Journal of Dairy Research

cambridge.org/dar

Omics technologies in mastitis: text mining and topic modelling analysis of global research trends

Necati Esener

Bahri Dagdas International Agricultural Research Institute, Konya, Türkiye

Research Article

Cite this article: Esener N (2026) Omics technologies in mastitis: text mining and topic modelling analysis of global research trends. *Journal of Dairy Research* **93**, 117–127. https://doi.org/10.1017/S0022029925101830

Received: 29 July 2025
Revised: 17 October 2025
Accepted: 22 November 2025
First published online: 18 December 2025

Keywords:
latent dirichlet allocation; mastitis; multiomics; text mining; topic analysis

Email: necati.esener@tarimorman.gov.tr

Abstract

Mastitis, an inflammation of the mammary gland, is a disease of significant clinical and economic importance. In recent years, advances in omics technologies have provided powerful tools to unravel the complex biological mechanisms underlying mastitis. These approaches encompass diverse fields such as genomics, proteomics, transcriptomics, metagenomics, metabolomics, epigenomics, lipidomics, glycomics, pharmacogenomics, foodomics, interactomics and exposomics. However, despite the rapid growth of omics research, the thematic structure of this literature has not been systematically examined. In this study, latent dirichlet allocation (LDA) was employed to perform topic modelling on publications related to omics and mastitis retrieved from Scopus and Web of Science. The LDA analysis revealed ten distinct topics, labelled according to the most frequent terms within each cluster: 'proteomics', 'pathogen genomics', 'differential expression', 'metabolism', 'genetic selection', 'disease economy', 'molecular diagnostics', 'microbiome', 'antimicrobial resistance' and 'genetic variation.' Among these, the topics of 'genomics', 'differential expression' and 'antimicrobial resistance' accounted for the highest number of publications, while 'metabolism' emerged more recently. All topics exhibited an increasing trend in publication volume over time, likely driven by the declining costs and greater accessibility of high-throughput omics technologies. This study provides a comprehensive thematic overview of omics research on mastitis, identifies key areas of emphasis and emerging directions, and highlights knowledge gaps that may inform future investigations and the development of targeted strategies for disease control and prevention.

Mastitis is a complex inflammatory condition of the mammary gland, primarily resulting from the invasion of pathogenic microorganisms into the mammary parenchyma (Bourganou *et al.*, 2024). It represents one of the most significant diseases in dairy production systems, leading to substantial economic losses due to decreased milk yield and quality. While mastitis is predominantly associated with livestock, particularly dairy cattle, it can also occur in humans (Contreras and Rodríguez, 2011). Given that bacterial pathogens are the primary causative agents, mastitis is predominantly treated with antimicrobials, making it one of the leading drivers of antimicrobial use on dairy farms (Ruegg, 2017). Mastitis should be addressed within a One Health framework, as it contributes to the emergence and dissemination of antimicrobial resistance genes, posing a potential threat to animal, human, and environmental health (Garcia *et al.*, 2019).

The field of omics encompasses various disciplines, including genomics, transcriptomics, proteomics, and metabolomics, each focusing on distinct biological components such as genomes, transcriptomes, proteomes and metabolomes (Vailati-Riboni *et al.*, 2017). These disciplines study the structure, function and dynamics of their respective biological entities, such as mastitis pathogens, aiming to understand their roles and interactions within cells, tissues and organisms (Rathore *et al.*, 2023; Vailati-Riboni *et al.*, 2017).

The genome encompasses the complete set of DNA in an organism, either the host (such as human or livestock) or the pathogen (Vailati-Riboni *et al.*, 2017). Genome-wide association studies (GWAS), quantitative trait loci (QTL) mapping and single-nucleotide polymorphism (SNP) analysis are fundamental genomic tools used to identify genetic variation and its impact on traits and diseases (Brajnik and Ogorevc, 2023).

The transcriptome comprises the complete set of RNA molecules – including messenger RNA (mRNA), noncoding RNAs (e.g. ribosomal RNA [rRNA], transfer RNA [tRNA]) and regulatory noncoding RNAs – synthesised within a cell, tissue or organism under defined conditions (Wang *et al.*, 2009). It reflects the actively transcribed genes at a specific time point, thereby offering a dynamic representation of the functional state and regulatory activity governing cellular metabolism (Vailati-Riboni *et al.*, 2017). Microarrays (Schena *et al.*, 1995) and RNA-seq (Voelkerding *et al.*, 2009) represent high-throughput technologies to measure transcriptomics profiles.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Hannah Dairy Research Foundation.

The proteome refers to the full complement of proteins expressed within a biological system (e.g. cell, tissue or organism) at a defined time point, encompassing their abundances, post-translational modifications and interaction networks (Chandramouli and Qian, 2009; Wasinger *et al.*, 1995). It provides a dynamic representation of the functional state of a system, reflecting responses to physiological conditions, developmental stages or environmental stimuli (Kim *et al.*, 2011). Proteomic profiling typically involves protein separation via techniques such as one- or two-dimensional polyacrylamide gel electrophoresis (1D/2D-PAGE) or liquid chromatography (LC), followed by mass spectrometry (MS)-based identification and quantification of peptides. While gel electrophoresis resolves proteins for visual comparison, MS provides sequence-level resolution, enabling the characterisation of post-translational modifications and protein isoforms (Aebersold and Mann, 2016).

The metabolome encompasses the complete set of small-molecule metabolites, including metabolic intermediates, hormones and signalling molecules present within a biological system at a specific time point (Patti *et al.*, 2012). These metabolites represent the end products of cellular processes and provide an instantaneous snapshot of the physiological state of the system, reflecting biochemical activity and responses to genetic, environmental or pathological stimuli (Rinschen *et al.*, 2019). Metabolomics, the systematic study of these metabolites, serves as a direct functional readout of cellular physiology (Patti *et al.*, 2012). Metabolomic profiling employs advanced analytical techniques such as nuclear magnetic resonance (NMR) spectroscopy and MS, often coupled with gas or liquid chromatography (GC-MS or LC-MS), to identify and quantify metabolites with high sensitivity and specificity (Alonso *et al.*, 2015).

Beyond those introduced above, a range of other omics technologies have emerged to provide complementary insights into biological systems. Epigenomics investigates genome-wide patterns of heritable chemical modifications, such as DNA methylation and histone acetylation, that regulate gene expression without altering the underlying DNA sequence (Fazzari and Greally, 2004). Metagenomics involves the large-scale analysis of genetic material recovered directly from environmental or host-associated microbial communities to study their composition, function and dynamics (Knight *et al.*, 2018). Lipidomics focuses on the systematic characterisation of lipid species and their metabolic pathways within biological systems, providing insights into lipid-mediated cellular processes (Yang and Han, 2016). Glycomics examines the structure, biosynthesis and functional roles of glycans in biological processes, including cell signalling and immune regulation (Aoki-Kinoshita, 2008). Pharmacogenomics studies how genetic variability influences individual responses to drugs, enabling personalised therapeutic strategies (Evans and Relling, 2004). Foodomics applies omics technologies to investigate food composition, safety and interactions with the consumer (Herrero *et al.*, 2012). Interactomics maps and analyses molecular interaction networks, such as protein–protein or protein–DNA complexes, to elucidate cellular regulatory mechanisms (Vidal *et al.*, 2011). Exposomics characterises cumulative environmental exposures throughout life and their association with disease risk (Wild, 2005).

Text-mining and topic-analysis approaches have been applied to various omics-related fields, including gene expression studies in cancer (Yalamanchili *et al.*, 2017), genomics research (De Filippis *et al.*, 2024; Vailati-Riboni et al., 2017) and microarray analyses (Ribeiro *et al.*, 2024). Despite substantial advances in omics technologies and their application to mastitis research, no systematic effort has been made to characterise the thematic structure of this body of literature using text-mining approaches. In particular, the integration of LDA topic modelling offers a powerful means of identifying underlying themes, emerging trends and research gaps within large collections of scientific texts. Accordingly, the present study aims to provide a comprehensive thematic overview of omics-related mastitis research by applying LDA-based topic modelling to publications retrieved from major scientific databases. The specific objectives are to (i) identify and categorise the principal research themes, (ii) determine long-standing and emerging areas of focus, (iii) highlight knowledge gaps and opportunities for future investigation and (iv) offer an integrated framework that can guide researchers, veterinarians, and policymakers in shaping the future trajectory of omics-driven mastitis research.

Material and methods

Dataset

A comprehensive literature search was conducted in two scientific databases: Web of Science (WoS) (Birkle *et al.*, 2020) and Scopus (Baas *et al.*, 2020). The search was performed separately in each database on 28 February 2025. For WoS, the keyword search was conducted using the title, abstract and keyword fields. To ensure consistency, the search in Scopus was performed using the Title and Abstract (tiab) field, which also screens the title, abstract, and keyword fields.

The keywords used in search were ('omic* technolog*' OR 'genomic*' OR 'epigenomic*' OR 'proteomic*' OR 'transcriptomic*' OR 'metabolomic*' OR 'metagenomic*' OR 'metataxonomic*' OR 'microbiomic*' OR 'metabonomic*' OR 'foodomic*' OR 'lipidomic*' OR 'glycomic*' OR 'pharmacogenomic*' OR 'toxicogenomic*' OR 'interactomic*' OR 'exposomic*' OR 'multiomic*' OR 'host-microbiomic*') AND (mastitis). The search was refined to include only original research articles written in English. No additional exclusion criteria were applied. The results from each database search were combined, and duplicates were removed. Furthermore, articles lacking essential information, such as abstracts, authors or titles were manually excluded. During this manual review, review articles and non-English publications were also removed. The remaining documents underwent further analysis, which included bibliometrics, text mining and topic analysis.

Bibliometrics analysis

The bibliometric analysis covered the annual scientific production, the most prolific countries, the most cited documents, the most relevant sources and affiliations. These analyses were conducted using the Bibliometrix library (Aria and Cuccurullo, 2017) in the R programming language (R Core Team, 2019).

Text mining

Text mining was performed in the R programming language using the following packages: 'tm', 'topicmodels', 'tidyverse', 'tidytext', 'ggplot2', 'broom' and 'wordcloud'.

To ensure consistency and improve the quality of text data for analysis, the following pre-processing steps were applied:

- Lowercasing: All text was converted to lowercase to eliminate case sensitivity and standardise word forms.

https://doi.org/10.1017/S0022029925101830 Published online by Cambridge University Press

- Punctuation Removal: All punctuation marks were removed to focus on meaningful textual content.
- Number Removal: Numerical values were eliminated to avoid potential biases from non-informative numerical data.
- Whitespace Normalisation: Excess whitespace, including multiple consecutive spaces and leading or trailing spaces, was removed for uniformity.
- Stopword Removal: A predefined list of common English stopwords was removed to filter out non-informative words. Additionally, a custom list of domain-specific stopwords (e.g. 'mastitis', 'genomic', 'study', 'result') was incorporated to eliminate words frequently occurring in the dataset but lacking analytical significance.
- Stemming: Words were reduced to their root forms using stemming techniques to consolidate different word variants.

After text pre-processing, a Document-Term Matrix (DTM) was constructed using the 'tm' package in R. The DTM represents the frequency of words across documents, enabling subsequent analysis. To enhance the relevance of term contributions, Term Frequency-Inverse Document Frequency (TF-IDF) weighting was applied. This transformation adjusts raw term frequencies by penalising commonly occurring words while amplifying the importance of rare yet meaningful terms, ensuring a more informative representation of textual data.

Topic analysis

LDA was employed to uncover latent thematic structures within the textual data. LDA is a probabilistic generative model that assumes each document is composed of a mixture of topics, where each topic is characterised by a distinct set of words. The implementation was carried out using the 'topicmodels' package in R.

The number of topics (K) is a crucial hyperparameter in LDA modelling, significantly influencing the interpretability and coherence of the extracted topics. Optimal K values were defined within the range of K values (between 2 and 20, with increments of 2), based on perplexity-based evaluation, coherence-based metrics and author interpretation.

Once the optimum value is determined, the probable topic names are estimated. Subsequently, for each identified topic, a comprehensive bibliometric analysis was performed to determine the first publication date, the most prolific country, and the primary journal of publication. As suggested by Trapanese *et al.* (2025), trend analysis was performed by the coefficient of determination (R^2) and considering p-value < 0.05 for statistical significance.

Intertopic distance maps were generated using the LDAvis tool (Sievert and Shirley, 2014), which applies multidimensional scaling to visualise topic distributions. Additionally, a heatmap illustrating topic clusters and temporal trends in omics studies on mastitis was constructed using the 'pheatmap' package (Kolde and Kolde, 2015) in R.

Results

Dataset

The keyword search retrieved 1,196 documents from the Scopus database and 1,051 documents from the WoS database. Applying language and document type filters (English language and research articles) reduced the dataset to 1,011 documents in Scopus and 929 documents in WoS. After merging both datasets, 708 duplicate records were identified and removed. A subsequent manual screening led to the exclusion of 14 additional documents. Following these pre-processing steps, a total of 1,218 documents remained for further analysis.

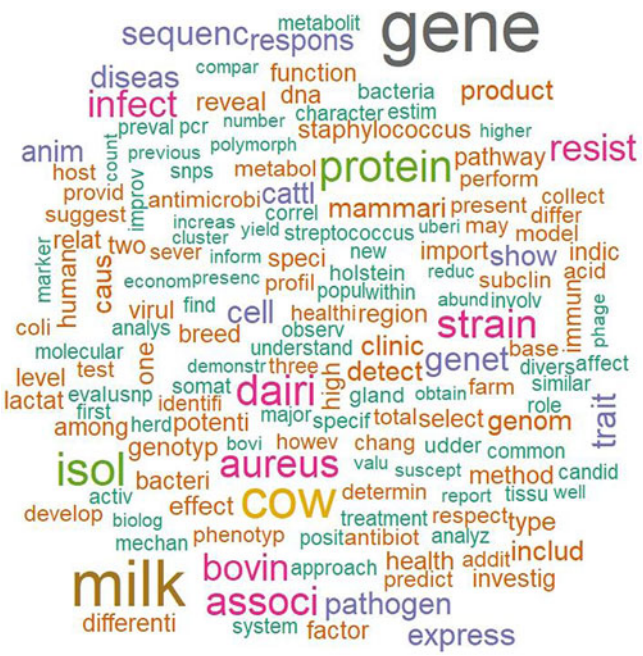


Figure 1. Word cloud of the most frequent terms. The figure illustrates the most commonly occurring words, with font size proportional to their frequency. The minimum frequency threshold for inclusion was set at 200 occurrences.

Text mining

After the pre-processing stage, the most frequent words were shown in Fig. 1. Terms such as 'gene', 'milk', 'cow', 'isol', 'protein' and 'dairi' appear in larger font sizes, highlighting their prominence in the documents. It is worth noting that stemming algorithms reduce words to their root forms by removing common suffixes. This process is rule-based and does not aim to produce linguistically meaningful or dictionary-valid words. As a result, stems may appear as 'non-words' or truncated forms.

Topic analysis

The optimal number of topics ($K = 6$, 10, and 12) was determined based on perplexity and coherence metrics (see Supplementary Figure 1 and 2). Among these, $K = 10$ was identified as the most suitable value, and topic names were assigned based on the top ten words associated with each topic. The identified topics related to omics studies on mastitis are shown in Fig. 2.

Table 1 shows ten research topics identified using LDA, along with the number of publications, the first year of publication, publication trends over time, and the most prolific country and journal for each topic. China was the most prolific country in all topics except Topic 1 (Proteomics), where the United States led with 18 publications. In Topic 6 (Disease Economy), China and the United States had an equal number of publications (11 each). Most research topics were first published in the early 1990s, except for Topic 4 (Metabolism) and Topic 6, which have been studied more

https://doi.org/10.1017/S0022029925101830 Published online by Cambridge University Press

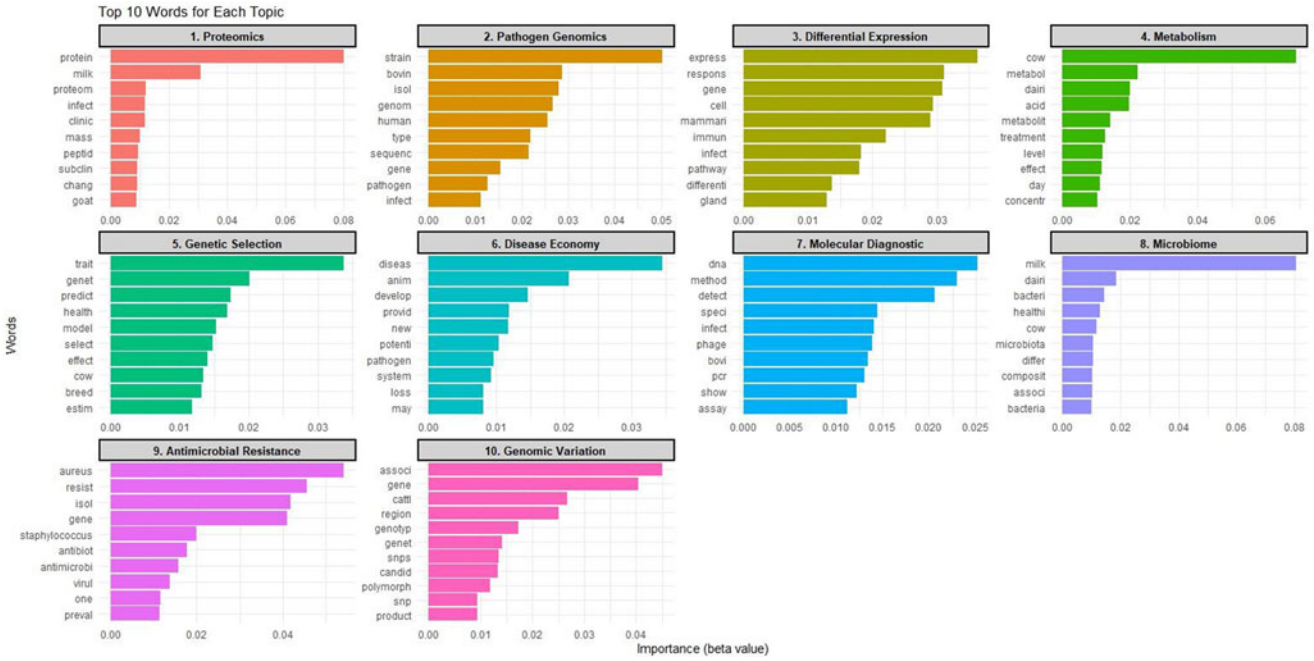


Figure 2. Top 10 words for each identified topic in omics studies on mastitis. The bar plots represent the most important words for each topic, ranked by their beta values derived from latent dirichlet allocation (LDA) modelling.

Table 1. Trends in research topics related to omics studies on mastitis. The table presents the number of publications, the first year of publication, the coefficient of determination (R^2) indicating the trend in publication volume, statistical significance (P-value), the most prolific country, and the most related journal for each topic

Topic	Number of publications (%)	First publication year	R^2 and trend direction	*P*-value	Most prolific country (number of publications)	Most related journal (number of publication)
1. Proteomics	117 (9.61%)	1994	0.53+	<0.001	USA (18)	*Journal of Dairy Science* (15)
2. Pathogen genomics	181 (14.86%)	1991	0.57+	<0.001	China (36)	*Journal of Dairy Science* & *BMC Genomics* (8)
3. Differential expression	154 (12.64%)	1992	0.69+	<0.001	China (25)	*Journal of Dairy Science* (20)
4. Metabolism	100 (8.21%)	2002	0.55+	<0.001	China (24)	*Frontiers in Microbiology* (8)
5. Genetic selection	108 (8.87%)	1993	0.42+	<0.001	China (32)	*Journal of Dairy Science* (9)
6. Disease economy	60 (4.93%)	1999	0.29+	<0.005	China & USA (11)	*Journal of Dairy Science* & *Veterinary Microbiology* (6)
7. Molecular diagnostics	108 (8.87%)	1990	0.44+	<0.001	China (26)	*Animals* (7)
8. Microbiome	105 (8.62%)	1992	0.41+	<0.001	China (21)	*Journal of Dairy Science* (9)
9. Antimicrobial resistance	151 (12.40%)	1991	0.53+	<0.001	China (30)	*Journal of Dairy Science* (20)
10. Genetic variation	134 (11.00%)	1991	0.49+	<0.001	China (28)	*Journal of Dairy Science* (11)

recently, with their first publications appearing in 2002 and 1999, respectively.

An increasing trend in publication volume was observed across all topics since their initial publication years. The Journal of Dairy Science was the most frequently cited, being the primary publication source for Topics 1, 2, 3, 5, 6, 8, 9 and 10. In Topic 2 (Pathogen Genomics) and Topic 6, BMC Genomics and Veterinary Microbiology, respectively, had publication counts comparable

https://doi.org/10.1017/S0022029925101830 Published online by Cambridge University Press

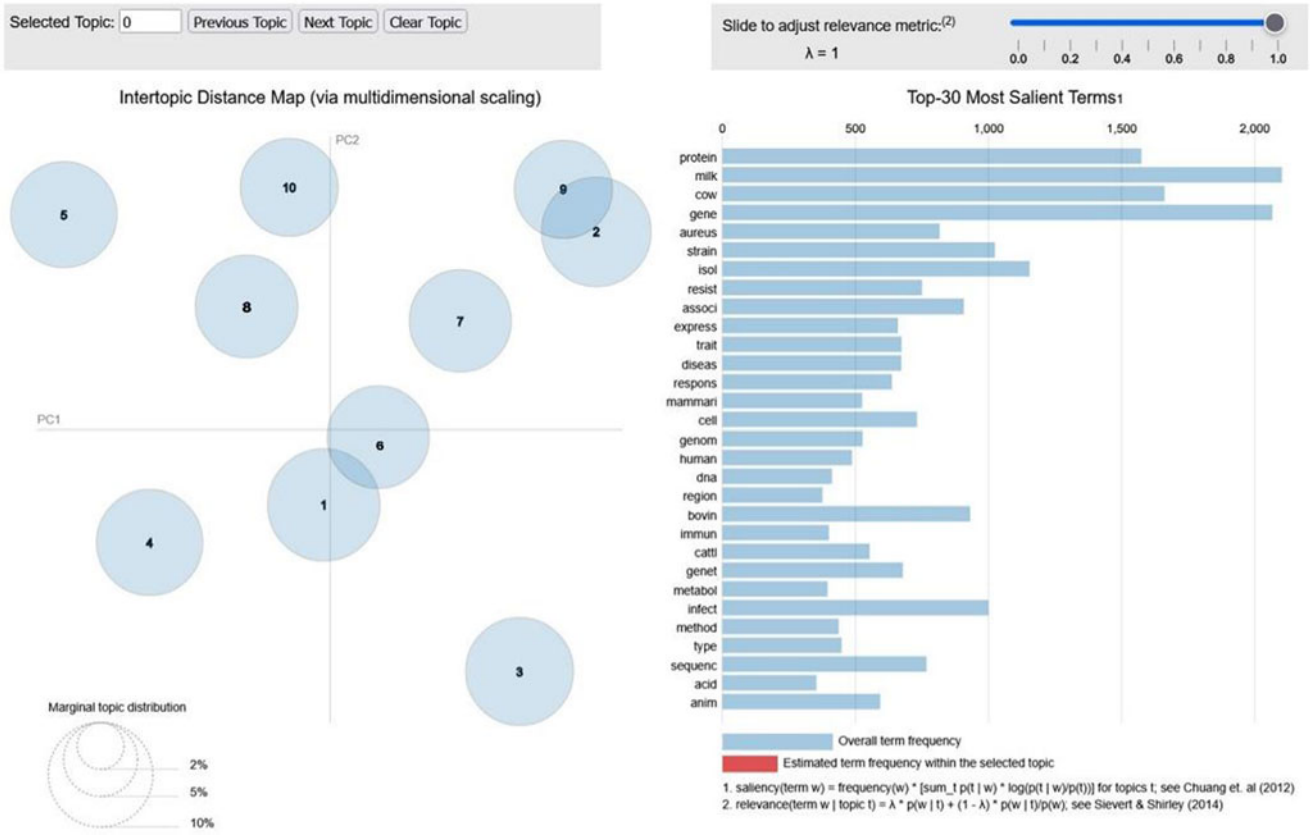


Figure 3. Visualisation of latent dirichlet allocation (LDA) topic modelling results. The left panel displays the intertopic distance map generated using multidimensional scaling (MDS), where each bubble represents a topic, and the bubble's size corresponds to the topic's prevalence across the corpus. Topics are positioned based on their similarity, with overlapping or closely located bubbles indicating related topics. The right panel shows the top 30 most salient terms across all topics, ranked by their overall term frequency in the corpus (blue bars).

to the Journal of Dairy Science. Frontiers in Microbiology and Animals were the most prolific journals for Topics 4 and 7, respectively.

Figure 3 presents the intertopic distance map of omics studies on mastitis based on key terms and their frequency. Among the identified topics, Topic 3 was the most distinct, positioned furthest from the others. In contrast, Topics 6 and 1, as well as Topics 2 and 9, exhibited significant overlap, indicating shared thematic elements. Additionally, the top 30 most salient terms for each topic, along with their proportional representation, are provided in Supplementary Figures 3–12.

Figure 4 illustrates topic clusters representing trends over time in omics studies on mastitis. Among the hierarchical clusters, Topic 6 was the most distant from the others. In contrast, Topics 1 and 8, Topics 3 and 10, Topics 7 and 9, as well as Topics 4 and 5 formed closer clusters. Several topics, including Topics 2, 3, 7, 9 and 10, exhibited a notable increasing trend in recent years, particularly between 2020 and 2025.

Discussion

The current study analysed the omics studies on mastitis by using text mining and topic analysis techniques. The omics studies on mastitis were initiated in the 1990s and have shown to increase over the years. This trend is attributed to the increasing accessibility of next-generation sequencing (NGS) technologies, driven by their growing popularity and decreasing costs (Shendure and Ji, 2008). Furthermore, the growing expertise in bioinformatics within veterinary science may have also contributed to this outcome. The increased trend of omics studies in veterinary medicine was seen in previous studies as well (Bourganou *et al.*, 2024; Esener, 2025).

The most frequent words included 'gene', 'protein', 'cow', 'milk', 'dairi', 'isol' and 'aureus'. The term gene must have been related to genomics and transcriptomics, as these are the most frequently used omics studies. Protein must show the importance of proteomics. As stated earlier, bovine mastitis is the biggest issue in the dairy industry; hence, 'dairi' and 'cow' must have been that frequent. Terms like 'isol', 'strain' and 'aureus' must be related to mastitis agents as they are mainly bacteria. The virulence and antimicrobial resistance genes of mastitis pathogens such as *Staphylococcus aureus* might have been studied intensively.

Omics approaches beyond genomics, transcriptomics, and proteomics appear to be underrepresented in the current literature on mastitis. For instance, Luo *et al.* (2023) employed an integrated analysis combining 16S rRNA sequencing and lipidomics to investigate microbiota diversity and milk quality, demonstrating that specific microbial populations are associated with the regulation of certain lipid profiles. Similarly, Hasankhani *et al.* (2023) explored the molecular interactome involved in intramammary inflammation caused by *Streptococcus uberis* in dairy cows. However, research explicitly addressing other omics disciplines – such as glycomics, pharmacogenomics, foodomics and exposomics – in the

https://doi.org/10.1017/S0022029925101830 Published online by Cambridge University Press

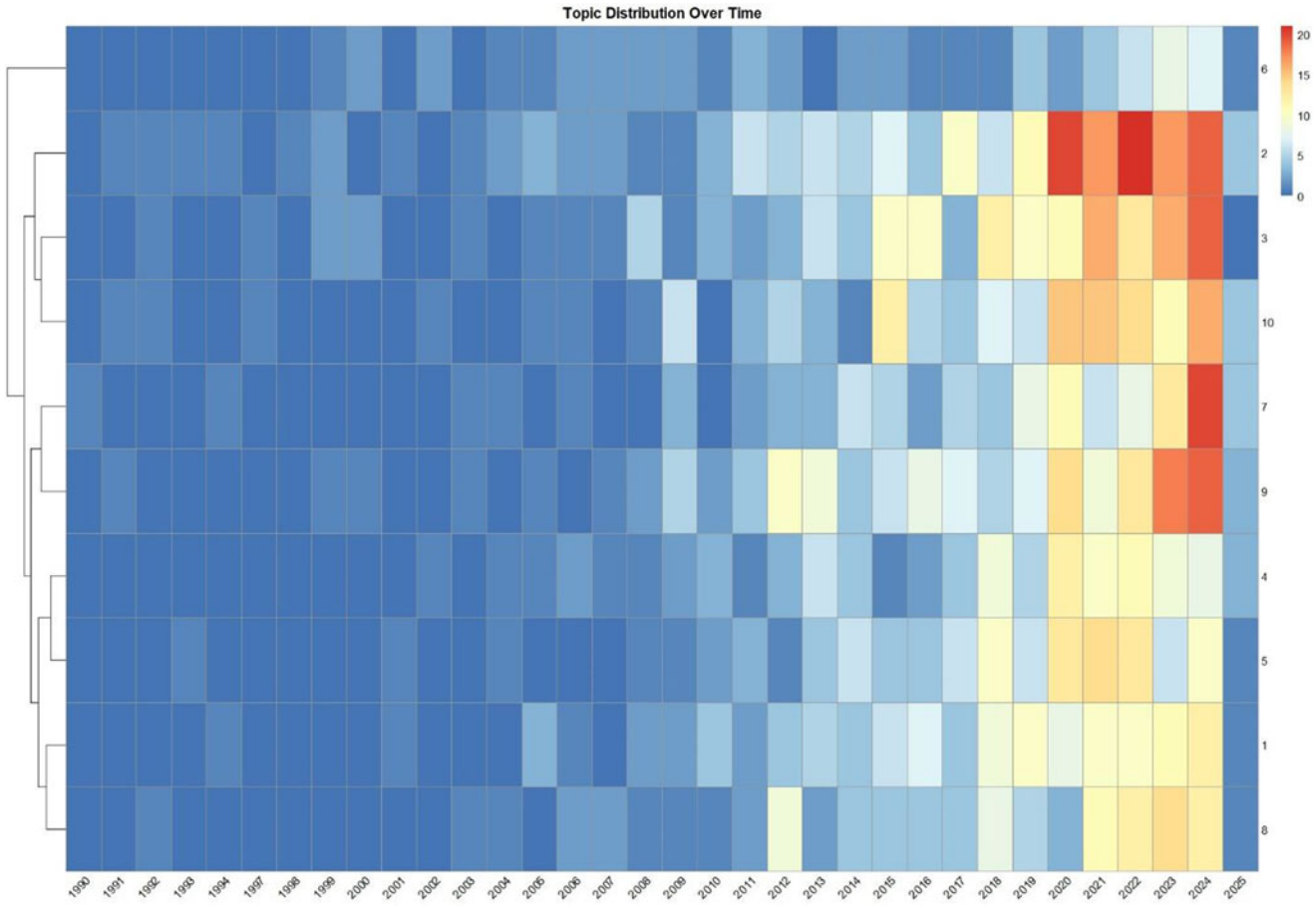


Figure 4. Topic clustering for trends over time in omics studies on mastitis. The heatmap illustrates the trend of topic prevalence across years from 1990 to 2025. The right y-axis represents topic numbers, while the x-axis shows publication years. The colour gradient ranges from blue (low prevalence) to red (high prevalence), indicating the relative frequency of each topic in a given year. Hierarchical clustering on the left highlights similarities between topics based on their temporal patterns.

context of mastitis was not identified, aside from a few instances where these terms appeared as keywords without direct investigation.

The keywords such as 'protein', 'proteom', 'peptid', 'mass' and 'spectrometri' showed that Topic 1 was related to 'Proteomics' (see Fig. 2). The studies found in this topic were mainly related to inflammatory responses in cows (Danielsen *et al.*, 2010), goats (Olumee-Shabon *et al.*, 2013), buffalos (Das *et al.*, 2008) and yaks (Li *et al.*, 2014) milk proteome. These studies mainly targeted to discover potential vaccine antigens for intramammary infections such as *Streptococcus uberis* (Collado *et al.*, 2016), *S. aureus* and *Escherichia coli*. Moreover, regulation of the proteins during antimicrobial-resistant pathogen infections such as Tylosin resistance of *Staphylococcus xylosus* (Liu *et al.*, 2019), during the treatment of the intramammary infection (Kang *et al.*, 2016) was also studied. This topic was the only one where China was not the most prolific country, but the USA. This also accords with the previous scientometrics of proteomic technologies on ruminant mastitis (Bourganou *et al.*, 2024).

Topic 2 had keywords such as 'strain', 'isol', 'genom', 'type' and 'sequence' that are terms related to mastitis pathogen genomics (see Fig. 2). Draft genomes of *Streptococcus agalactiae* (Murungi *et al.*, 2022) and *E. coli* (Rahman Mohammad *et al.*, 2023) were among the studies on this topic. Moreover, whole genome sequencing (WGS) studies of *S. aureus* (Greening *et al.*, 2021), *Mycoplasma arginini* (Gioia *et al.*, 2023), *S. agalactiae* (Oliveira *et al.*, 2024), *E. coli, S. uberis* (Silva *et al.*, 2021), *Actinobaculum suis* (Amigo *et al.*, 2019) were also listed in this topic. Park *et al.* (2022) studied pangenome analysis of *S. aureus*. Furthermore, comparative genomic studies of *Enterococcus faecium* and *Enterococcus faecalis* (Apostolakos *et al.*, 2022) were performed. Nonetheless, this group was characterised by the presence of not only pathogenic bacterial genomics but also the potential probiotic *Pediococcus pentosaceus* (Rahman *et al.*, 2025) and the zoonotic fungal species *Candida tropicalis* (Jiang *et al.*, 2024). This topic has the highest number of papers (181 articles). This is plausible because genomics represents the foundational milestone of the omics era and remains the most extensively developed and established discipline within the omics sciences (Nehme *et al.*, 2018). Moreover, genomics for mastitis can be studied from two perspectives, either host or pathogen; the pathogen genome is more convenient (i.e. smaller genomes and lower costs) to study (Donkor, 2013).

Topic 9 was found to be overlapping with Topic 2 on the intertopic distance map, where keywords such as 'aureus', 'resist', 'isol', 'gene', 'staphylococcus', 'antibiot', 'antimicrobial' and 'virul' were the main terms (see Fig. 2). Therefore, it was defined as 'Antimicrobial resistance'. The observed overlap between Topic 2 ('pathogen genomics') and Topic 9 ('antimicrobial resistance') in the intertopic distance map (see Fig. 3) is expected, as both topics are intrinsically connected through the genomic study of

https://doi.org/10.1017/S0022029925101830 Published online by Cambridge University Press

pathogenic microorganisms. Research efforts have predominantly concentrated on antimicrobial resistance and virulence genes in *Staphylococcus* spp. (Rodríguez *et al.*, 2023), especially *S. aureus* (Zaatout *et al.*, 2019) and *Streptococcus* spp. including *S. pyogenes, S. agalactiae, S. porcinus* and *S. uberis* (Vidal Amaral *et al.*, 2022). Other studies screened the antimicrobial resistance genes in *E. coli* (Alawneh *et al.*, 2020), *Klebsiella pneumaniae* (Y Yang *et al.*, 2019) and *Brucella* spp. (Grützke *et al.*, 2021). Hoque *et al.* (2020) screened the antimicrobial resistance profiles of different pathological conditions of mastitis at the microbiome level.

Topic 3 included studies that explored differentially expressed genes and proteins. According to the intertopic distance map (see Fig. 3), this topic appeared to be the most distant from the others, indicating the most unique terms in this topic. In a study with ewes, different somatic cell scores (high and low) and energy balance groups (negative and positive) were compared by using RNA-seq data in terms of mastitis resistance (Bouvier-Muller *et al.*, 2018). There were examples of studies comparing the expression differences of both genes and proteins between healthy and mastitic dairy cows (Olumee-Shabon *et al.*, 2013; Sun *et al.*, 2019). Although these studies were mainly with *E. coli* infection (Moyes *et al.*, 2016), there were also other mastitis pathogens such as *Prototheca* spp. and *S. agalactiae* (Bisutti *et al.*, 2023; Zhang *et al.*, 2018). Moreover, (Chen *et al.*, 2018) compared the differentially expressed proteins from the healthy and *S. aureus*-infected tissues of the same individual dairy cows. In another study, Wang *et al.* (2020) compared the epigenetic changes in the expression of *S. aureus*-induced and healthy mammary gland quarters. In this topic, there were also studies enabling to understand mastitis resistance in depth, such as detecting the differentially expressed long non-coding RNAs in dairy cows (Asselstine *et al.*, 2024). Furthermore, low susceptibility to intramammary infections and low prevalence of mastitis in donkeys were also studied by using RNA-seq (Wang *et al.*, 2024).

The relevant terms in Topic 4 suggest that this topic is best described under the umbrella of 'metabolism' (see Fig. 2). The studies related to mammary gland metabolism of both ovine (Chiaradia *et al.*, 2013) and bovine (Ying *et al.*, 2021) regarding mastitis were listed in this topic. A study on Canadian dairy cows demonstrated distinct differences in blood metabolomic profiles between mastitic and healthy individuals, highlighting their potential application in the early identification of animals susceptible to mastitis (Haxhiaj *et al.*, 2022). Furthermore, metabolomic differences by applying alternative treatment techniques for women with mastitis were also shown (Vázquez-Fresno *et al.*, 2014). There were also studies related to the carbon and energy metabolism of mastitis pathogens listed in this topic (Li *et al.*, 2024). This topic appeared to be the most recent, as indicated by the timing of the first publications, suggesting that metabolite-focused studies on mastitis were initiated later compared to other omics disciplines. This aligns with the historical timeline of omics technologies, as the term 'metabolome' was first introduced in 1998 (Arıkan and Muth, 2023).

In Topic 5, researches performed with techniques such as GWAS and WGS of cattle were included. These studies revealed the genes related to milk quality and production (Liu *et al.*, 2020), mastitis resistance (F Yang *et al.*, 2019) and dairy-related other traits (Marete *et al.*, 2018). Local cattle breeds were also studied based on production-wise important traits (Schmidtmann *et al.*, 2021). A study conducted on the Holstein breed compared mastitis-resistant and susceptible dairy cows, identifying 117 candidate SNPs and 27 QTLs that may serve as potential markers for selective breeding programs (Kurz *et al.*, 2019). In a comprehensive study with 16 US dairy farms, breeding values of dairy-related diseases, including mastitis were predicted by using the Bayesian regression model (Lopes *et al.*, 2020).

Topic 6 was found to be slightly overlapping with Topic 1, as shown on the intertopic distance map (see Fig. 3). Relevant keywords for this topic, such as 'disease', 'anim' and 'pathogen' etc. might explain mastitis and its agents. Keywords such as 'loss', 'econom' and 'industry' might be related to the cost of the disease in the dairy industry. The articles clustered on this topic support this claim as there were mainly studies mentioning the economic losses in dairy farms due to mastitis (Fenske *et al.*, 2022; Goldstone *et al.*, 2016; Opsal *et al.*, 2006; Taverna *et al.*, 2007). Another study, which offered potential Multiple Loci Variable Number of Tandem Repeats Analysis for *S. uberis* as a low-cost typing technique was located on this topic (Gilbert *et al.*, 2006). Topic 6 has the least number of papers (60 articles) among all these topics. The reason for this might be that although mastitis is a well-known issue for the dairy sector, there were limited omics studies combined with its economy.

Topic 7 was named as 'Molecular Diagnostic' based on the frequent use of 'dna', 'method', 'detect', 'pcr', 'assay' and 'molecular'. In this topic, researchers have used several molecular techniques, such as restriction fragment length polymorphism (RFLP) (Musser *et al.*, 1990), pulsed-field gel electrophoresis (Fijałkowski *et al.*, 2013), polymerase chain reaction (PCR; Matthews *et al.*, 1994), short (Rahman *et al.*, 2024) and long-read sequencing technologies (White *et al.*, 2024), for the detection, identification and typing of mastitis pathogens.

The frequent keywords such as 'milk', 'dairy', 'bacteria', 'microbiota' and 'composition' in Topic 8 lead to 'Microbiome'. Some examples of research related to this topic were milk microbiome characterisation for dairy ewes (Castro *et al.*, 2019) and cows (Gao *et al.*, 2020). Moreover, some studies screened specific mastitis agents in the bulk milk samples (Patel *et al.*, 2017; Xu *et al.*, 2021). A few examples of hindgut (Vasco *et al.*, 2024) and ruminal (Chang *et al.*, 2023) microbiota studies were also found on this topic. Moreover, there were not only animal microbiome studies but also breast microbiota that were compared between non-puerperal mastitis patients and healthy females (Zhu *et al.*, 2022).

Topic 10 was related to polymorphism studies. Unlike Topic 5, there were mainly studies related to mastitis pathogens clustered. Some of the studies were as follows: *S. aureus* was characterised by random amplified polymorphic DNA-PCR (RAPD-PCR) (Myllys *et al.*, 1997), fluorescence amplified fragment length polymorphism (fAFLP) (Trovó Fabiano *et al.*, 2005) and other typing techniques (Vautor *et al.*, 2009), *S. uberis* was differentiated by RFLP (Jayarao *et al.*, 1991), Coagulase-Negative Staphylococci was isolated by RFLP-PCR (Melo *et al.*, 2018). Studies using SNPs for *S. aureus* (Huang *et al.*, 2024), *S. agalactiae* (Wang *et al.*, 2015) and *Mycoplasma bovis* (Kumar *et al.*, 2020) were also among the clustered articles on this topic.

Although two major bibliographic databases, Scopus and Web of Science, were utilised in this study, a key limitation was the exclusion of other relevant databases, such as PubMed and Cochrane, which may have resulted in the omission of omics-related mastitis studies indexed exclusively in those sources. Another constraint was that only articles published in English were considered, potentially limiting the inclusion of valuable regional research published in other languages. Despite a meticulous manual screening process, it is possible that some studies that did not fully meet the inclusion criteria – such as review articles or papers not directly addressing the omics-based investigation of mastitis (e.g. studies

https://doi.org/10.1017/S0022029925101830 Published online by Cambridge University Press

related to dairy products rather than mastitis) – remained in the dataset. Moreover, a critical limitation stems from the broader issue of publication bias, where studies reporting statistically significant findings are more likely to be published, while non-significant results are underreported (Rosenthal, 1979), potentially skewing the thematic distribution observed in this type of analysis.

In conclusion, this study provides a comprehensive overview of omics-related research on mastitis through the application of bibliometric analysis integrated with text-mining techniques. The use of LDA enabled the classification of omics approaches applied to mastitis into ten distinct thematic topics. The quantitative findings offer valuable insights into current research trends and highlight existing gaps in the literature. The results suggest that advancements in high-throughput technologies, such as NGS, have contributed to the growing prevalence of omics studies in this field. As the costs associated with sequencing technologies continue to stabilise, the accessibility of such methods is likely to increase, facilitating broader research efforts, particularly in developing countries. Furthermore, there remains a need for additional omics research focusing on mastitis in livestock species beyond dairy cattle, such as sheep and goats, as well as in the context of human mastitis.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0022029925101830.

Acknowledgements. I would like to thank Bahri Dagdas International Research Institute and the General Directorate of Agricultural Research and Policies (TAGEM). I extend my gratitude to the authors of the works used in this bibliometric analysis; without their valuable contributions, this research would not have been possible.

Funding statement. The author declares that no financial support was received for the research, authorship and/or publication of this article.

Competing interest. The author declares no conflicts of interest.

Data availability statement. The data were retrieved from the Scopus and WoS databases. However, the data supporting the findings of this study will be made available upon reasonable request from the corresponding author.

Author contribution. **NE**: Conceptualisation, Methodology, Formal analysis, Visualisation, Project administration, Writing – Original Draft, Writing – Review & Editing.

Ethics approval. This study is a bibliometric analysis of the existing literature based on published articles; therefore, ethical approval is not required.

References

Aebersold R and Mann M (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* **537**(7620), 347–355. https://doi.org/10.1038/nature19949

Alawneh JI, Vezina B, Ramay HR, Al-Harbi H, James AS, Soust M, Moore RJ and Olchowy TWJ (2020) Survey and Sequence Characterization of Bovine Mastitis-Associated Escherichia coli in Dairy Herds. *Frontiers in Veterinary Science* **7**, 582297. https://doi.org/10.3389/fvets.2020.582297

Alonso A, Marsal S and Julià A (2015) Analytical methods in untargeted metabolomics: state of the art in 2015. *Frontiers in Bioengineering and Biotechnology* **3**, 23.

Amigo CR, Moreno LZ, Ferreira TSP, Gomes VTM, Poor AP, de Oliveira CH, Parra BM, Silva APS and Moreno AM (2019) Phenotypic, molecular and genomic characterization of Actinobaculum suis isolated from swine in Brazil. *Anaerobe* **56**, 27–33. https://doi.org/10.1016/j.anaerobe.2019.01.004

Aoki-Kinoshita KF (2008) An Introduction to Bioinformatics for Glycomics Research. *PLOS Computational Biology* **4**(5), e1000075. https://doi.org/10.1371/journal.pcbi.1000075

Apostolakos I, Tsigkrimani M, Paramithiotis S and Mataragas M (2022) Whole-Genome Sequencing and Comparative Genomic Analysis of Enterococcus spp. Isolated from Dairy Products: genomic Diversity, Functional Characteristics, and Pathogenic Potential. *Applied Sciences*. https://doi.org/10.3390/app12199620

Aria M and Cuccurullo C (2017) bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics* **11**(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Arıkan M and Muth T (2023) Integrated multi-omics analyses of microbial communities: a review of the current state and future directions. *Molecular Omics* **19**(8), 607–623.

Asselstine V, Medrano JF, Muniz MMM, Mallard BA, Karrow NA and Cánovas A (2024) Novel lncRNA regulatory elements in milk somatic cells of Holstein dairy cows associated with mastitis. *Communications Biology* **7**(1), 98. https://doi.org/10.1038/s42003-024-05764-y

Baas J, Schotten M, Plume A, Côté G and Karimi R (2020) Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies* **1**(1), 377–386.

Birkle C, Pendlebury DA, Schnell J and Adams J (2020) Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies* **1**(1), 363–376.

Bisutti V, Mach N, Giannuzzi D, Vanzin A, Capra E, Negrini R, Gelain ME, Cecchinato A, Ajmone-Marsan P and Pegolo S (2023) Transcriptome-wide mapping of milk somatic cells upon subclinical mastitis infection in dairy cattle. *Journal of Animal Science and Biotechnology* **14**(1), 93. https://doi.org/10.1186/s40104-023-00890-9

Bourganou MV, Chatzopoulos DC, Lianou DT and Tsangaris GT (2024) Fthenakis GC and Katsafadou AI Scientometrics Evaluation of Published Scientific Papers on the Use of Proteomics Technologies in Mastitis Research in Ruminants. *Pathogens*. https://doi.org/10.3390/pathogens13040324

Bouvier-Muller J, Allain C, Enjalbert F, Farizon Y, Portes D, Foucras G and Rupp R (2018) Somatic cell count-based selection reduces susceptibility to energy shortage during early lactation in a sheep model. *Journal of Dairy Science* **101**(3), 2248–2259. https://doi.org/10.3168/jds.2017-13479

Brajnik Z and Ogorevc J (2023) Candidate genes for mastitis resistance in dairy cattle: a data integration approach. *Journal of Animal Science and Biotechnology* **14**(1), 10.

Castro I, Alba C, Aparicio M, Arroyo R, Jiménez L, Fernández L, Arias R and Rodríguez JM (2019) Metataxonomic and immunological analysis of milk from ewes with or without a history of mastitis. *Journal of Dairy Science* **102**(10), 9298–9311.

Chandramouli K and Qian P-Y (2009) Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Human Genomics and Proteomics: HGP* **2009**, 239204.

Chang H, Wang X, Zeng H, Zhai Y, Huang N, Wang C and Han Z (2023) Comparison of ruminal microbiota, metabolomics, and milk performance between Montbéliarde× Holstein and Holstein cattle. *Frontiers in Veterinary Science* **10**, 1178093.

Chen Z, Xia H, Shen H, Xu X, Arbab AAI, Li M, Zhang H, Mao Y and Yang Z (2018) Pathological features of staphylococcus aureus induced mastitis in dairy cows and isobaric-tags-for-relative-and-absolute-quantitation proteomic analyses. *Journal of Agricultural and Food Chemistry* **66**(15), 3880–3890. https://doi.org/10.1021/acs.jafc.7b05461

Chiaradia E, Valiani A, Tartaglia M, Scoppetta F, Renzone G, Arena S, Avellini L, Benda S, Gaiti A and Scaloni A (2013) Ovine subclinical mastitis: proteomic analysis of whey and milk fat globules unveils putative diagnostic biomarkers in milk. *Journal of Proteomics* **83**, 144–159. https://doi.org/10.1016/j.jprot.2013.03.017

Collado R, Prenafeta A, González-González L, Pérez-Pons JA and Sitjà M (2016) Probing vaccine antigens against bovine mastitis caused by Streptococcus uberis. *Vaccine* **34**(33), 3848–3854. https://doi.org/10.1016/j.vaccine.2016.05.044

Contreras GA and Rodríguez JM (2011) Mastitis: comparative etiology and epidemiology. *Journal of Mammary Gland Biology and Neoplasia* **16**(4), 339–356.

Danielsen M, Codrea MC, Ingvartsen KL, Friggens NC, Bendixen E and Røntved CM (2010) Quantitative milk proteomics–Host responses to lipopolysaccharide-mediated inflammation of bovine mammary gland. *Proteomics* **10**(12), 2240–2249.

Das H, Swamy N, Sahoo G, Ahmed SU and More T (2008) Beta-defensin antibiotic peptides in the innate immunity of the Buffalo: in vivo and in vitro studies. *Alternatives to Laboratory Animals* **36**(4), 429–440. https://doi.org/10.1177/026119290803600404

De Filippis GM, Rinaldi AM, Russo C and Tommasino C (2024, December) Advanced topic modeling in genomics: towards personalized dietary recommendations through BERTopic analysis. In *International Conference on Information Integration and Web Intelligence* 3–17. Cham: Springer Nature Switzerland.

Donkor ES (2013) Sequencing of bacterial genomes: principles and insights into pathogenesis and development of antibiotics. *Genes (Basel)* **4**(4), 556–572. https://doi.org/10.3390/genes4040556

Esener N (2025) Exploring trends in reproductive system microbiome research in farm animals: a bibliometric approach. *Research in Veterinary Science* **186**, 105583.

Evans WE and Relling MV (2004) Moving towards individualized medicine with pharmacogenomics. *Nature* **429**(6990), 464–468. https://doi.org/10.1038/nature02626

Fazzari MJ and Greally JM (2004) Epigenomics: beyond CpG islands. *Nature Reviews Genetics* **5**(6), 446–455. https://doi.org/10.1038/nrg1349

Fenske L, Noll I, Blom J, Ewers C, Semmler T, Fawzy A and Eisenberg T (2022) A dominant clonal lineage of Streptococcus uberis in cattle in Germany. *Antonie van Leeuwenhoek* **115**(7), 857–870. https://doi.org/10.1007/s10482-022-01740-w

Fijałkowski K, Masiuk H, Czernomysy-Furowicz D, Karakulska J, Nawrotek P, Paszkowska A, Struk M and Giedrys-Kalemba S (2013) Superantigen gene profiles, genetic relatedness and biological activity of exosecretions of Staphylococcus aureus isolates obtained from milk of cows with clinical mastitis. *Microbiology and Immunology* **57**(10), 674–683. https://doi.org/10.1111/1348-0421.12088

Gao J, Liu Y-C, Wang Y, Li H, Wang X-M, Wu Y, Zhang D-R, Gao S and Qi Z-L (2020) Impact of yeast and lactic acid bacteria on mastitis and milk microbiota composition of dairy cows. *AMB Express* **10**(1), 22. https://doi.org/10.1186/s13568-020-0953-8

Garcia SN, Osburn BI and Cullor JS (2019) A one health perspective on dairy production and dairy food safety. *One Health* **7**, 100086. https://doi.org/10.1016/j.onehlt.2019.100086

Gilbert FB, Fromageau A, Lamoureux J and Poutrel B (2006) Evaluation of tandem repeats for MLVA typing of Streptococcus uberis isolated from bovine mastitis. *BMC Veterinary Research* **2**(1), 33. https://doi.org/10.1186/1746-6148-2-33

Gioia G, Severgnini M, Cremonesi P, Castiglioni B, Freeman J, Sipka A, Santisteban C, Wieland M, Gallardo VA, Scott JG, Moroni P and Addis MF (2023) Genomic characterization of Mycoplasma arginini isolated from a housefly on a dairy farm and comparison with isolates from bovine milk and lung tissue. *Microbiology Spectrum* **11**(3), e03010–03022. https://doi.org/10.1128/spectrum.03010-22

Goldstone RJ, Harris S and Smith DGE (2016) Genomic content typifying a prevalent clade of bovine mastitis-associated Escherichia coli. *Scientific Reports* **6**(1), 30115. https://doi.org/10.1038/srep30115

Greening SS, Zhang J, Midwinter AC, Wilkinson DA, McDougall S, Gates MC and French NP (2021) The genetic relatedness and antimicrobial resistance patterns of mastitis-causing staphylococcus aureus strains isolated from New Zealand dairy cattle. *Veterinary Sciences*. https://doi.org/10.3390/vetsci8110287

Grützke J, Gwida M, Deneke C, Brendebach H, Projahn M, Schattschneider A, Hofreuter D, El-Ashker M, Malorny B and Al Dahouk S (2021) Direct identification and molecular characterization of zoonotic hazards in raw milk by metagenomics using Brucella as a model pathogen. *Microbial Genomics* **7**(5), 000552.

Hasankhani A, Bakherad M, Bahrami A, Shahrbabak HM, Pecho RDC and Shahrbabak MM (2023) Integrated analysis of inflammatory mRNAs, miRNAs, and lncRNAs elucidates the molecular interactome behind bovine mastitis. *Scientific Reports* **13**(1), 13826. https://doi.org/10.1038/s41598-023-41116-2

Haxhiaj K, Li Z, Johnson M and Dunn SM (2022) Wishart DS and Ametaj BN blood metabolomic phenotyping of dry cows could predict the high milk somatic cells in early lactation – preliminary results. *Dairy*. https://doi.org/10.3390/dairy3010005

Herrero M, Simó C, García-Cañas V, Ibáñez E and Cifuentes A (2012) Foodomics: MS-based strategies in modern food science and nutrition. *Mass Spectrometry Reviews* **31**(1), 49–69. https://doi.org/10.1002/mas.20335

Hoque MN, Istiaq A, Rahman MS, Islam MR, Anwar A, Siddiki AMAMZ, Sultana M, Crandall KA and Hossain MA (2020) Microbiome dynamics and genomic determinants of bovine mastitis. *Genomics* **112**(6), 5188–5203. https://doi.org/10.1016/j.ygeno.2020.09.039

Huang L, Zhu L, Yan J, Lin Y, Ding D, He L, Li Y, Ying Y, Shen L and Jiang Y (2024) Genomic characterization and outbreak investigations of methicillin-resistant Staphylococcus aureus in a county-level hospital in China. *Frontiers in Microbiology* **15**, 1387855.

Jayarao BM, Doré Jr JJ, Baumbach GA, Matthews KR and Oliver SP (1991) Differentiation of Streptococcus uberis from Streptococcus parauberis by polymerase chain reaction and restriction fragment length polymorphism analysis of 16S ribosomal DNA. *Journal of Clinical Microbiology* **29**(12), 2774–2778. https://doi.org/10.1128/jcm.29.12.2774-2778.1991

Jiang C, Fang W, Chen S, Guo X, Gao X, Liu P, Hu G, Li G, Mai W and Liu P (2024) Genetic framework sequencing analysis of Candida tropicalis in dairy cow mastitis and study of pathogenicity and drug resistance. *BMC Microbiology* **24**(1), 428. https://doi.org/10.1186/s12866-024-03522-y

Kang SJ, Cho YI, Kim KH and Cho ES (2016) Proteomic analysis to elucidate the antibacterial action of silver ions against bovine mastitis pathogens. *Biological Trace Element Research* **171**, 101–106.

Kim Y, Atalla H, Mallard B, Robert C and Karrow N (2011) Changes in Holstein cow milk and serum proteins during intramammary infection with three different strains of Staphylococcus aureus. *BMC Veterinary Research* **7**, 1–13.

Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, Gonzalez A, Kosciolek T, McCall L-I, McDonald D, Melnik AV, Morton JT, Navas J, Quinn RA, Sanders JG, Swafford AD, Thompson LR, Tripathi A, Xu ZZ, Zaneveld JR, Zhu Q, Caporaso JG and Dorrestein PC (2018) Best practices for analysing microbiomes. *Nature Reviews Microbiology* **16**(7), 410–422. https://doi.org/10.1038/s41579-018-0029-9

Kolde R and Kolde MR (2015) Package 'pheatmap'. *R Package* **1**(7), 790.

Kumar R, Register K, Christopher-Hennings J, Moroni P, Gioia G, Garcia-Fernandez N, Nelson J, Jelinski MD, Lysnyansky I and Bayles D (2020) Population genomic analysis of Mycoplasma bovis elucidates geographical variations and genes associated with host-types. *Microorganisms* **8**(10), 1561.

Kurz JP, Yang Z, Weiss RB, Wilson DJ, Rood KA, Liu GE and Wang Z (2019) A genome-wide association study for mastitis resistance in phenotypically well-characterized Holstein dairy cattle using a selective genotyping approach. *Immunogenetics* **71**(1), 35–47. https://doi.org/10.1007/s00251-018-1088-9

Li X, Ding XZ, Wan YL, Liu YM and Du GZ (2014) Comparative proteomic changes of differentially expressed whey proteins in clinical mastitis and healthy yak cows. *Genetics and Molecular Research* **13**(3), 6593–6601.

Li Y, Chen W, Ma J, Huang G, Li G, He Q, Kong X, Tang L, Chen J, Ding W and Zhang Z (2024) Rhein against Staphylococcus xylosus by interfering with respiratory metabolism and inducing oxidative stress. *Current Research in Food Science* **8**, 100718. https://doi.org/10.1016/j.crfs.2024.100718

Liu A, Lund MS, Boichard D, Karaman E, Fritz S, Aamand GP, Nielsen US, Wang Y and Su G (2020) Improvement of genomic prediction by integrating additional single nucleotide polymorphisms selected from imputed whole genome sequencing data. *Heredity* **124**(1), 37–49. https://doi.org/10.1038/s41437-019-0246-7

Liu X, Wang J, Chen M, Che R, Ding W, Yu F, Zhou Y, Cui W, Xiaoxu X and God'spower B-O (2019) Comparative proteomic analysis reveals drug resistance of Staphylococcus xylosus ATCC700404 under tylosin stress. *BMC Veterinary Research* **15**, 1–11.

Lopes F, Rosa G, Pinedo P, Santos JEP, Chebel RC, Galvao KN, Schuenemann GM, Bicalho RC, Gilbert RO, Rodrigez-Zas S, Seabury CM

https://doi.org/10.1017/S0022029925101830 Published online by Cambridge University Press

and Thatcher W (2020) Genome-enable prediction for health traits using high-density SNP panel in US Holstein cattle. *Animal Genetics* **51**(2), 192–199. https://doi.org/10.1111/age.12892

Luo Y, Kong Z, Yang B, He F, Huan C, Li J and Yi K (2023) Relationship between microflora changes and mammary lipid metabolism in dairy cows with mastitis. *Animals*. https://doi.org/10.3390/ani13172773

Marete A, Lund MS, Boichard D and Ramayo-Caldas Y (2018) A system-based analysis of the genetic determinism of udder conformation and health phenotypes across three French dairy cattle breeds. *PLoS One* **13**(7), e0199931.

Matthews KR, Kumar SJ, O'Conner SA, Harmon RJ, Pankey JW, Fox LK and Oliver SP (1994) Genomic fingerprints of Staphylococcus aureus of bovine origin by polymerase chain reaction-based DNA fingerprinting. *Epidemiology and Infection* **112**(1), 177–186. https://doi.org/10.1017/S095026880005754X

Melo DA, Motta CC, Rojas A, Soares BS, Coelho IS, Coelho SMO and Souza MMS (2018) Characterization of Coagulase-Negative Staphylococci and pheno-genotypic beta lactam resistance evaluation in samples from bovine Intramammary infection. *Arquivo Brasileiro de Medicina Veterinária E Zootecnia* **70**, 368–374.

Moyes KM, Sørensen P and Bionaz M (2016) The impact of intramammary Escherichia coli challenge on liver and mammary transcriptome and cross-talk in dairy cows during early lactation using RNAseq. *PLOS ONE* **11**(6), e0157480. https://doi.org/10.1371/journal.pone.0157480

Murungi E, Masila E, Ogali I, Langat N, Onywera R, Malonza V, Inguyesi C, Onyambu F, Wesonga H and Maichomo M (2022) Draft genome sequence of Streptococcus agalactiae KALRO-LC1 strain isolated from a mastitis-infected camel in Laikipia County, Kenya. *Microbiology Resource Announcements* **11**(10), e00910–00922. https://doi.org/10.1128/mra.00910-22

Musser JM, Schlievert PM, Chow AW, Ewan P, Kreiswirth BN, Rosdahl VT, Naidu AS, Witte W and Selander RK (1990) A single clone of Staphylococcus aureus causes the majority of cases of toxic shock syndrome. *Proceedings of the National Academy of Sciences* **87**(1), 225–229. https://doi.org/10.1073/pnas.87.1.225

Myllys V, Ridell J, Björkroth J, Biese I and Pyörälä S (1997) Persistence in bovine mastitis of Staphylococcus aureus clones as assessed by random amplified polymorphic DNA analysis, ribotyping and biotyping. *Veterinary Microbiology* **57**(2-3), 245–251.

Nehme A, Awada Z, Kobeissy F, Mazurier F and Zibara K (2018) coupling large-scale omics data for deciphering systems complexity. In Rajewsky N, Jurga S and Barciszewski J (eds.), *Systems Biology*. Cham: Springer International Publishing, pp. 153–172.

Oliveira LM, Simões LC, Crestani C, Costa NS, Pantoja JC, Rabello RF, Teixeira LM, Khan UB, Bentley S, Jamrozy D, Pinto TD and Zadoks RN (2024) Long-term co-circulation of host-specialist and host-generalist lineages of group B Streptococcus in Brazilian dairy cattle with heterogeneous antimicrobial resistance profiles. *Antibiotics*. https://doi.org/10.3390/antibiotics13050389

Olumee-Shabon Z, Swain T, Smith EA, Tall E and Boehmer JL (2013) Proteomic analysis of differentially expressed proteins in caprine milk during experimentally induced endotoxin mastitis. *Journal of Dairy Science* **96**(5), 2903–2912.

Opsal MA, Våge DI, Hayes B, Berget I and Lien S (2006) Genomic organization and transcript profiling of the bovine toll-like receptor gene cluster TLR6-TLR1-TLR10. *Gene* **384**, 45–50. https://doi.org/10.1016/j.gene.2006.06.027

Park S, Jung D, O'Brien B, Ruffini J, Dussault F, Dube-Duquette A, É D, Lucier J-F, Malouin F and Dufour S (2022) Comparative genomic analysis of Staphylococcus aureus isolates associated with either bovine intramammary infections or human infections demonstrates the importance of restriction-modification systems in host adaptation. *Microbial Genomics* **8**(2), 000779.

Patel K, Joshi C, Nauriyal D and Kunjadiya A (2017) Genotypic identification of methicillin resistance and virulence factors in Staphylococcus spp. from bovine mastitis milk. *Comparative Clinical Pathology* **26**(6), 1355–1361. https://doi.org/10.1007/s00580-017-2540-1

Patti GJ, Yanes O and Siuzdak G (2012) Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology* **13**(4), 263–269.

R Core Team (2019) *R: a Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rahman MH, El Zowalaty ME, Falgenhauer L, Khan MFR, Alam J, Popy NN and Rahman MB (2024) Draft genome sequences of clinical mastitis-associated Enterococcus faecalis and Enterococcus faecium carrying multiple antimicrobial resistance genes isolated from dairy cows. *Journal of Global Antimicrobial Resistance* **38**, 111–115. https://doi.org/10.1016/j.jgar.2024.05.011

Rahman MM, Siddique N, Gilman MAA, Hasnat S, Haider MG, Talukder AK, Rahman A, Islam T, Das ZC, Hossain MA and Hoque MN (2025) Genomic and in vitro analysis of Pediococcus pentosaceus MBBL4 Implicated Its Therapeutic use against mastitis pathogens and as a potential probiotic. *Probiotics and Antimicrobial Proteins*. https://doi.org/10.1007/s12602-025-10484-8

Rahman Mohammad H, El Zowalaty Mohamed E, Falgenhauer L, Khan Mohammad Ferdousur R, Alam J, Popy Najmun N and Rahman MB (2023) Draft genome sequences of two clinical mastitis-associated Escherichia coli strains, of sequence type 101 and novel sequence type 13054, isolated from dairy cows in Bangladesh. *Microbiology Resource Announcements* **12**(8), e00166–00123. https://doi.org/10.1128/mra.00166-23

Rathore D, Marino MJ and Nita-Lazar A (2023) Omics and systems view of innate immune pathways. *Proteomics* **23**(13-14), 2200407.

Ribeiro GAS and Melgaço BR, Márcio dCR and Costa NL (2024) From bibliometrics to text mining: exploring feature selection methods in microarray research. *Communications in Statistics - Simulation and Computation* 1–17. https://doi.org/10.1080/03610918.2024.2331083

Rinschen MM, Ivanisevic J, Giera M and Siuzdak G (2019) Identification of bioactive metabolites using activity metabolomics. *Nature Reviews Molecular Cell Biology* **20**(6), 353–367. https://doi.org/10.1038/s41580-019-0108-4

Rodríguez MF, Gomez AP and Ceballos-Garzon A (2023) Antimicrobial resistance profiles of Staphylococcus Isolated from cows with subclinical mastitis: do strains from the environment and from humans contribute to the dissemination of resistance among bacteria on dairy farms in Colombia? *Antibiotics*. https://doi.org/10.3390/antibiotics12111574

Rosenthal R (1979) The file drawer problem and tolerance for null results. *Psychological Bulletin* **86**(3), 638.

Ruegg PL (2017) A 100-Year Review: mastitis detection, management, and prevention. *Journal of Dairy Science* **100**(12), 10381–10397. https://doi.org/10.3168/jds.2017-13023

Schena M, Shalon D, Davis RW and Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**(5235), 467–470.

Schmidtmann C, Schönherz A, Guldbrandtsen B, Marjanovic J, Calus M, Hinrichs D and Thaller G (2021) Assessing the genetic background and genomic relatedness of red cattle populations originating from Northern Europe. *Genetics Selection Evolution* **53**(1), 23. https://doi.org/10.1186/s12711-021-00613-6

Shendure J and Ji H (2008) Next-generation DNA sequencing. *Nature Biotechnology* **26**(10), 1135–1145. https://doi.org/10.1038/nbt1486

Sievert C and Shirley K (2014, June) LDAvis: a method for visualizing and interpreting topics. *In Proceedings of the workshop on interactive language learning, visualization, and interfaces*, 63–70.

Silva NCC, Yang Y, Rodrigues MX, Tomazi T and Bicalho RC (2021) Whole-genome sequencing reveals high genetic diversity of Streptococcus uberis isolated from cows with mastitis. *BMC Veterinary Research* **17**(1), 321. https://doi.org/10.1186/s12917-021-03031-4

Sun Y, Li L, Li C, Wang G and Xing G (2019) Gene microarray integrated with iTRAQ-based proteomics for the discovery of NLRP3 in LPS-induced inflammatory response of bovine mammary epithelial cells. *Journal of Dairy Research* **86**(4), 416–424. https://doi.org/10.1017/S0022029919000761

Taverna F, Negri A, Piccinini R, Zecconi A, Nonnis S, Ronchi S and Tedeschi G (2007) Characterization of cell wall associated proteins of a Staphylococcus aureus isolated from bovine mastitis case by a proteomic approach. *Veterinary Microbiology* **119**(2), 240–247.

https://doi.org/10.1017/S0022029925101830 Published online by Cambridge University Press

Trapanese L, Bifulco G, Macchio AC, Aragona F, Purrone S, Campanile G and Salzano A (2025) Precision Livestock Farming applied to the dairy sector: 50 years of history with a text mining and topic analysis approach. *Smart Agricultural Technology* **10**, 100827.

Trovó Fabiano TL, Lemos MV and Givisiez PE (2005) Fluorescent amplified fragment length polymorphism genotyping of human and animal Staphylococcus aureus isolates from dairy farms with manual milking. *Veterinary Microbiology* **109**(1-2), 57–63. https://doi.org/10.1016/j.vetmic.2005.03.009

Vailati-Riboni M, Palombo V and Loor JJ (2017) What are omics sciences? *Periparturient Diseases of Dairy Cows: a Systems Biology Approach*, 1–7. Cham: Springer International Publishing.

Vasco KA, Hansen ZA, Schilmiller AL, Bowcutt B, Carbonell SL, Ruegg PL, Quinn RA, Zhang L and Manning SD (2024) Untargeted metabolomics and metagenomics reveal signatures for intramammary ceftiofur treatment and lactation stage in the cattle hindgut. *Frontiers in Molecular Biosciences* **11**, 1364637.

Vautor E, Cockfield J, Le Marechal C, Le Loir Y, Chevalier M, Robinson DA, Thiery R and Lindsay J (2009) Difference in virulence between Staphylococcus aureus isolates causing gangrenous mastitis versus subclinical mastitis in a dairy sheep flock. *Veterinary Research* **40**(6), 56.

Vázquez-Fresno R, Llorach R, Marinic J, Tulipani S, Garcia-Aloy M, Espinosa-Martos I, Jiménez E, Rodríguez JM and Andres-Lacueva C (2014) Urinary metabolomic fingerprinting after consumption of a probiotic strain in women with mastitis. *Pharmacological Research* **87**, 160–165. https://doi.org/10.1016/j.phrs.2014.05.010

Vidal Amaral JR, Jucá Ramos RT, Almeida Araújo F, Bentes Kato R, Figueira Aburjaile F, de Castro Soares S, Góes-Neto A, Matiuzzi da Costa M, Azevedo V, Brenig B, de Oliveira S S and Soares Rosado A (2022) Bacteriocin producing Streptococcus agalactiae strains isolated from bovine mastitis in Brazil. *Microorganisms*. https://doi.org/10.3390/microorganisms10030588

Vidal M, Cusick Michael E and Barabási A-L (2011) Interactome networks and human disease. *Cell* **144**(6), 986–998. https://doi.org/10.1016/j.cell.2011.02.016

Voelkerding KV, Dames SA and Durtschi JD (2009) Next-Generation Sequencing: from basic research to diagnostics. *Clinical Chemistry* **55**(4), 641–658. https://doi.org/10.1373/clinchem.2008.112789

Wang M, Liang Y, Ibeagha-Awemu EM, Li M, Zhang H, Chen Z, Sun Y, Karrow NA, Yang Z and Mao Y (2020) Genome-Wide DNA Methylation analysis of mammary gland tissues from Chinese Holstein cows with staphylococcus aureus induced mastitis. *Frontiers in Genetics* **11**, 550515.

Wang R, Li L, Huang Y, Luo F, Liang W, Gan X, Huang T, Lei A, Chen M and Chen L (2015) Comparative genome analysis identifies two large deletions in the genome of highly-passaged attenuated Streptococcus agalactiae strain YM001 compared to the parental pathogenic strain HN016. *BMC Genomics* **16**(1), 897. https://doi.org/10.1186/s12864-015-2026-y

Wang X, Fei Y, Shao Y, Liao Q, Meng Q, Chen R and Deng L (2024) Transcriptome analysis reveals immune function-related mRNA expression in donkey mammary glands during four developmental stages. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **49**, 101169. https://doi.org/10.1016/j.cbd.2023.101169

Wang Z, Gerstein M and Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**(1), 57–63. https://doi.org/10.1038/nrg2484

Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL and Humphery-Smith I (1995) Progress with gene-product mapping of the Mollicutes: mycoplasma genitalium. *Electrophoresis* **16**(1), 1090–1094.

White RT, Bakker S, Burton M, Castro ML, Couldrey C, Dyet K, Eustace A, Harland C, Hutton S and Macartney-Coxson D (2024) Rapid identification and subsequent contextualization of an outbreak of methicillin-resistant Staphylococcus aureus in a neonatal intensive care unit using nanopore sequencing. *Microbial Genomics* **10**(7), 001273.

Wild CP (2005) Complementing the Genome with an 'Exposome': the Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiology, Biomarkers & Prevention* **14**(8), 1847–1850. https://doi.org/10.1158/1055-9965.EPI-05-0456

Xu S, Liu Y, Gao J, Zhou M, Yang J, He F, Kastelic JP, Deng Z and Han B (2021) Comparative genomic analysis of Streptococcus dysgalactiae subspecies dysgalactiae Isolated from bovine mastitis in China. *Frontiers in Microbiology* **12**, 751863.

Yalamanchili HB, Kho SJ and Raymer ML (2017) Latent dirichlet allocation for classification using gene expression data. In: *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE).*

Yang F, Chen F, Li L, Yan L, Badri T, Lv C, Yu D, Zhang M, Jang X and Li J (2019) Three novel players: PTK2B, SYK, and TNFRSF21 were identified to be involved in the regulation of bovine mastitis susceptibility via GWAS and post-transcriptional analysis. *Frontiers in Immunology* **10**, 1579.

Yang K and Han X (2016) Lipidomics: techniques, applications, and outcomes related to biomedical sciences. *Trends in Biochemical Sciences* **41**(11), 954–969. https://doi.org/10.1016/j.tibs.2016.08.010

Yang Y, Higgins Catherine H, Rehman I, Galvao Klibs N, Brito Ilana L, Bicalho Marcela L, Song J, Wang H and Bicalho Rodrigo C (2019) Genomic diversity, virulence, and antimicrobial resistance of Klebsiella pneumoniae strains from cows and humans. *Applied and Environmental Microbiology* **85**(6), e02654–02618. https://doi.org/10.1128/AEM.02654-18

Ying YT, Yang J, Tan X, Liu R, Zhuang Y, Xu JX and Ren WJ (2021) Escherichia coli and Staphylococcus aureus differentially regulate Nrf2 pathway in bovine mammary epithelial cells: relation to distinct innate immune response. *Cells* **10**(12). https://doi.org/10.3390/cells10123426

Zaatout N, Ayachi A, Kecha M and Kadlec K (2019) Identification of staphylococci causing mastitis in dairy cattle from Algeria and characterization of Staphylococcus aureus. *Journal of Applied Microbiology* **127**(5), 1305–1314. https://doi.org/10.1111/jam.14402

Zhang H, Jiang H, Fan Y, Chen Z, Li M, Mao Y, Karrow NA, Loor JJ, Moore S and Yang Z (2018) Transcriptomics and iTRAQ-proteomics analyses of bovine mammary tissue with Streptococcus agalactiae-Induced mastitis. *Journal of Agricultural and Food Chemistry* **66**(42), 11188–11196. https://doi.org/10.1021/acs.jafc.8b02386

Zhu J, Wu J, Liang Z, Mo C, Qi T, Liang S, Lian T, Qiu R, Yu X, Tang X and Wu B (2022) Interactions between the breast tissue microbiota and host gene regulation in nonpuerperal mastitis. *Microbes and Infection* **24**(3), 104904. https://doi.org/10.1016/j.micinf.2021.104904

https://doi.org/10.1017/S0022029925101830 Published online by Cambridge University Press