Comparative genomic analysis of ovine and other host associated isolates of Staphylococcus aureus exhibit the important role of mobile genetic elements and virulence factors in host adaptation

Alessandra Lima, Ana Carolina Barbosa Caetano, Raquel Hurtado Castillo, Roselane Gonçalves dos Santos, Diego Lucas Neres Rodrigues, Thiago de Jesus Sousa, Rodrigo Bentes Kato, Marcus Vinicius Canário Viana, Anne Cybelle Pinto Gomide, Flavia Figueira Aburjaile, et al.

Gene 855 (2023) 147131

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Comparative genomic analysis of ovine and other host associated isolates of *Staphylococcus aureus* exhibit the important role of mobile genetic elements and virulence factors in host adaptation

Alessandra Lima [a], Ana Carolina Barbosa Caetano [a], Raquel Hurtado Castillo [a], Roselane Gonçalves dos Santos [a], Diego Lucas Neres Rodrigues [a], Thiago de Jesus Sousa [a], Rodrigo Bentes Kato [a], Marcus Vinicius Canário Viana [a], Anne Cybelle Pinto Gomide [a], Flavia Figueira Aburjaile [f], Sandeep Tiwari [a,b,h], Arun Jaiswal [a], Alfonso Gala-García [c], Núbia Seyffert [b], Thiago Luiz de Paula Castro [b], Bertram Brenig [g], Mateus Matiuzzi da Costa [i], Elaine Maria Seles Dorneles [e], Yves Le Loir [d], Vasco Azevedo [a,*]

[a] *Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil*
[b] *Programa de Pós-graduação em Imunologia, Instituto De Biologia, Universidade Federal da Bahia, Salvador, BA, Brasil.*
[c] *Faculdade de Odontologia, Universidade Federal da Bahia, Salvador, BA, Brazil*
[d] *Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), Paris, France*
[e] *Departamento de Medicina Veterinária, Faculdade de Zootecnia e Medicina Veterinária, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil*
[f] *Departamento de Medicina Veterinária Preventiva, Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil*
[g] *Institute of Veterinary Medicine, University of Göttingen, Burckhardtweg 2, Göttingen, Germany*
[h] *Programa de Pós-graduação em Microbiologia, Instituto De Biologia, Universidade Federal da Bahia, Salvador, BA, Brasil*
[i] *Laboratório de Microbiologia e Imunologia Animal, Campus Ciências Agrárias, Universidade Federal do Vale do São Francisco (UNIVASF), Petrolina, Pernambuco, Brazil*

ARTICLE INFO

Keywords:
Staphylococcus aureus
Comparative genomics
Mastitis
Pangenomics
Virulence genes

ABSTRACT

Staphylococcus aureus is the main etiological agent of mastitis in small ruminants worldwide. This disease has a difficult cure and possible relapse, leading to significant economic losses in production, milk quality and livestock. This study performed comparative genomic analyses between 73 *S. aureus* genomes from different hosts (human, bovine, pig and others). This work isolated and sequenced 12 of these genomes from ovine. This study contributes to the knowledge of genomic specialization and the role of specific genes in establishing infection in ovine mastitis-associated *S. aureus*. The genomes of *S. aureus* isolated from sheep maintained a higher representation when grouped with clonal complexes 130 and 133. The genomes showed high genetic similarity, the species pan-genome consisting of 4200 genes (central = 2008, accessory = 1559 and unique = 634). Among these, 277 unique genes were related to the genomes isolated from sheep, with 39.6 % as hypothetical proteins,

Abbreviations: list: AMR, Antimicrobial Resistance; ANI, Average Nucleotide Identity; ANIM, Average Nucleotide Identity Mummer; ARG-ANNOT, Antibiotic Resistance Gene-Annotation; BLASTN, Basic Local Alignment Search Tool Nucleotide; BLAZ, PC1 Beta-Lactamase; BPGA, Bacterial Pan Genome Analysis Tool; BRIG, Blast Ring Image Generator; BV-BRC, Bacterial and Viral Bioinformatics Resource Center; CAPES, Agencies Coordination for The Improvement Of Higher Education Personnel; CARD, The Comprehensive Antibiotic Resistance Database; CC, Clonal Complex; CCs, Clonal Complexes; CHIPS, Chemotaxis Inhibitory Protein of Staphylococcus; *cna*, Collagen-Binding Protein; *coa*, Coagulase; COG, Cluster of Orthologous Groups; *entb*, Staphylococcal Enterotoxin B; FAPEMIG, Minas Gerais Research Funding Foundation; GC, Genomic Content; GEIs, Genomic Islands; GIPSY, Genomic Island Prediction Software; HYSA, Hyaluronate Lyase Precursor; INRAE, Institut National de Recherche pour L'agriculture, L'alimentation et L'environnement; KEGG, Kyoto Encyclopedia Of Genes And Genomes; LB, Luria Bertani; LUKS-PV, Panton-Valentine Leukocidin; MATE, Multidrug And Toxic Compound Extrusion; MEGARES, An Antimicrobial Database for High-Throughput Sequencing; MFS, Major Facilitator Superfamily; MLSA, Multilocus Sequence Analysis; MLST, Multi-Locus Sequence Typing; MRSA, Methicillin-Resistant *Staphylococcus aureus*; MSSA, Methicillin-Susceptible *Staphylococcus Aureus*; NCBI, National Center for Biotechnology Information; PAIs, Pathogenicity Islands; RECOM, Omics Science Network; *S. aureus*, *Staphylococcus aureus*; *sak*, Staphylokinase; SCCmec, Chromosome Cassette *mec*; *scn*, Staphylococcal Complement Inhibitor; *selk*, Staphylococcal Enterotoxin K; SIs, Symbiosis Islands; SNP, Single Nucleotide Polymorphism; VFDB, Virulence Factors of Pathogenic Bacteria; vWpb, Secreted Von Willebrand Factor-Binding Protein Precursor.

* Corresponding author.
E-mail address: vascoariston@gmail.com (V. Azevedo).

https://doi.org/10.1016/j.gene.2022.147131
Received 15 September 2022; Received in revised form 1 December 2022; Accepted 16 December 2022
Available online 17 December 2022

6.4 % as phages, 6.4 % as toxins, 2.9 % as transporters, and 44.7 % as related to other proteins. Furthermore, at the pathogen level, they showed 80 genes associated with virulence factors and 19 with antibiotic resistance shared in almost all isolates. Although *S. aureus* isolated from ovine showed susceptibility to antimicrobials *in vitro*, ten genes were predicted to be associated with antibiotic inactivation and efflux pump, suggesting resistance to gentamicin and penicillin. This work may contribute to identifying genes acquired by horizontal transfer and their role in host adaptation, virulence, bacterial resistance, and characterization of strains affecting ovine.

1. Introduction

Mastitis is an inflammation of the mammary gland, most often due to a bacterial infection, responsible for substantial economic losses in milk production worldwide (Halasa et al., 2009; Ruegg, 2017). That disease may vary from subclinical to clinical and can become chronic in some cases. Occasionally, mastitis can also become systemic. This infection can be caused by different bacterial species, among which *Staphylococcus aureus* is one of the most prevalent (Oviedo-Boyso et al., 2007; Le Maréchal, Hernandez, et al., 2011; Haag, Fitzgerald and Penadés, 2019a). However, antibiotics are often inefficient in curing the infection caused by *S. aureus;* in addition, their disease is prone to resurgence, as bacteria may remain viable inside the epithelial cells of the mammary gland (Haag, Fitzgerald and Penadés, 2019a; Oget, Tosser-Klopp and Rupp, 2019; Astrup, Pedersen and Farre, 2022). Additionally, some differences between the ovine and bovine epithelium lead to different interactions during treatment, such as the distribution of serotonin receptor subtypes in the mammary gland (Suárez-Trujillo et al., 2019; Hughes, 2020).

S. aureus strains are commonly categorized into clonal complexes (CCs) according to Multi-locus Sequence Typing (MLST). The MLST method has extensively been used to characterize *S. aureus* isolates and showed host-specificity within CCs, such as CC133 for ovine isolates (Smith et al., 2005; Aires-de-Sousa et al., 2007a; Zakour and Loir, 2007; Smyth et al., 2009; Merz, Stephan and Johler, 2016; Hoekstra et al., 2019). In addition, previous studies also showed host-specific variations in terms of genome content (mobile genetic elements) and differential expression of regulatory genes in the *S. aureus* species (Alves et al., 2009; Guinane, ben Zakour, et al., 2010; Richardson et al., 2018).

Comparative genomic analyses are conducted to understand better the basis of intraspecific evolution and pathogenesis of microorganisms since bacteria can have specific genetic characteristics of strains such as single nucleotide polymorphism (SNP) and mobile genetic elements (e.g., islands of pathogenicity, bacteriophages, plasmids, transposons, and integrative elements) (Dobrindt, 2001; ben Zakour et al., 2008). The mobile elements can carry copies of genes that confer additional characteristics in the recipient genome, such as resistance to antibiotics, heavy metals, or greater pathogenicity. The comparison allows the identification of specificities in hosts, as the *seh* gene was identified only in cow isolates. The *cap8* genotype is present in almost any type of host in a study involving isolates of mastitis *S. aureus* in cattle and goats (Acosta et al., 2018; Sheppard, Guttman and Fitzgerald, 2018; Oget, Tosser-Klopp and Rupp, 2019).

At the genomic sequence level, although the number of complete *S. aureus* genomes deposited in the NCBI genome database has increased considerably in recent years, most are related to human isolates. More recently, only three ovine strains (ED133, NCTC1803, and NCTC9555) have been isolated from mastitis (Guinane, ben Zakour, et al., 2010). The information about peculiarities of the genome of small ruminant isolates is rare, and there is no knowledge of the genomic characteristics that may be involved in the onset of mastitis in sheep (Aires-de-Sousa et al., 2007b; Peton et al., 2014; Bosi et al., 2016a; Oget, Tosser-Klopp and Rupp, 2019).

This study contributes to the knowledge of genomic specialization and the role of specific genes in establishing infection in ovine mastitis-associated *S. aureus*. Comparative genomic analyses were performed among 73 *S. aureus* genomes from different hosts (human, bovine, ovine, pig and others). In addition, this works isolated and sequenced 12 of these genomes collected in France in cases of ovine mastitis.

2. Material and methods

2.1. Ovine S. aureus strains and genomic DNA extraction

Twelve *S. aureus* strains were isolated on a Luria Bertani (LB) agar plate from different forms of mastitis in flocks of French sheep (Table 1). The genomic DNA was extracted after cultivating the strain on 30 mL of LB broth at 37 °C for 12 h. Every culture was centrifuged for bacterial pellet formation, and the supernatant was discarded. The pellet was suspended in 600 μL of solution (Tris-HCI pH 7.0, 0.5 M EDTA pH 8.0, NaCl 5 M, and distilled H2O sufficient to obtain 50 mL) and transferred to a tube of 2 mL containing glass beads (VK01) (Bertin Technologies). Next, 3 μL of RNase A solution (20 mg/mL) was added before subjecting the bacteria to mechanical lysis. Two homogenization cycles of 15 *sec* each, at 6,500 rpm, were performed using Precellys 24 (Bertin Technologies). Subsequently, 1 mL of phenol/chloroform/isoamyl alcohol (25:24:1) solution was added to the tube, and the mixture was homogenized and centrifuged at 13,000 rpm 7 min. Next, the upper aqueous phase of the mixture was transferred to a new tube, and the second round of phenol/chloroform/isoamyl alcohol purification was performed. Next, the upper aqueous phase was recovered and mixed with 1 mL of chloroform. Following centrifugation at 13,000 rpm for 7 min, the upper aqueous phase was transferred to a new tube. Next, 1 mL of ethyl alcohol, 40 μL of 3 M NaAc, and 4 μL of 20 mg/mL glycogen were added. The mixture was placed at −20 °C overnight following gentle inversion for DNA precipitation. Following centrifugation at 13,000 rpm, for 15 min, the supernatant was discarded. Next, 1 mL of 70 % ethyl alcohol was added to rehydrate the DNA pellet. The second round of 70 % ethyl alcohol wash was performed using centrifugation at 13,000 rpm for 15 min. Then, the DNA pellet was dried at 60 °C. Finally, the DNA precipitate was suspended in 50 μL of sterilized ultra-pure water. DNA quantity and quality assessments were conducted using NanoDropTM 2000 (Thermo ScientificTM), Qubit Fluorometer (Thermo ScientificTM), and 1 % agarose gel electrophoresis.

Table 1
S. aureus strains isolated from ovine mastitis sequenced for the present study.

Strain	Mastitis range	Isolation Year	Locality isolation (France)
O268	Clinical	1998	Pyrénées-Atlantiques
O17	Clinical	2003	Alpes Maritime
O322	Clinical	2008	Alpes de Haute-Provence
O326	Clinical	2008	Alpes de Haute-Provence
O217	Chronic	2002	Aveyron
O11*	Gangrenous	2002	Southeast of France
O408	Gangrenous	2010	Alpes de Haute-Provence
O46*	Subclinical	2002	Southeast of France
O82	Subclinical	2002	Unspecified
O267	Subclinical	1998	Pyrénées-Atlantiques
O331	Subclinical	2008	Alpes de Haute-Provence
O55	Subclinical	2002	Unspecified

* The O11 and O46 strains were not considered in the DNA extraction, since their genomes were previously sequenced.

2.2. Sequencing, assembly, and annotation of the ovine S. Aureus genomes

Sequencing of all total DNA the strains were performed using Hiseq 2500 platform, 150 bp, paired end (Illumina, San Diego, CA, USA). The two genomes previously sequenced (O11 and 046) was obtained with using an Illumina Genome Analyzer GAII (Le Maréchal, Hernandez, et al., 2011). The reads quality was checked with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

SPAdes, version 3.9.1 (Bankevich et al., 2012), was used for the *ab initio* assembly of all genomes, including O46 and O11. CONTIGuator version 2.7 (Galardini et al., 2011) generated scaffolding. The gaps resulting from assemblies were manually filled using the CLC Genomics Workbench software, version 7.0 (https://www.qiagenbioinformatics.com/products/clc-genomics-workbench). Reads were mapped against the reference genome to generate the consensus sequences for gap filling. All genomes were submitted to an automatic annotation using the PROKKA version 3.0 (Seemann, 2014).

2.3. Genome average nucleotide identity analysis in the selected S. aureus strains

To determine the genome similarity among the ovine strains and the strains isolated from various hosts, we used the Average Nucleotide Identity (ANI) analysis, specifically ANIm, based on MUMmer calculations (Konstantinidis, Ramette and Tiedje, 2006; Kim et al., 2014a). JSpeciesWS Online Service Taxonomic Thresholds (available at jspecies.ribohost.com/jspeciesws/) (Richter et al., 2016) were used with BLASTn (Altschul et al., 1990). This analysis method performs pairwise comparisons through the identity and similarity values implemented (Goris et al., 2007). The similarity heatmap was screened using Morpheus software (available at https://software.broadinstitute.org/morpheus/).

3. Molecular typing and phylogenetic analysis of the selected *S. aureus* strains

To determine the Sequence Type (ST) of each selected *S. aureus* strain, the allelic prediction was performed using MLST (Multi-locus sequence typing) server version 1.8 (Larsen et al., 2012). The sequence typing of *S. aureus* is routinely performed with allelic variants of seven housekeeping genes (*arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi,* and *yqiL*) (Enright et al., 2000a), and the combination of these variants defines a specific sequence type. It compares allelic diversity based on approximately 500 bp internal gene fragments (available at cge.CBS.dtu.dk//services/MLST). PHYLOViZ, version 1.1 (Francisco et al., 2012), was used to form a minimum spanning tree (MST) through the goeBURST algorithm (Francisco et al., 2009) and clonal complex (CC) formation (available at https://online.phyloviz.net/index).

3.1. Phylogenomic analyses of the selected S. aureus strains

The phylogenomic tree was generated based on the nucleotide alignment of the core genome from seventy-three *Staphylococcus aureus* strains predicted by the Roary pangenome pipeline (Page et al., 2015). FastTree (Price, Dehal and Arkin, 2010) and iTOL version 3 (https://itol.embl.de/) were used to construct the maximum likelihood tree.

3.2. Pangenome analysis of ovine mastitis-associated S. Aureus genes and proteins

Bacterial Pan Genome Analysis Tool (BPGA) version 1.3 (Chaudhari, Gupta and Dutta, 2016a) determines the ovine's core, accessory, and singleton genes. In addition, a group of complete genomes isolated from other hosts was added to the dataset for comparative purposes of the presence and absence of genes between hosts. This work considered two groups of genomes for pangenome analysis. The first group comprised twelve ovine strains (isolated from ovine, sequenced and assembled from herds in France), and three more ovine strains (ED133, NCTC1803, and NCTC9555) genomes sequences were downloaded from NCBI for the analysis. Moreover, genomes sequences of fifty-eight strains of *S. aureus* isolated from different hosts were retrieved from the NCBI database. In total this work, seventy-three genomes were considered for comparative genomic analyses. These strains were selected based on the diversity of hosts and the availability of complete and high-quality genomes, thus seeking their greatest population representability (**Table S1**).

3.3. Genome plasticity analyses in the selected S. aureus strains

The Genomic (GEIs), Pathogenicity (PAIs), and Symbiosis (SIs) Island predictions were performed using GIPSy (Genomic Island prediction Software), version 1.1.2 (Soares et al., 2016), with the genome of *Staphylococcus warneri* strain 16A (CP031269.1), retrieved from the NCBI genome database, used as the reference. BLAST Ring Image Generator (BRIG), version 0.95 (Alikhan et al., 2011), was used to generate a circular image of all ovine genomes with the predicted prophages, GEIs, and PAIs. Similar to the two clades (CC133 and CC130), strains 0268 and 082 were used (**Table S1**). The PHASTER (Arndt et al., 2016) was used to predict and annotate prophages in all strains' genomes. Considering the *S. aureus* methicillin resistance presence, we tried to identify the cassette chromosome *mec* (SCCmec) elements (Zafalon et al., 2018). A search for SCCmec IV, a cassette chromosome carrying the *mecA* or *mecC* gene, was performed by SCCmecfinder version 1.2 (Kaya et al., 2018). For this analysis, the thresholds of 90 % and 60 % for minimum identity and minimum length were used to compare the genome against the reference database (Kaya et al., 2018).

3.4. Resistance genes analysis in the ovine S. aureus strains

The Abricate (https://github.com/tseemann/abricate) tool was used to mass screen the genomes to find resistance: NCBI AMRFinderPlus (Feldgarden et al., 2019) CARD (Jia et al., 2017), Resfinder (Feldgarden et al., 2019), ARG-NNOT (Gupta et al., 2014) and MEGARES 2.00 (Doster et al., 2019) and for virulence genes, using the Virulence Factors of Pathogenic Bacteria (VFDB) (Chen et al., 2016).

4. Results

4.1. Genomic statistics from ovine S. aureus strains

The chromosomal genomes of the study's fifteen ovine *S. aureus* strains ranged from 2.7 to 2.8 Mb in size, with approximately 33 % of GC content. The PROKKA annotation predicted between 2,465 and 2,755 gene coding sequences. All genomes showed at least one region associated with prophages, with a maximum of seven regions on some chromosomes Table 2.

4.2. Phylogenomic analyses provide better clustering of ovine S. aureus strains

Phylogenomic analyses evaluated the clustering of *S. aureus* strains from different hosts on a genomic scale. The phylogenomic tree built showed a bootstrap between 0.9 and 1.0, as shown in Fig. 1, represented by the taxon's internal circles. It is possible to observe the grouping of populations by their type of clonal complex. The strains belonging to CC130 (O82, O46, O326, O11, and O408) formed a single clade (Fig. 1). The phylogenomic tree shows a common ancestor for all lineages proven to be ovine. The strains of CC130, CC133, CC425, CC59, and CC30 comprise five distinct clades. Two ovine clusters were formed from two major clonal complexes with similarities ranging from 97 to 100 % (CC130 and CC133). The phylogenomic analysis also revealed that these clusters did not group three strains from ovine (O217, O55, and O331).

Table 2
General genomic features of fifteen ovine *S. aureus* strains analyzed.

	Strain	Genomes length (bp)	GC content (%)	CDS Predicted	Number of Phages	NCBI Accession Number
1	B119	2,768,115	32.82	2597	3	CP038460
2	O11	2,770,352	32.72	2610	3	CP024649
3	O17	2,772,340	32.89	2598	4	CP032051
4	O217	2,764,228	32.86	2572	4	CP038461
5	O267	2,864,632	32.93	2755	6	CP034102
6	O268	2,841,948	32.91	2707	7	CP038612
7	O326	2,778,005	32.85	2616	3	CP032481
8	O331	2,695,389	32.91	2465	1	CP038269
9	O408	2,778,443	32.85	2618	3	CP038270
10	O46	279,141	32.80	2644	3	CP025395
11	O55	2,794,042	32.89	2650	4	CP038268
12	O82	2,761,328	32.89	2593	3	CP038819
13	NCTC1803	2,836,670	32.90	2692	5	LR134305
14	NCTC9555	2,843,795	32.94	2685	6	LR134090
15	ED133	2,832,478	32.92	2701	6	CP001996

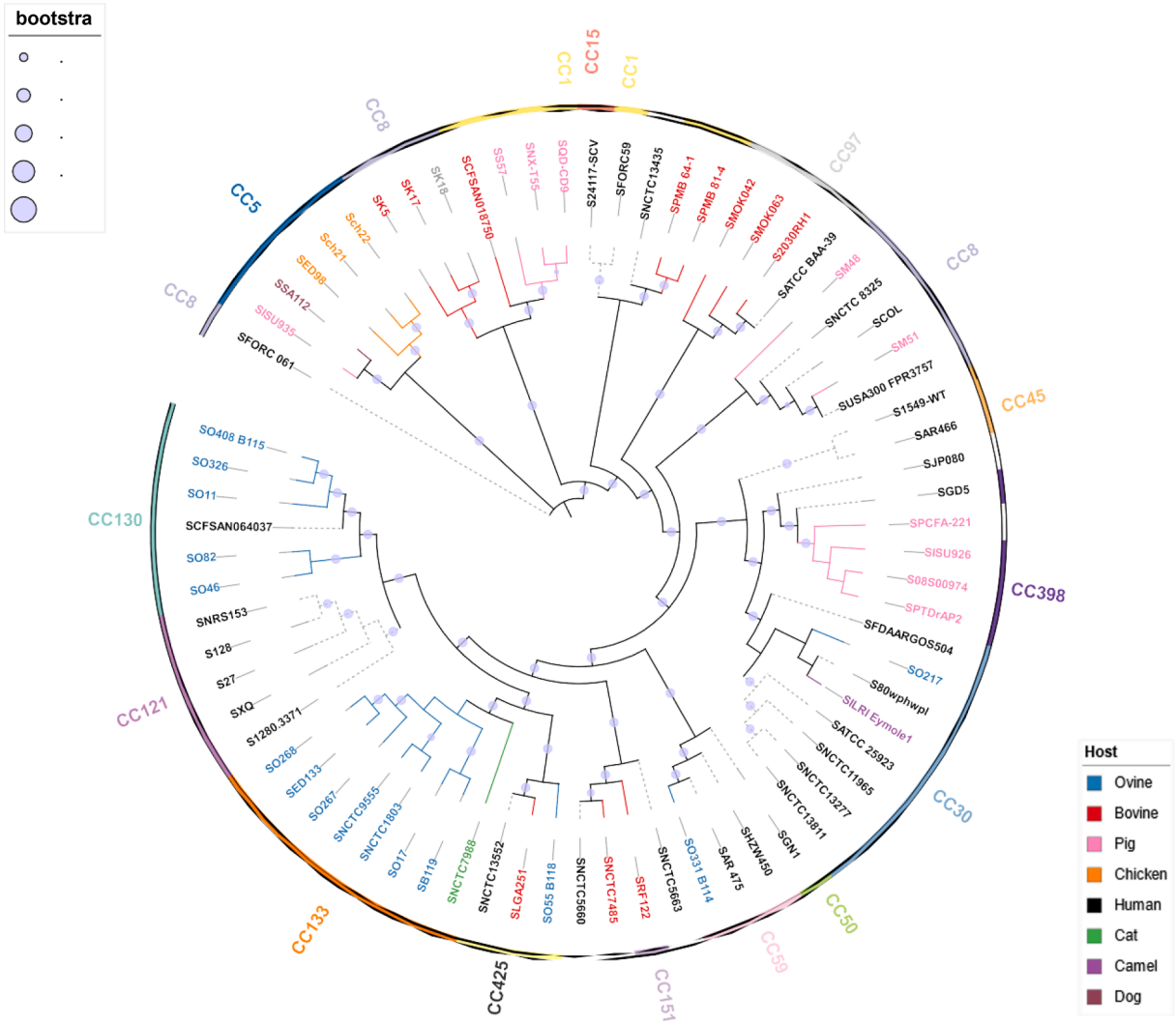

Fig. 1. Molecular phylogenomic analysis based on the nucleotide alignment of the core *Staphylococcus aureus* strain genomes. The hosts are in the following colours, and each complex clonal (CC) is highlighted with a coloured line.

However, strain O55 (CC425) was grouped with the strains isolated from cattle and humans. The strain O331 (CC59) is isolated from humans, and the strain O217 (CC30) forms a single clade with the strains from humans and camels.

4.3. ANI analysis evidence ovine S. aureus genomic relatedness

ANI predictions were conducted using all selected *S. aureus* strains to understand how the ovine strains genetically correlate with each other and with the strains from the other hosts. The genome of ovine *S. aureus*

strain ED133 was used as the reference for the ovine sequence. All genomes presented a high degree of similarity with the genome of the reference strain ED133 (ANI > 97 %). Most ovine *S. aureus* genomes were clustered in the highest similarity group (Fig. 2). The genomes NCTC9555, B119, O17, O267, O268, NCTC1803, and NCTC7988, presented the most significant similarities (nearly 100 %) with ED133. Interestingly, NCTC7988 was the non-ovine strain with the most considerable similarity with the ED133 reference, corroborating the previous phylogenomic clustering in this work (Fig. 1). In this analysis, all ovine genomes presented more than 98 % similarity regarding the ED133 genome, except for O217, which showed a minor similarity among sheep strains. In the same way, this strain was not grouped with any ovine genomes in MLST analysis. This study showed that it belongs to CC30, which is important in human and animal infections, and suggests no-host-specialization by this strain. On the other hand, genomes isolated from milk or mastitis in cows (LGA251 and RF122) present a similar level of 98 % regarding ED133 (isolated from ovine).

4.4. Molecular typing and phylogenetic analysis revealed large clonal complexes

The MLST analysis revealed that most of the sheep strains considered in this study had a cluster of two large clonal complexes (CC), CC130 and CC133. The exceptions are strain O55, which was grouped in CC425, together with the strains cattle (LGA251) and human (NCTC13552). The strain O217, which comprises CC30 together with the strains isolated from humans (ATCC25923, 80wphwpl, SFAARGOS_504, NCTC11965, NCTC13277, and NCTC13811), and camel (ILRI_Emoyle1/1). The strain O331 shares the clonal complex, only the human strains (SAR475 and SHZW450). All alleles predicted in these analyses are shown in **Table S1**.

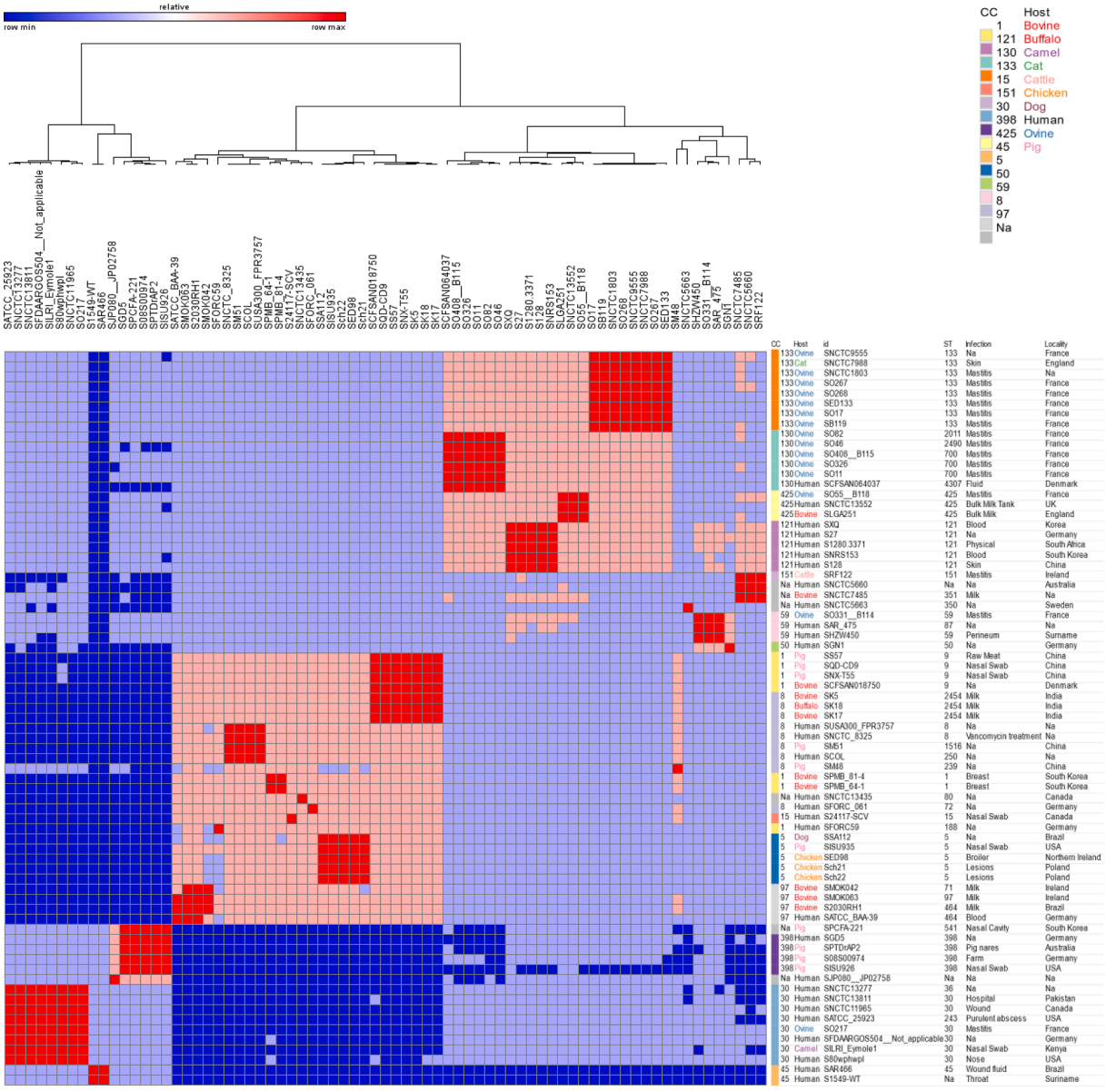

Fig. 2. The study investigates the heatmap of pairwise ANIb values of the 73 *Staphylococcus aureus* genomes. Red a high similarity; Blue a low similarity and light colors (blue and red) have an intermediate similarity. The value of similarity level intensity is based on the ANI analysis by JSpeciesWS. Metadata for cluster correlation: Sequence type (ST), Clonal complex (CC), Host, Sample, and Isolation locality. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.5. Identification of core, accessory, and exclusive genes in selected S. aureus strains

The 73 *S. aureus* genomes were separated into two groups for the comparative genomics analysis. The first group consisted of genomes isolated from the ovine host, and the second group was composed of genomes isolated from other hosts (Table 2). Using 50 % sequence identity as the cutoff value, the BPGA yielded 4200 distinct gene clusters. Out of which 2008 clusters respectively contain at least one gene from each genome and thus construct the core genome, 1559 of these genes are accessory genes (i.e., genes that are present in at least two genomes but not in all the genomes) and 634 unique genes (exclusive genes) of *S. aureus*. The genome statistics, such as the gene family distribution and new gene distribution of the *S. aureus* pangenome, are shown in Fig. 3. In addition, the number of genes unique or exclusively present and absent in each strain of the *S. aureus* dataset is shown in the last two columns of **Table S2**.

At the level of functional genome analysis, the genes belonging to each group's core, accessory, and exclusive genomes were categorized into cellular processes and functions by Cluster of Orthologous Groups (COG) distribution (Fig. 3**c**). Most of the significant spotted genes were associated with cellular metabolism, mainly the transport and metabolism of amino acids and unknown proteins. The accessory and unique genes were classified in greater representativeness with the following COG functional groups, replication (information storage) and recombination and repair (processing activities), general function, defense mechanisms, transcription, and unknown proteins. Out of 4200 gene clusters, BPGA could map 1866 (44.5 %) to KEGG pathways. From a perspective, KEGG assignments from BPGA showed an overall higher representation of metabolism-related pathways (Fig. 3**d**) and **Fig S4** (more detailed version).

The highest values of the number of exclusively absent genes in strains SMOK042 (bovine), SATCC_BAA-39 (human), and SM51(pig). The first two genomes belong to the CC97 and the third to CC8. When compared in this scenario against other hosts, the genomes of the ovine-associated strains did not show exclusively absent genes. Among the 634 unique genes seen, 28 genes were exclusive to ovine-associated strains; we highlight the species SO55, SNCTC9555 (18 and 6), and the SB119, SO46, SO82, and SO268 strains, with only one exclusive gene in each genome in TableS2. When these sequences are characterized *in silico* prediction, primarily hypothetical proteins, phage proteins, and only four proteins with related functions, such as DNA binding, cytosolic resistance, hyperosmolarity, and centrosome B protein, they are related to molecular function, cellular components, and biological processes.

Therefore, analyzing the inter-host (ovine) in subsets was necessary. In the subset analysis conducted, the pangenome profile analysis for fifteen ovine hosts and fifty-eight other hosts of the *S. aureus* dataset may help identify host-specific or virulence-specific genes in *S. aureus*. The results achieved for gene family distribution in these two subsets are shown in Table S3 and Table S4. We observed the presence of 277 unique genes in sheep strains. The genomes that presented the most exclusive genes were the genes corresponding to the complex clonal singletons, the SO55, SO217, and SO331 (72, 54, and 41) exclusive genes, respectively. The functional characterization resulted in products such as hypothetical proteins (39.6 %), phages (6.4 %), transporter (2.9 %), toxins (6.4 %), and others (44.7 %). A very high percentage of exclusive ovine *S. aureus* genes encode proteins with unknown functions, evidencing the need for further studies. Also, genes encode proteins involved in different cellular functions, such as iron acquisition proteins, transferases, DNA helicases, and coagulases. Interestingly the exclusive ovine genes are found in prophages and mobile genetic elements, suggesting that many of the genetic traits related to the ovine host specificity might be horizontally disseminated.

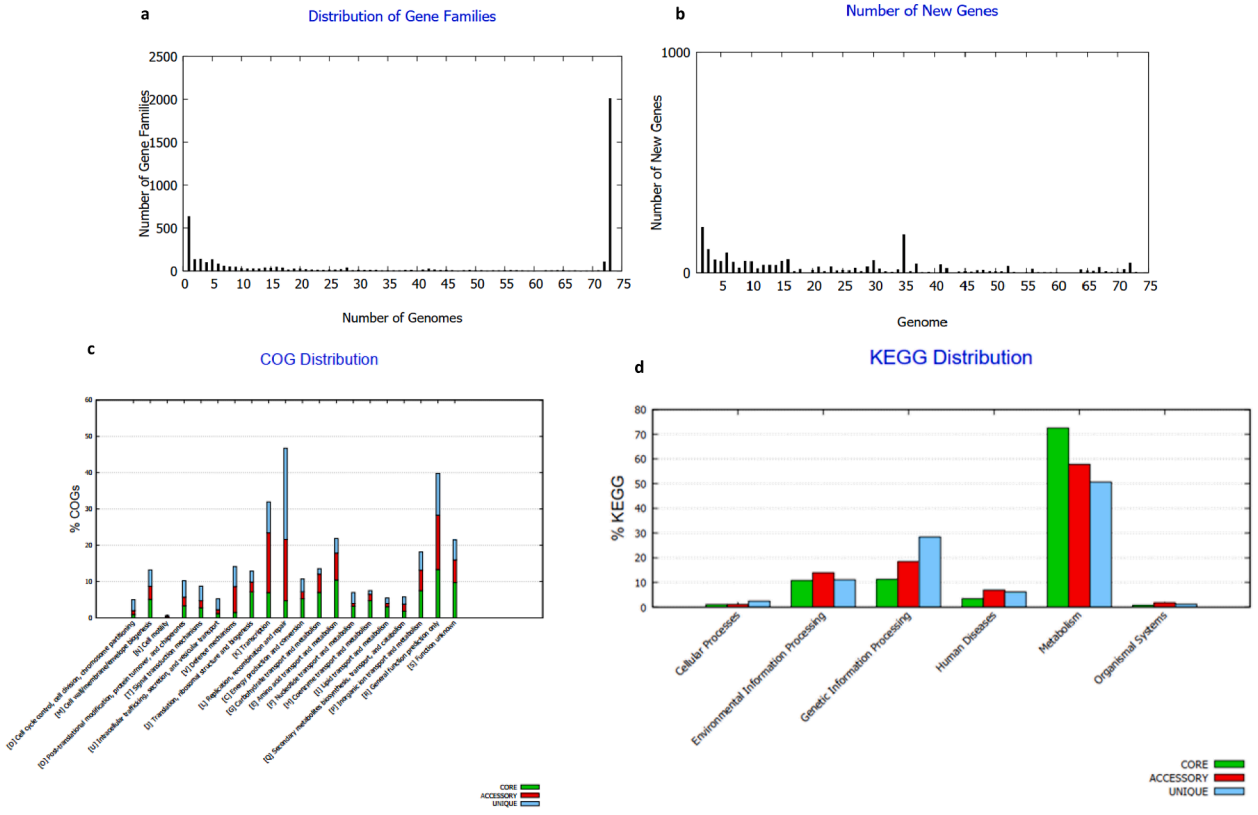

Fig. 3. Graphical representations of profiles of the genes assessed in the BPGA comparative analysis of all 73 selected *S. aureus* strains. (**a**) The gene family frequency spectrum. (**b**) New gene family distribution after sequential addition of each genome to the analysis. (**c**) COG distribution of core, accessory, and unique genes. (**d**) KEGG distribution of core, accessory, and unique genes.

4.6. Virulence genes in the S. aureus strains

Eighty genes were observed, considered encoders of the virulence factors, and 57 to 72 genes were present in all seventy-three genomes. All virulence proteins predicted using the VFDB database are listed in (Fig. 4**a & 4b**).

The exciting thing is the sharing of genes related to the type of clonal complex, regardless of its type of host. However, there are some signatures regarding the suborder, as in the case of ruminants and small ruminants (Fig. 4**b**). When analyzed at the ovine subgroup level, we observed 59 to 70 virulence genes. CC130 and CC133 (large ovine clusters) have a repertoire of similar genes compared with singletons: CC30, CC59, and CC425. Some genes will be clustered according to the CC (twenty-nine genes), genes related to enzymes, *chp* (chemotaxis inhibitory protein of *Staphylococcus* - CHIPS) and *sak* (staphylokinase), and toxins such as the *lukS-PV* gene (Panton-Valentine leukocidin), at the level of clonal complexes shared with ovines, are seen only in strains isolated from humans. Some of these genes share only an ovine strain, such as the *cna* (collagen-binding protein) and the *scn* (staphylococcal complement inhibitor) genes, in the SO217 and SO331 strains, respectively. Three genes involved in the secretion system, Type VII, *esaC*, *essC,* and *esxB* relationship, were predicted only in CC130 and CC133 compared to other complexes shared with them. The overall comparison level (n = 73) had a significant presence and absence when comparing the genes associated with the secretion system (Fig. 4**b**).

Toxin.

Stress protein Secretion system Iron uptake Immune evasion.

Enzyme Capsular Adherence.

Pig.

4.7. Resistance genes in the ovine S. Aureus strains

The analysis of the presence and absence of resistance genes was performed via ABRicate with the support of the five databases and other relevant data available in PATRIC, such as AMR phenotypes for genomes, AMR genes, and AMR regions, aimed at predicting resistance to some antibiotic (Fig. 5 **and** Fig. 6). ABRicate predicted 37 genes related to some antimicrobial resistance for the 73 genomes. PATRIC is classified into eight classes of antibiotic drugs (ciprofloxacin, clindamycin, erythromycin, gentamicin, methicillin, penicillin, tetracycline, and trimethoprim-sulfamethoxazole) from a database containing 97 antibiotics. Analyzing the repertoire of resistance-related genes for ovines, each database yielded different predictions, with some of these genes varying in nomenclature. MEGARES 2.00 (18), CARD (10) ARG-NNOT (8), NCBI AMRFinderPlus (5) and Resfinder (2). All ovine strains did not show the presence of the chromosomal cassette SCCmec, referring to phenotypic resistance to methicillin; however, some genomes showed resistance to penicillin and gentamicin *in silico* analysis (Fig. 5). The *bla* operon genes (*blaI*, *blaR*, and *blaZ*) were predicted only in isolates SO217 (CC30) and SO55 (CC425). The *fosB* gene was present only in B119, ED133, NCTC1803, NCTC9555, O17, O267, O268, O217, and O55. Eight genomes showed the same nucleotide variations in the *GlpT* (A100V) and *murA* (E291D and T396N) genes, and both mutations confer resistance to fosfomycin. Only the SO217 genome showed specific variations in *GlpT* (A100V and V213I) and *murA* (D278E and E291D) (Table 3).

4.8. Prophages

In this study, in the 73 bacteria of *S. aureus* species, 299 phage-like elements were detected; 114 were classified as intact prophages, 133 as incomplete, and 52 were questionable using the PHASTER tool. The distribution of completeness among hosts can be visualized in Fig. 7 and Table S5. All genomes had at least one predicted prophage in the chromosomal region and were considered lysogenic bacteria (bacteria containing the integrated prophage) or poly-lysogenic bacteria (bacteria containing more than one integrated prophage). Table S5 can see more details about these regions, such as the number of predicted regions in each genome, totally predicted proteins, region position, most common phage, and GC% content. Of the prophages considered intact or complete (1 1 4), the smallest was found in *S. aureus* O408_B115 (ovine) with 17.8 kb, and the largest was in M48 (pig) with 134.2 kb. In the 52 considered questionable, the lowest was found in NCTC13811, 8 kb, and the highest in O46 (ovine), 61.4.5 kb. For the prophages considered

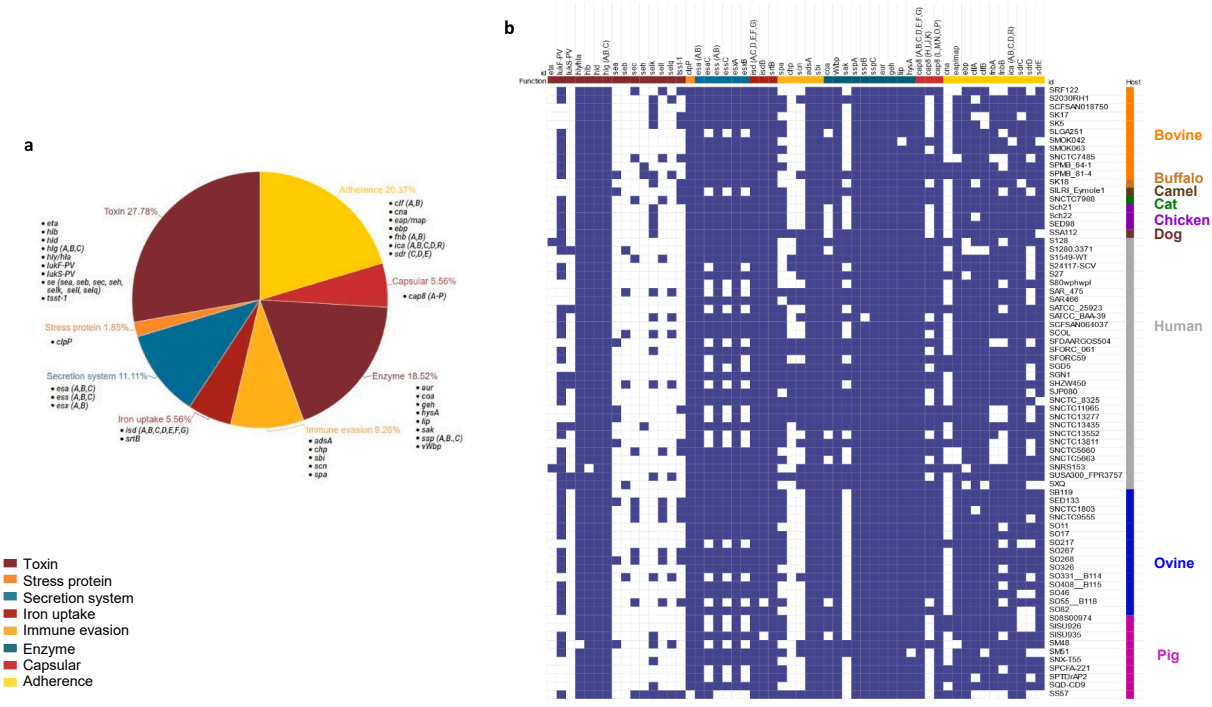

Fig. 4. Visualization of core virulome genes of the *S. aureus* strains according to the VFDB prediction (**a**) Graphical representations of virulence factor classes of these genes. (**b**) Heatmap with presence and absence associated with virulence factors that *S. aureus* correlates to strains and hosts.

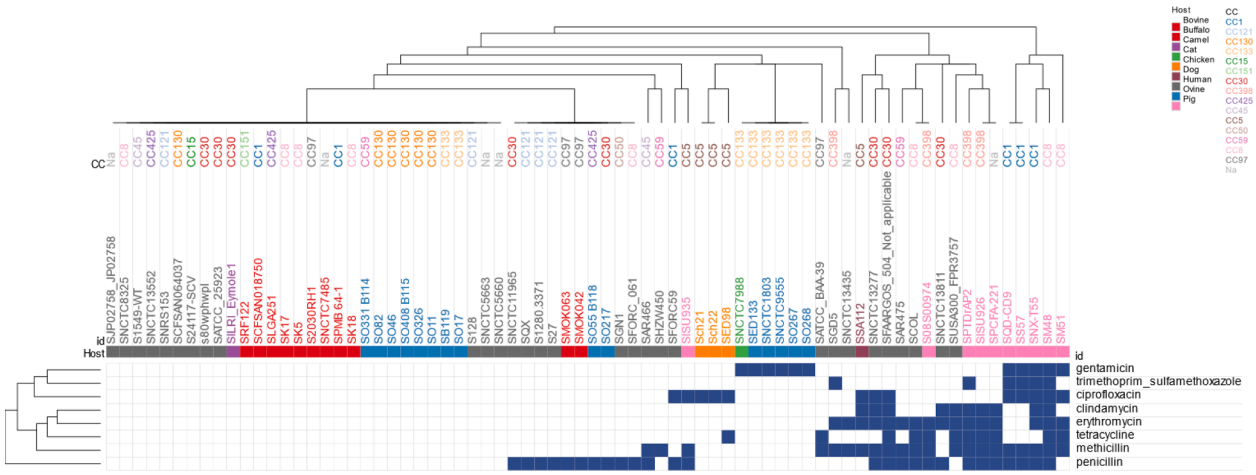

Fig. 5. Visualizing antimicrobial resistance (antibiotic) of the *S. aureus* strains according to the PATRIC prediction. Heatmap with presence (blue) and absence associated with resistance genes that *S. aureus* correlates to strains, hosts, and complex clones (CC). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

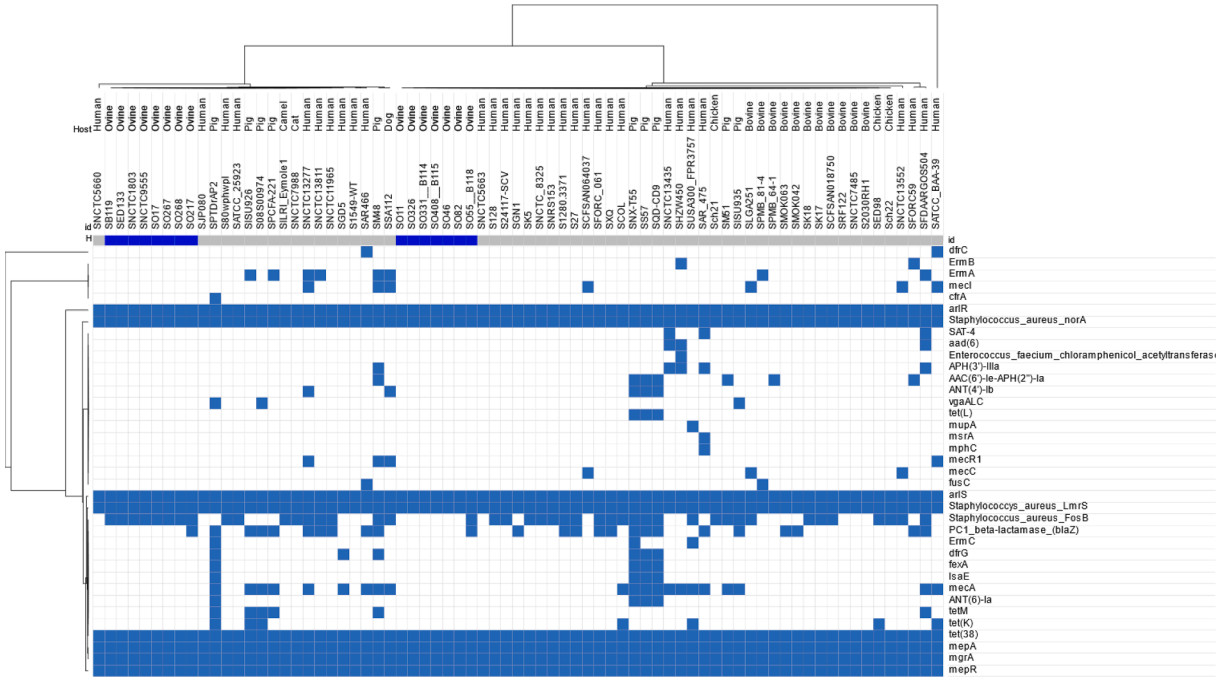

Fig. 6. Visualization of core resistome genes of the *S. aureus* strains according to the CARD prediction. Heatmap with presence (blue) and absence associated with resistance genes that *S. aureus* correlates to strains and hosts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

incomplete (degenerate prophages), the smallest size was M48 (pig) with 5.8 Kb and the largest MOK063 (bovine) 45.5 Kb. At the ruminant level, ovine showed approximately twice as many prophages classified as intact (33), while in bovine, only 12 can be observed as a percent in Fig. 7. All ovine genomes showed regions referring to prophages, and 68 regions were identified (intact = 33, incomplete = 28 and questionable = 7). Each isolate showed at least three to seven regions; only strain O331_B114 showed one region (PHAGE_Staphy_PT1028_NC_007045), and even then, with questionable completeness. The most common phages (20 types) were identified as bacteriophage sequences for *Staphylococcus phage* (16) and other types such as *Paenibacillus phage* (1) *Clostridium phage* (1), *Streptococcus prophage* (1) and *Thermus phage* (1). Only PHAGE_Staphy_PT1028_NC_007045 was predicted in all ovine isolates and PHAGE_Staphy_phiPV83_NC_002486 in almost all except O217, O331_B114 and O55_B118.

All prophages had their ORFs predicted for virulence factors (128 genes) using the VFDB; 349 genes were identified in 67 genomes, and only six strains (CFSAN018750, K17, LGA251, K18, 08S00974, ED98, NCTC13552) did not identify virulence factors in their prophage sequences Table S6. In ruminants (ovine and bovine), virulence genes associated with a toxin (32 and 30), immune evasion (13 and 3), and enzyme (12 and 24), respectively, were identified. It is noted that there was an absence of genes related to adherence and secretion system carried by prophages in these hosts in Fig. 8. The complete list of virulence genes predicted in prophage-related sequences can be seen in Table S6.

In the circular comparative genome mapping representation (Figs. 9 and 10), it is possible to visualize the regions related to prophages and

Table 3
Resistance genes identified by *in silico* prediction against ABRIcate of fifteen ovine *S. aureus* strains.

AMR Gene	SNP	Detection Criteria	AMR Gene Family	Drug Class	Resistance Mechanism
mepA			multidrug and toxic compound		
mepR			extrusion (MATE) transporter	glycylcycline, tetracycline antibiotic	
arlR					
arlS				fluoroquinolone antibiotic, acridine dye	
Staphylococcus aureus **norA**				fluoroquinolone antibiotic	
Staphylococcus aureus **LmrS**		protein homolog model	major facilitator superfamily (MFS) antibiotic efflux pump	macrolide antibiotic, aminoglycoside antibiotic, oxazolidinone antibiotic, diaminopyrimidine antibiotic, phenicol antibiotic	antibiotic efflux
tet (38)				tetracycline	
PC1 beta-lactamase (blaZ)			blaZ beta-lactamase	penam	
Staphylococcus aureus **FosB**			fosfomycin thiol transferase		antibiotic inactivation
GlpT*	A100V		antibiotic-resistant GlpT	fosfomycin	
murA*	E291D, T396N	protein variant model	antibiotic-resistant murA transferase		antibiotic target alteration

* *Staphylococcus aureus* GlpT /murA with mutation conferring resistance to fosfomycin.

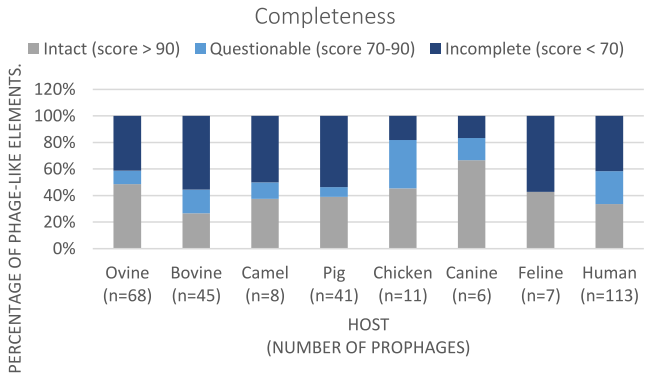

Fig. 7. . Distribution of the prophage-like elements within *S. aureus* in a different host. The values of the number of features found (n) are shown below each host's name. Completeness is divided into three stages represented by different colours (intact = grey, questionable = light blue and incomplete = dark blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

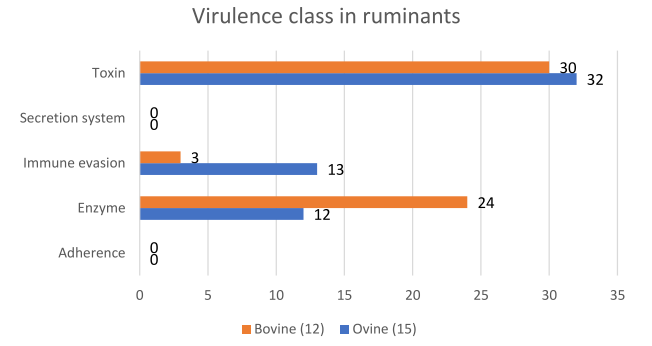

Fig. 8. Distribution of virulence class predicted with VFDB tool in prophage sequences in ruminants. Axis × is composed of the number of predicted genes. Axis y is categorized into all predicted virulence classes. The small ruminant (ovine) is the blue bar, while the orange bar represents the large ruminant (bovine). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

predicted virulence genes positioned along the genomes.

5. Discussion

S. aureus is an important pathogen of ovine mastitis (Le Maréchal et al., 2011; Smith et al., 2014). This infection is difficult to control since *S. aureus* can spread within the herd and become resistant to antibiotic treatment (Oviedo-Boyso et al., 2007). In addition, the number of ovine-associated *S. aureus* genomes available in the NCBI database and genomic studies involving these strains are scarce. Therefore, more information about *S. aureus* genomic characteristics is necessary to efficiently develop new strategies to control infection (Peton et al., 2014).

The isolates from France are of different herds, cities, and years, yet they share a high similarity. For example, the strains SAR475 and O331 are grouped in the MLST analysis. Considering that SAR475 was isolated from humans and O331 from ovine mastitis, this result shows that these two strains are closely related. Additionally, LGA251 and O55 are once more grouped, reinforcing the close phylogenetic relationship between these two strains isolated from different hosts and countries. These two strains present mammary gland tropism since LGA251 was isolated from a cow's bulk milk and O55 from ovine mastitis. In turn, the grouping of O217, IRLI_Emoyle1/1, and ATCC25923 is not supported in the phylogenomic approach, in contrast to the result provided by the MLST analysis. This result suggests that although these three strains belong to the same clonal complex (CC30), they do not share an extremely high level of similarities in their proteomes. Two ovine clusters were formed in the phylogenomic analysis with similarities ranging from 97 to 100 %. It is possible to observe clonal behavior among them compared with strains from the other hosts. However, these clusters did not group three strains (O217, O55, and O331) from ovine. O217 was grouped with genomes isolated from humans, agreeing with the CC formation. The same clustering agreement is observed for O55, which belongs to the same clonal complex of LGA251.

The ANI method is another way to measure evolutionary relatedness among closely related bacterial strains through the identity and similarity values of the total genome sequence (Konstantinidis, Ramette and Tiedje, 2006; Kim et al., 2014b). In this analysis, all ovine genomes presented more than 98 % similarity regarding the ED133 genome, except for O217, which showed a minor similarity among sheep strains. In the same way, this strain was not grouped with any ovine genomes in MLST analysis. This study showed that it belongs to CC30, which is important in human and animal infections, and suggests no-host-specialization by this strain. On the other hand, genomes isolated from

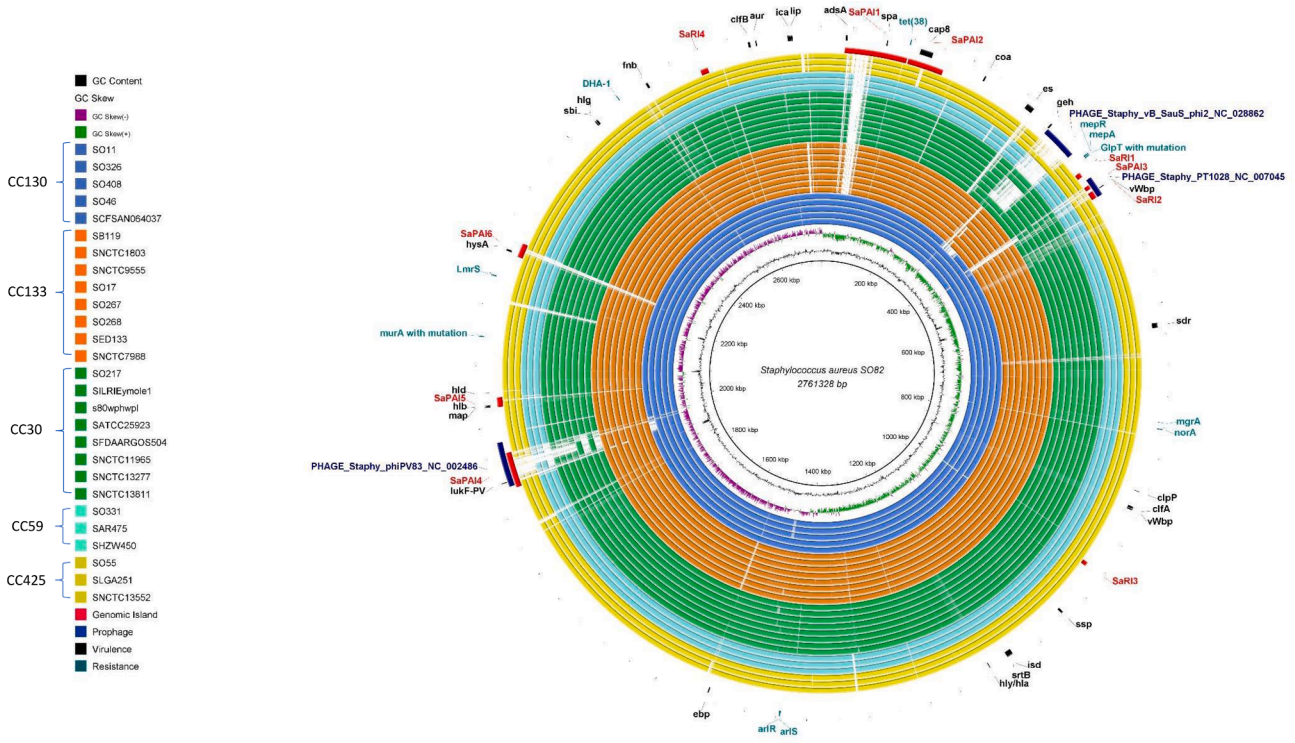

Fig. 9. Comparative circular genome maps the representation of the BRIG software with 28 genomes of *Staphylococcus aureus* isolates of ovines and other hosts. The O82 strain was the reference in the central position, with the first three inner rings showing its size, GC content, and GC skew, mapping all prophages, genes associated with resistance, virulence, and pathogenic islands found in the O82 strain.

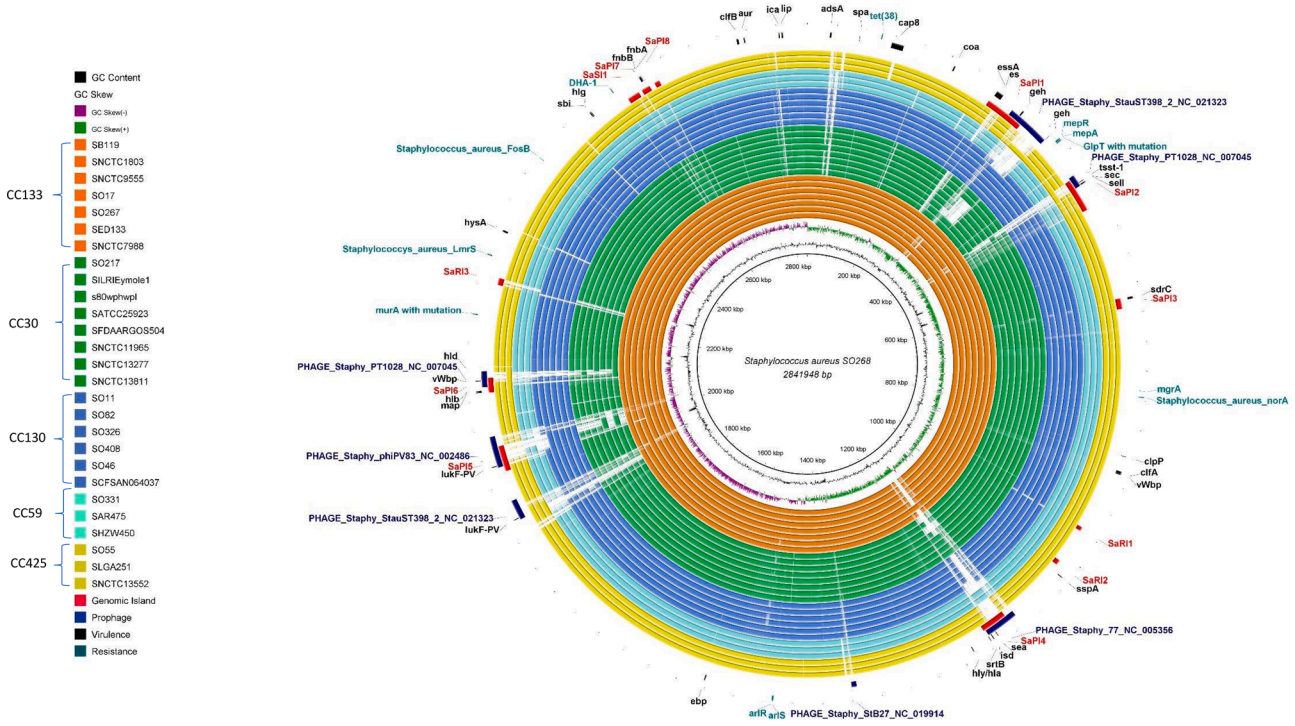

Fig. 10. Comparative circular genome maps the representation of the BRIG software with 28 genomes of *Staphylococcus aureus* isolates of ovines and other hosts. The O268 strain was the reference in the central position, with the first three inner rings showing its size, GC content, and GC skew, mapping all prophages, genes associated with resistance and virulence, and pathogenic islands found in the O268 strain.

milk or mastitis in cows (LGA251 and RF122) present a similar level of 98 % regarding ED133 (isolated from ovine). Previous studies reported that some CCs originated from humans and acquired genetic adaptation to infect ruminants (Guinane, Ben Zakour, et al., 2010; Sakwinska et al., 2011). Furthermore, the evolutionary separation of some ovine and bovine strains is unclear in the present study.

Additionally, three of the selected ovine strains belonged to the CC30 (STs 30 and 243). Studies have associated these STs with human, bovine, and ovine infections in Europe and Asia (Rabello et al., 2007; Monecke, Slickers and Ehricht, 2008; Smith et al., 2014; Haag, Fitzgerald and Penadés, 2019b). The strains ED133, O268, O17, O322, and O267 present the same ST and belong to the CC133, which is associated with intramammary infections in ovine (Smyth et al., 2009; Smith et al., 2014). Interestingly, only strain O331 presents the ST59. *S. aureus* ST59 has already been isolated from bovine milk (Hata et al., 2010; Richardson et al., 2018) and dairy goat mastitis (Chu et al., 2012). Therefore, the strain O331 is likely an atypical clone among the ovine strains considered in this study. Studies have shown that ST59 isolates originated in the United States and have spread to Asia. Furthermore, ST59 *S. aureus* has been isolated in Oceania and northern Europe (Tristan et al., 2007; Mediavilla et al., 2012; Smith et al., 2014). Further studies are necessary to track the origin of the ST59 isolates in France. A previous study suggested that the Multilocus Sequence Analysis (MLSA) can accurately determine the evolutionary relationships among the strains of *Flavobacterium columnare* (Kayansamruaj et al., 2017). However, a comparative genomics approach can provide a much higher strain typing resolution than the MLST analysis since more genes are considered (Hall, Ehrlich and Hu, 2010). In the MLST-based phylogenetic tree generated, the group of CC130 is divided into two different clades, one formed by the STs 2490 (O46) and 2011 (O82) and the other formed by ST700 (Fig. 1). In turn, the phylogenomic analysis based on the amino acid variation in the proteomes of the *S. aureus* strains resulted in the grouping of all strains belonging to the CC130, with a bootstrap value of 1. However, the phylogenomic analysis also reinforced that STs 2490 and 2011 are more closely related than ST700. This result indicates that the determination of phylogeny benefitted from a proteome-scale analysis compared to the limited number of gene *loci* considered in the MLST-based phylogenetic analysis.

A very conserved synteny level is observed on the ovine *S. aureus* strains. However, a few regions of DNA inversion were found. Differences caused by gene acquisition from phages and DNA recombination events likely explain the size variation among the ovine strains' genomes. One example involves PHAGE_Staphy_phiPV83_NC_002486, which was absent only in the O217, O331_B114 and O55_B118 genomes (Figs. 9 and 10). To evaluate the clonal relationship of the ovine *S. aureus* strains, MLST analysis was conducted. If at least 5 out of the seven alleles (*arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi,* and *yqiL*) in the different loci are shared between two *S. aureus* isolates, these isolates form a single clonal complex (Maiden et al., 1998; Enright et al., 2000b; Tettelin et al., 2005). The CC130, which had already been associated with ovine mastitis (Guinane, Ben Zakour, et al., 2010; Smith et al., 2014 Haag, Fitzgerald and Penadés, 2019b), encompassed 5 of the ovine strains considered in this study (STs 2411, 2490, and 700).

The genomic diversity of multiple strains can be uncovered by determining the core, accessory, and single genomes (Bosi et al., 2016b; Chaudhari, Gupta and Dutta, 2016b). The pangenome curve almost reached a plateau, indicating that it is open, and the *S. aureus* pangenome will probably close soon (Fig. S1). All strains used in this comparative analysis have an average of resulting genes predicted by the PATRIC server. The analysis of proteins exclusively shared among genomes of *S. aureus* involved in a single form of mastitis resulted in predicting hypothetical and phage-related proteins. However, exclusive genes are shared by at least two mastitis group strains and not by all strains (Table S4). Therefore, further studies are necessary to elucidate the molecular basis of *S. aureus* virulence that might lead to the development of different forms of mastitis. Additionally, specific traits such as susceptibility to ovine and interference with the environment and intestinal microbiota dysbiosis could contribute to developing a determined form of mastitis caused by *S. aureus* (Vonaesch, Anderson and Sansonetti, 2018; Oikonomou, Addis, Chassard, Nader-Macias, Grant, Delbès, Bogni, le Loir, et al., 2020).

The comparative genomics performed using BPGA software yielded 4200 genes. In turn, the core genome is formed by 2008 genes among all strains. It represents 48 % of the *S. aureus* genes average predicted in this study and a core genome 1,7-fold proportionally higher than a previous study of comparative genomic analysis in 64 *S. aureus* strains, which resulted in a core genome of 56 % on average of the predicted genes (Bosi et al., 2016b). The core genome analysis was also performed among ovine and other host strains, resulting in 2149 and 2119 genes, respectively. It represents 79 % (ovine strains) and 81 % (other hosts) of the average genes predicted by the PATRIC server. This study's number of core genes suggests low diversification in the core genome of *S. aureus* strains used for comparison. The functions of core genes were classified by COG categories, playing a role in amino acid transport and metabolism. These data are expected because the core genome belongs to the group of housekeeping functions (Tettelin et al., 2005 Medini et al., 2005).

The accessory genes predicted in this present study (n = 1559) are primarily involved in replication, recombination, and repair. These genes play a role in lateral gene transfer events from mobile genetic elements, such as transposons and bacteriophages (Tettelin et al., 2005Bosi et al., 2016b). Of the exclusive genes found in ovine genomes (n = 277), the central part consists of hypothetical proteins (39.6 %), which are proteins of unknown function, and can contribute to many activities in the genome. For instance, a functional assignment of hypothetical proteins in the *S. aureus* study predicted these as binding proteins, helicases, transporters, and virulence factors (Prava, G and Pan, 2018). Mobile genetic elements represent 25 % of the exclusive ovine genes. They are essential in bacterial diversity due to their capacity to transduce host genes and confer novel genetic information for bacteria, such as genes involved in virulence (Kwan et al., 2005). Some genes are predicted to play a role in virulence and are carried by phage, *entB* (Staphylococcal enterotoxin B), *selk* (Staphylococcal enterotoxin K), and *vWpb* (Secreted von Willebrand factor-binding protein precursor). Genes encoding cell wall proteins represent 3 % of the genes found in all ovine genomes. These proteins are essential for adherence (Silhavy, Kahne and Walker, 2010) and antibiotic resistance (Assis, Nedeljković and Dessen, 2017).

Staphylococci are significant causes of drug-resistant infections, and phages that infect and destroy these organisms have enormous therapeutic potential (Nayeemul Bari and Hatoum-Aslan, 2019). The five prophages (intact = 1 and incomplete = 4) were found in only ovine strains, not in strains from other host groups. The PHAGE_Staphy_42E_NC_007052, present only in NCTC9555, was predicted to be an intact phage with a PHASTER score of 150 and produces 69 proteins. The phage gene set is functionally classified into integrase, tail, portal, head and capsule. Of note, (Kwan et al., 2005) showed that gene transfer between *S. aureus* phages is more prevalent than between *S. aureus* and other species. However, four incomplete regions were found only in ovine, more commonly in other species different from *S. aureus.* The PHAGE_Strept_EJ_1_NC_005294 (O217) in the *Myoviridae* morphology family and found an atypical *Streptococcus pneumoniae* (Romero, López and García, 2004), the PHAGE_Thermu_OH2_NC_021784 (O268), in *Geobacillus kaustophilus* (Doi et al., 2013) and the PHAGE_Clostr_phiMMP03_NC_028959 (NCTC9555) in *Clostridioides difficile* (Rashid et al., 2016). The protocols described in Nayeemul Bari and Hatoum-Aslan, 2019, hoped to advance the basic understanding of *Staphylococcus* phages while allowing the development of more powerful phage-based antimicrobials. Likewise, 32 genes were predicted to be toxins precursors in phage regions. Toxins have an essential role in pathogenicity; in this context, *S. aureus* exotoxins are a leading cause of gastroenteritis in humans from the consumption of contaminated food, principally raw milk and raw milk cheese, which can infect animals(Balaban and Rasooly, 2000; Le Loir, Baron and Gautier, 2003; Oikonomou, Addis, Chassard, Nader-Macias, Grant, Delbès, Bogni, le Loir, et al., 2020; van den Brom et al., 2020). Additionally, exfoliative toxins were found in *S. aureus* isolated from cows, and these proteins are an agent of a scalded-skin syndrome in humans (Vautor

et al., 2009; Que and Moreillon, 2014). In addition, the exclusive ovine genes were compared to the VFDB database through BLASTp. They resulted from 14 predicted proteins, such as the coagulase (*coa*), secreted by *S. aureus* and causes clotting in the host's plasma. As a result of mechanisms of escape from the immune system (Salaberry et al., 2015; Javid et al., 2018), hyaluronate lyase precursor (*hysA*) plays a role in subcutaneous infection (Makris et al., 2004; Ibberson et al., 2014) and toxins (Le Maréchal, Seyffert, et al., 2011; Monistero et al., 2018). It was previously reported that different levels of iron metabolism, transcriptional regulators, and exoprotein production could contribute to ovine mastitis severity, and different levels of toxin expression were related to the pathogenic potential of genomes isolated from bovine mastitis (ben Zakour et al., 2008; Le Maréchal et al., 2011).

These strains were predicted to be sensitive to methicillin. However, the horizontal acquisition of resistance genes could increase the relevance of ovine mastitis in human infection with methicillin-resistant *S. aureus*. Although the SCCmec chromosome cassette was not found in the genomes of the ovine strains considered in this study, these genomes present multiple chromosomal factors that can contribute to the methicillin resistance phenotype, such as the *fmtB* gene (Komatsuzawa et al., 1997). It has already been shown that the *fmtB* gene indirectly affects methicillin resistance, but further biochemical studies are necessary to elucidate the role of *fmtB* in this resistance mechanism (Komatsuzawa et al., 2000).

6. Conclusion

The comparative genomics among the ovine and other host-associated isolates of *S. aureus* showed genetic differences. Our results suggest that the accessory genes encoding virulence factors and unknown proteins that might be essential in establishing infection are exclusively found in the ovine genomes. Although it is possible to observe that the *S. aureus* genomes are very similar at the genomic level, independent of the host, clonal complexes CC130 and CC133, CC425, CC59, and CC30 are associated with small ruminants, where CC130 and CC133 are considered representative of the host, ovine. As for the invasion and permanence of the bacteria in the cell, it showed a vast repertoire of virulence-associated genes. All strains showed genes related to antimicrobial resistance. However, at the level of antibiotic resistance, some highly resistant representatives were found predicted *in silico*, as observed in the genomes isolated from swine. The cassette chromosome *mec* was not detected in strains isolated from the ovine. However, it showed genes associated with resistance in specific genomes, which may be associated with the reservoir of resistance genes for other hosts. This work shows new evidence of genomic specialization in *S. aureus* associated with ovine mastitis.

Funding

This research received no external funding.

Data availability statement

Genome sequences are available in the NCBI database.

Declaration 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.

Data availability

Data will be made available on request.

Acknowledgements

We thank the agencies Coordination for the Improvement of Higher Education Personnel (CAPES),CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and Minas Gerais Research Funding Foundation (FAPEMIG). Our thanks to the Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), Post Graduate Program in Genetics and Post Graduate Program in Bioinformatics at the Federal University of Minas Gerais, and the Omics Science Network (RECOM).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2022.147131.

References

Acosta, A.C. et al., 2018. 'Frequency of Staphylococcus aureus virulence genes in milk of cows and goats with mastitis', Pesquisa Veterinária Brasileira, 38(11), pp. 2029–2036. Available at: 10.1590/1678-5150-pvb-5786.

Aires-de-Sousa, M. et al., 2007. Characterization of Staphylococcus aureus isolates from buffalo, bovine, ovine, and caprine milk samples collected in Rio de Janeiro State, Brazil. Appl. Environ. Microbiol. 73(12), pp. 3845–3849. Available at: 10.1128/AEM.00019-07.

Alikhan, N.F. et al., 2011. 'BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons, BMC Gen. 12(1), p. 402. Available at: 10.1186/1471-2164-12-402.

Altschul, S.F. et al., 1990. 'Basic local alignment search tool', J. Mol. Biol. 215(3), 403–410. Available at: 10.1016/S0022-2836(05)80360-2.

Alves, P.D.D. et al., 2009. Molecular characterisation of Staphylococcus aureus strains isolated from small and large ruminants reveals a host rather than tissue specificity. Vet. Microbiol. 137(1–2), 190–195. Available at: 10.1016/j.vetmic.2008.12.014.

Arndt, D. et al.,. 2016.'PHASTER: a better, faster version of the PHAST phage search tool', Nucl. Acids Res. 44(W1) W16–W21. Available at: 10.1093/nar/gkw387.

Assis, L.M., Nedeljković, M., Dessen, A., 2017. New strategies for targeting and treatment of multi-drug resistant Staphylococcus aureus. Drug Resistance Updates, 31, pp. 1–14. Available at: 10.1016/j.drup.2017.03.001.

Astrup, L.B., Pedersen, K., Farre, M., 2022. Microbiological diagnoses on clinical mastitis—comparison between diagnoses made in veterinary clinics versus in laboratory applying MALDI-TOF MS. Antibiotics 11 (2), 271. Available at: 10.3390/antibiotics11020271.

Balaban, N., Rasooly, A., 2000. 'Staphylococcal enterotoxins', Int. J. Food Microbiol. 61 (1) (2000) 1–10. Available at: 10.1016/S0168-1605(00)00377-9.

Bankevich, A. et al., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Computat. Biol. 19(5) (2012) 455–477. Available at: 10.1089/cmb.2012.0021.

ben Zakour, N.L. et al., 2008. Genome-wide analysis of ruminant staphylococcus aureus reveals diversification of the core genome. J. Bacteriol. 190(19), 6302–6317. Available at: 10.1128/JB.01984-07.

Bosi, E. et al., 2016a.'Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity. Proc. Natl. Acad. Sci. 113(26). Available at: 10.1073/pnas.1523199113.

Chaudhari, N.M., Gupta, V.K., Dutta, C., 2016a. BPGA- an ultra-fast pan-genome analysis pipeline. Scientific Reports, 6(1), p. 24373. Available at: 10.1038/srep24373.

Chaudhari, N.M., Gupta, V.K., Dutta, C., 2016b. BPGA-an ultra-fast pan-genome analysis pipeline. Scientific Reports, 6(March), pp. 1–10. Available at: 10.1038/srep24373.

Chen, L. et al., 2016.'VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on', Nucl. Acids Res. 44(D1), D694–D697. Available at: 10.1093/nar/gkv1239.

Chu, C. et al., 2012. 'Genetically divergent methicillin-resistant Staphylococcus aureus and sec-dependent mastitis of dairy goats in Taiwan', BMC Vet. Res., 8(1), 39. Available at: 10.1186/1746-6148-8-39.

Dobrindt, U., 2001. Whole genome plasticity in pathogenic bacteria. Curr. Opin. Microbiol. 4(5), 550–557. Available at: 10.1016/S1369-5274(00)00250-2.

Doi, K. et al., 2013. Draft Genome Sequence of Geobacillus kaustophilus GBlys, a Lysogenic Strain with Bacteriophage ϕOH2. Genome Announcements, 1(4). Available at: 10.1128/genomeA.00634-13.

Doster, E. et al., 2019.'MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data'. Nucleic Acids Res. [Preprint]. Available at: 10.1093/nar/gkz1010.

Enright, M.C. et al., 2000. 'Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of Staphylococcus aureus'. J. Clin. Microbiol. 38(3), 1008–1015. Available at: 10.1128/JCM.38.3.1008-1015.2000.

Feldgarden, M. et al., 2019a. 'Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates'. Antimicrobial Agents Chemotherapy, 63(11). Available at: 10.1128/AAC.00483-19.

Francisco, A.P. et al., 2009. 'Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach'. BMC Bioinform. 10(1), 152. Available at: 10.1186/1471-2105-10-152.

Francisco, A.P. et al., 2012. 'PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods'. BMC Bioinformat.13(1), p. 87. Available at: 10.1186/1471-2105-13-87.

Galardini, M. et al., 2011. 'CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes.', Source Code Biol. Med., 6(1), 11. Available at: 10.1186/1751-0473-6-11.
Goris, J. et al., 2007. 'DNA–DNA hybridization values and their relationship to whole-genome sequence similarities'. Int. J. Systemat. Evolut. Microbiol. 57(1), 81–91. Available at: 10.1099/ijs.0.64483-0.
Guinane, C.M., ben Zakour, N.L., et al, 2010. 'Evolutionary genomics of Staphylococcus aureus reveals insights into the origin and molecular basis of ruminant host adaptation.'. Genome Biol. Evolut. 2, 454–66. Available at: 10.1093/gbe/evq031.
Gupta, S.K. et al., 2014.'ARG-ANNOT, a New bioinformatic tool to discover antibiotic resistance genes in bacterial genomes, Antimicrobial Agents Chemotherapy, 58(1), 212–220. Available at: 10.1128/AAC.01310-13.
Haag, A.F., Fitzgerald, J.R., Penadés, J.R., 2019. 'Staphylococcus aureus in Animals', Microbiology Spectrum. Edited by V.A. Fischetti et al., 7(3). Available at: 10.1128/microbiolspec.GPP3-0060-2019.
Halasa, T. et al., 2009. 'Stochastic bio-economic model of bovine intramammary infection', Livestock Sci. 124(1–3), pp. 295–305. Available at: 10.1016/j.livsci.2009.02.019.
Hall, B.G., Ehrlich, G.D., Hu, F.Z., 2010. 'Pan-genome analysis provides much higher strain typing resolution than multi-locus sequence typing'. Microbiology 156(4), 1060–1068. Available at: 10.1099/mic.0.035188-0.
Hata, E. et al., 2010. 'Genetic Variation among Staphylococcus aureus Strains from Bovine Milk and Their Relevance to Methicillin-Resistant Isolates from Humans'. J. Clin. Microbiol. 48(6), 2130–2139. Available at: 10.1128/JCM.01940-09.
Hoekstra, J. et al., 2019.'Differences between Staphylococcus aureus lineages isolated from ovine and caprine mastitis but not between isolates from clinical or subclinical mastitis. J. Dairy Sci. 102(6), 5430–5437. Available at: 10.3168/jds.2018-16196.
Hughes, K., 2020. 'Comparative mammary gland postnatal development and tumourigenesis in the sheep, cow, cat and rabbit: Exploring the menagerie', Seminars Cell Dev. Biol. [Preprint], (xxxx). Available at: 10.1016/j.semcdb.2020.09.010.
Ibberson, C.B. et al., 2014. 'Staphylococcus aureus Hyaluronidase Is a CodY-Regulated Virulence Factor'. Infect. Immunity. Edited by A. Camilli, 82(10), pp. 4253–4264. Available at: 10.1128/IAI.01710-14.
Javid, F. et al., 2018. Molecular typing of Staphylococcus aureus based on coagulase gene. Vet. World 11(4), 423–430. Available at: 10.14202/vetworld.2018.423-430.
Jia, B. et al., 2017, 'CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database., Nucleic Acids Res. 45(D1), D566–D573. Available at: 10.1093/nar/gkw1004.
Kaya, H. et al., 2018. 'SCCmecFinder, a web-based tool for typing of staphylococcal cassette chromosome mec in staphylococcus aureus using whole-genome sequence data.mSphere. Edited by B.M. Limbago, 3(1). Available at: 10.1128/mSphere.00612-17.
Kayansamruaj, P. et al., 2017. 'Comparative genome analysis of fish pathogen Flavobacterium columnare reveals extensive sequence diversity within the species. Infect. Genet. Evolut. 54, . 7–17. Available at: 10.1016/j.meegid.2017.06.012.
Kim, M. et al., 2014.'Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes', International Journal of Systematic and Evolutionary Microbiology, 64 (Pt_2), pp. 346–351. Available at: 10.1099/ijs.0.059774-0.
Komatsuzawa, H. et al., 1997. Cloning and characterization of the fmt gene which affects the methicillin resistance level and autolysis in the presence of triton X-100 in methicillin-resistant Staphylococcus aureus. Antimicrobial Agents Chemother. 41 (11), 2355–2361. Available at: 10.1128/AAC.41.11.2355.
Komatsuzawa, H. et al., 2000. 'Tn551-mediated insertional inactivation of the fmtB gene encoding a cell wall-associated protein abolishes methicillin resistance in Staphylococcus aureus. J. Antimicrobial Chemother. 45(4), 421–431. Available at: 10.1093/jac/45.4.421.
Konstantinidis, K.T., Ramette, A., Tiedje, J.M., 2006.'The bacterial species definition in the genomic era', Philos. Trans. Roy. Soc. B: Biol. Sci. 361(1475) (2006) 1929–1940. Available at: 10.1098/rstb.2006.1920.
Kwan, T. et al., 2005. 'The complete genomes and proteomes of 27 Staphylococcus aureus bacteriophages. Proc. Natl. Acad. Sci. 102(14), 5174–5179. Available at: 10.1073/pnas.0501140102.
Larsen, M.V. et al., 2012. 'Multilocus sequence typing of total-genome-sequenced bacteria. Jo. Clin. Microbiol. 50(4), 1355–1361. Available at: 10.1128/JCM.06094-11.
Le Loir, Y., Baron, F., Gautier, M., 2003. Staphylococcus aureus and food poisoning. Genet. Mol. Res. : GMR 2 (1), 63–76.
Le Maréchal, C. et al., 2011. Molecular basis of virulence in staphylococcus aureus mastitis. PLoS ONE, 6(11). Available at: 10.1371/journal.pone.0027354.
le Maréchal, C., Hernandez, D., et al., 2011. 'Genome Sequences of Two Staphylococcus aureus Ovine Strains That Induce Severe (Strain O11) and Mild (Strain O46) Mastitis. J. Bacteriol. 193(9), 2353–2354. Available at: 10.1128/JB.00045-11.
le Maréchal, C., Seyffert, N., et al., 2011. 'Molecular Basis of Virulence in Staphylococcus aureus Mastitis', *PLoS ONE*. Edited by M.J. Horsburgh, 6(11), e27354. Available at: 10.1371/journal.pone.0027354.
Maiden, M.C.J. et al., 1998. 'Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Nat. Acad. Sci. 95(6), pp. 3140–3145. Available at: 10.1073/pnas.95.6.3140.
Makris, G. et al., 2004. 'The hyaluronate lyase of Staphylococcus aureus – a virulence factor?', Microbiol. 150(6), 2005–2013. Available at: 10.1099/mic.0.26942-0.
Mediavilla, J.R. et al., 2012. 'Global epidemiology of community-associated methicillin resistant Staphylococcus aureus (CA-MRSA). Curr. Opin. Microbiol. 15(5), 588–595. Available at: 10.1016/j.mib.2012.08.003.
Medini, D. et al., 2005. The microbial pan-genome', Current Opinion in Genetics and Development, 15(6), 589–594. Available at: 10.1016/j.gde.2005.09.006.
Merz, A., Stephan, R., Johler, S., 2016. Staphylococcus aureus isolates from goat and sheep milk seem to be closely related and differ from isolates detected from bovine milk. Front. Microbiol. 7 (MAR), 1–7. Available at: 10.3389/fmicb.2016.00319.
Monecke, S., Slickers, P., Ehricht, R., 2008. Assignment of Staphylococcus aureus isolates to clonal complexes based on microarray analysis and pattern recognition. FEMS Immunol. Med. Microbiol. 53 (2), 237–251. Available at: 10.1111/j.1574-695X.2008.00426.x.
Monistero, V., et al., 2018. Staphylococcus aureus isolates from bovine mastitis in eight countries: Genotypes, detection of genes encoding different toxins and other virulence genes. Toxins 10 (6), 1–8. Available at: 10.3390/toxins10060247.
Nayeemul Bari, S.M., Hatoum-Aslan, A., 2019. 'CRISPR–Cas10 assisted editing of virulent staphylococcal phages', in, pp. 385–409. Available at: 10.1016/bs.mie.2018.10.023.
Oget, C., Tosser-Klopp, G., Rupp, R., 2019. Genetic and genomic studies in ovine mastitis. Small Rumin. Res. 176, 55–64. Available at: 10.1016/j.smallrumres.2019.05.011.
Oikonomou, G., Addis, M.F., Chassard, C., Nader-Macias, M.E.F., Grant, I., Delbès, C., Bogni, C.I., le Loir, Y., *et al.* (2020) 'Milk Microbiota: What Are We Exactly Talking About?', *Frontiers in Microbiology*, 11. Available at: 10.3389/fmicb.2020.00060.
Oviedo-Boyso, J., et al., 2007. Innate immune response of bovine mammary gland to pathogenic bacteria responsible for mastitis. J. Infect. 54 (4), 399–409. Available at: 10.1016/j.jinf.2006.06.010.
Page, A.J., et al., 2015. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics 31 (22), 3691–3693. Available at: 10.1093/bioinformatics/btv421.
Peton, V., et al., 2014. Fine-tuned characterization of Staphylococcus aureus Newbould 305, a strain associated with mild and chronic mastitis in bovines. Vet. Res. 45 (1), 106. Available at: 10.1186/s13567-014-0106-7.
Prava, J., G, P. and Pan, A. (2018) 'Functional assignment for essential hypothetical proteins of Staphylococcus aureus N315. Int. J. Biol. Macromol., 108, 765–774. Available at: 10.1016/j.ijbiomac.2017.10.169.
Price, M.N., Dehal, P.S., Arkin, A.P., 2010. 'FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments', *PLoS ONE*. Edited by A.F.Y. Poon, 5(3), p. e9490. Available at: 10.1371/journal.pone.0009490.
Que, Y.A., Moreillon, P., 2014. Staphylococcus aureus (Including Staphylococcal Toxic Shock Syndrome). Eighth Edi, Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. Eighth Edi. Elsevier Inc. Available at: 10.1016/B978-1-4557-4801-3.00196-X.
Rabello, R.F., et al., 2007. Multilocus sequence typing of Staphylococcus aureus isolates recovered from cows with mastitis in Brazilian dairy herds. J. Med. Microbiol. 56 (11), 1505–1511. Available at: 10.1099/jmm.0.47357-0.
Rashid, S., et al., 2016. Two novel myoviruses from the North of Iraq reveal insights into clostridium difficile phage diversity and biology. Viruses 8 (11), 310. Available at: 10.3390/v8110310.
Richardson, E.J., et al., 2018. Gene exchange drives the ecological success of a multi-host bacterial pathogen. Nat. Ecol. Evol. 2 (9), 1468–1478. Available at: 10.1038/s41559-018-0617-0.
Richter, M., et al., 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32 (6), 929–931. Available at: 10.1093/bioinformatics/btv681.
Romero, P., López, R., García, E., 2004. 'Genomic organization and molecular analysis of the inducible prophage EJ-1, a mosaic myovirus from an atypical pneumococcus'. Virology, 322(2), pp. 239–252. Available at: 10.1016/j.virol.2004.01.029.
Ruegg, P.L., 2017. A 100-Year Review: Mastitis detection, management, and prevention. J. Dairy Sci. 100 (12), 10381–10397. Available at: 10.3168/jds.2017-13023.
Sakwinska, O., et al., 2011. Staphylococcus aureus Host Range and Human-Bovine Host Shift. Appl. Environ. Microbiol. 77 (17), 5908–5915. Available at: 10.1128/AEM.00238-11.
Salaberry, S.R.S., et al., 2015. Virulence factors genes of Staphylococcus spp. isolated from caprine subclinical mastitis. Microb. Pathog. 85, 35–39. Available at: 10.1016/j.micpath.2015.05.007.
Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30 (14), 2068–2069. Available at: 10.1093/bioinformatics/btu153.
Sheppard, S.K., Guttman, D.S., Fitzgerald, J.R., 2018. Population genomics of bacterial host adaptation. Nat. Rev. Genet. 1. Available at: 10.1038/s41576-018-0032-z.
Silhavy, T.J., Kahne, D., Walker, S., 2010. The Bacterial Cell Envelope. Cold Spring Harb. Perspect. Biol. 2 (5), a000414–a. Available at: 10.1101/cshperspect.a000414.
Smith, E.M., et al., 2005. Multilocus sequence typing of intercontinental bovine staphylococcus aureus isolates. J. Clin. Microbiol. 43 (9), 4737–4743. Available at: 10.1128/JCM.43.9.4737-4743.2005.
Smith, E.M., et al., 2014. Global distribution and diversity of ovine-associated Staphylococcus aureus. Infect. Genet. Evol. 22, 208–215. Available at: 10.1016/j.meegid.2013.09.008.
Smyth, D.S., et al., 2009. Molecular genetic typing reveals further insights into the diversity of animal-associated Staphylococcus aureus. J. Med. Microbiol. 58 (10), 1343–1353. Available at: 10.1099/jmm.0.009837-0.
Soares, S.C., et al., 2016. GIPSy: genomic island prediction software. J. Biotechnol. 232, 2–11. Available at: 10.1016/j.jbiotec.2015.09.008.
Suárez-Trujillo, A., et al., 2019. Short communication: Differences in distribution of serotonin receptor subtypes in the mammary gland of sheep, goats, and cows during lactation and involution. J. Dairy Sci. 102 (3), 2703–2707. Available at: 10.3168/jds.2018-15328.
Tettelin, et al., 2005a. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc. Natl. Acad. Sci. 102 (39), 13950–13955. Available at: 10.1073/pnas.0506758102.

Tristan, A., et al., 2007. Global Distribution of Panton-Valentine Leukocidin–positive Methicillin-resistant Staphylococcus aureus, 2006. Emerg. Infect. Dis. 13 (4), 594–600. Available at: 10.3201/eid1304.061316.

van den Brom, R. et al., 2020. 'Zoonotic risks of pathogens from sheep and their milk borne transmission'. Small Ruminant Res. 189(May), p. 106123. Available at: 10.1016/j.smallrumres.2020.106123.

Vautor, E., et al., 2009. Difference in virulence between Staphylococcus aureus isolates causing gangrenous mastitis versus subclinical mastitis in a dairy sheep flock. Vet. Res. 56. Available at: 10.1051/vetres/2009039.

Vonaesch, P., Anderson, M., Sansonetti, P.J., 2018. 'Pathogens, microbiome and the host: emergence of the ecological Koch's postulates', FEMS Microbiol. Rev. [Preprint], (January 2018). Available at: 10.1093/femsre/fuy003.

Zafalon, L.F., et al., 2018. Relationship between virulence factor genes in coagulase-negative Staphylococcus spp. and failure of antimicrobial treatment of subclinical mastitis in sheep. Pesquisa Veterinária Brasileira 38 (4), 579–585. Available at: 10.1590/1678-5150-pvb-4984.

Zakour, N. Ben, Loir, Y., 2007. 'Designing primers for whole genome PCR scanning using the software package GenoFrag'. 348–368. Available at: 10.1007/978-1-59745-528-2_18.