grouped by SCC or lactation period, highlighting a significant
660 degree of heterogeneity. However, four distinct clusters of pathogen-negative samples
661 are evident in the heatmap, characterized by a notable lack of metabolic pathway
662 enrichment. Additionally, clusters of samples identified as positive for *S. aureus* and
663 *E. faecalis* were also discernible.

664 Focusing on pathways that were differentially expressed (q-value < 0.25) based
665 on somatic cell content, we observed an enrichment of 134 microbial MetaCyc
666 pathways for samples with high SCC and four MetaCyc pathways for samples with
667 low SCC **(Figure 8BC)**. When pathways were stratified by bacterial species within the

high SCC group, many pathways were common across several bacterial species, including Staphylococci and Streptococci, and were largely associated with central metabolism. Notably, the pathway PWY-6507 (PWY-6507: 4-deoxy-L-threo-hex-4-enopyranuronate degradation) was significantly enriched in *E. faecalis*. In contrast, the four pathways enriched in the low SCC group (PWY-7185, PWY-7184, PWY-6545, and PWY-7210) were primarily related to pyrimidine biosynthesis.

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* **(Figure 9)**.

For *S. aureus*, 37 pathways were significantly identified, with 36 pathways enriched in this group and one pathway (ASPASN-PWY) more abundant in the reference group. For *S. chromogenes*, 33 pathways were detected, including 32 enriched in this group and one (POLYISOPRENSYN.PWY) enriched in the reference group. In *E. faecalis*, 38 pathways were identified, of which 29 were enriched in this group, while 9 were highly represented in the negative group: PWY-1861, PWY-1042, CITRULBIO-PWY, PWY-7200, SULFATE.CYS.PWY, HISDEG.PWY, SO4ASSIM.PWY, PWY-6703, and PWY1ZNC.1. For *S. epidermidis*, only pathways enriched in the negative group were found, including PWY66-429, PEPTIDOGLYCANSYN.PWY, PWY-6124, PWY-7234, PENTOSE.P.PWY, PWY-3841, PWY-6123, PWY-6122, PWY-6277, and PWY-6121. Lastly, for *S. uberis*, all seven identified pathways were enriched in this group. Overall, these findings highlight distinct metabolic profiles associated with specific pathogens, underscoring the utility of pathway-based analyses to better understand pathogen-host interactions and their metabolic underpinnings in mastitis.

Based on the previously identified pathways with a q-value < 0.25, a Venn diagram was constructed to identify unique and shared MetaCyc pathways **(Figure 10)**. For *S. aureus*, 14 unique MetaCyc pathways were identified (PWY-5484, PWY-5005, ARGSYNBSUB.PWY, GLYCOLYSIS, PWY-5840, PWY-7111, PWY-7953, PWY-3001, PWY-7977, FERMENTATION.PWY, PWY-5189, P125-PWY, PWY-5918, and RUMP.PWY). The enrichment of these pathways demonstrated a versatile metabolic profile, encompassing central carbon metabolism, amino acid and nucleotide

biosynthesis, stress tolerance, and virulence factors, highlighting the adaptability of *S. aureus* and the potential for pathogenicity in various host environments. *S. chromogenes* exhibited unique metabolic pathways associated with amino acid biosynthesis (PWY-7383, glutamate and glutamine biosynthesis), lipid metabolism (PWY-6936, fatty acid biosynthesis), and energy production (PWY-6163, fermentation), 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 (GLYCOGENSYNTH.PWY), lactose and galactose utilization (P461-PWY), and nucleotide sugar metabolism (PWY-5973), as well as processes involved in cell wall biosynthesis (PWY-621), 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 (glycolysis – PWY66-429), cell-wall structural maintenance (peptidoglycan biosynthesis – PEPTIDOGLYCANSYN-PWY), stress tolerance (polyamine biosynthesis – PWY-6122), and nutrient acquisition (siderophore biosynthesis – PWY-7234**;** and arginine pathways – PWY-6123). These features contribute to its resilience and success as a commensal and opportunistic pathogen. Finally, *E. faecalis* displayed a metabolic profile enriched in pathways related to amino acid biosynthesis (PWY-5971, PWY.7357), nucleotide metabolism (PWY.6317, P164.PWY), and lipid-related processes (PWY0.1297, PWY0.301), as well as pathways supporting secondary metabolite production (PWY0.1477) and cofactor biosynthesis (PWY.7242, PWY.7791), underscoring its metabolic flexibility and potential resilience in diverse environmental conditions.

3.5. Genome-resolved metagenomic data analysis

3.5.1. MAG quality assessment

The genome-centric approach applied to shotgun reads from milk microbiomes facilitated the reconstruction of 142 metagenome-assembled genomes (MAGs), of which 26 were obtained using a co-assembly method and 116 using 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 >

735 10%) **(Supplementary Table 3)**. MAGs with contamination higher than 10% were
736 excluded from subsequent analyses.
737 Taxonomic assignments were consolidated at the species level when always
738 possible. For high-quality MAGS, species were annotated as *Corynebacterium bovis*
739 (n = 1), *Corynebacterium kroppenstedtii* (n= 6), *Enterococcus faecalis* (n =9),
740 *Staphylococcus aureus* (n = 5), *Staphylococcus chromogenes* (n = 4), *Staphylococcus*
741 *epidermidis* (n = 2), *Staphylococcus haemolyticus* (n = 2), *Streptococcus equinus* (n =
742 2), and *Streptococcus uberis* (n = 7). For medium-quality MAGs, *Corynebacterium*
743 *bovis* (n = 5), *Corynebacterium kroppenstedtii* (n = 12), *Corynebacterium urogenitale*
744 (n = 1), *Dermatophilus* spp. (n = 2), *Enterococcus faecalis* (n = 10), *Leuconostoc*
745 *mesenteroides* (n = 1), *Staphylococcus aureus* (n = 2), *Staphylococcus borealis* (n =
746 2), *Staphylococcus chromogenes* (n = 3), *Staphylococcus epidermidis* (n = 3),
747 *Staphylococcus hominis* (n = 2), *Streptococcus bovis* (n = 1), *Streptococcus*
748 *dysgalactiae* (n = 2), *Streptococcus equinus* (n = 1), *Streptococcus gallolyticus* (n =
749 1), and *Streptococcus uberis* (n = 2).
750 The highest diversity of MAGs were found among the low-quality ones, such as
751 *Aerococcaceae* (n = 2), *Cellulomonas shaoxiangyii* (n = 1), *Corynebacterium bovis* (n
752 = 2), *Corynebacterium glutamicum* (n = 2), *Corynebacterium kroppenstedtii* (n = 4),
753 *Corynebacterium rouxii* (n = 1), *Corynebacterium urogenitale* (n = 1), *Corynebacterium*
754 *vitaeruminis* (n = 1), *Enterococcus faecalis* (n = 2), *Enterococcus faecium* (n = 1),
755 *Facklamia miroungae* (n = 1), *Kocuria atrinae* (n = 2), *Kocuria carniphila* (n = 1),
756 *Lysobacter* (n = 1), *Micrococcaceae* bacterium (n = 1), *Peptoniphilus* (n = 1),
757 *Peptoniphilus nemausensis* (n = 1), *Ruminococcus flavefaciens* (n = 1), *Ruoffia*
758 *tabacinasalis* (n = 1), *Staphylococcus aureus* (n = 1), *Staphylococcus chromogenes*
759 (n = 4), *Staphylococcus haemolyticus* (n = 1), *Staphylococcus hominis* (n = 1),
760 *Streptococcaceae* (n = 1), *Streptococcus equinus* (n = 1), *Streptococcus uberis* (n =
761 1), and *Trichococcus flocculiformis* (n = 1).
762 After genome dereplication and the exclusion of MAGs with contamination
763 levels exceeding 10% (four MAGs assigned to the family *Aerococcaceae* and one to
764 the family *Micrococcaceae*), 26 MAGs were selected as representative of the
765 metagenomes. These MAGs were subsequently used to map the reads and estimate
766 their relative abundances across the samples **(Figure 11)**. The MAG assigned as
767 *Cutibacterium acnes* was identified as a putative contaminant in our dataset and
768 removed from downstream analysis.

To explore the functional potential of the microbiome in greater detail, we predicted microbial genes from the MAGs of interest using the NCBI Prokaryotic Genome Annotation Pipeline, enabling robust downstream comparisons. The resulting protein sequences were annotated using EggNOG, and the COG categories were normalized to the total number of proteins per MAG. Our analysis focused on the bacterial families *Streptococcaceae*, *Aerococcaceae*, *Staphylococcaceae*, and *Corynebacteriaceae* due to their significant representation within the dataset.

As shown in **Figure 12**, the top 20 COG categories primarily comprised proteins associated with metabolism (E: Amino acid transport and metabolism, P: Inorganic ion transport and metabolism, G: Carbohydrate transport and metabolism, C: Energy production and conversion, F: Nucleotide transport and metabolism, H: Coenzyme transport and metabolism, I: Lipid transport and metabolism, and Q: Secondary metabolites biosynthesis, transport, and catabolism), cellular processes and signaling (M: Cell wall/membrane/envelope biogenesis, V: Defense mechanisms, O: Post-translational modification, protein turnover, and chaperones, D: Cell cycle control, cell division, chromosome partitioning, T: Signal transduction mechanisms, and U: Intracellular trafficking, secretion, and vesicular transport), and information storage and processing (L: Replication, recombination, and repair, J: Translation, ribosomal structure, and biogenesis, and K: Transcription). Notably, a substantial proportion of proteins across all bacterial families fell into the “function unknown” group (COG category S), highlighting the large fraction of poorly characterized proteins within these genomes.

We aimed to determine which functional categories displayed statistically significant differences in the abundance of annotated proteins across the bacterial families. In total, four COG categories exhibited significant variation. For COG category G (Carbohydrate transport and metabolism), the families *Aerococcaceae* and *Streptococcaceae* demonstrated a higher abundance of orthologs compared to *Corynebacteriaceae* and *Staphylococcaceae*. In COG category Q (Secondary metabolites biosynthesis, transport, and catabolism), *Corynebacteriaceae* displayed the highest abundance of orthologs among the four families. For COG category C (Energy production and conversion), the family *Staphylococcaceae* was enriched compared to *Corynebacteriaceae* and *Streptococcaceae*, although no significant difference was observed when compared to *Aerococcaceae*. Finally, for COG category U (Intracellular trafficking, secretion, and vesicular transport), the family

803 *Streptococcaceae* exhibited a greater abundance of orthologs compared to
804 *Staphylococcaceae*. Overall, the classification of COGs into functional categories
805 provided valuable insights into the metabolic preferences of these bacterial families,
806 highlighting differences in their functional potential and adaptation to specific
807 ecological niches.
808

809 **3.5.2. Multi-Locus Sequence Typing (MLST), Virulence Factors (VFs) and**
810 **Antimicrobial Resistance Genes (AMRGs)**

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

817 The distribution of virulence factors (VFs) across multiple bacterial taxa found
818 in milk samples was analyzed to assess their biosynthetic and pathogenic potential
819 **(Figure 13)**. With regards to *Corynebacterium* spp., *C. bovis* exhibited a predominant
820 focus on regulatory factors, representing 51.85% of its VF genes, followed by stress
821 survival and other factors (18.52% each), and nutritional/metabolic factors at 11.11%.
822 Similarly, *C. kroppenstedtii* showed a strong emphasis on regulatory factors (73.33%),
823 with other VF categories accounting for the remaining 26.67%. In *C. urogenitale*,
824 regulation (66.67%) and other factors (33.33%) were the only VF categories identified.
825 Bacterial species annotated as *E. faecalis* showed a broad distribution of VF
826 categories, with adherence genes comprising the largest fraction (43.90%).
827 Exoenzymes (15.70%), immune modulation (12.50%), and stress survival (12.50%)
828 were also well-represented. Other categories included biofilm formation (11.34%),
829 nutritional/metabolic factors (3.78%), and effector delivery systems (0.29%).

830 For *Staphylococcus* spp., *S. aureus* showed dominant VFs categories
831 associated to immune modulation (24.53%) and exoenzymes (14.34%), followed by
832 nutritional/metabolic factors (13.58%), exotoxins (12.08%), adherence (11.70%), and
833 effector delivery systems (11.70%). Other categories included biofilm formation
834 (6.79%) and stress survival (5.28%). *S. chromogenes* was predominantly associated
835 with immune modulation (51.02%), followed by stress survival (25.51%) and
836 exoenzymes (13.27%). Adherence (7.14%) and biofilm formation (3.06%) were less

837 frequent. In *S. epidermidis*, exoenzymes were most abundant (36.08%), followed by
838 adherence (22.68%), stress survival (19.59%), and immune modulation (18.56%).
839 Exotoxins accounted for a minor proportion (3.09%). *S. hominis* was heavily skewed
840 towards immune modulation (77.27%), with smaller contributions from stress survival,
841 adherence, and exoenzymes (13.64%, 4.55%, and 4.55%, respectively).

842 In species of *Streptococcus*, adherence, immune modulation, and stress
843 survival were equally represented (33.33% each) for *S. bovis*. *S. dysgalactiae*
844 demonstrated a focus on adherence (50%), followed by exoenzymes and stress
845 survival (14.29% each), with other categories such as immune modulation, invasion,
846 and nutritional/metabolic factors at 7.14% each. For *S. equinus*, exotoxins dominated
847 (71.11%), with smaller contributions from immune modulation (13.33%), adherence
848 (8.89%), and stress survival (6.67%). *S. gallolyticus* exhibited a strong preference for
849 immune modulation (63.64%), with stress survival and adherence making up 27.27%
850 and 9.09%, respectively. Finally, *S. uberis* showed a varied distribution, with immune
851 modulation (46.67%) and adherence (28%) being the most common categories,
852 followed by stress survival (13.33%), nutritional/metabolic factors (6.67%),
853 exoenzymes (4%), and exotoxins (1.33%).

854 Since the VFDB database lacks information on members of the *Aerococcaceae*
855 family (as it only includes experimentally verified virulence factors), the prediction of
856 virulence factors for microorganisms associated with this taxon was performed using
857 category V (Defense mechanisms) from the COG database as a reference. This
858 analysis identified six protein-coding genes that may enable bacteria of this taxon to
859 colonize the bovine mammary gland and enhance their pathogenic potential: hrtA
860 (putative hemin import ATP-binding protein), hsdM (site-specific DNA-
861 methyltransferase, adenine-specific), ylbB (efflux ABC transporter, permease protein),
862 yheH (ABC transporter transmembrane region), ywjA (ABC transporter
863 transmembrane region), and XK27_06785 (putative ABC transport system ATP-
864 binding protein, ABC.CD.A).

865 In the analysis of antimicrobial resistance (AMR) genes, a total of 144 genes
866 were annotated using three databases: ResFinder, CARD, and ARG-ANNOT **(Figure**
867 **14)**. The top five antibiotic classes represented in the dataset were fluoroquinolones
868 (15%), macrolides (10.2%), tetracyclines (9.6%), aminoglycosides (7.6%), and
869 rifamycins (5.8%).

When stratified by bacterial species, distinct patterns of AMR gene distribution were observed. *Corynebacterium* spp. exhibited a high abundance of genes associated with rifamycins (20.4%), macrolides (14.6%), penams (14.0%), aminoglycosides (12.7%), and aminocoumarins (12.1%). For *Staphylococcus* spp., the most abundant AMR genes were associated with fluoroquinolones (18.2%), acridine dyes (11.2%), tetracyclines (10.8%), aminoglycosides (8.3%), and phenicols (7.9%). Similarly, *Streptococcus* spp. displayed a predominance of genes linked to fluoroquinolones (24.2%), macrolides (14.2%), beta-lactams (11.1%), rifamycins (10.4%), and tetracyclines (6.6%). 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%). *Dermatophilus* spp. demonstrated a high prevalence of genes for aminocoumarins (22.9%), fluoroquinolones (20%), macrolides (14.3%), ciprofloxacin (8.6%), and oleandomycin (8.6%). In *Kocuria atrinae*, AMR genes were predominantly annotated to rifamycins (35.7%) and fluoroquinolones (28.6%), with additional annotations to aminocoumarins, macrolides, oleandomycin, oxazolidinones, and peptides (each at 7.1%). Lastly, *Leuconostoc mesenteroides* had only one annotated AMR gene, which was associated with rifamycins. 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.

3.5.3. 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 translationally modified Peptides (RiPPs) and (unmodified) bacteriocins. In total, 27 gene clusters were predicted **(Figure 15A)**. Bacterial species annotated as *S. aureus* the highest number of different classes such as 159.2; Pediocin (only predicted in the MAG H249_bin.1, 100% identity with bacteriocin immunity protein from *Pediococcus pentosaceus*), 185.1;BacCH91, putative_bacteriocin (Lactococcin_972 – PF09683), 314.1; Auto_Inducing_Peptide_I, 491.1; Lantibiotic, 96.1; putative_lantibiotic (BsaA2), and Sactipeptides. Four areas of interest were identified in MAGs assigned to *E. faecalis* (62.3;enterolysin_A, 63.3; Enterolysin_A, Lasso_peptide, and Sactipeptides), *S. chromogenes* (145.1; Subtilosin_A , 314.1;Auto_Inducing_Peptide_I, 315.1;Auto_Inducing_Peptide_II, and Sactipeptides), *S. uberis* (157.1; uberolysin, 163.2; Penocin_A,

904 42.2;Bovicin_255_peptide, and 61.3;Dysgalacticin), and *C. kroppenstedtii*
905 (189.2;protease-activatedantimicrobialprotein(PAMP), 42.2;Bovicin_255_peptide,
906 84.1;Nukacin_A_(NukacinISK-1), and 97.2;Enterocin_X_chain_beta). Three gene
907 clusters were identified in *S. epidermidis* (Sactipeptides,
908 314.1;Auto_Inducing_Peptide_I, and 64.2;Delta-lysinI), *Dermatophilus* spp.
909 (Sactipeptides, 64.1;Microbisporicin_(NAI-107), and 93.3;Zoocin_A), *S. hominis*
910 (Sactipeptides, 314.1;Auto_Inducing_Peptide_I, and
911 315.1;Auto_Inducing_Peptide_II), and *S. dysgalactiae* (13.1;Butyrivibriocin,
912 219.1;Streptolysin, and 58.1;Macedocin). Two gene clusters were identified in *S.*
913 *equinus* (42.2;Bovicin_255_peptide and 84.1;Nukacin_A_(NukacinISK-1)) and *S.*
914 *haemolyticus* (315.1;Auto_Inducing_Peptide_II and Sactipeptides). Only one gene
915 cluster was predicted for *S. gallolyticus* (LAPs), *S. bovis* (159.2;Pediocin), *S. borealis*
916 (Sactipeptides), and *L. mesenteroides* (97.2;Enterocin_X_chain_beta).
917 Genome mining using AntiSMASH identified 322 biosynthetic gene clusters
918 (BGCs) across all analyzed MAGs. These BGCs were classified into 26 distinct types
919 of biosynthetic pathways associated with the production of secondary/specialized
920 metabolites. The ten most abundant pathways were non-ribosomal peptide synthetase
921 (NRPS) (12.7%), terpene (11.1%), non-ribosomal peptide siderophore (NI-
922 siderophore) (10.8%), type III polyketide synthase (T3PKS) (10.2%), ribosomally
923 synthesized and post-translationally modified peptides (RiPP-like) (9.6%), cyclic
924 lactone autoinducer (9.3%), opine-like metallophore (5.4%), NRPS-like (4.8%),
925 aminopolycarboxylic acid (3.9%), and type I polyketide synthase (T1PKS) (3.0%).
926 When the analysis was stratified by bacterial species and limited to hits with
927 similarity associated with known clusters, several species-specific putative
928 metabolites were identified **(Figure 15B)**. In *C. bovis*, the annotated metabolites
929 included Œμ-Poly-L-lysine, 5-isoprenylindole-3-carboxylate β-D-glycosyl ester (here
930 abbreviated as 5-IPI-β-D-Gly), SF2575, and rhizomide A/B/C. For *C. kroppenstedtii*,
931 the identified metabolites were phenazine SA/SB/SC, [S,S]-ethylenediamine-
932 disuccinate ([S,S]-EDDS), trichrysobactin/cyclic
933 trichrysobactin/chrysobactin/dichrysobactin (here abbreviated as
934 TCB/cTCB/CB/DCB), griseobactin, and ε-Poly-L-lysine. *C. urogenitale* showed the
935 presence of SF2575 and carotenoid. Metabolites detected in *Dermatophilus* spp.
936 included carotenoid, hedamycin, microtermolide A, rakicidin A/B, and ε-Poly-L-lysine.
937 The compound ε-Poly-L-lysine was the sole metabolite identified in *K. atrinae*. For *S.*

938 *bovis*, [S,S]-EDDS was annotated, while macedocin was specific to *S. dysgalactiae*.
939 *S. equinus* displayed streptococcin A-FF22, whereas mutanocyclin/leuvalin/tyrvalin
940 and streptide were annotated for *S. gallolyticus*. In *S. uberis*, the detected putative
941 metabolites included uberolysin, ubericin K, and iturin. *S. aureus* was associated with
942 aureusimine A/B/C, hyicin 3682, staphyloferrin A, staphyloferrin B, staphylopine, and
943 kijanimicin. For *S. borealis*, staphyloferrin A was identified, while *S. chromogenes*
944 yielded staphyloferrin A, staphylopine, and kijanimicin. Staphyloferrin A and
945 staphylopine were predicted in *S. epidermidis*, and *S. haemolyticus* showed
946 staphyloferrin A. Lastly, in *S. hominis*, kijanimicin and staphyloferrin A were annotated.
947 The AntiSMASH analysis revealed a group of bioactive compounds shared among
948 several Staphylococci species, including staphyloferrin A, staphylopine, and
949 kijanimicin. In contrast, unique metabolic signatures were observed in specific
950 species, such as *S. aureus*, *C. kroppenstedtii*, *S. uberis*, and *C. bovis*. These findings
951 suggest a potential common production of certain bioactive compounds, here
952 predominantly identified among Staphylococci, while other microorganisms, such as
953 *C. kroppenstedtii*, appear to exhibit a more species-specific metabolite spectrum.
954

955 **4. Discussion**

956 Bovine mastitis is widely recognized as the most economically significant disease
957 affecting dairy cattle [66]. Its multifactorial nature poses substantial challenges to
958 effective management through zootechnical and veterinary interventions[67]. In Norway,
959 where the disease is relatively well-controlled and possesses one of the lowest
960 antimicrobial use in livestock in the world, studies on microbial ecology offer valuable
961 insights into potential taxa associated with udder health, contributing to advancements
962 in prevention and management strategies [68].

963 Somatic cell count (SCC) is globally used as a proxy for diagnosing intramammary
964 infections and monitoring mastitis control throughout lactation. Elevated SCC values
965 are typically associated with a recent or ongoing infection, often confirmed by a
966 positive bacterial culture or, more recently, by PCR targeting specific taxa [69]. The
967 beginning and end of lactation are generally considered critical periods for udder
968 health due to increased susceptibility to infections. In this study, the lactation period
969 significantly influenced SCC levels, with lower counts observed during early and mid-
970 lactation, while higher counts were recorded near the dry-off period in 2022 and late
971 lactation in 2023, which is in agreement with the scientific literature [70].

Mastitis-causing pathogens are a primary driver of elevated SCC in the udder, with various bacterial species implicated. These pathogens are conventionally classified into two main groups: major and minor mastitis-causing pathogens [3]. In our study, major pathogens such as *S. aureus*, *Enterococcus* spp., and environmental streptococci were prominent, whereas minor pathogens like *S. chromogenes*, *S. haemolyticus*, *S. epidermidis*, and *C. bovis* were less frequently reported. Apart from the elevated culturing and detection of *E. faecalis* in milk samples from the mid-lactation period of 2023, our findings align with a recent prevalence study of udder pathogens in Norwegian dairy cows between 2019 and 2020 [71]. The high occurrence of *E. faecalis* coincides with the period when animals transition from freestalls to pasture. As reported elsewhere, *Enterococcus* is considered both a commensal organism and an opportunistic pathogen, possessing a remarkable ability to survive under adverse conditions. It is commonly found as part of the normal gut microbiota in animals, which may facilitate its dissemination under specific environmental or management practices [72].

Bacteriological culturing of milk samples aids in guiding mastitis treatment decisions by identifying the causative bacteria of intramammary infections. However, as not all mastitis-causing bacteria require treatment, this approach, from a microbial ecology perspective, overlooks the presence of a substantial fraction of microorganisms within a given sample due to the fastidious nature of some microbial taxa [73]. In our study, two groups of samples classified as negative (i.e., absence of mastitis-causing pathogens according to the gold standard) were identified. The first group (n = 72) exhibited high colony counts (30–300 CFUs) but had somatic cell counts (SCC) below 100,000 cells/mL. The second group (n = 57) also showed high colony counts but had a median SCC of 2.9×10^5 cells/mL. Although these microorganisms are not routinely identified, they may contribute to elevated SCC levels. For example, the genus *Lactococcus* has emerged as a potential group of emerging mastitis pathogens, yet it is rarely documented in conventional culturing reports [74,75].

In an effort to overcome the limitations of culture-dependent methods and complement them, numerous studies have utilized metataxonomic analysis to explore the microbial composition of bovine milk in both health and disease [21,76–78]. In this longitudinal study, we conducted a comprehensive investigation of the microbial composition at the quarter level in over 300 milk samples, aiming to uncover microbial

succession patterns associated with bovine mastitis. The lower diversity indices observed in samples with high somatic cell counts are a common finding in ecological studies of the bovine mammary gland. This reduction in diversity reflects the dominance of specific microbial taxa, which disrupt the homeostasis of the milk microbial community, resulting in a dysbiotic state within the affected quarter [19,23,76].

In this study, we observed that both the composition and relative abundances of bacterial taxa at the phylum and genus levels varied across lactation periods. Specifically, an enrichment of *Proteobacteria* during the early lactation period (EL-2023) was associated with a lower frequency of samples exhibiting high somatic cell counts. This phylum is linked to cases of clinical mastitis, such as those caused by *Escherichia coli*. However, *Proteobacteria* are also frequently observed at high relative abundances in healthy milk samples. In contrast, cases of subclinical mastitis are more commonly associated with the phylum *Firmicutes*, including bacterial species such as *Staphylococcus* spp. [17]. The increase in the phylum *Actinobacteriota* in the milk microbiota during the mid- and late-lactation periods may be attributed to dietary changes and seasonal factors, as it aligns with rising temperatures and the transition to a pasture-based system. Recent studies suggest that the ruminal microbiota and climatic conditions can significantly impact udder health [79,80].

The microbial diversity observed at the genus level in this study aligns with findings from previous research conducted on the same herd and globally, with taxa such as *Staphylococcus*, *Corynebacterium*, *Streptococcus*, *Enterococcus*, *Aerococcus*, *Kocuria* and *Romboutsia* among the most prevalent ones [28,81,82]. To further investigate the structure and dynamics of the milk microbiome community, we applied a dimensionality reduction method commonly used in human microbiome studies to explore associations with parameters of interest [83]. This approach provided valuable insights, particularly regarding sample classification, bridging the gap between culturomics and culture-independent methods. By combining these approaches, we increased the detection of specific pathogens in samples initially classified as negative, highlighting the potential of this methodology to enhance microbial detection and classification. Conventional bacteriological examination remains the most widely used method for diagnosing bovine intramammary infections, however, it is estimated that over 30% of milk samples yield negative culture results or fail to identify the causative agent [84,85]. According to Smistad et al. (2023) [71], the proportion of samples classified as culture-negative in subclinical mastitis cases can exceed 20% in Norway,

while culture-negative samples from clinical mastitis cases account for approximately 10%. Overall, culture-negative mastitis is often associated with bacterial species present below the detection threshold of culture methods or with fastidious microorganisms [11,86].

The application of dimensionality reduction enabled the evaluation of the relative abundance of key bacterial genera within each cluster, offering valuable insights into their prevalence across various samples. *Staphylococcus*, *Corynebacterium*, and *Aerococcus* were consistently distributed among the clusters, bringing out their widespread occurrence and variability in the analyzed milk samples. This observation confirms the significant prevalence of these taxa in the milk microbiome, irrespective of whether the sample originates from a healthy udder or a mastitis case. *Staphylococcus* species stand out for their high contagious potential within herds and can be present in both healthy and mastitic quarters, with their prevalence being highly strain-dependent. Although the specific role of *Corynebacterium* spp. in udder health remains unclear [87], it has been hypothesized that natural infections may confer protection against major mastitis pathogens, such as *S. aureus*. This is consistent with our findings and may be explained by the production of antimicrobial compounds or modulation of the udder immune system, as reported in previous studies [88–90]. 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 [26,91].

Within the bacterial genera evaluated in this study, the remarkable diversity of *Corynebacterium* and *Staphylococcus* species observed through taxonomic analysis stands out, aligning with findings from the scientific literature. According to Seshadri et al. (2022) [90], only 30%–50% of the projected actinobacterial phylogenetic diversity has genomic representation via isolates or metagenome-assembled genomes. This limited representation may account for the high number of amplicon sequence variants annotated as uncultured bacteria within the *Corynebacterium* genus in our study, highlighting the need for further metagenomic research utilizing shotgun sequencing to improve taxonomic resolution. In a study designed to survey and compare the microbial diversity of external and teat canal epithelial microbiomes using different sequencing approaches, Dean et al. (2021)[92] reported that the most diverse operational taxonomic units across all sampling locations were assigned to the *Corynebacterium* genus. Among these, *C. efficiens*, *C. marinum*, *C. maris*, and *C.*

1074 *xerosis* constituted the majority of species-level identifications (mean relative
1075 abundance exceeding 1%). Except for *C. bovis*, many *Corynebacterium* spp. present
1076 in milk samples can also be isolated from the housing environment and milking-related
1077 niches [93], reflecting the ubiquity of these microorganisms in the dairy environment.

1078 The majority of the ASVs annotated within the *Staphylococcus* genus belong to the
1079 Non-aureus Staphylococci and Mammaliicocci (NASM) group, one of the most
1080 frequently isolated bacterial groups from milk samples [94]. In this study, NASM species
1081 such as *S. borealis*, *S. chromogenes*, *S. hominis*, *S. epidermidis*, *S. haemolyticus*, *S.*
1082 *xylosus*, *S. saprophyticus*, and *S. simulans* were identified across the milk samples.
1083 Compared to coagulase-positive staphylococci, these species generally exhibit lower
1084 virulence but can occasionally cause mastitis [95,96]. Moreover, they are often
1085 overlooked in routine laboratory diagnostics, as recently reported for *S. borealis* and
1086 *S. rostri* [97]. Nonetheless, NASM has garnered increasing attention due to its genetic
1087 diversity and its ability to antagonize colonization by *S. aureus* [98,99]. This antagonistic
1088 effect is attributed to the production of various antibacterial molecules, such as
1089 bacteriocins, which may play a critical role in shaping microbial dynamics and udder
1090 health [100].

1091 We observed that the composition of milk microbiota varied both at the cow level
1092 and the quarter level. When an intramammary infection was established by
1093 *Staphylococcus*, a persistent dysbiosis was detected in subsequent samples, a
1094 phenomenon not observed for *Streptococcus*. While a considerable number of cross-
1095 sectional studies have examined the bovine mammary gland [68,77,101], there is limited
1096 knowledge regarding the dynamics of milk microbiota at the quarter level in
1097 longitudinal studies over a complete lactation cycle. Our findings align with the
1098 scientific literature, which indicates that once *Staphylococcus* establishes dominance
1099 within a quarter, it becomes challenging for the cow to eliminate it [21]. This aligns with
1100 the wide variety of mechanisms deployed by *S. aureus*, such as biofilm formation, the
1101 emergence of small colony variants, and its capability to invade professional and
1102 nonprofessional cells. These mechanisms protect *S. aureus* from the cow's innate and
1103 adaptive immune responses as well as from antimicrobial treatments [102]. In contrast,
1104 our findings regarding *Streptococcus* are consistent with a recent study by Urrutia-
1105 Angulo et a. (2024) [19] which reported a microbiome profile suggesting full recovery
1106 from dysbiosis caused by *Streptococcus* one and three months after the diagnosis of
1107 clinical mastitis. Based on the ASVs identified for the genus *Streptococcus* in this

1108 study, there is a high prevalence of sequences attributed to *S. uberis*, a primarily
1109 environmental pathogen [103]. This bacterium can cause either persistent subclinical
1110 inflammation or severe inflammation associated with transient mastitis, depending on
1111 the strain involved [104]. These results highlight distinct microbiota dynamics depending
1112 on the causative pathogen and underscore that *S. aureus* remains a major concern.

1113 Through shotgun sequencing and genome-centric metagenomics, we aimed to
1114 gain deeper insights into the functional potential of the microbial community and to
1115 reconstruct genomes to better understand microbial interactions within the udder.
1116 Despite progress in this field, a significant gap remains in the scientific literature, which
1117 can be addressed through more studies incorporating shotgun sequencing to
1118 complement metataxonomic analyses at the quarter level. One major challenge is the
1119 contamination of eukaryotic DNA from mammalian cells during metagenomic DNA
1120 extraction. Combined with the low microbial biomass typically present in milk samples,
1121 this reduces the proportion of microbial reads, resulting in poor resolution of the
1122 microbial community. Our research group recently developed a method to address this
1123 issue [23], although further optimization is required for milk samples with low somatic
1124 cell counts.

1125 The microbial diversity observed from the reconstructed microbial genomes in the
1126 metagenomic sequencing data included the most prevalent bacterial genera identified
1127 through amplicon sequencing (*Staphylococcus*, *Corynebacterium*, *Streptococcus*,
1128 *Enterococcus*, *Aerococcus*, *Kocuria*, and *Romboutsia*). Notably, a diverse array of
1129 draft genomes was detected, particularly within the genera *Staphylococcus* and
1130 *Corynebacterium*. Of particular interest is the high abundance of the MAG *C.*
1131 *kroppenstedtii* HC-IA_H069.1. This species belongs to the *C. kroppenstedtii*-like
1132 group, which includes *C. parakroppenstedtii* and *C. pseudokroppenstedtii* [105]. Species
1133 of this group have gained recognition for their involvement in breast abscesses and
1134 granulomatous mastitis in women [106]. However, the role of *C. kroppenstedtii*-like
1135 strains in bovine udder health and disease remains unclear and warrants further
1136 investigation. Species of the genus *Corynebacterium* are well-known for their ability to
1137 produce secondary metabolites and are routinely used as cell factories, exemplified
1138 by *C. glutamicum* [107]. A recent study conducted by Liu et at. (2024) [108] reports the
1139 discovery of a novel glycolipid involved in the development of granulomatous lobular
1140 mastitis. This pathogenic factor called corynekropbactin is secreted by *C.*
1141 *parakroppenstedtii* and may chelate iron, cause the death of mammary cells and other

1142 mammary-gland-colonizing bacteria, and increase the levels of inflammatory
1143 cytokines. Studies exploring the adaptation of this species to the bovine mammary
1144 gland, as well as comparative genomics analyses with human isolates, are needed to
1145 better understand its potential implications in the bovine mammary gland.

1146 The MAGs *Facklamia* bacterium HC-CA013, *Kocuria atrinae* HC-CA018, *S.*
1147 *equinus* HC-IA_356.3, *Staphylococcus* bacterium HC-IA_319.2, *S. chromogenes* HC-
1148 IA_H042.1, *Peptoniphilus* bacterium HC-CA024, *Corynebacterium* bacterium HC-
1149 CA026, and *Corynebacterium* bacterium HC-IA_212.3 were found at higher coverage
1150 in milk samples with low somatic cell counts. The genera *Facklamia* and *Kocuria* have
1151 been reported as part of the milk microbiota of healthy bovine quarters and human
1152 precolostrum [109,110]. *Facklamia* belongs to the family *Aerococcaceae*, which comprises
1153 nine genera; recently, some species of *Facklamia* were reclassified, such as *F.*
1154 *tabacinasalis*, now *Ruoffia tabacinasalis* [111]. Meanwhile, *Kocuria* is a member of the
1155 family *Micrococcaceae*, and the genus was established by subdividing the genus
1156 *Micrococcus* [112]. Metabolically, *Aerococcaceae* are classified as lactic acid bacteria
1157 and can produce lactic acid as the primary product of carbohydrate fermentation [113], a
1158 pathway highlighted in this study by the enrichment of COG category G (Carbohydrate
1159 transport and metabolism). Their metabolism reflects adaptation to low-oxygen,
1160 nutrient-rich environments, such as milk. Although these microorganisms are well-
1161 suited to milk, studies exploring the crosstalk between the identified MAGs and their
1162 roles in bovine udder health remain largely unexplored.

1163 A comprehensive understanding of the genetic variation among mastitis-causing
1164 pathogens in bovine populations, including temporal trends, is essential for developing
1165 effective mastitis control strategies. This is particularly important because the severity
1166 of the disease and its impact on milk production are, at least in part, influenced by the
1167 genetic and phenotypic characteristics of specific pathogenic microorganisms. In this
1168 study, we conducted an *in silico* multilocus sequence typing (MLST) analysis on
1169 metagenomic samples and the metagenome-assembled genomes (MAGs) derived
1170 from these datasets. Among the lineages identified, all *S. aureus* sequences were
1171 associated with the ST133, which has been recognized as a globally distributed
1172 sequence type predominantly associated with livestock such as large and small
1173 ruminants [114,115]. Regarding *E. faecalis*, all sequences analyzed in this study were
1174 assigned to clonal complex 40 (ST40), the most prevalent clonal type of this species.
1175 ST40 is globally distributed, with isolates originating from various sources, including

1176 livestock, humans, and environmental reservoirs [116]. This lineage is of particular
1177 significance due to its widespread presence in clinical and non-clinical settings. It is
1178 frequently isolated from both commensal and pathogenic contexts, demonstrating its
1179 adaptability across different environments. *S. dysgalactiae* ST302 has been previously
1180 identified in cows, with evidence suggesting it is well-adapted to this host [117]. In
1181 contrast, limited information is available in the scientific literature regarding
1182 *Streptococcus uberis* ST1409. However, it appears closely related to sequence type
1183 1405 and has been reported to be sensitive to penicillin, tetracycline, and erythromycin
1184 [118]. The ST100 identified in *S. epidermidis* has been described as having an animal
1185 origin and is reported to cause mastitis in both ewes and bovines [119]. In contrast,
1186 limited information is available regarding ST575. However, one study identified an
1187 antimicrobial compound known as epifadin, produced by nasal *S. epidermidis* strain
1188 IVK83 belonging to the ST575, demonstrating the ability to eliminate *S. aureus* [120]. For
1189 *S. chromogenes*, ST1 has been reported as one of the most predominant sequence
1190 types in a study investigating differences in the occurrence of intramammary infections
1191 in freshly calved heifers in Sweden [121]. In contrast, for ST59 and ST104, further
1192 research is needed to elucidate their clinical significance, antimicrobial resistance
1193 profiles, and associated virulence factors.
1194 To successfully colonize a host, bacteria possess an array of proteins and other
1195 components collectively known as virulence factors [122], although this terminology for
1196 commensal bacteria is not appropriate and should be replaced by a more appropriate
1197 term such as “niche factors” [123]. Interestingly, many of these factors are not exclusively
1198 specialized for host-pathogen interactions but play broader roles in adapting to diverse
1199 ecological niches and for cooperation in complex biological ecosystems [124]. The
1200 analysis of virulence factors across bacterial taxa in milk samples revealed distinct
1201 patterns. According to Tauch & Burkovski (2015) [125], many members within the genus
1202 *Corynebacterium* developed into rare pathogens “by chance” and most of the species
1203 share a more general adhesion mechanism. In our study, regulatory factors such as
1204 iron-dependent repressor (IdeR) and PhoP seem to play an important role in niche
1205 adaptation. IdeR in *Mycobacterium tuberculosis* and its homolog, the diphtheria toxin
1206 repressor (DtxR) in *Corynebacterium diphtheriae*, are crucial iron-dependent
1207 transcriptional regulators [126,127]. IdeR controls iron acquisition, storage, and virulence-
1208 related genes in *M. tuberculosis* [128], whereas PhoP gene is the response regulator of
1209 the PhoPR two-component systems, which plays a crucial role in the virulence of *C.*

1210 *pseudotuberculosis* [129]. In *E. faecalis*, adherence genes constituted the largest fraction
1211 of virulence factors in this study. Key colonization factors included the collagen
1212 adhesin Ace, the endocarditis antigen EfaA, and the endocarditis- and biofilm-
1213 associated pili (Ebp), all of which are commonly present in *E. faecalis* isolates [130,131].
1214 The Esp surface protein is more frequently found in isolates from urinary tract
1215 infections and bacteremia [131]. Ebp pili play a major role in adherence to fibrinogen and
1216 a minor role in collagen binding, whereas Ace serves as the primary collagen adhesin
1217 [132]. Species of *Staphylococcus* employ diverse immune evasion strategies to persist
1218 within host organisms. In this study, immune modulation was identified as the
1219 predominant category of virulence factors among *Staphylococcus* species, particularly
1220 genes involved in capsule formation (*cap8*), *adsA*, and *sbi*. The *cap8* gene cluster,
1221 consisting of 16 open reading frames (*cap8A-P*), is responsible for the synthesis of
1222 type 8 capsular polysaccharides, with 11 genes directly contributing to its production
1223 [133]. Capsular polysaccharides 8, which is predominant among *S. aureus* isolates from
1224 humans, can elicit a robust inflammatory response. However, these capsular types are
1225 rapidly eliminated compared to capsule-negative *S. aureus*, which are more efficiently
1226 internalized by mammary gland epithelial cells and capable of establishing chronic
1227 infections in the host [134]. Other virulence factors, such as staphylococcal protein A
1228 (SpA), staphylococcal binder of immunoglobulin (Sbi), and adenosine synthase A
1229 (AdsA), also play crucial roles in immune evasion [135,136]. Sbi, which is both secreted
1230 and cell-envelope associated, enhances bacterial survival in human blood and
1231 prevents neutrophil-mediated opsonophagocytosis and modulates B-cell activation
1232 [137,138]. Collectively, these mechanisms enable *S. aureus* to evade host immune
1233 defenses effectively, facilitating persistent and recurrent infections. *Streptococcus*
1234 species possess various virulence factors that contribute to their pathogenicity. Biofilm
1235 production is a prevalent trait, observed in over 70% of *Streptococcus* isolates [139]. *S.*
1236 *agalactiae* strains frequently carry genes like *cyl* genes, *cfa*/*cfb*, and *hylB* [140]. The
1237 lactoferrin-binding protein is a crucial virulence factor found in *S. dysgalactiae*, *S.*
1238 *uberis* and *S. agalactiae*, with differing effects on cell invasion [141–143]. Notably, *S.*
1239 *agalactiae* strains carrying the *lmb* gene showed enhanced invasion ability and were
1240 associated with reduced antibiotic treatment efficacy [140].
1241 Many of the identified virulence factors are directly linked to the high recalcitrance
1242 associated with antibiotic use in dairy farming. More recently, shotgun metagenomic
1243 sequencing has shown to be a valuable tool to quantify the dynamics of AMR

transmission at the human–livestock interface [144], as well as implications for antimicrobial stewardship in livestock [145]. Norway stands out for its notably low rates of antibiotic use and bovine mastitis treatment globally, with benzylpenicillin procaine being the first-line therapy for mastitis [71]. Currently, antimicrobial treatments are primarily reserved for moderate to severe cases of mastitis during lactation and for addressing subclinical mastitis caused by *S. aureus*, *S. dysgalactiae*, *S. uberis*, or *S. agalactiae* at dry-off [71]. *Corynebacterium* species exhibit widespread antibiotic resistance, particularly to macrolides, lincosamides, and streptogramins [146,147]. Additionally, resistance to β-lactams, clindamycin, erythromycin, azithromycin, ciprofloxacin, and gentamicin is also common among various *Corynebacterium* species. Most strains remain susceptible to vancomycin, minocycline, and linezolid [148]. Given the high prevalence and diversity of *Corynebacterium* species in milk samples, this may be linked to their common resistance to beta-lactams [149,150], potentially contributing to the persistence of these microorganisms within the herd enrolled in this study. Of food safety concern, *S. aureus* has developed resistance to various antibiotics through mechanisms including enzymatic inactivation, target alteration, antibiotic trapping, and efflux pumps [151]. Exposure to biocides and dyes can lead to overexpression of efflux pump genes like *mepA*, *mdeA*, *norA*, and *norC*, contributing to multidrug resistance [152]. Staphylococci of animal origin possess numerous resistance genes conferring resistance to multiple antibiotic classes, including fluoroquinolones, tetracyclines, aminoglycosides, and phenicols. These genes are often carried on mobile genetic elements, facilitating their spread among bacterial populations [153]. Species of *Streptococcus* have also developed resistance to multiple classes of antibiotics in both humans and various animal species, presenting significant challenges for effective treatment [154,155]. Key resistance mechanisms include genetic modifications, alterations in target sites, and the activity of efflux pumps. [156]. The dissemination of resistance genes among streptococci is primarily mediated by transposons and integrative conjugative elements [155]. While streptococci are typically regarded as sensitive to benzylpenicillin and are not routinely tested in Norway, there have been reports of structural modifications in penicillin-binding proteins caused by point mutations in *S. pneumoniae*, *S. agalactiae*, and *S. dysgalactiae* subsp. *equisimilis*. These changes have been linked to reduced susceptibility to β-lactam antibiotics [157–159]. Macrolide resistance in *Streptococcus* arises through three primary mechanisms: ribosomal modifications (both post- and

pre-transcriptional), active antibiotic expulsion via efflux pumps, and target protection [160], and studies have frequently linked this resistance to the presence of the *ermB* gene in *S. agalactiae*, *S. dysgalactiae*, and *S. uberis* [161,162]. Although fluoroquinolone resistance remains low, it is increasing due to genomic mutations and plasmid-encoded genes, while tetracycline resistance is mediated by ribosomal protection genes, with resistance determinants for macrolides and tetracyclines often co-occurring on mobile genetic elements [163,164]. *E. faecalis* has been identified as a significant pathogen in bovine mastitis and isolates can exhibit high resistance to tetracycline and erythromycin, with resistance rates of 78.6-87.7% and 28.7-79.0%, respectively [165,166]. The most prevalent resistance genes include *tetK*, *tetL*, and *tetM* for tetracycline, as well as *ermB* and *ermC* for erythromycin, although the latter were not identified in our study [165,166].

Bacteriocins produced by LAB play a significant role in shaping microbial communities and maintaining microbiome homeostasis [167]. In the context of bovine mastitis, LAB isolated from the mammary glands of cows have demonstrated potential as probiotic strains for prevention [168]. However, it remains debatable whether the use of LAB aligns with the biology of the mammary gland, as both the bacterial strain and the site of application are critical factors for ensuring effective action without triggering a strong pro-inflammatory response [169]. Given that coagulase-negative staphylococci are more naturally adapted to the mammary gland, they could represent a viable alternative for the prevention or treatment of mastitis. In this study, several gene clusters were identified across different bacterial species, reflecting the diverse range of molecules utilized by microorganisms in the mammary gland for both competition and pathogenesis. *S. aureus* can express lantibiotics like BsaA2, which provides a competitive advantage in community-acquired strains [170]. Nukacins are lantibiotics produced by *Staphylococci* and show promising potential for controlling bovine mastitis [171,172]. *Staphylococcus* species also utilize peptide-based quorum sensing systems, such as the accessory gene regulator (*agr*) system, which employs auto-inducing peptides (AIPs) to regulate various cellular processes [173]. Enterolysin A, a bacteriocin produced by *E. faecalis*, exhibits antimicrobial activity against Gram-positive bacteria, such as *S. aureus* and *Listeria monocytogenes* [174,175]. Another toxin, the *E. faecali*s cytolysin, is a two-peptide lytic system active against both prokaryotic and eukaryotic, belongs to the lantibiotic family, and requires two post-translationally modified peptides for its activity cells [176,177]. The expression of the cytolysin is carefully

controlled by a quorum-sensing mechanism, potentially balancing its defensive function against predators with maintaining stealth [177,178]. *S. agalactiae* strains isolated from bovine mastitis cases in Brazil demonstrated antimicrobial activity against other *Streptococcus* species, with the bacteriocin zoocin A identified in their genomes [179]. Similarly, bacteriocin-producing strains of *Lactobacillus fermentum* and *S. bovis* isolated from raw milk in Thailand exhibited inhibitory effects against *S. dysgalactiae* [180].

Staphylococcus species produce various secondary metabolites that play crucial roles in microbiome interactions and nutrient acquisition. Staphyloferrin A, a highly hydrophilic siderophore, was isolated from *Staphylococcus hyicus* DSM 20459 and found to be active in 37 other staphylococci [181]. This compound, along with staphyloferrin B, helps scavenge scarce iron in the environment. Another metallophore, staphylopine, is produced by staphylococci to acquire transition metal ions [182]. Siderophores play a pivotal role in iron acquisition for bacterial pathogens associated with bovine mastitis, enabling their survival in iron-restricted environments. For example, *S. aureus* can utilize a variety of exogenous iron sources, including siderophores such as ferrichrome and desferrioxamine, as well as heme-containing proteins [183]. The ability to acquire iron through these mechanisms significantly enhances the competitive fitness of *Staphylococcus* species across diverse ecological niches, contributing to their adaptability and pathogenic potential. A genome-wide CRISPR interference sequencing (CRISPRi-seq) screen carried out by Mårli et al. (2024) [184] identified key fitness determinants essential for *S. aureus* growth and survival in milk. The study revealed several genes with differential fitness, highlighting specific adaptations for thriving in this environment, with metal acquisition emerging as particularly critical. In this study, ε-Poly-l-lysine, a bioactive secondary metabolite with potent antimicrobial properties, was identified across various *Corynebacterium* species. This compound is widely utilized in several Asian countries as a food preservative and is also implicated in shaping the microbial dynamics of cheese and skin ecosystems [185]. The indole-type compound 5-isoprenylindole-3-carboxylate β-D-glycosyl ester has its bioactivity exploited in developing new drugs and it is widely found in actinobacteria [186]. The compound [S,S]-ethylenediamine-disuccinate ([S,S]-EDDS) was predicted in *C. kroppenstedtii*. According to Costa et a. 2008[187] , metal complexes of the biodegradable ligand EDDS exhibit antimicrobial activity against fungi and bacteria. Another two noteworthy compounds predicted for *C. kroppenstedtii*

using AntiSMASH analysis are a triscatecholamide siderophore and a catechol-peptide structure named griseobactin. These compounds may play a crucial role in iron acquisition under low-iron conditions in the bovine mammary gland, where iron availability varies depending on udder health, lactation stage, and daily milk production [188,189]. Phenazines, versatile secondary metabolites predicted for *C. kroppenstedtii*, are commonly produced by various bacteria, including *Pseudomonas*, *Burkholderia*, and *Streptomyces* species [190,191]. These compounds exhibit broad-spectrum antibiotic activity and play a significant role in enhancing the ecological fitness and pathogenic potential of the producing strains [192,193]. Overall, understanding the potential production of bacteriocins and bioactive compounds may lead to novel anti-infective strategies and microbiome modulation approaches to tackle bovine mastitis.

5. Conclusion

Bovine mastitis continues to present a significant challenge to global dairy production, highlighting the need for innovative strategies to manage this complex, multi-etiological disease. This longitudinal study offers valuable insights into the udder microbiome dynamics across lactation stages using both metataxonomic and shotgun metagenomic approaches. Our results reveal substantial variations in somatic cell count and microbiota composition, with dominant species identified in high SCC samples. Through dimensionality reduction, we were able to identify health-promoting bacteria and examine microbial dynamics at the quarter level, both before and after pathogen establishment, demonstrating that *Staphylococcus* remains a primary concern in mastitis pathogenesis. Metagenomic analyses provided a deeper understanding by uncovering pathogen-specific metabolic pathways and reconstructing 142 metagenome-assembled genomes. These findings offer crucial insights into pathogen adaptation, virulence, and resistance mechanisms. The protective roles of *Aerococcus*, *Kocuria*, and Coryneform bacteria warrant further investigation as potential health-promoting bacteria. Additionally, the identification of bacteriocin and biosynthetic gene clusters opens potential pathways for the development of targeted antimicrobial therapies. In conclusion, this study enhances our understanding of the bovine udder microbiome and its role in mastitis pathophysiology. These insights provide a foundation for microbiome-informed strategies that can improve the health and productivity of dairy herds.

Acknowledgments

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.

Declarations of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI in scientific writing

No generative AI has been used in this article.

References

1. Morales-Ubaldo, A. L. *et al.* Bovine mastitis, a worldwide impact disease: Prevalence, antimicrobial resistance, and viable alternative approaches. *Veterinary and Animal Science* vol. 21 Preprint at https://doi.org/10.1016/j.vas.2023.100306 (2023).
2. Diseases of the Mammary Gland. *Vet Med* 1904–2001 (2017) doi:10.1016/B978-0-7020-5246-0.00020-6.
3. Cobirka, M., Tancin, V. & Slama, P. Epidemiology and Classification of Mastitis. *Animals (Basel)* **10**, 2212 (2020).
4. Milk — Enumeration of somatic cells — Part 1: Microscopic method (Reference method). Preprint at https://www.iso.org/standard/40446.html (2008).
5. Adkins, P. R. F. & Middleton, J. R. Methods for Diagnosing Mastitis. *Veterinary Clinics of North America: Food Animal Practice* **34**, 479–491 (2018).
6. 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, Vol. 11, Page 271* **11**, 271 (2022).
7. 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).

1414 8. Bradley, A. J., Leach, K. A., Breen, J. E., Green, L. E. & Green, M. J. Survey of
1415 the incidence and aetiology of mastitis on dairy farms in England and Wales.
1416 *Veterinary Record* **160**, 253–258 (2007).
1417 9. Langhorne, C. *et al.* Bacterial culture and susceptibility test results for clinical
1418 mastitis samples from Australia's subtropical dairy region. *J Dairy Sci* **107**, 1151–1163
1419 (2024).
1420 10. Yadav, D. *et al.* Next-Generation sequencing transforming clinical practice and
1421 precision medicine. *Clinica Chimica Acta* **551**, 117568 (2023).
1422 11. Oikonomou, G. *et al.* Microbiota of Cow's Milk; Distinguishing Healthy, Sub-
1423 Clinically and Clinically Diseased Quarters. *PLoS One* **9**, e85904 (2014).
1424 12. Sakwinska, O. & Bosco, N. Host microbe interactions in the lactating mammary
1425 gland. *Front Microbiol* **10**, 448318 (2019).
1426 13. Cheema, S. K., Li, R. & Cameron, S. J. S. Culturomics as a tool to better
1427 understand the human milk microbiota and host–microbiota interactions. *Microbiota*
1428 *and Host* **1**, (2023).
1429 14. Oikonomou, G., Machado, V. S., Santisteban, C., Schukken, Y. H. & Bicalho,
1430 R. C. Microbial Diversity of Bovine Mastitic Milk as Described by Pyrosequencing of
1431 Metagenomic 16s rDNA. *PLoS One* **7**, e47671 (2012).
1432 15. Srithanasuwan, A., Pangprasit, N., Mektrirat, R., Suriyasathaporn, W. &
1433 Chuammitri, P. Divergent Immune Responses to Minor Bovine Mastitis-Causing
1434 Pathogens. *Vet Sci* **11**, 262 (2024).
1435 16. Hoque, M. N. *et al.* Metagenomic deep sequencing reveals association of
1436 microbiome signature with functional biases in bovine mastitis. *Sci Rep* **9**, 13536
1437 (2019).
1438 17. Hoque, M. N. *et al.* Microbiome dynamics and genomic determinants of bovine
1439 mastitis. *Genomics* **112**, 5188–5203 (2020).
1440 18. Oikonomou, G. *et al.* Milk Microbiota: What Are We Exactly Talking About?
1441 *Frontiers in Microbiology* vol. 11 Preprint at https://doi.org/10.3389/fmicb.2020.00060
1442 (2020).
1443 19. Urrutia-Angulo, L. *et al.* Unravelling the complexity of bovine milk microbiome:
1444 insights into mastitis through enterotyping using full-length 16S-metabarcoding. *Anim*
1445 *Microbiome* **6**, 1–14 (2024).

1446 20. Bhar, S. & Bose, T. Evidence from metagenomic study indicate that subclinical
1447 mastitis may have a different pathological origin than clinical mastitis. *bioRxiv*
1448 2024.05.24.595548 (2024) doi:10.1101/2024.05.24.595548.
1449 21. Winther, A. R. *et al.* Longitudinal dynamics of the bovine udder microbiota. *Anim*
1450 *Microbiome* **4**, 26 (2022).
1451 22. Winther, A. R., da Silva Duarte, V. & Porcellato, D. Metataxonomic analysis and
1452 host proteome response in dairy cows with high and low somatic cell count: a quarter
1453 level investigation. *Vet Res* **54**, (2023).
1454 23. Duarte, V. da S. & Porcellato, D. Host DNA depletion methods and genome-
1455 centric metagenomics of bovine hindmilk microbiome. *mSphere* **9**, (2024).
1456 24. Farkaš, V. *et al.* Biomarkers for subclinical bovine mastitis: a high throughput
1457 TMT-based proteomic investigation. *Vet Res Commun* **48**, 2069–2082 (2024).
1458 25. Satheesan, L. *et al.* Comparative Profiling of Milk Somatic Cells Proteomes
1459 Revealed Key Players in Mammary Immune Mechanisms During Mastitis in Tropical
1460 Sahiwal (Bos indicus) Cows. *Proteomics Clin Appl* **18**, e202400054 (2024).
1461 26. Jung, D., Park, S., Kurban, D., Dufour, S. & Ronholm, J. The occurrence of
1462 Aerococcus urinaeequi and non-aureus staphylococci in raw milk negatively correlates
1463 with Escherichia coli clinical mastitis . *mSystems* (2024)
1464 doi:10.1128/MSYSTEMS.00362-24.
1465 27. Caballero-Flores, G., Pickard, J. M. & Núñez, G. Microbiota-mediated
1466 colonization resistance: mechanisms and regulation. *Nature Reviews Microbiology*
1467 *2022 21:6* **21**, 347–360 (2022).
1468 28. Porcellato, D., Meisal, R., Bombelli, A. & Narvhus, J. A. A core microbiota
1469 dominates a rich microbial diversity in the bovine udder and may indicate presence of
1470 dysbiosis. *Sci Rep* **10**, (2020).
1471 29. Hope, A. Laboratory Handbook on Bovine Mastitis. *Aust Vet J* **78**, (2000).
1472 30. Smistad, M., Sølverød, L., Inglingstad, R. A. & Østerås, O. Distribution of
1473 somatic cell count and udder pathogens in Norwegian dairy goats. *J Dairy Sci* **104**,
1474 (2021).
1475 31. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome
1476 data science using QIIME 2. *Nature Biotechnology* Preprint at
1477 https://doi.org/10.1038/s41587-019-0209-9 (2019).
1478 32. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina
1479 amplicon data. *Nat Methods* (2016) doi:10.1038/nmeth.3869.

1480 33. Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based
1481 on fast Fourier transform. *Nucleic Acids Res* (2002) doi:10.1093/nar/gkf436.
1482 34. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved
1483 data processing and web-based tools. *Nucleic Acids Res* **41**, (2013).
1484 35. Davis, N. M., Proctor, Di. M., Holmes, S. P., Relman, D. A. & Callahan, B. J.
1485 Simple statistical identification and removal of contaminant sequences in marker-gene
1486 and metagenomics data. *Microbiome* **6**, 1–14 (2018).
1487 36. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for
1488 genome-resolved metagenomic data analysis. *Microbiome* **6**, (2018).
1489 37. Krueger, F. Trim Galore!: A wrapper tool around Cutadapt and FastQC to
1490 consistently apply quality and adapter trimming to FastQ files. *Babraham Institute*
1491 (2015).
1492 38. Li, D. *et al.* MEGAHIT v1.0: A fast and scalable metagenome assembler driven
1493 by advanced methodologies and community practices. *Methods* **102**, 3–11 (2016).
1494 39. Kang, D. D. *et al.* MetaBAT 2: An adaptive binning algorithm for robust and
1495 efficient genome reconstruction from metagenome assemblies. *PeerJ* **2019**, (2019).
1496 40. Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: An automated binning
1497 algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**,
1498 (2016).
1499 41. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition.
1500 *Nat Methods* **11**, (2014).
1501 42. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W.
1502 CheckM: Assessing the quality of microbial genomes recovered from isolates, single
1503 cells, and metagenomes. *Genome Res* **25**, (2015).
1504 43. Beghini, F. *et al.* Integrating taxonomic, functional, and strain-level profiling of
1505 diverse microbial communities with biobakery 3. *Elife* **10**, (2021).
1506 44. Bowers, R. M. *et al.* Minimum information about a single amplified genome
1507 (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea.
1508 *Nature Biotechnology* vol. 35 Preprint at https://doi.org/10.1038/nbt.3893 (2017).
1509 45. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2:
1510 memory friendly classification with the genome taxonomy database. *Bioinformatics*
1511 **38**, (2022).

1512 46. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 - Approximately maximum-
1513 likelihood trees for large alignments. *PLoS One* (2010)
1514 doi:10.1371/journal.pone.0009490.
1515 47. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the
1516 phylogenetic tree display and annotation tool. *Nucleic Acids Res* **52**, W78–W82
1517 (2024).
1518 48. Aroney, S. T. N. *et al.* CoverM: Read coverage calculator for metagenomics.
1519 *Zenodo* (2024) doi:10.5281/ZENODO.10531254.
1520 49. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids*
1521 *Res* **44**, 6614–6624 (2016).
1522 50. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and
1523 phylogenetically annotated orthology resource based on 5090 organisms and 2502
1524 viruses. *Nucleic Acids Res* **47**, D309–D314 (2019).
1525 51. Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas,
1526 J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain
1527 Prediction at the Metagenomic Scale. *Mol Biol Evol* **38**, 5825–5829 (2021).
1528 52. Quijada, N. M., Rodríguez-Lázaro, D., Eiros, J. M. & Hernández, M. TORMES:
1529 an automated pipeline for whole bacterial genome analysis. *Bioinformatics* **35**, 4207–
1530 4212 (2019).
1531 53. Florensa, A. F., Kaas, R. S., Clausen, P. T. L. C., Aytan-Aktug, D. & Aarestrup,
1532 F. M. ResFinder – an open online resource for identification of antimicrobial resistance
1533 genes in next-generation sequencing data and prediction of phenotypes from
1534 genotypes. *Microb Genom* **8**, 000748 (2022).
1535 54. Alcock, B. P. *et al.* CARD 2023: expanded curation, support for machine
1536 learning, and resistome prediction at the Comprehensive Antibiotic Resistance
1537 Database. *Nucleic Acids Res* **51**, D690–D699 (2023).
1538 55. Gupta, S. K. *et al.* ARG-annot, a new bioinformatic tool to discover antibiotic
1539 resistance genes in bacterial genomes. *Antimicrob Agents Chemother* **58**, 212–220
1540 (2014).
1541 56. Liu, B., Zheng, D., Zhou, S., Chen, L. & Yang, J. VFDB 2022: a general
1542 classification scheme for bacterial virulence factors. *Nucleic Acids Res* **50**, D912–
1543 D917 (2022).
1544 57. Jolley, K. A. & Maiden, M. C. J. BIGSdb: Scalable analysis of bacterial genome
1545 variation at the population level. *BMC Bioinformatics* **11**, 1–11 (2010).

58. Van Heel, A. J. *et al.* BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res* **46**, W278 (2018).

59. 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).

60. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* (2013) doi:10.1371/journal.pone.0061217.

61. 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**, (2011).

62. Wickham, H., François, R., Henry, L. & Müller, K. dplyr: A Grammar of Data Manipulation. R package version. *Media* Preprint at (2019).

63. Barnett, D., Arts, I. & Penders, J. microViz: an R package for microbiome data visualization and statistics. *J Open Source Softw* **6**, (2021).

64. Mallick, H. *et al.* Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol* **17**, (2021).

65. Van Der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).

66. 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).

67. Ruegg, P. L. A 100-Year Review: Mastitis detection, management, and prevention. *J Dairy Sci* **100**, 10381–10397 (2017).

68. 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).

69. Ramuada, M., Tyasi, T. L., Gumede, L. & Chitura, T. A practical guide to diagnosing bovine mastitis: a review. *Frontiers in Animal Science* **5**, 1504873 (2024).

70. Hagnestam-Nielsen, C., Emanuelson, U., Berglund, B. & Strandberg, E. Relationship between somatic cell count and milk yield in different stages of lactation. *J Dairy Sci* **92**, 3124–3133 (2009).

71. Smistad, M., Bakka, H. C., Sølverød, 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**, (2023).

72. Paschoalini, B. R. *et al.* The Emergence of Antimicrobial Resistance and Virulence Characteristics in Enterococcus Species Isolated from Bovine Milk. *Antibiotics* **12**, (2023).

73. Angelopoulou, A. *et al.* Bovine mastitis is a polymicrobial disease requiring a polydiagnostic approach. *Int Dairy J* **99**, 104539 (2019).

74. Rodrigues, M. X., Lima, S. F., Higgins, C. H., Canniatti-Brazaca, S. G. & Bicalho, R. C. The Lactococcus genus as a potential emerging mastitis pathogen group: A report on an outbreak investigation. *J Dairy Sci* **99**, 9864–9874 (2016).

75. Lin, Y. *et al.* Comparative Genomic Analyses of Lactococcus garvieae Isolated from Bovine Mastitis in China. *Microbiol Spectr* **11**, (2023).

76. Alessandri, G. *et al.* Metataxonomic analysis of milk microbiota in the bovine subclinical mastitis. *FEMS Microbiol Ecol* **99**, 1–10 (2023).

77. 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).

78. Oultram, J. W. H., Ganda, E. K., Boulding, S. C., Bicalho, R. C. & Oikonomou, G. A metataxonomic approach could be considered for cattle clinical mastitis diagnostics. *Front Vet Sci* **4**, (2017).

79. Hu, X. *et al.* The Rumen Microbiota Contributes to the Development of Mastitis in Dairy Cows. *Microbiol Spectr* **10**, (2022).

80. Nguyen, T. T., Wu, H. & Nishino, N. An investigation of seasonal variations in the microbiota of milk, feces, bedding, and airborne dust. *Asian-Australas J Anim Sci* **33**, 1858 (2019).

81. Tong, J., Zhang, H., Zhang, Y., Xiong, B. & Jiang, L. Microbiome and Metabolome Analyses of Milk From Dairy Cows With Subclinical Streptococcus agalactiae Mastitis—Potential Biomarkers. *Front Microbiol* **10**, 474819 (2019).

82. Derakhshani, H. *et al.* Invited review: Microbiota of the bovine udder: Contributing factors and potential implications for udder health and mastitis susceptibility. *J Dairy Sci* **101**, 10605–10625 (2018).

1612 83. Xu, X., Xie, Z., Yang, Z., Li, D. & Xu, X. A t-SNE Based Classification Approach
1613 to Compositional Microbiome Data. *Front Genet* **11**, 620143 (2020).
1614 84. Schwaiger, K. *et al.* Hot topic: Bovine milk samples yielding negative or
1615 nonspecific results in bacterial culturing—The possible role of PCR-single strand
1616 conformation polymorphism in mastitis diagnosis. *J Dairy Sci* **95**, 98–101 (2012).
1617 85. Taponen, S., Salmikivi, L., Simojoki, H., Koskinen, M. T. & Pyörälä, S. Real-time
1618 polymerase chain reaction-based identification of bacteria in milk samples from bovine
1619 clinical mastitis with no growth in conventional culturing. *J Dairy Sci* **92**, 2610–2617
1620 (2009).
1621 86. Kuehn, J. S. *et al.* Bacterial Community Profiling of Milk Samples as a Means
1622 to Understand Culture-Negative Bovine Clinical Mastitis. *PLoS One* **8**, e61959 (2013).
1623 87. Lücken, A., Woudstra, S., Wente, N., Zhang, Y. & Krömker, V. Intramammary
1624 infections with Corynebacterium spp. in bovine lactating udder quarters. *PLoS One*
1625 **17**, e0270867 (2022).
1626 88. Silva, V. M. *et al.* Milk lymphocyte profile and macrophage functions: new
1627 insights into the immunity of the mammary gland in quarters infected with
1628 Corynebacterium bovis. *BMC Vet Res* **17**, 1–8 (2021).
1629 89. Lam, T. J. G. M. *et al.* Effect of natural infection with minor pathogens on
1630 susceptibility to natural infection with major pathogens in the bovine mammary gland.
1631 *Am J Vet Res* **58**, (1997).
1632 90. Seshadri, R. *et al.* Expanding the genomic encyclopedia of Actinobacteria with
1633 824 isolate reference genomes. *Cell Genomics* **2**, 100213 (2022).
1634 91. Park, S. *et al.* A longitudinal census of the bacterial community in raw milk
1635 correlated with Staphylococcus aureus clinical mastitis infections in dairy cattle. *Anim*
1636 *Microbiome* **4**, 1–13 (2022).
1637 92. Dean, C. J. *et al.* Investigating the cow skin and teat canal microbiomes of the
1638 bovine udder using different sampling and sequencing approaches. *J Dairy Sci* **104**,
1639 644–661 (2021).
1640 93. Woudstra, S. *et al.* Reservoirs of Corynebacterium spp. in the Environment of
1641 Dairy Cows. *Pathogens* **12**, 139 (2023).
1642 94. Ruiz-Romero, R. A. & Vargas-Bello-Pérez, E. Non-aureus staphylococci and
1643 mammaliicocci as a cause of mastitis in domestic ruminants: current knowledge,
1644 advances, biomedical applications, and future perspectives – a systematic review.

Veterinary Research Communications vol. 47 Preprint at https://doi.org/10.1007/s11259-023-10090-5 (2023).

95. Król, J. *et al.* Staphylococcus borealis – A newly identified pathogen of bovine mammary glands. *Vet Microbiol* **286**, 109876 (2023).

96. De Buck, J. *et al.* Non-aureus Staphylococci and Bovine Udder Health: Current Understanding and Knowledge Gaps. *Front Vet Sci* **8**, 658031 (2021).

97. Penati, M. *et al.* An update on non-aureus staphylococci and mammaliicocci in cow milk: unveiling the presence of Staphylococcus borealis and Staphylococcus rostri by MALDI-TOF MS. *Vet Res Commun* **48**, 2555–2561 (2024).

98. Toledo-Silva, B., Oliveira, A. C. D., Souza, F. N., Haesebrouck, F. & De Vliegher, S. Metabolites of non-aureus staphylococci affect the ability of Staphylococcus aureus to adhere to and internalize into bovine mammary epithelial cells. *Vet Res* **54**, 100 (2023).

99. Mahmmod, Y. S., Klaas, I. C., Svennesen, L., Pedersen, K. & Ingmer, H. Communications of Staphylococcus aureus and non-aureus Staphylococcus species from bovine intramammary infections and teat apex colonization. *J Dairy Sci* **101**, 7322–7333 (2018).

100. Carson, D. A., Barkema, H. W., Naushad, S. & De Buck, J. Bacteriocins of non-aureus staphylococci isolated from bovine milk. *Appl Environ Microbiol* **83**, (2017).

101. 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).

102. Zaatout, N., Ayachi, A. & Kecha, M. Staphylococcus aureus persistence properties associated with bovine mastitis and alternative therapeutic modalities. *Journal of Applied Microbiology* vol. 129 Preprint at https://doi.org/10.1111/jam.14706 (2020).

103. Kabelitz, T., Aubry, E., van Vorst, K., Amon, T. & Fulde, M. The Role of Streptococcus spp. in Bovine Mastitis. *Microorganisms* **9**, 1497 (2021).

104. Srithanasuwan, A., Schukken, Y. H., Pangprasit, N., Chuammitri, P. & Suriyasathaporn, W. Different cellular and molecular responses of Bovine milk phagocytes to persistent and transient strains of Streptococcus uberis causing mastitis. *PLoS One* **19**, e0295547 (2024).

105. Luo, Q. *et al.* Classification of 27 Corynebacterium kroppenstedtii -Like Isolates Associated with Mastitis in China and Descriptions of C. parakroppenstedtii sp. nov. and C. pseudokroppenstedtii sp. nov . *Microbiol Spectr* **10**, (2022).

106. Stevenson, D. R. *et al.* Corynebacterium kroppenstedtii breast abscesses in context, a retrospective cohort study. *J Med Microbiol* **71**, 001616 (2022).

107. Milke, L., Kallscheuer, N., Kappelmann, J. & Marienhagen, J. Tailoring Corynebacterium glutamicum towards increased malonyl-CoA availability for efficient synthesis of the plant pentaketide noreugenin. *Microb Cell Fact* **18**, 1–12 (2019).

108. Liu, R. *et al.* Corynebacterium parakroppenstedtii secretes a novel glycolipid to promote the development of granulomatous lobular mastitis. *Signal Transduction and Targeted Therapy 2024 9:1* **9**, 1–17 (2024).

109. Shinozuka, Y. *et al.* Examination of the microbiota of normal cow milk using MinION™ nanopore sequencing. *J Vet Med Sci* **83**, 1620 (2021).

110. Ruiz, L. *et al.* Microbiota of human precolostrum and its potential role as a source of bacteria to the infant mouth. *Scientific Reports 2019 9:1* **9**, 1–13 (2019).

111. Fotedar, R. *et al.* Reclassification of Facklamia ignava, Facklamia sourekii and Facklamia tabacinasalis as Falseniella ignava gen. nov., comb. nov., Hutsoniella sourekii gen. nov., comb. nov., and Ruoffia tabacinasalis gen. nov., comb. nov., and description of Ruoffia halotolerans sp. nov., isolated from hypersaline Inland Sea of Qatar. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* **114**, 1181–1193 (2021).

112. Stackebrandt, E., Koch, C., Gvozdiak, O. & Schumann, P. Taxonomic dissection of the genus Micrococcus: Kocuria gen. nov., Nesterenkonia gen. nov., Kytococcus gen. nov., Dermacoccus gen. nov., and Micrococcus cohn 1872 gen. emend. *Int J Syst Bacteriol* **45**, 682–692 (1995).

113. Stackebrandt, E. The family aerococcaceae. *The Prokaryotes: Firmicutes and Tenericutes* **9783642301209**, 3–6 (2014).

114. Naushad, S., Nobrega, D. B., Naqvi, S. A., Barkema, H. W. & De Buck, J. Genomic Analysis of Bovine Staphylococcus aureus Isolates from Milk To Elucidate Diversity and Determine the Distributions of Antimicrobial and Virulence Genes and Their Association with Mastitis. *mSystems* **5**, (2020).

115. Merz, A., Stephan, R. & Johler, S. Staphylococcus aureus isolates from goat and sheep milk seem to be closely related and differ from isolates detected from bovine milk. *Front Microbiol* **7**, 184387 (2016).

116. Zischka, M. *et al.* Comprehensive molecular, genomic and phenotypic analysis of a major clone of Enterococcus faecalis MLST ST40. *BMC Genomics* **16**, 175 (2015).

117. Porcellato, D. *et al.* Whole genome sequencing reveals possible host species adaptation of Streptococcus dysgalactiae. *Scientific Reports 2021 11:1* **11**, 1–13 (2021).

118. Myrenås, M., Fasth, C., Waller, K. P. & Pedersen, K. Genomic analyses of Streptococcus uberis reveal high diversity but few antibiotic resistance genes. *Vet Microbiol* **300**, 110319 (2025).

119. Vasileiou, N. G. C. *et al.* Antimicrobial Agent Susceptibility and Typing of Staphylococcal Isolates from Subclinical Mastitis in Ewes. *https://home.liebertpub.com/mdr* **25**, 1099–1110 (2019).

120. Torres Salazar, B. O. *et al.* Commensal production of a broad-spectrum and short-lived antimicrobial peptide polyene eliminates nasal Staphylococcus aureus. *Nature Microbiology 2023 9:1* **9**, 200–213 (2023).

121. Persson Waller, K., Lundberg, Börjesson, S. & Nyman, A. K. Intramammary infections and risk factors in freshly calved heifers in Swedish dairy herds. *J Dairy Sci* **106**, 7893–7907 (2023).

122. Sharma, A. K. *et al.* Bacterial Virulence Factors: Secreted for Survival. *Indian J Microbiol* **57**, 1–10 (2017).

123. Hill, C. Virulence or Niche Factors: What's in a Name? *J Bacteriol* **194**, 5725 (2012).

124. 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).

125. Tauch, A. & Burkovski, A. Molecular armory or niche factors: virulence determinants of Corynebacterium species. *FEMS Microbiol Lett* **362**, 185 (2015).

126. 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**, (1995).

127. Marcos-Torres, F. J., Juniar, L. & Griese, J. J. The molecular mechanisms of the bacterial iron sensor IdeR. *Biochemical Society Transactions* vol. 51 Preprint at https://doi.org/10.1042/BST20221539 (2023).

128. Chou, C. J. *et al.* Functional studies of the Mycobacterium tuberculosis iron-dependent regulator. *Journal of Biological Chemistry* **279**, (2004).

129. Tiwari, S. *et al.* C. pseudotuberculosis Phop confers virulence and may be targeted by natural compounds. *Integrative Biology* **6**, 1088–1099 (2014).

130. Kafil, H. S., Mobarez, A. M. & Moghadam, M. F. Adhesion and virulence factor properties of Enterococci isolated from clinical samples in Iran. *Indian J Pathol Microbiol* **56**, (2013).

131. Creti, R. *et al.* Survey for virulence determinants among Enterococcus faecalis isolated from different sources. *Journal of Medical Microbiology* vol. 53 Preprint at https://doi.org/10.1099/jmm.0.05353-0 (2004).

132. Nallapareddy, S. R., Singh, K. V., Sillanpää, J., Zhao, M. & Murray, B. E. Relative contributions of Ebp Pili and the collagen adhesin ace to host extracellular matrix protein adherence and experimental urinary tract infection by Enterococcus faecalis OG1RF. *Infect Immun* **79**, (2011).

133. Sau, S. *et al.* The Staphylococcus aureus allelic genetic loci for serotype 5 and 8 capsule expression contain the type-specific genes flanked by common genes. *Microbiology (N Y)* **143**, (1997).

134. Tuchscherr, L. P. N. *et al.* Capsule-negative Staphylococcus aureus induces chronic experimental mastitis in mice. *Infect Immun* **73**, 7932–7937 (2005).

135. Kim, H. K., Thammavongsa, V., Schneewind, O. & Missiakas, D. Recurrent infections and immune evasion strategies of Staphylococcus aureus. *Current Opinion in Microbiology* vol. 15 Preprint at https://doi.org/10.1016/j.mib.2011.10.012 (2012).

136. Darisipudi, M. N. & Bröker, B. M. How S. aureus blinds the inflammasome to escape immune control. *EBioMedicine* **71**, 103549 (2021).

137. Smith, E. J., Visai, L., Kerrigan, S. W., Speziale, P. & Foster, T. J. The Sbi protein is a multifunctional immune evasion factor of Staphylococcus aureus. *Infect Immun* **79**, (2011).

138. Dunphy, R. W. *et al.* Staphylococcal Complement Evasion Protein Sbi Stabilises C3d Dimers by Inducing an N-Terminal Helix Swap. *Front Immunol* **13**, 892234 (2022).

139. Kaczorek, E., Małaczewska, J., Wójcik, R. & Siwicki, A. K. Biofilm production and other virulence factors in Streptococcus spp. isolated from clinical cases of bovine mastitis in Poland. *BMC Vet Res* **13**, (2017).

140. Torres, G., Macias, D., Reyes-Velez, J., Rios-Agudelo, P. & Caraballo-Guzmán, A. Streptococcus agalactiae virulence factors isolated from bovine mastitis and antibiotic treatment response. *J Appl Microbiol* **134**, (2023).

141. O'Halloran, F., Beecher, C., Chaurin, V., Sweeney, T. & Giblin, L. Lactoferrin affects the adherence and invasion of Streptococcus dysgalactiae ssp. dysgalactiae in mammary epithelial cells. *J Dairy Sci* **99**, 4619–4628 (2016).

142. Moshynskyy, I. *et al.* Characterization of a bovine lactoferrin binding protein of Streptococcus uberis. *Microb Pathog* **35**, 203–215 (2003).

143. Rainard, P. Binding of bovine lactoferrin to Streptococcus agalactiae. *FEMS Microbiol Lett* **98**, 235–239 (1992).

144. 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. *Clinical Microbiology and Infection* **26**, 1612 (2020).

145. Rubiola, S. *et al.* Shotgun metagenomic sequencing of bulk tank milk filters reveals the role of Moraxellaceae and Enterobacteriaceae as carriers of antimicrobial resistance genes. *Food Research International* **158**, 111579 (2022).

146. Ortiz-Pérez, A. *et al.* High frequency of macrolide resistance mechanisms in clinical isolates of Corynebacterium species. *Microbial Drug Resistance* **16**, (2010).

147. Olender, A. Antibiotic resistance and detection of the most common mechanism of resistance (MLSB) of opportunistic Corynebacterium. *Chemotherapy* **59**, (2014).

148. 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**, (2018).

149. 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).

150. Feßler, A. T. & Schwarz, S. Antimicrobial Resistance in Corynebacterium spp., Arcanobacterium spp., and Trueperella pyogenes. *Microbiol Spectr* **5**, (2017).

151. Brdová, D., Ruml, T. & Viktorová, J. Mechanism of staphylococcal resistance to clinically relevant antibiotics. *Drug Resistance Updates* **77**, 101147 (2024).

152. Huet, A. A., Raygada, J. L., Mendiratta, K., Seo, S. M. & Kaatz, G. W. Multidrug efflux pump overexpression in Staphylococcus aureus after single and multiple in vitro exposures to biocides and dyes. *Microbiology (N Y)* **154**, (2008).

153. Schwarz, S. *et al.* Antimicrobial Resistance among Staphylococci of Animal Origin. *Microbiol Spectr* **6**, (2018).

154. Samanta, I. & Bandyopadhyay, S. Streptococcus. *Antimicrobial Resistance in Agriculture* 217–232 (2020) doi:10.1016/B978-0-12-815770-1.00017-1.

155. Haenni, M., Lupo, A. & Madec, J.-Y. Antimicrobial Resistance in Streptococcus spp . *Microbiol Spectr* **6**, (2018).

156. Alves-Barroco, C., Rivas-García, L., Fernandes, A. R. & Baptista, P. V. Tackling Multidrug Resistance in Streptococci – From Novel Biotherapeutic Strategies to Nanomedicines. *Front Microbiol* **11**, 579916 (2020).

157. Yu, D., Guo, D., Zheng, Y. & Yang, Y. A review of penicillin binding protein and group A Streptococcus with reduced-β-lactam susceptibility. *Front Cell Infect Microbiol* **13**, 1117160 (2023).

158. Fuursted, K. *et al.* Description and characterization of a penicillin-resistant Streptococcus dysgalactiae subsp. equisimilis clone isolated from blood in three epidemiologically linked patients. *Journal of Antimicrobial Chemotherapy* **71**, (2016).

159. Metcalf, B. J. *et al.* Short-read whole genome sequencing for determination of antimicrobial resistance mechanisms and capsular serotypes of current invasive Streptococcus agalactiae recovered in the USA. *Clinical Microbiology and Infection* **23**, (2017).

160. Berbel, D., González-Díaz, A., López de Egea, G., Càmara, J. & Ardanuy, C. An Overview of Macrolide Resistance in Streptococci: Prevalence, Mobile Elements and Dynamics. *Microorganisms* **10**, 2316 (2022).

161. Kamińska, D. *et al.* Macrolide and lincosamide resistance of Streptococcus agalactiae in pregnant women in Poland. *Scientific Reports 2024 14:1* **14**, 1–8 (2024).

162. Loch, I. M., Glenn, K. & Zadoks, R. N. Macrolide and lincosamide resistance genes of environmental streptococci from bovine milk. *Vet Microbiol* **111**, 133–138 (2005).

163. Malhotra-Kumar, S., Lammens, C., Piessens, J. & Goossens, H. Multiplex PCR for simultaneous detection of macrolide and tetracycline resistance determinants in streptococci. *Antimicrob Agents Chemother* **49**, (2005).

164. Cherazard, R. *et al.* Antimicrobial Resistant Streptococcus pneumoniae: Prevalence, Mechanisms, and Clinical Implications. *Am J Ther* **24**, (2017).

165. Yang, F. *et al.* Short communication: Antimicrobial resistance and virulence genes of Enterococcus faecalis isolated from subclinical bovine mastitis cases in China. *J Dairy Sci* **102**, (2019).

166. Goksel, E. *et al.* Distribution of antibiotic resistance genes in enterococcus spp. isolated from mastitis bovine milk. *Acta Vet Brno* **66**, (2016).

167. 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).

168. Espeche, M. C., Otero, M. C., Sesma, F. & Nader-Macias, M. E. F. Screening of surface properties and antagonistic substances production by lactic acid bacteria isolated from the mammary gland of healthy and mastitic cows. *Vet Microbiol* **135**, 346–357 (2009).

169. Rainard, P. & Foucras, G. A Critical Appraisal of Probiotics for Mastitis Control. *Front Vet Sci* **5**, (2018).

170. Daly, K. M. *et al.* Production of the Bsa lantibiotic by community-acquired Staphylococcus aureus strains. *J Bacteriol* **192**, (2010).

171. 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**, (2024).

172. Ceotto, H. *et al.* Nukacin 3299, a lantibiotic produced by Staphylococcus simulans 3299 identical to nukacin ISK-1. *Vet Microbiol* **146**, 124–131 (2010).

173. Thoendel, M., Kavanaugh, J. S., Flack, C. E. & Horswill, A. R. Peptide signaling in the Staphylococci. *Chem Rev* **111**, (2011).

174. Nilsen, T., Nes, I. F. & Holo, H. Enterolysin A, a cell wall-degrading bacteriocin from Enterococcus faecalis LMG 2333. *Appl Environ Microbiol* **69**, (2003).

175. 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).

176. Coburn, P. S. & Gilmore, M. S. The Enterococcus faecalis cytolysin: A novel toxin active against eukaryotic and prokaryotic cells. *Cellular Microbiology* vol. 5 Preprint at https://doi.org/10.1046/j.1462-5822.2003.00310.x (2003).

177. Cox, C., Coburn, P. & Gilmore, M. Enterococcal Cytolysin: A Novel Two Component Peptide System that Serves as a Bacterial Defense Against Eukaryotic and Prokaryotic Cells. *Curr Protein Pept Sci* **6**, (2005).

178. Van Tyne, D., Martin, M. J. & Gilmore, M. S. Structure, function, and biology of the Enterococcus faecalis cytolysin. *Toxins* vol. 5 Preprint at https://doi.org/10.3390/toxins5050895 (2013).

179. Amaral, J. R. V. *et al.* Bacteriocin Producing Streptococcus agalactiae Strains Isolated from Bovine Mastitis in Brazil. *Microorganisms* **10**, (2022).

180. Chaimanee, V., Sakulsingharoj, C., Deejing, S., Seetakoses, P. & Niamsup, P. Screening and characterisation of bacteriocin-producing bacteria capable of inhibiting the growth of bovine mastitis. *Maejo International Journal of Science and Technology* **3**, (2009).

181. Meiwes, J. *et al.* Isolation and characterization of staphyloferrin A, a compound with siderophore activity from Staphylococcus hyicus DSM 20459. *FEMS Microbiol Lett* **67**, (1990).

182. Torres Salazar, B. O., Heilbronner, S., Peschel, A. & Krismer, B. Secondary Metabolites Governing Microbiome Interaction of Staphylococcal Pathogens and Commensals. *Microbial Physiology* vol. 31 Preprint at https://doi.org/10.1159/000517082 (2021).

183. Diarra, M. S., Petitclerc, D. & Lacasse, P. Response of Staphylococcus aureus isolates from bovine mastitis to exogenous iron sources. *J Dairy Sci* **85**, (2002).

184. Mårli, M. T. *et al.* Genome-wide analysis of fitness determinants of Staphylococcus aureus during growth in milk. *bioRxiv* 2024.10.31.621336 (2024) doi:10.1101/2024.10.31.621336.

185. 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).

186. P, V. *et al.* Streptomyces sp. VITGV100: An endophyte from Lycopersicon esculentum as new source of indole type compounds. *Biochem Syst Ecol* **105**, (2022).

187. Costa, N. J. S. *et al.* Antimicrobial Activity of Ethylenediaminedisuccinate Metal Complexes. Short Communication. *Chem Biodivers* **5**, 2156–2159 (2008).

188. Hyvönen, P. *et al.* Concentrations of bovine lactoferrin and citrate in milk during experimental endotoxin mastitis in early-versus late-lactating dairy cows. *Journal of Dairy Research* **77**, (2010).

189. 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**, (2010).

190. Mavrodi, D. V. *et al.* Diversity and evolution of the Phenazine Biosynthesis Pathways. *Appl Environ Microbiol* **76**, (2010).

191. Hendry, S. *et al.* Functional Analysis of Phenazine Biosynthesis Genes in Burkholderia spp. *Appl Environ Microbiol* **87**, (2021).

192. Guttenberger, N., Blankenfeldt, W. & Breinbauer, R. Recent developments in the isolation, biological function, biosynthesis, and synthesis of phenazine natural products. *Bioorganic and Medicinal Chemistry* vol. 25 Preprint at https://doi.org/10.1016/j.bmc.2017.01.002 (2017).

193. Fitzpatrick, D. A. Lines of evidence for horizontal gene transfer of a phenazine producing operon into multiple bacterial species. *J Mol Evol* **68**, (2009).

1922 **Captions**

1923 **Figure 1 –** Raincloud plot **(A)** illustrating somatic cell count (log10(SCC)) across
1924 different lactation periods (DO-2022, EL-2023, ML-2023, and LL-2023). Individual
1925 samples are represented by dots, color-coded as high SCC (red) or low SCC (light
1926 blue). Side histograms depict data distribution, while central bars indicate the median
1927 and interquartile range. Statistical comparisons between lactation periods are denoted
1928 by significance levels: ***$p < 0.001$, **$p < 0.01$, and NS (not significant). Density plot
1929 **(B)** showing the relationship between SCC and bacterial count (log10(IBC)) across
1930 lactation periods, stratified by bacterial species identified using MALDI-ToF. Density
1931 curves represent distinct pathogens (*E. faecalis*, *S. aureus*, *S. chromogenes*, *S.*
1932 *epidermidis*, and *S. uberis*) and samples classified as "negative", highlighting
1933 variations in bacterial prevalence and load among lactation periods.

1934

1935 **Figure 2 –**Bubble chart **(A)** displaying the relationship between somatic cell count
1936 (SCC) and individual bacterial count (IBC) across samples, categorized by bacterial
1937 species. Bubble sizes represent CFU/mL on blood agar, while colors indicate bacterial
1938 species (*E. faecalis*, *S. aureus*, *S. chromogenes*, *S. epidermidis*, *S. uberis*), and milk
1939 samples classified as "negative" for specific mastitis-causing pathogens. Box plots
1940 comparing Shannon **(B)** and Simpson **(C)** diversity indices between high and low SCC
1941 groups across different time points and lactation periods. Statistical significance is
1942 denoted as ***$p < 0.001$, **$p < 0.01$, and *$p < 0.05$. Volatility plot of Shannon **(D)** and
1943 Simpson **(E)** diversity indices over time, showing trends across lactation periods (DO-
1944 2023, EL-2023, ML-2023, and LL-2023). Solid and dashed lines represent high and
1945 low SCC groups, respectively, with shaded areas indicating 95% confidence intervals
1946 with group-wise comparisons highlighted.

1947

1948 **Figure 3 –** Non-metric multidimensional scaling (NMDS) plot based on unweighted
1949 **(A)** and weighted **(B)** UniFrac distances, illustrating microbial community composition
1950 across different lactation periods (DO-2022, EL-2023, ML-2023, LL-2023) and SCC
1951 levels (high and low). Points represent individual samples, with colors indicating SCC
1952 levels and shapes representing lactation periods. Marginal density plots along the
1953 axes depict the distribution of samples by MDS dimensions.

1954

Figure 4 – Box plots **(A–F)** showing the distribution of the relative abundance of the six dominant genera (*Corynebacterium*, *Staphylococcus*, *Enterococcus*, *Streptococcus*, *Lactococcus*, and *Aerococcus*) 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).

Figure 5 – Distribution of bacterial count **(A)**, somatic cell count **(B)**. and alpha diversity indices (**C** and **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).

Figure 6 – Stacked bar plots illustrating the relative abundance of the most abundant amplicon sequence variants (ASVs) across microbiota clusters. Each bar represents the proportional composition of ASVs within a specific cluster, with colors corresponding to distinct ASVs as shown in the legend.

Figure 7 – 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.

Figure 8 – Heatmap **(A)** showing the most abundant MetaCyc pathways detected across 73 metagenomic samples analyzed using HUMAnN3. Hierarchical clustering was applied to both samples and pathways, with samples annotated by lactation period, somatic cell count, and pathogen (identified by MALDI-Tof). Pathway abundance is represented by a red-to-blue color gradient, where darker red indicates higher abundance and darker blue indicates lower abundance. Bar plots **(B)** display the significant MetaCyc pathways (q-value < 0.25) associated with high and low SCC groups, with coefficients categorized as positive (blue) or negative (red) to illustrate their direction of association. Boxplots **(C)** depict the relative abundance of significant MetaCyc pathways stratified by SCC levels, highlighting pathway abundance differences between the two groups.

Figure 9 – Abundance distribution of significant MetaCyc pathways stratified by bacterial species: (A) *S. aureus*, (B) *S. chromogenes*, (C) *S. epidermidis*, (D) *E. faecalis*, and (E) *S. uberis*. Pathway abundances are shown as boxplots, grouped by species, with pathways color-coded according to their coefficients. Positive and negative coefficients are represented by gradients of blue and red, respectively, highlighting pathways significantly associated with each bacterial species. An empty white dot indicates the mean abundance for each pathway.

Figure 10 – 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.

Figure 11 – Phylogenetic tree and relative abundance of the species. From left to right tree shows pie charts with species's 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.

Figure 12 – 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: *Aeroococcaceae* **(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).

Figure 13 – 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.

Figure 14 – Dot plot showing the distribution of antibiotic resistance genes across metagenome-assembled genomes. Dot sizes are proportional to the frequency of each gene as annotated by three databases: ResFinder, CARD, and ARG-ANNOT. Each dot is color-coded by MAG, highlighting the presence and prevalence of specific ARGs across different MAGs. Larger dots indicate higher consensus across the databases regarding the detection of that gene.

Figure 15 – Chord diagram illustrating the relationships between bacteriocin **(A)** gene clusters predicted using BAGEL4 across the different MAGs. Chord diagram depicting the diversity and distribution of secondary metabolite gene clusters identified with antiSMASH **(B)** across the same MAGs. Connections highlight shared features or

2056 pathways among the identified clusters, providing insights into the biosynthetic
2057 capabilities within each taxon
2058
2059

Figures

Figure 1

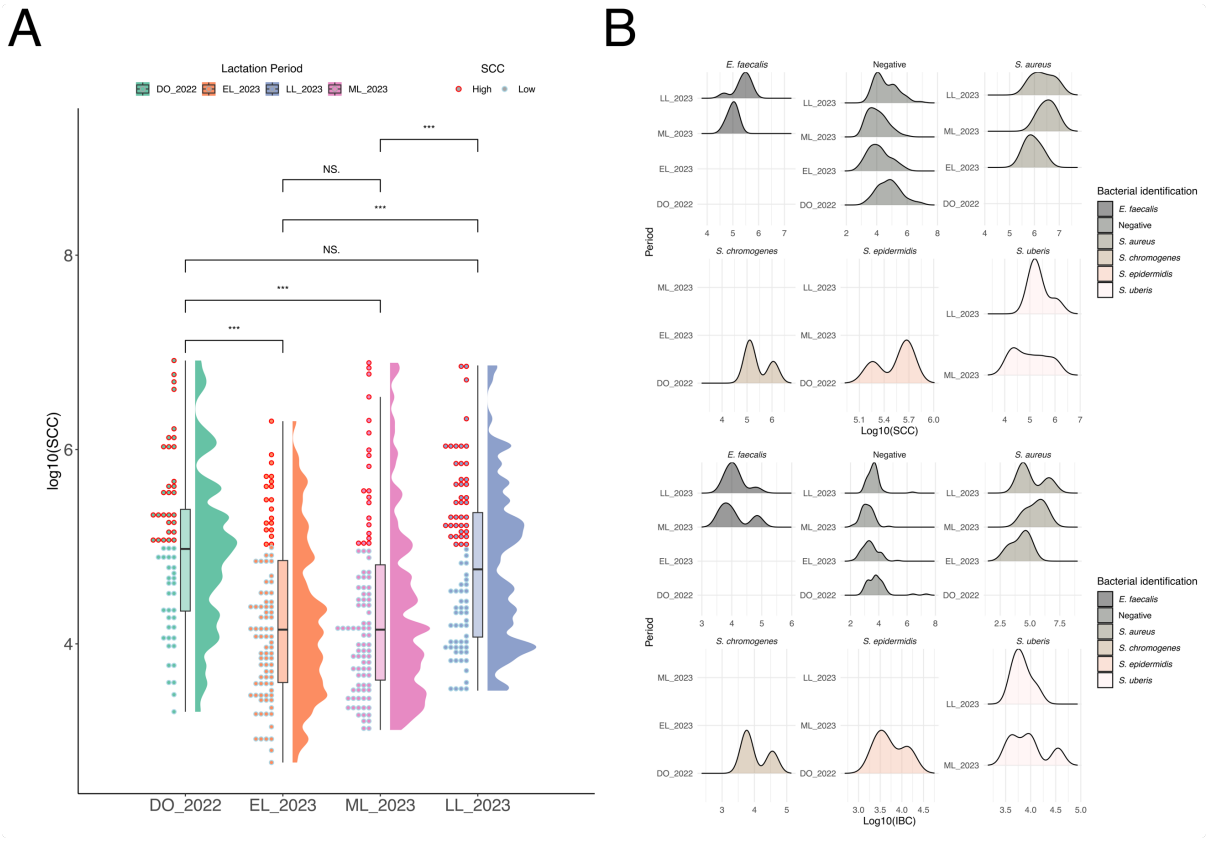


Figure 2

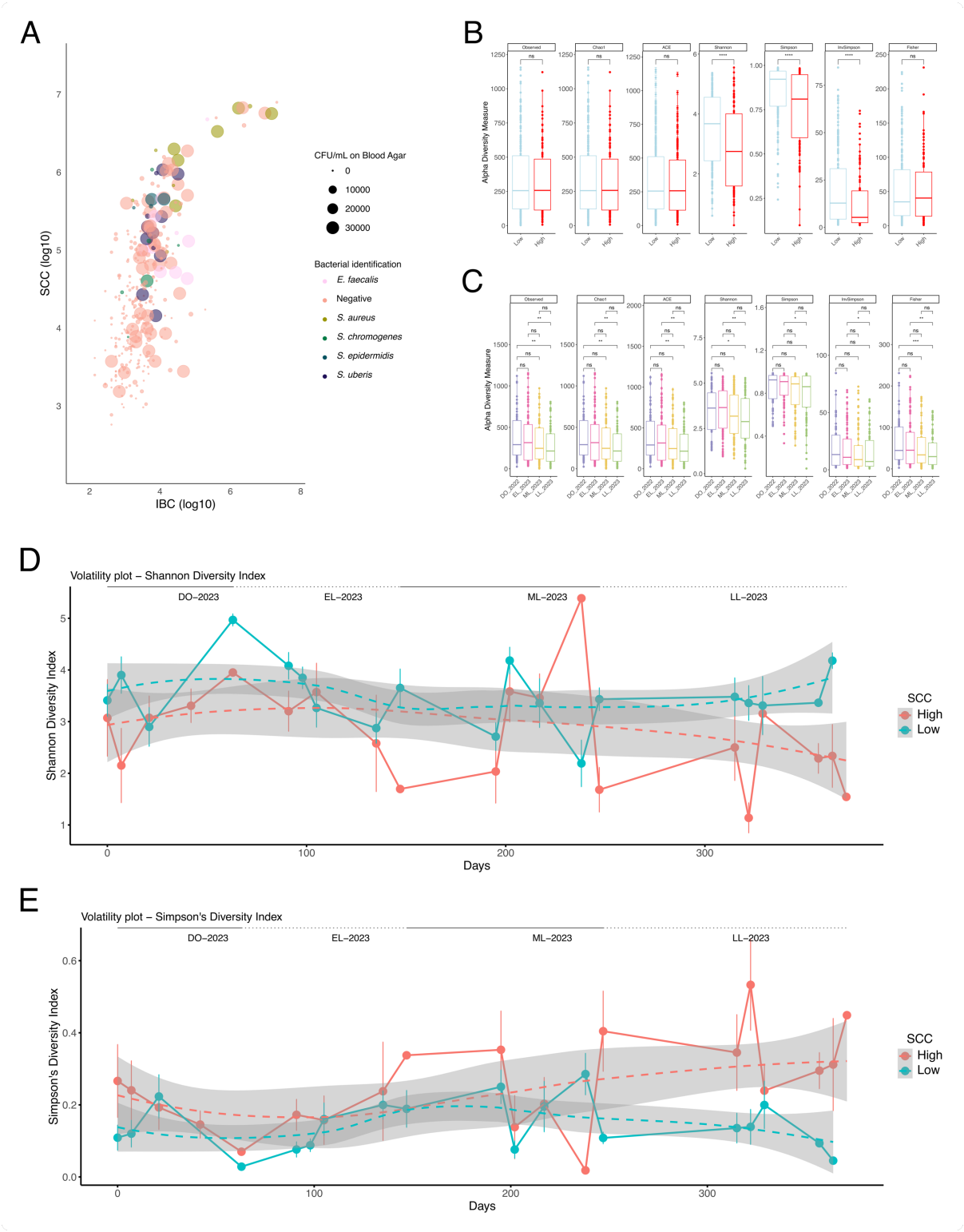

A
SCC (log10)
IBC (log10)
CFU/mL on Blood Agar
0
10000
20000
30000
Bacterial identification
E. faecalis
Negative
S. aureus
S. chromogenes
S. epidermidis
S. uberis
B
Alpha Diversity Measure
Observed
Chao1
ACE
Shannon
Simpson
InvSimpson
Fisher
Low
High
C
Alpha Diversity Measure
DO_2022
EL_2023
ML_2023
LL_2023
D
Volatility plot – Shannon Diversity Index
DO–2023
EL–2023
ML–2023
LL–2023
Shannon Diversity Index
Days
SCC
High
Low
E
Volatility plot – Simpson's Diversity Index
DO–2023
EL–2023
ML–2023
LL–2023
Simpson's Diversity Index
Days
SCC
High
Low

Figure 3

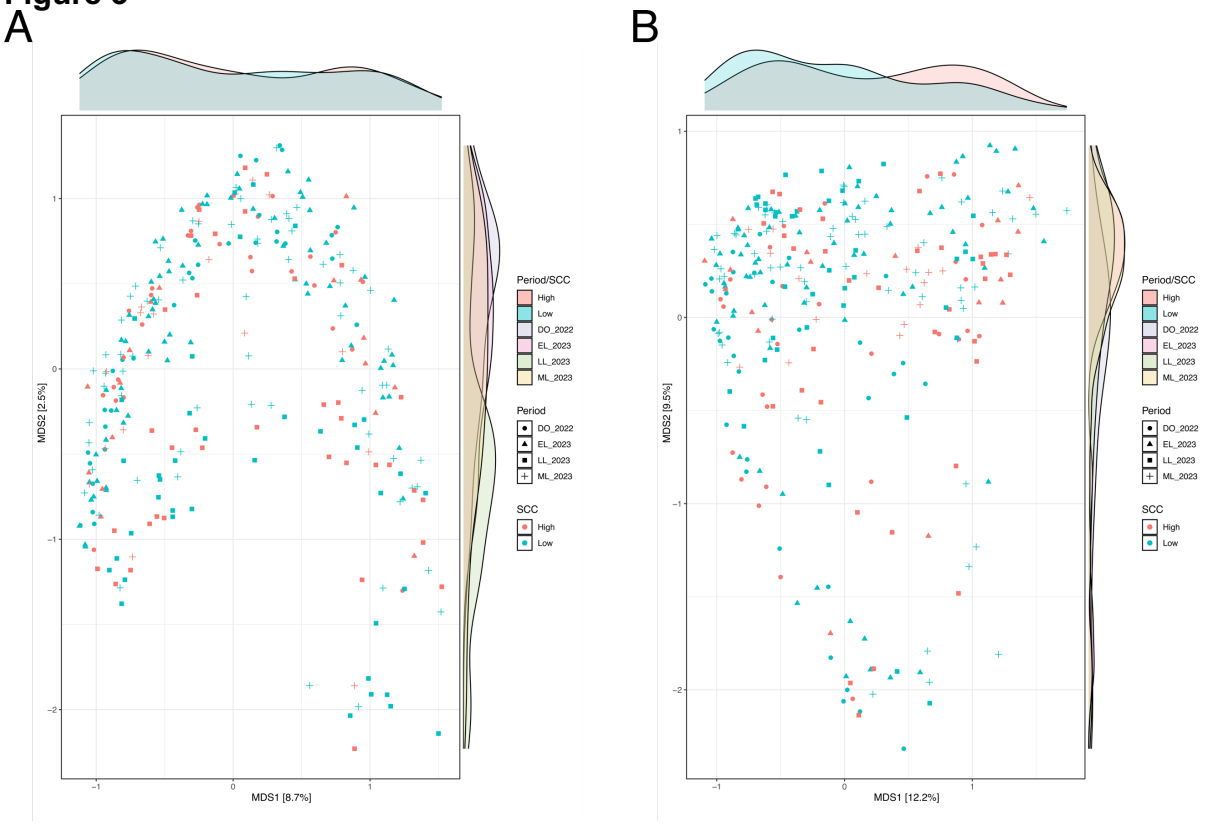

A
B
MDS1 [8.7%]
MDS2 [2.5%]
MDS1 [12.2%]
MDS2 [9.5%]
Period/SCC
High
Low
DO_2022
EL_2023
LL_2023
ML_2023
Period
DO_2022
EL_2023
LL_2023
ML_2023
SCC
High
Low

Figure 4

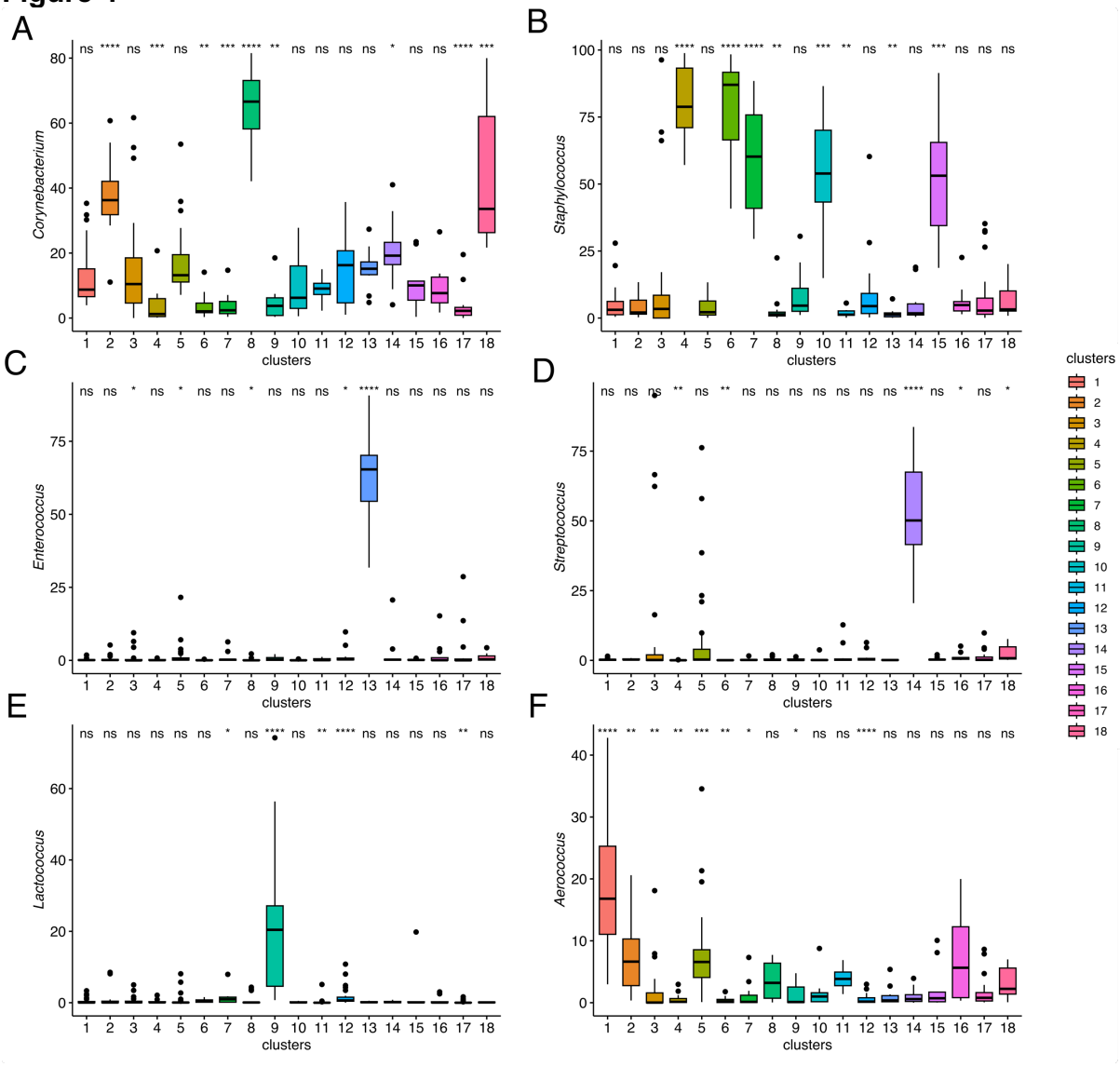

A
B
C
D
E
F
Corynebacterium
Staphylococcus
Enterococcus
Streptococcus
Lactococcus
Aerococcus
clusters
clusters
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

Figure 5

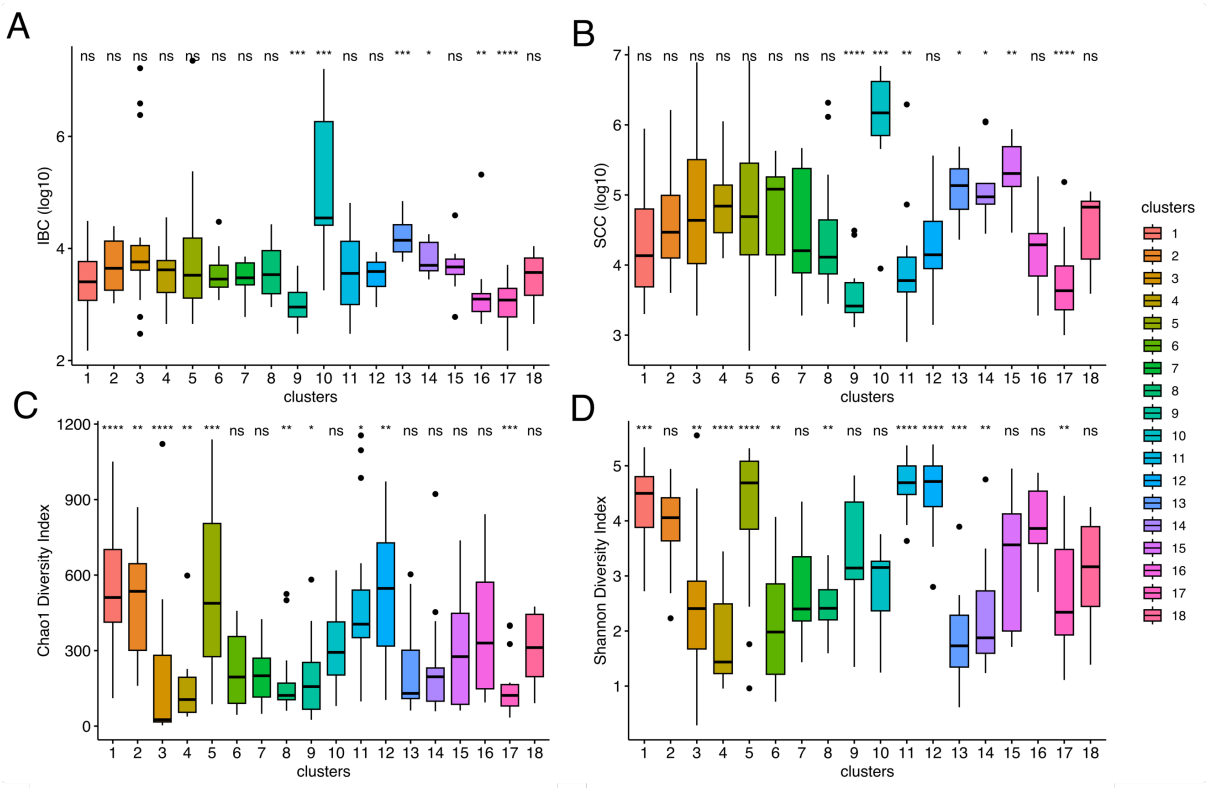

A
IBC (log10)
clusters
B
SCC (log10)
clusters
C
Chao1 Diversity Index
clusters
D
Shannon Diversity Index
clusters
clusters
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

Figure 6

Cluster 4 – Staphylococcus
Cluster 6 – Staphylococcus
Cluster 7 – Staphylococcus
Cluster 10 – Staphylococcus
Cluster 15 – Staphylococcus
Cluster 2 – Corynebacterium
Cluster 8 – Corynebacterium
Cluster 18 – Corynebacterium
Cluster 3 – Streptococcus
Cluster 14 – Streptococcus
Relative Abundance
ASV

Figure 7

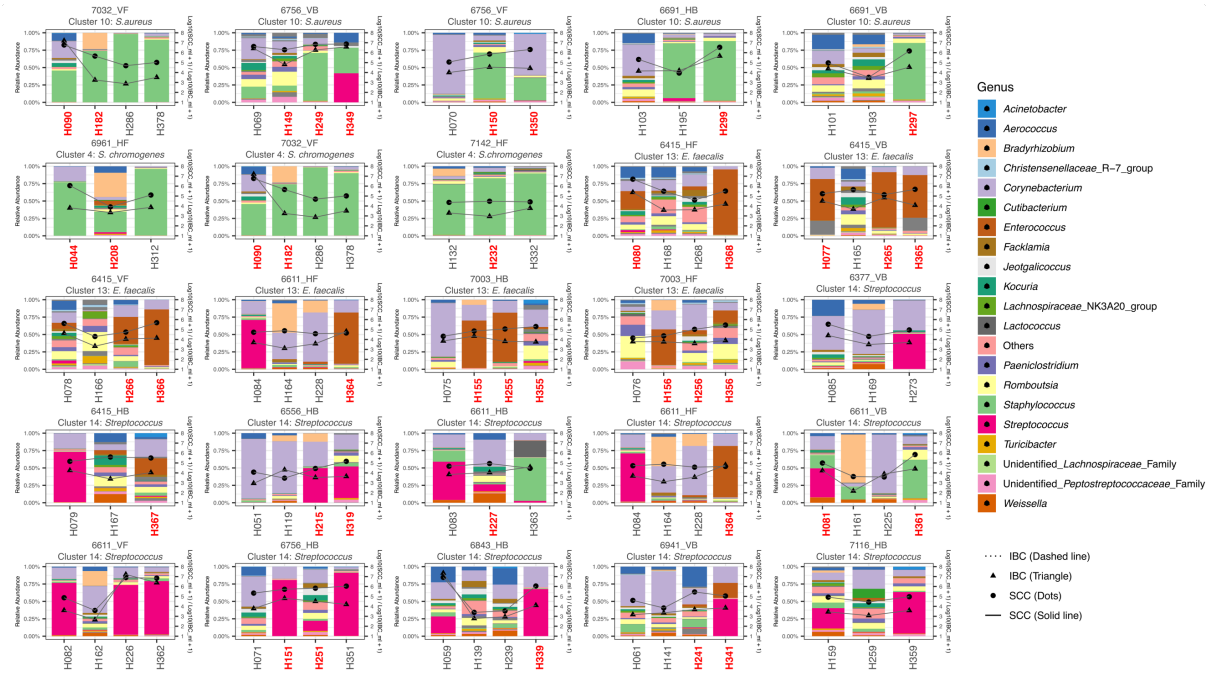

7032_VF
Cluster 10: S.aureus
H090 H182 H286 H378
6756_VB
Cluster 10: S.aureus
H069 H149 H249 H349
6756_VF
Cluster 10: S.aureus
H070 H150 H350
6691_HB
Cluster 10: S.aureus
H103 H195 H299
6691_VB
Cluster 10: S.aureus
H101 H193 H297
6961_HF
Cluster 4: S. chromogenes
H044 H208 H312
7032_VF
Cluster 4: S. chromogenes
H090 H182 H286 H378
7142_HF
Cluster 4: S.chromogenes
H132 H232 H332
6415_HF
Cluster 13: E. faecalis
H080 H168 H268 H368
6415_VB
Cluster 13: E. faecalis
H077 H165 H265 H365
6415_VF
Cluster 13: E. faecalis
H078 H166 H266 H366
6611_HF
Cluster 13: E. faecalis
H084 H164 H228 H364
7003_HB
Cluster 13: E. faecalis
H075 H155 H255 H355
7003_HF
Cluster 13: E. faecalis
H076 H156 H256 H356
6377_VB
Cluster 14: Streptococcus
H085 H169 H273
6415_HB
Cluster 14: Streptococcus
H079 H167 H367
6556_HB
Cluster 14: Streptococcus
H051 H119 H215 H319
6611_HB
Cluster 14: Streptococcus
H083 H227 H363
6611_HF
Cluster 14: Streptococcus
H084 H164 H228 H364
6611_VB
Cluster 14: Streptococcus
H081 H161 H225 H361
6611_VF
Cluster 14: Streptococcus
H082 H162 H226 H362
6756_HB
Cluster 14: Streptococcus
H071 H151 H251 H351
6843_HB
Cluster 14: Streptococcus
H059 H139 H239 H339
6941_VB
Cluster 14: Streptococcus
H061 H141 H241 H341
7116_HB
Cluster 14: Streptococcus
H159 H259 H359
Relative Abundance
Genus
Acinetobacter
Aerococcus
Bradyrhizobium
Christensenellaceae_R-7_group
Corynebacterium
Cutibacterium
Enterococcus
Facklamia
Jeotgalicoccus
Kocuria
Lachnospiraceae_NK3A20_group
Lactococcus
Others
Paeniclostridium
Romboutsia
Staphylococcus
Streptococcus
Turicibacter
Unidentified_Lachnospiraceae_Family
Unidentified_Peptostreptococcaceae_Family
Weissella
IBC (Dashed line)
IBC (Triangle)
SCC (Dots)
SCC (Solid line)

Figure 8

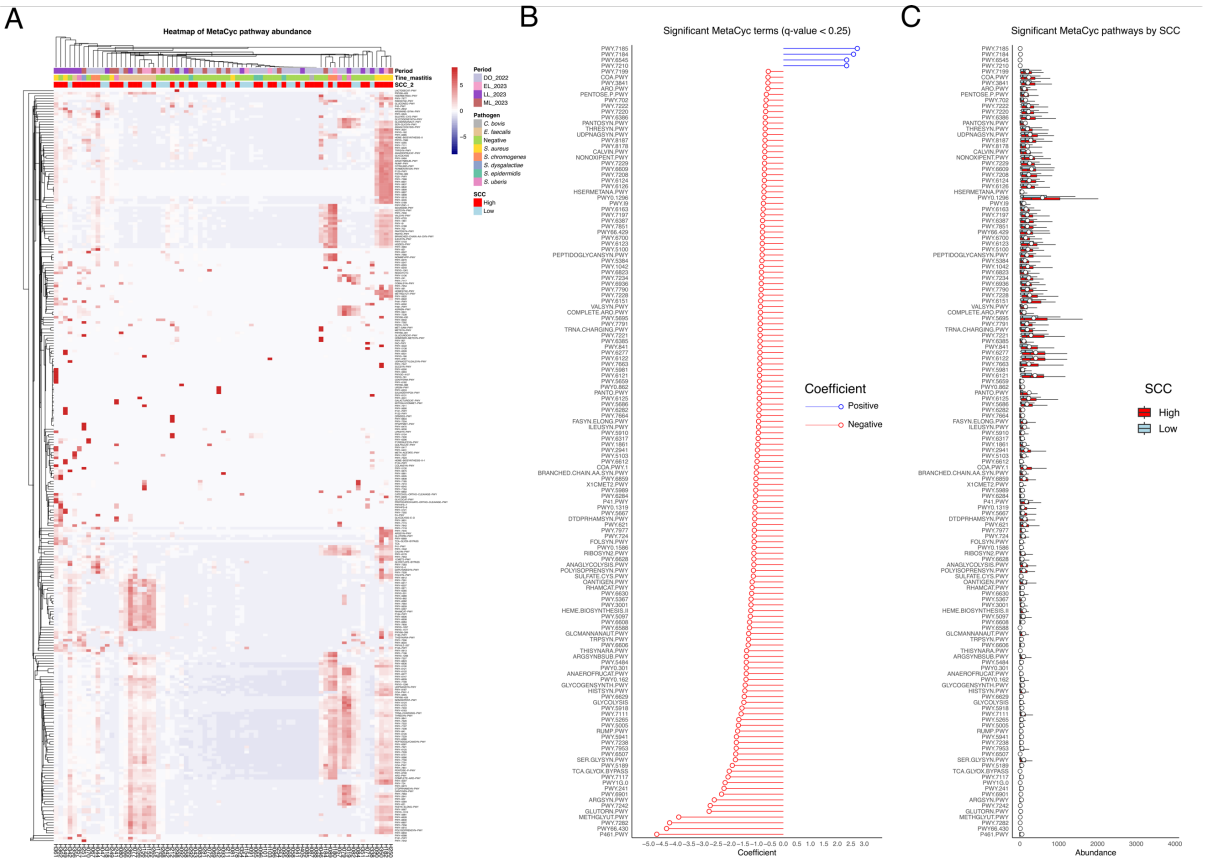
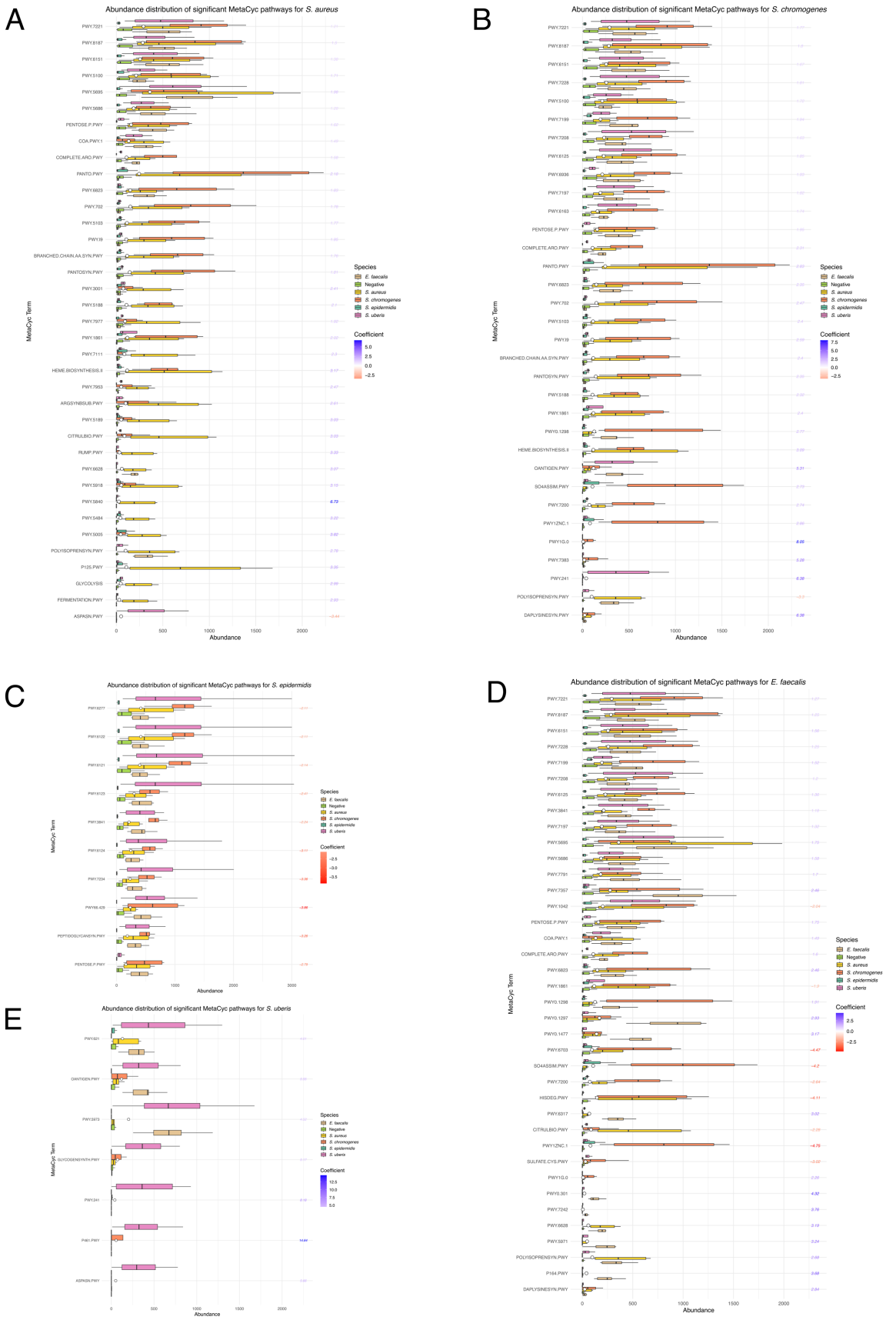
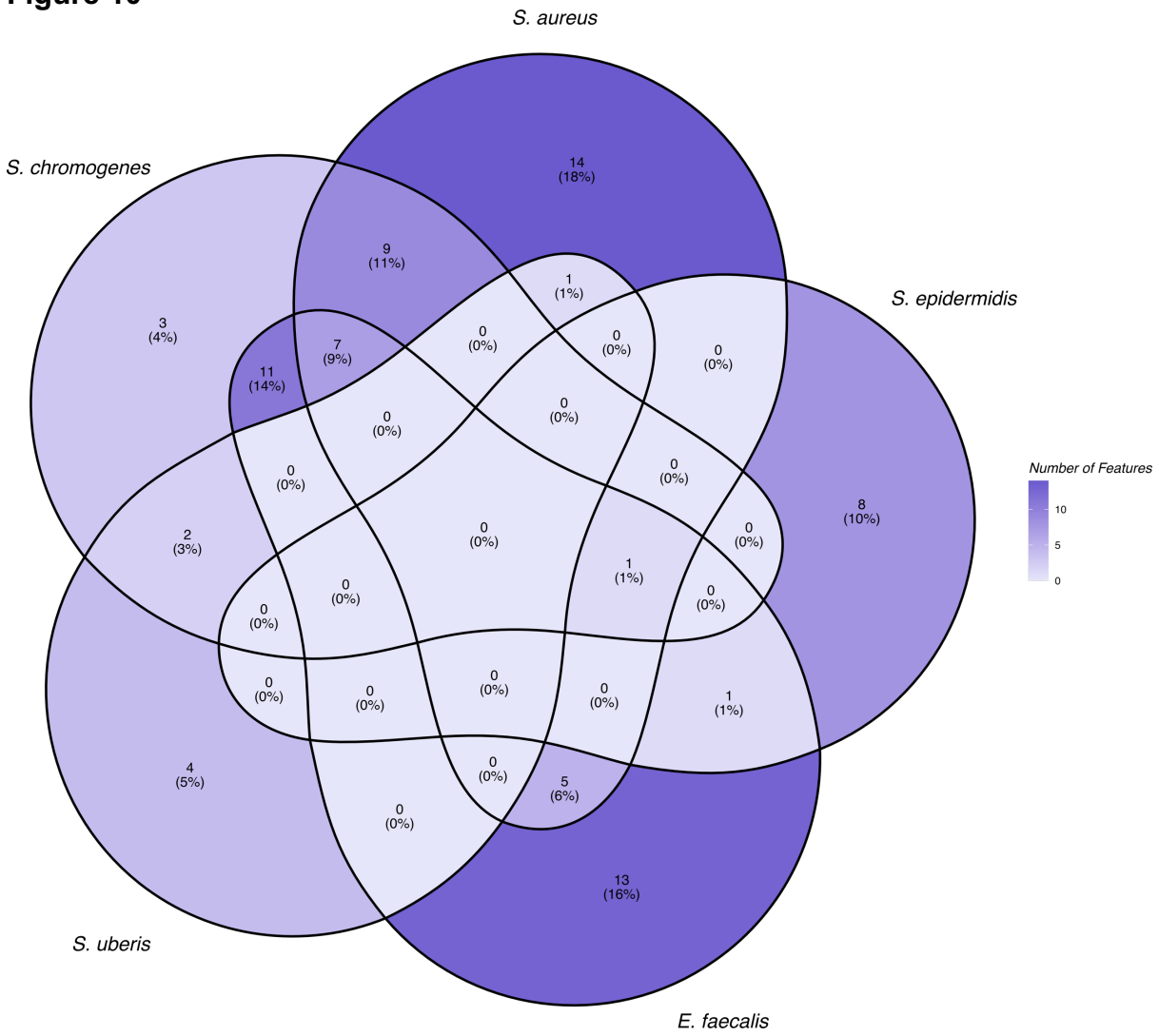

A
Heatmap of MetaCyc pathway abundance
Period
Time_mastitis
SCC_2
Period
DO_2022
EL_2023
LL_2023
ML_2023
Pathogen
C. bovis
E. faecalis
Negative
S. aureus
S. chromogenes
S. dysgalactiae
S. epidermidis
S. uberis
SCC
High
Low
B
Significant MetaCyc terms (q-value < 0.25)
Coefficient
Positive
Negative
Coefficient
C
Significant MetaCyc pathways by SCC
SCC
High
Low
Abundance

Figure 9

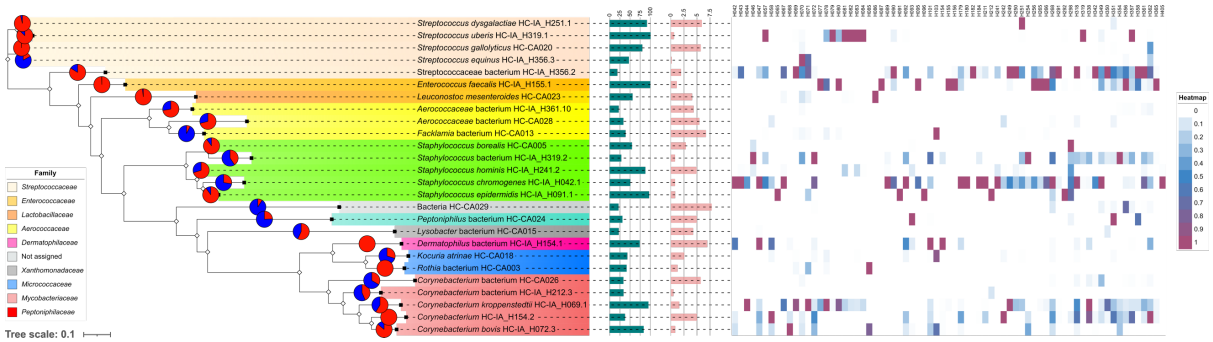

A
Abundance distribution of significant MetaCyc pathways for S. aureus
B
Abundance distribution of significant MetaCyc pathways for S. chromogenes
C
Abundance distribution of significant MetaCyc pathways for S. epidermidis
D
Abundance distribution of significant MetaCyc pathways for E. faecalis
E
Abundance distribution of significant MetaCyc pathways for S. uberis
MetaCyc Term
Abundance
Species
E. faecalis
Negative
S. aureus
S. chromogenes
S. epidermidis
S. uberis
Coefficient

2104 **Figure 10**

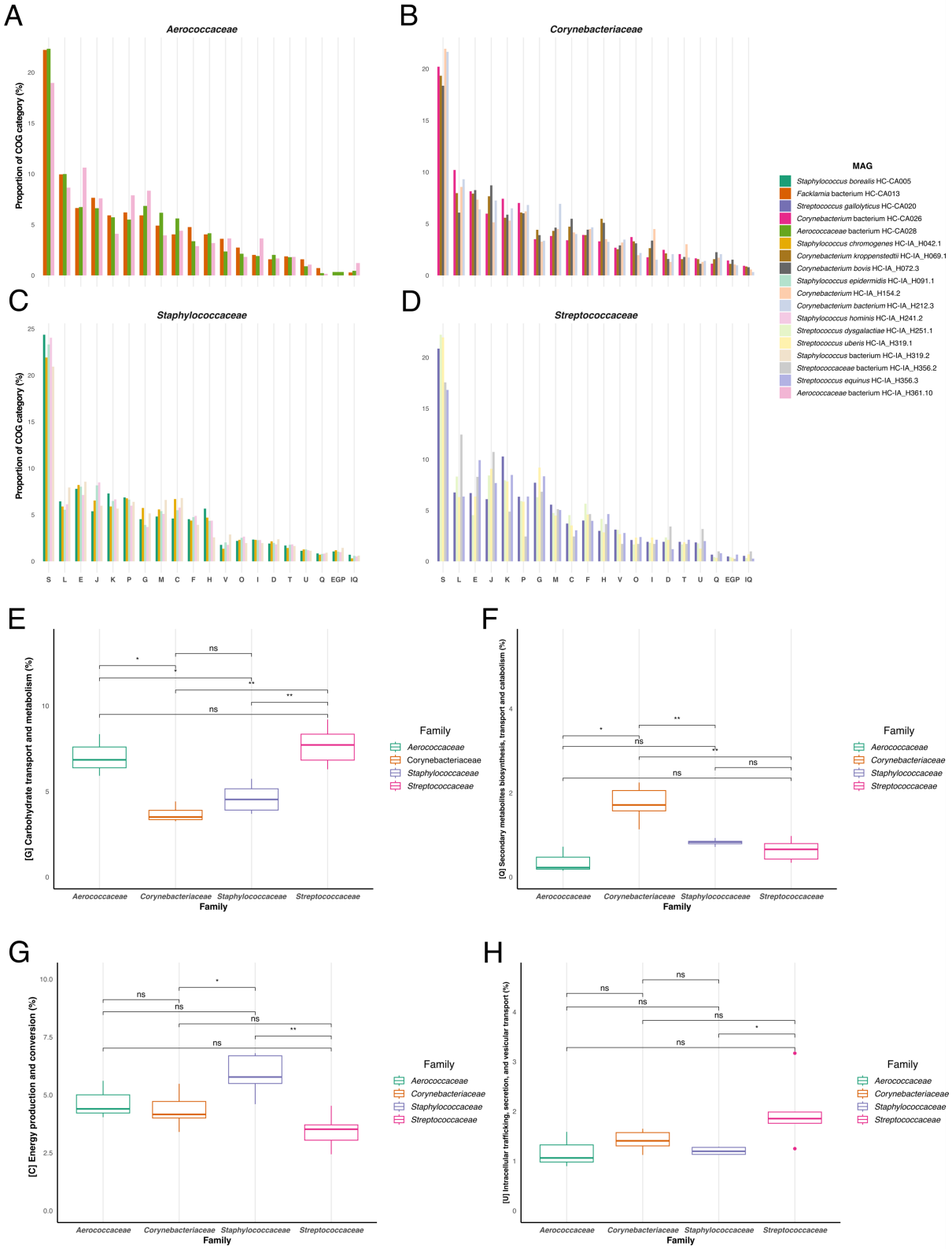


2105
2106
2107

Figure 11


Streptococcus dysgalactiae HC-IA_H251.1
Streptococcus uberis HC-IA_H319.1
Streptococcus gallolyticus HC-CA020
Streptococcus equinus HC-IA_H356.3
Streptococcaceae bacterium HC-IA_H356.2
Enterococcus faecalis HC-IA_H155.1
Leuconostoc mesenteroides HC-CA023
Aerococcaceae bacterium HC-IA_H361.10
Aerococcaceae bacterium HC-CA028
Facklamia bacterium HC-CA013
Staphylococcus borealis HC-CA005
Staphylococcus bacterium HC-IA_H319.2
Staphylococcus hominis HC-IA_H241.2
Staphylococcus chromogenes HC-IA_H042.1
Staphylococcus epidermidis HC-IA_H091.1
Bacteria HC-CA029
Peptoniphilus bacterium HC-CA024
Lysobacter bacterium HC-CA015
Dermatophilus bacterium HC-IA_H154.1
Kocuria atrinae HC-CA018
Rothia bacterium HC-CA003
Corynebacterium bacterium HC-CA026
Corynebacterium bacterium HC-IA_H212.3
Corynebacterium kroppenstedtii HC-IA_H069.1
Corynebacterium HC-IA_H154.2
Corynebacterium bovis HC-IA_H072.3
Family
Streptococcaceae
Enterococcaceae
Lactobacillaceae
Aerococcaceae
Dermatophilaceae
Not assigned
Xanthomonadaceae
Micrococcaceae
Mycobacteriaceae
Peptoniphilaceae
Tree scale: 0.1
Heatmap

2114 **Figure 12**



2115
2116
2117

Figure 13

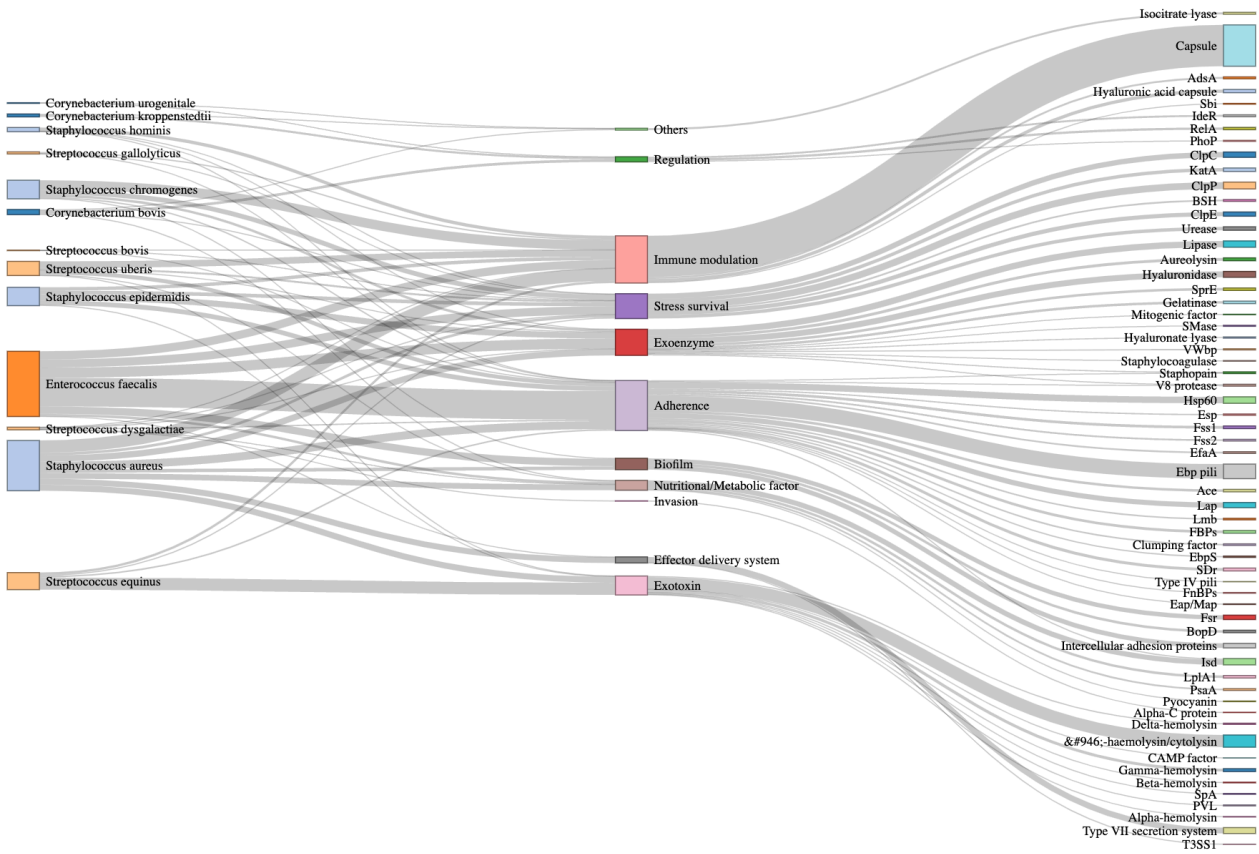
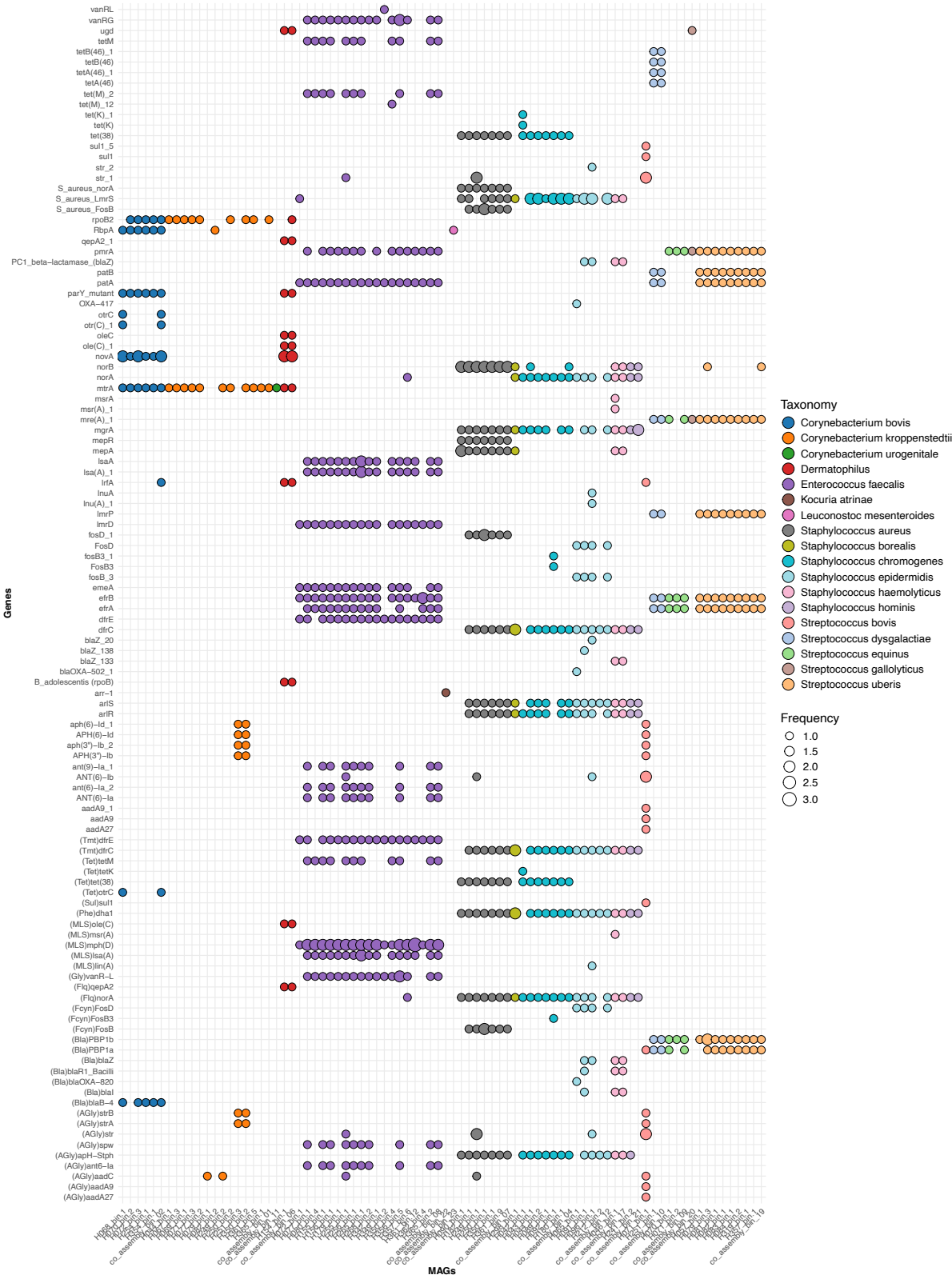

Corynebacterium urogenitale
Corynebacterium kroppenstedtii
Staphylococcus hominis
Streptococcus gallolyticus
Staphylococcus chromogenes
Corynebacterium bovis
Streptococcus bovis
Streptococcus uberis
Staphylococcus epidermidis
Enterococcus faecalis
Streptococcus dysgalactiae
Staphylococcus aureus
Streptococcus equinus
Others
Regulation
Immune modulation
Stress survival
Exoenzyme
Adherence
Biofilm
Nutritional/Metabolic factor
Invasion
Effector delivery system
Exotoxin
Isocitrate lyase
Capsule
AdsA
Hyaluronic acid capsule
Sbi
IdeR
RelA
PhoP
ClpC
KatA
ClpP
BSH
ClpE
Urease
Lipase
Aureolysin
Hyaluronidase
SprE
Gelatinase
Mitogenic factor
SMase
Hyaluronate lyase
VWbp
Staphylocoagulase
Staphopain
V8 protease
Hsp60
Esp
Fss1
Fss2
EfaA
Ebp pili
Ace
Lap
Lmb
FBPs
Clumping factor
EbpS
SDr
Type IV pili
FnBPs
Eap/Map
Fsr
BopD
Intercellular adhesion proteins
Isd
LplA1
PsaA
Pyocyanin
Alpha-C protein
Delta-hemolysin
β-haemolysin/cytolysin
CAMP factor
Gamma-hemolysin
Beta-hemolysin
SpA
PVL
Alpha-hemolysin
Type VII secretion system
T3SS1

Figure 14


Genes
MAGs
Taxonomy
Corynebacterium bovis
Corynebacterium kroppenstedtii
Corynebacterium urogenitale
Dermatophilus
Enterococcus faecalis
Kocuria atrinae
Leuconostoc mesenteroides
Staphylococcus aureus
Staphylococcus borealis
Staphylococcus chromogenes
Staphylococcus epidermidis
Staphylococcus haemolyticus
Staphylococcus hominis
Streptococcus bovis
Streptococcus dysgalactiae
Streptococcus equinus
Streptococcus gallolyticus
Streptococcus uberis
Frequency
1.0
1.5
2.0
2.5
3.0

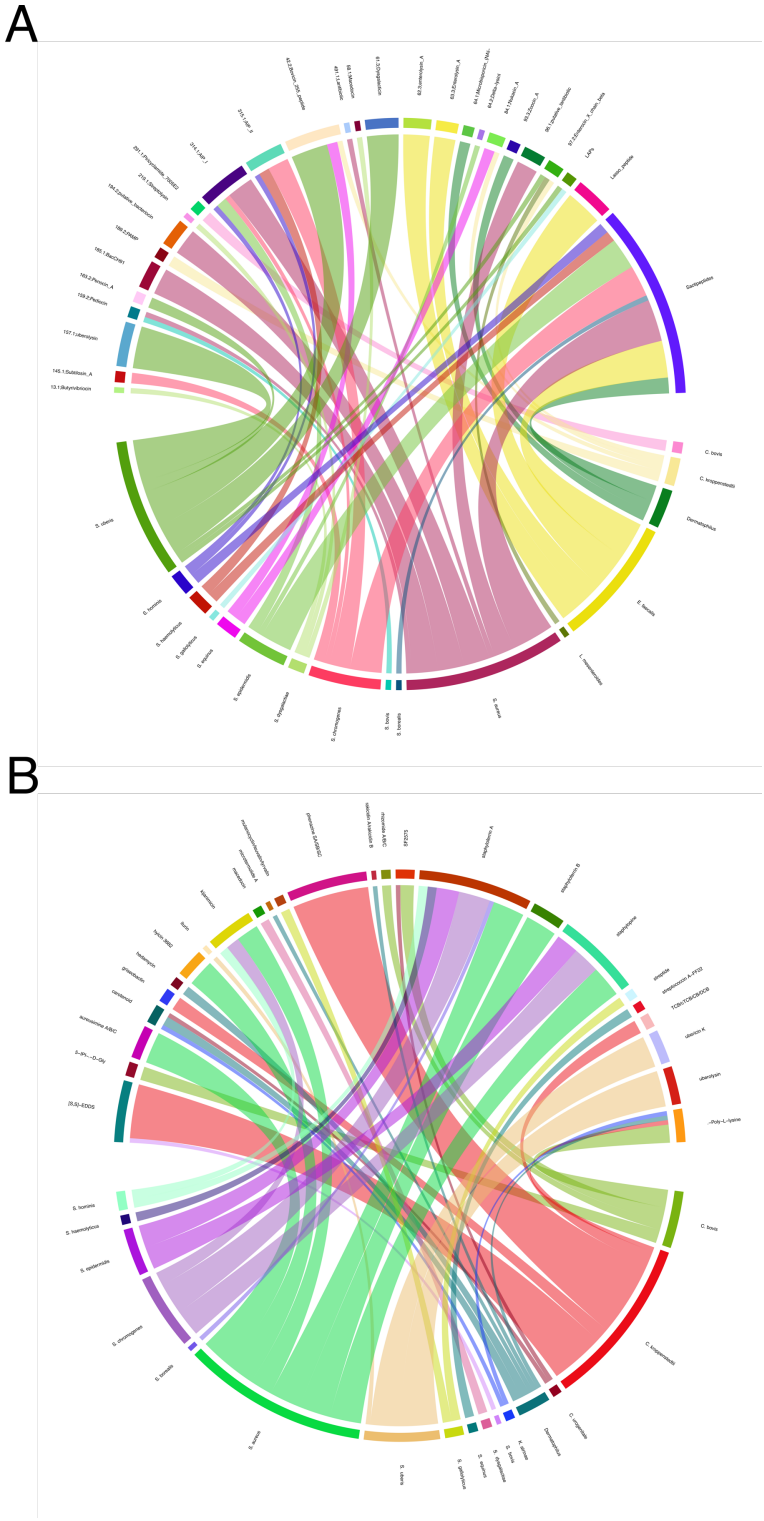



Figure 15