UNIVERSIDADE FEDERAL DE MINAS GERAIS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

Dissertação

Análise genômica comparativa de linhagens de *Staphylococcus aureus* isoladas de diferentes formas de mastite em ovinos

DISCENTE: **Ana Carolina Barbosa Caetano**

ORIENTADOR: **Prof. Dr. Thiago Luiz de Paula Castro**

CO-ORIENTADOR: **Prof. Dr. Vasco Ariston de Carvalho Azevedo**

BELO HORIZONTE

2018

Ana Carolina Barbosa Caetano

Análise genômica comparativa de linhagens de *Staphylococcus aureus* isoladas de diferentes formas de mastite em ovinos

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo programa Interunidades de Pós-Graduação em Bioinformática, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais.

ORIENTADOR: **Prof. Dr. Thiago Luiz de Paula Castro**

CO-ORIENTADOR: **Prof. Dr. Vasco Ariston de Carvalho Azevedo**

BELO HORIZONTE

2018

Universidade Federal de Minas Gerais
Instituto de Ciências Biológicas
Programa Interunidades de Pós-Graduação em Bioinformática da UFMG

ATA DA DEFESA DE DISSERTAÇÃO

Ana Carolina Barbosa Caetano

51/2018
entrada
2º/2016
CPF:
114.650.756-90

Às quatorze horas do dia **07 de agosto de 2018**, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: **"Análise Genômica Comparativa de Linhagens de Staphylococcus aureus Isoladas de Diferentes Formas de Mastite em Ovinos"**, requisito para obtenção do grau de Mestre em **Bioinformática.** Abrindo a sessão, o Presidente da Comissão, **Dr. Thiago Luiz de Paula Castro**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuidas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Dr. Thiago Luiz de Paula Castro	UFBA	074603546 97	Aprovada
Dr. Vasco Ariston de Carvalho Azevedo	UFMG	283 [illegible] 1205 44	Aprovado
Dra. Núbia Seyffert	UFBA	000234720-86	Aprovada
Dr. Aristóteles Góes Neto	UFMG	541348815-20	APROVADO
Dr. José Miguel Ortega	UFMG	05950265-07	Aprovado

Pelas indicações, a candidata foi considerada: APROVADA
O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.
Belo Horizonte, 07 de agosto de 2018.

Dr. Thiago Luiz de Paula Castro - Orientador _____
Dr. Vasco Ariston de Carvalho Azevedo - Coorientador _____
Dra. Núbia Seyffert _____
Dr. Aristóteles Góes Neto _____
Dr. José Miguel Ortega _____

Agradecimento

À minha família e ao meu namorado por todo o apoio e compreensão. Sem vocês minha caminhada teria sido muito mais árdua!

Ao LGCM por todas as contribuições dadas à minha formação, e pelas amizades que aqui construí;

Ao profº. Thiago Castro por ser tão paciente na orientação, e ao profº Vasco Azevedo por ter confiado em meu trabalho.

Meus sinceros agradecimentos a todos que de alguma forma colaboraram para que esse sonho pudesse se tornar realidade.

Resumo

Staphylococcus aureus é o maior agente etiológico de mastite em pequenos ruminantes de todo o mundo. Mastite consiste em uma infecção na glândula mamária que é de difícil cura e possível reincidência, levando a grandes perdas econômicas na produção de leite e na criação de animais de rebanho. Pouco se sabe sobre características genômicas de linhagens de *S. aureus* isoladas de mastite em pequenos ruminantes, principalmente em ovinos. Estudos dos genomas desse microrganismo podem contribuir para melhor compreensão de traços envolvidos na especialização por hospedeiro e além disso, contribuir para novas estratégias de tratamento e controle da mastite ovina. Neste estudo, análises de genômica comparativa entre 12 linhagens isoladas de mastite de ovelhas da França e 11 genomas isolados de diferentes hospedeiros, disponíveis no banco de dados do NCBI, foram realizadas. Como resultado, 3,964 diferentes genes foram encontrados nos genomas e 2,969 genes são compartilhados entre eles, destes, 859 são genes acessórios. Os 2,110 genes que fazem parte do genoma central estão envolvidos no metabolismo celular, de acordo com as análises de distribuição de COG. Além disso, os grupos de genes acessórios e genes exclusivos encontrados nas linhagens de ovinos estão incluídos na categoria de sinalização celular. Profagos, ilhas de patogenicidade e ilhas genômicas foram preditas e carreiam importantes fatores de virulência bacteriano. Um agrupamento de linhagens ovinas pode ser observado na análise de Identidade Média de Nucleotídeos. Além disso, as análises de Tipagem de Sequências Multilocus revelaram dois grandes grupos clonais de *S. aureus* ovino, exceto por três linhagens que podem ser consideradas atípicas. Este resultado está de acordo com as análises filogenéticas. Em adição, algumas linhagens isoladas de ovelha e vaca podem estar evolutivamente relacionadas. Estas análises de genômica comparativa podem contribuir para a identificação dos genes adquiridos por transferência horizontal, bem como o papel destes na adaptação do hospedeiro e virulência bacteriana, e caracterização das linhagens que acometem ovinos.

Palavras- chave: Genômica comparativa, mastite, ovinos, genoma central, complexo clonal.

Abstract

Staphylococcus aureus is the major etiological agent of mastitis in small ruminants, worldwide. Mastitis is a mammary gland inflammation that is difficult to cure and prone to resurgence, leading to great economical losses in milk production and herd animals. To date, little is known about the genomic features specific to *S. aureus* strains isolated from mastitis in small ruminants, such as ovines. The study of the genomes of these microorganisms might contribute to a better understanding of the bacterial traits involved in the host specialization, supporting the development of new treatment and control strategies for ovine mastitis. In this study, comparative genomic analyses between 12 strains isolated from ovine mastitis in France and 11 genomes, isolated from various hosts, retrieved from the NCBI database, were performed. Results revealed 3,964 different genes found in the genomes and 2,969 genes shared between the groups, out of these, 859 are accessory genes. The 2,110 genes, which comprise the core genome, are involved in the cellular metabolism, according to the COG distribution analysis. Beyond the cellular metabolism category, the groups of accessory genes and genes exclusively found in the ovine strains include the cell signaling category. Prophages, pathogenicity and genomic island were predicted and carries important bacterial virulence factors. A cluster of ovine strains was formed in the Average Nucleotide Identity analysis. Besides that, the Multilocus Sequence Typing analysis revealed two big clonal- related group of *S. aureus* ovine, except for three strains that can be considered atypical. This result is in agreement with the phylogenetic analysis. In addition, some sheep and cattle strains can be evolutionary related. These comparative genomic analyses may contribute to the identification of traits acquired through lateral gene transfer, as well as the roles of these traits in host adaptation and virulence, and characterization of *S. aureus* ovine strains.

Key words: Comparative genomics, mastitis, ovines, core genome, clonal complex.

Lista de Figuras

Figura 1: Etapas do sequenciamento de DNA na plataforma Illumina. 17

Fig. 1: MLST phylogenetic analysis of all selected *S. aureus* strains, based on alleles of the genes *arcC, aroE, glpF, gmk, pta, tpi, yqiL*……..…..…31

Fig. 2.: Molecular phylogenomic analysis based on presence/absence of amino acids from *S. aureus* genomes……..……..…..32

Fig. 3: Phylogenomic tree and heatmap of the 23 *S. aureus* strains, generated using Genenees..…….....33

Fig. 4.: Fig. 4.: ANI similarities (%) of the 23 selected *S. aureus* genomes with the reference ovine strain ED133. ……...34

Fig. 5: Venn diagrams presenting the numbers of genes grouped according to the comparative genomic analysis between the ovine *S. aureus* strains and the strains isolated from other hosts ……...35

Fig. 6: Graphical representations of COG distribution profiles of the genes assessed in the BPGA comparative analysis of all 23 selected *S. aureus* strains. …….....…….. 36

Fig. 7: Categorization of the 281 genes exclusively found in the 13 ovine S. aureus strains according to the BPGA prediction and to the RAST annotation.…….37

Fig. 8: Circular genome map presenting all prophages found in the strain O268……...39

Fig. 9: Circular genome map representing all genomic and pathogenic islands found in the O82 strain …….. 41

Fig. S1: Synteny analysis of *S. aureus* strains isolated from ovines…….. 58

Lista de Tabelas

Table 1: *S. aureus* strains isolated from ovine mastitis used in the present study…….. 23

Table 2: *S. aureus* strains, isolated from different hosts, whose genomes were were retrieved from the NCBI database…….. 24

Table 3: General genomic features of twelve ovine *S. aureus* strains analyzed…… 28

Table 4: Sequence types and clonal complexes of the the ovine *S. aureus* strains used in this study…… 29

Table 5: Number of PAIs and GEIs predicted in every *S. aureus* genome considered in this study, using GIPSy……. 40

Table S1: Alleles prediction of the seven housekeeping genes considered in MLST analyses…….54

Table S2: Virulence proteins of exclusive genes from ovine *S. aureus* found in VFDB database……..…….55

Table S3: Shared proteins among *S. aureus* ovine genomes isolated of the same mastitis form……..…….....57

Table S4: Prophages found in ovine *S. aureus* genomes…….... 59

Table S5: Features of prophages predicted in the genome *S. aureus* O268... 61

Lista de siglas e termos em inglês

°C	Graus Celsius
Chip	Dispositivo microeletrônico
DNA	Ácido desoxirribonucleico
EDTA	Sal dehidratado dissódico
Forward	Fita senso do DNA
Flow cell	Superfície de clonagem
GC	Guanina e Citosina
h	Horas
H_2O	Água
Heatmap	Mapa de calor
In silico	Abordagem computacional
µl	Microlitro
µm	Micrômetro
Mb	Megabases
min.	Minutos
NaAc	Acetato de sódio
NCBI	Centro Nacional para Informação Biotecnológica
NaCl	Cloreto de Sódio
NGS	Sequenciamento de nova geração
Paired- end	Sequenciamento de ambas as extremidades da fita de DNA
pb	Pares de base
PCR	Do inglês, reação em cadeia da polimerase
Reads	Leituras provenientes do sequenciamento
Reverse	Fita consenso do DNA
rpm	Rotações por minuto
Single read	Sequenciamento de apenas uma extremidade da fita de DNA
Software	Programa
Tris- HCl	Tris-hidrocloreto

Sumário

1. Introdução geral 11
1.1. Mastite em ovinos e *S. aureus* 11
1.2. *Staphylococcus aureus*: aspectos microbiológicos 13
1.3. Genômica comparativa para a análise de bactérias patogênicas 15
1.4. Plataformas de sequenciamento de nova geração 16
2. Justificativa 18
3. Objetivos 19
3.1. Objetivo Geral 19
3.2. Objetivos Específicos 19
4. Manuscrito completo preparado para submissão à Frontiers in Microbiology 20
5. Perspectivas 63
6. Referências Bibliográficas 64

1. Introdução geral

1.1. Mastite em ovinos e *S. aureus*

A mastite é uma inflamação que acomete as glândulas mamárias de ruminantes de pequeno e grande porte, sendo de difícil cura e de provável reincidência. A severidade dessa infecção pode variar de subclínica a gangrenosa, e eventualmente se tornar crônica, ou resultar em uma infecção sistêmica (Maréchal *et al*., 2011; Peton *et al*., 2014). Esta doença causa impactos consideráveis na ovinocultura, levando a perdas econômicas que resultam desde a diminuição da produção leiteira até a morte dos animais (BRADLEY, 2002; EFSA Panel On Animal Health and Welfare (AHAW), 2014; Maréchal *et al*., 2011).

Quando assintomática, a mastite é considerada subclínica e caracterizada pelo aumento do número de células somáticas (leucócitos e células epiteliais) no leite, o que leva à diminuição da qualidade desse produto (Bradley, 2002; Oviedo-Boyso *et al.*, 2007; Brito & Brito, 1998). A mastite subclínica pode se tornar crônica, podendo, ainda, alternar entre as fases clínica e subclínica (Oviedo-Boyso *et al*., 2007). A mastite é considerada clínica quando há sinais visíveis no leite, como floculação e coágulos. Na mastite clínica, além do quarto mamário apresentar inchaço e secreções anormais, o animal pode apresentar febre, perda de apetite e desidratação (Crist *et al*., 1997). Em casos mais graves, como o da mastite gangrenosa, há necrose da glândula mamária, o que pode levar os animais a óbito (Vautor *et al*., 2009).

Staphylococcus aureus é o maior agente causador de mastite em ruminantes no mundo. Em ovelhas, este é o principal patógeno causador de mastite clínica (EFSA PANEL ON ANIMAL HEALTH AND WELFARE (AHAW), 2014; Peton *et al*., 2014). Esta bactéria é de difícil controle e invade rapidamente todos os tipos celulares da glândula mamária, podendo infectar outros ruminantes do rebanho e outros quartos mamários não-afetados do mesmo animal (Oviedo-Boyso *et al*., 2007). A prevenção contra novas infecções é realizada principalmente por meio da terapia da vaca seca, que consiste na infusão de antibióticos na glândula mamária do animal, na última

ordenha do período de lactação (Lents *et al*., 2008; Brito & Brito, 1998). Embora a antibioticoterapia seja uma das principais abordagens terapêuticas recomendadas contra a mastite, baixas taxas de cura têm sido relatadas (Barkema *et al.*, 2006; Da Silva *et al*., 2004; Roy & Keefe, 2012). A higiene do local de ordenha, dos objetos e do úbere do animal, durante e imediatamente após a ordenha, podem contribuir para a diminuição da taxa de infecção por *S. aureus* no rebanho (Da Silva *et al*., 2004).

Ainda não foram elucidadas as características genéticas bacterianas e/ou fatores do hospedeiro que podem contribuir para o estabelecimento das diferentes formas de mastite, portanto, novos estudos são necessários (Vautor *et al*., 2009). A mastite em bovinos é amplamente estudada, porém há poucos estudos sobre a mastite ovina. Há dados que estimam a prevalência da mastite subclínica em 5-30% dos pequenos ruminantes, e a incidência anual da mastite clínica é de 5%, podendo apresentar variações nesse índice (Vautor *et al*., 2003; Contreras *et al*., 2007).

1.2. *Staphylococcus aureus*: aspectos microbiológicos

S. aureus é uma bactéria Gram-positiva que apresenta 0.5-1.5µm de diâmetro. As células possuem forma de cocos, são imóveis, não formadoras de esporos e anaeróbias facultativas, sendo capazes de realizar tanto respiração aeróbia quanto fermentação em anaerobiose. Esse microrganismo é coagulase e catalase positivo, oxidase negativo, tolera até 10% de NaCl no meio de cultura, e sobrevive a uma faixa de temperatura de 7°C a 48,5°C (Harris; *et al.*, 2002; Murray, 1995). *S. aureus* pode colonizar e infectar diversos hospedeiros, causando pneumonia e infecções de pele em humanos, além de infecções como a mastite, em animais de rebanho. Essa bactéria também constitui um grande problema de saúde pública pela disseminação de clones resistentes a antibióticos, causando epidemias e surtos pelo mundo (Feil *et al*., 2003; Holden *et al*., 2013).

Durante a infecção em ovinos, a persistência de *S. aureus* no interior das células epiteliais da glândula mamária pode ser um dos fatores responsáveis pela dificuldade de eliminação do patógeno (Oviedo-Boyso *et al*., 2007). Além disso, a dose de antibióticos e a duração mínima do tratamento podem não ser respeitadas (Brito & Brito, 1998). Quando realizado de maneira inadequada, o tratamento da mastite pode tornar as bactérias resistentes aos antibióticos e propiciar a contaminação de hospedeiros humanos, através de alimentos de origem animal (Bradley, 2002).

Para estabelecer o processo infeccioso na mastite, a bactéria enfrenta diversas barreiras no organismo hospedeiro. A primeira delas se dá no canal do teto, onde o músculo do esfíncter mantém o orifício fechado, isolando o interior da glândula mamária. Além disso, o canal do teto é recoberto por queratina, dificultando a internalização da bactéria durante o processo infeccioso. Já foi relatado que a destruição dessa estrutura está relacionada com o aumento da incidência da mastite. Além das barreiras físicas que encontra, a bactéria precisa resistir ao recrutamento de neutrófilos, à ativação de linfócitos e produção de citocinas (Myllys *et al*., 1994; Oviedo-Boyso *et al*., 2007).

Por outro lado, *S. aureus* produz uma ampla gama de proteínas de superfície que promovem aderência às células do hospedeiro e destruição dos tecidos. Dentre elas, encontram-se as chamadas proteínas de ligação, que ajudam na evasão do sistema imune e participam da captação de ferro, as coagulases, e as toxinas que desempenham papel fundamental na intoxicação alimentar de humanos que consomem leite e derivados do animal contaminado (Gordon & Lowy, 2008; Le Loir *et al.*, 2003; Oviedo-Boyso *et al*., 2007). Genes envolvidos em fatores de virulência e resistência a antibióticos podem ser encontrados tanto no cromossomo quanto em elementos genéticos móveis, como os plasmídeos (Novick, 1990). Existem, ainda, relatos de mastite causada por linhagens hipervirulentas de *S. aureus* (Smith *et al*., 2014; Maréchal *et al*., 2011).

1.3. Genômica co mparativa para a análise de bactérias patogênicas

O primeiro genoma de bactéria patogênica inteiramente sequenciado foi da espécie *Haemophilus influenzae* (Fleischmann *et al*., 1995). Desde então, a genômica tem transformado o entendimento da interação bactéria-bactéria e bactéria-hospedeiro. Com a introdução das tecnologias de sequenciamento de nova geração (NGS, do inglês *Next-generation sequencing*), estudos de evolução e genômica comparativa em larga escala têm sido realizados de forma a contribuir para o conhecimento de novos determinantes de virulência bacterianos e resistência a drogas, bem como de funções atribuídas a determinados segmentos do genoma, como regiões gênicas regulatórias. Além disso, esses estudos têm contribuído para uma melhor compreensão da epidemiologia genômica em surtos bacterianos hospitalares, como já realizado com *S. aureus* (Deleo & Chambers, 2009; Delsuc, 2003; Harris *et al*., 2010; Loman & Pallen, 2015; Metzker, 2010).

O genoma de *S. aureus* consiste em um cromossomo circular de aproximadamente 2,8Mb, possuindo uma quantidade média de 2872 genes e conteúdo GC de cerca de 32,8% (Lowy, 1998; NCBI 2018). Os primeiros genomas dessa espécie foram sequenciados no ano de 2001, e pertencem às linhagens Mu50 e N315, isoladas de humanos (Kuroda *et al*., 2001). Atualmente, mais de oito mil genomas de *S. aureus* estão depositados na base de dados do NCBI (*National Center for Biotechnology Information*) e, destes, cerca de 189 estão depositados como completos. Apesar disso, dentre os genomas completos de linhagens com informações sobre o hospedeiro de origem, apenas o genoma da linhagem ED133 (Guinane *et al*., 2010) está associado a um caso de mastite ovina. Genomas completos são de extrema importância para a genômica comparativa, pois possuem toda a sequência do genoma e contribuem para a correta representação dos genes (Puranik *et al.*, 2014).

1.4. Plataformas de sequenciamento de nova geração

Algumas plataformas NGS realizam o sequenciamento com base na tecnologia da PCR, como os sequenciadores GS-FLX 454 da Roche (Roche Diagnostics Corp, Branford, CT, USA), Illumina Genome Analyzer (Illumina Inc., San Diego, CA, USA), ABI SOLiD System (Life Technologies Corp., Carlsbad, CA, USA) e Ion Personal Genome Machine (Life Technologies, South San Francisco, CA, USA). Para o sequenciamento na plataforma Illumina, após a fragmentação aleatória da sequência de DNA, adaptadores (sequências universais específicas de cada plataforma de sequenciamento) são ligados aos fragmentos de DNA e então amplificados pelo método de PCR, formando a biblioteca de DNA (Mardis, 2013; Illumina, 2016). Essa biblioteca é adicionada ao *chip* de sequenciamento (*flow cell*), o qual possui, em sua superfície, oligos que são complementares aos adaptadores dos fragmentos de DNA. Logo, cada fragmento se liga ao respectivo oligo e é amplificado pela chamada amplificação de ponte, formando grupos para o sequenciamento. O sequenciamento é realizado por síntese, no qual a DNA polimerase catalisa a incorporação de nucleotídeos fluorescentes à cadeia de DNA que está sendo sintetizada. Os nucleotídeos são identificados por excitação fluoróforа, na medida em que são incorporados à cadeia de DNA. O número de ciclos da corrida do sequenciamento determina o tamanho da *read* que será gerada. Além disso, essa plataforma utiliza a tecnologia *paired-end*, que consiste na leitura de ambas as direções dos fragmentos de DNA (*forward* e *reverse*), o que produz grande vantagem em comparação ao dado *single-read*, pois as sequências geradas têm maior acurácia e capacidade de detectar inserções e deleções (Berglund *et al.*, 2011; Illumina, 2016). As etapas do sequenciamento de DNA na plataforma Illumina estão sumarizadas na Figura 1**.**

Outras tecnologias, porém, são baseadas no sequenciamento de molécula única, e dispensam a etapa de amplificação, que normalmente antecede o sequenciamento. Estas plataformas incluem HeliScope (Helicos Bioscience Corp., Cambridge, MA, USA) e PacBio RS SMRT System (Pacific Biosciences Menlo Park, CA, USA) (Kaur & Malik, 2013). Soma-se, a esse grupo, o sequenciador MiniION da Oxford Nanopore Technologies, lançado em

2015, que tem como características o tamanho reduzido e a capacidade de portabilidade. Este sequenciador é acoplado ao computador como um *pen drive*, e é controlado por um *software* proprietário (Jain *et al*., 2015).

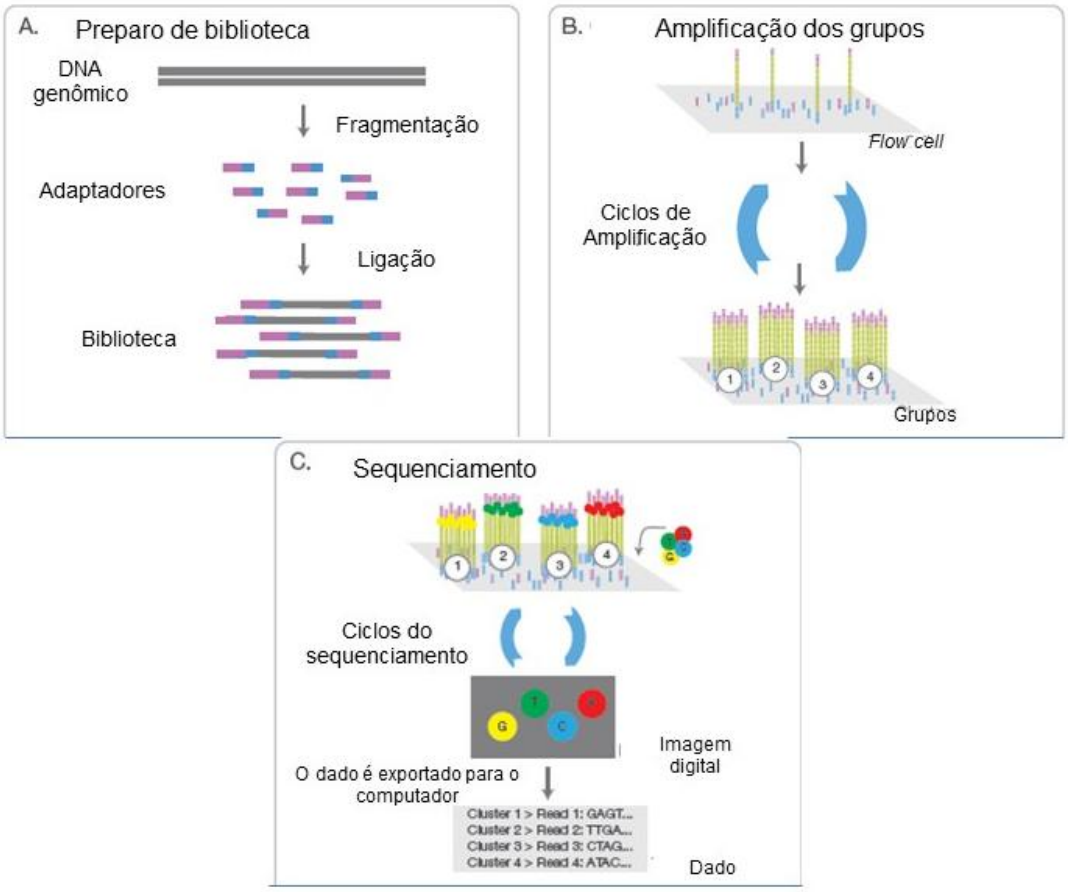

Figura 1: Etapas do sequenciamento de DNA na plataforma Illumina. **A)** Preparação de bibliotecas de DNA. Primeiro, a amostra de DNA total é fragmentada e, depois, as extremidades dos fragmentos são ligadas aos adaptadores. **B)** Os fragmentos com adaptadores ligam-se nos oligonucleotídeos existentes na *flow cell*, por complementariedade de bases. Cada fragmento é amplificado e dá origem a um grupo (*cluster*) de fragmentos clonais. **C)** Nucleotídeos com propriedades fluorescentes são adicionados e a emissão de fluorescência é captada para gerar a sequência de DNA.

Fonte: Adaptado de www.illumina.com/technology/next-generation-sequencing.html

2. Justificativa

S. aureus é o maior agente causador de mastite em animais ruminantes de todo o mundo. Esta infecção é frequente em animais de produção leiteira, leva a consideráveis perdas econômicas e é de difícil cura. O tratamento da mastite se dá principalmente com o uso de antibióticos, o que pode levar ao aumento da resistência bacteriana e ocasionar a reincidência da infecção (Barkema *et al.*, 2006; Da Silva *et al.*, 2004; Le Maréchal *et al*., 2011; Roy & Keefe, 2012). A melhor compreensão dos determinantes moleculares de virulência e patogenicidade de *S. aureus* poderá contribuir para o desenvolvimento de novas terapias não apenas contra a mastite, mas também contra outras doenças causadas por este agente em diferentes organismos hospedeiros.

Genomas de diversos isolados de *S. aureus* têm sido sequenciados e depositados no banco de dados do NCBI, sendo a maioria de origem humana. (dado disponível em: https://www.ncbi.nlm.nih.gov/). Dentre os isolados com genoma completo e depositado, somente a linhagem ED133 (Número de acesso CP001996) (Guinane *et al.,* 2010) tem origem ovina. Com isso, apesar dos esforços dedicados ao conhecimento do processo infeccioso da mastite e estratégias de prevenção e tratamento, pouco se sabe sobre as características exclusivas dos genomas de linhagens de *S. aureus* isoladas de pequenos ruminantes (Mavrogianni *et al*., 2006; Peton *et al*., 2014). Estudos prévios sugerem a existência de genes, em *S. aureus*, que podem estar envolvidos na associação específica com o hospedeiro ovino (Ben Zakour *et al*., 2008; Alves *et al.*, 2009). Diante desse cenário, o presente estudo se propõe avaliar as características genômicas das linhagens isoladas de ovinos a fim de analisar a especificidade pelo hospedeiro, a relação evolutiva entre os isolados e características que podem estar associadas à virulência, a fim de contribuir para novas estratégias de controle e tratamento dessa infecção. A contribuição desse trabalho também pode ir além da genômica, uma vez que as linhagens de ovinos, aqui utilizadas, estarão montadas, anotadas e disponíveis em banco de dados públicos a fim de contribuir para novos estudos envolvendo outras abordagens ômicas.

3. Objetivos

3.1. Objetivo Geral

Estudar as características genômicas de linhagens de *S. aureus* causadoras de mastite ovina, confrontando os genomas de treze isolados ovinos com um grupo de genomas de linhagens isoladas de outros hospedeiros animais.

3.2. Objetivos Específicos

- Obter as sequências genômicas completas e anotadas das linhagens O55, O17, O82, O217, O267, O268, O322, O326, O331, O408, O11 e O46 de *S. aureus*, isoladas de ovinos;

- Selecionar genomas completos de linhagens de *S. aureus* isoladas de ovinos e outros hospedeiros, depositados em bancos de dados;

- Proceder com a tipagem molecular *in silico* das linhagens de *S. aureus* utilizadas no estudo, por meio da identificação de *Sequence Types* (STs) e complexos clonais (CCs);

- Avaliar como as linhagens de *S. aureus* isoladas de ovinos se relacionam filogeneticamente entre si e com as demais linhagens selecionadas de outros hospedeiros;

- Definir um padrão de segregação filogenômica capaz de distinguir grupos de linhagens oriundos de ovinos de grupos de linhagens encontradas em outros hospedeiros;

- Avaliar os elementos genômicos, incluindo genes, ilhas genômicas e profagos, presentes apenas nas linhagens ovinas de *S. aureus*, em comparação com as linhagens de outros hospedeiros;

- Verificar a presença de genes que possam contribuir para o tropismo de hospedeiro nas linhagens ovinas de *S. aureus*, incluindo potenciais marcadores de virulência e patogenicidade.

4. Manuscrito completo preparado para submissão à Frontiers in Microbiology

O Manuscrito completo "**Comparative genomic analysis of ovine mastitis *Staphylococcus aureus***", referente às análises realizadas neste trabalho, foi preparado para submissão à revista "Frontiers in Microbiology", que é voltada para pesquisas que envolvem todo o espectro da microbiologia.

Genome host-specificity in *Staphylococcus aureus* isolated from ovine mastitis

Ana Carolina Barbosa Caetano[1], Raquel Hurtado Castillo[1], Roselane Gonçalves dos Santos[1], Doglas Parise[1], Marcus Vinícius Canário Viana[1], Rodrigo Profeta[1], Sandeep Tiwari[1], Anne Cybelle Pinto Gomide[1], Alfonso Gala-García[1], Yves Le-loir[4], Núbia Seyffert[1,3], Vasco Azevedo[1], Thiago Luiz de Paula Castro[1,2]*

[1]Cellular and Molecular Genetics Laboratory, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil
[2]Institute of Biology, Federal University of Bahia, Salvador, BA, Brazil
[3]Institute of Health Sciences, Federal University of Bahia, Salvador, BA, Brazil
[4]STLO, INRA, Agrocampus Ouest, F-35042 Rennes, France

Correspondence: thiago.luiz@ufba.br

Introduction

Mastitis is an infectious condition of the mammary gland that is responsible for economic losses in milk production and animal health. Mastitis may vary from subclinical to clinical, and can become chronic if infection persists for a long time. Occasionally, mastitis can also become systemic. This infection can be caused by different species of bacteria; however, *Staphylococcus aureus* is the major causer of mastitis in ruminants worldwide (Aires-de-Sousa et al., 2007; Oviedo-Boyso et al., 2007; Halasa et al., 2009; Maréchal et al., 2011). Similar to cattle, treatment against mastitis in small ruminants is done using antibiotics (Bergonier et al., 2003). However, the use of antibiotics is often inefficient to cure the infection caused by *S. aureus* and the disease is prone to resurgence, as bacteria may remain viable inside the epithelial cells of the mammary gland (Maréchal et al., 2011; Oviedo-Boyso et al., 2007).

S. aureus strains are commonly categorized into clonal complexes (CCs), according to the Multilocus Sequence Typing (MLST) of seven housekeeping genes (*arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi* and *yqiL*) (Enright et al., 2000). The MLST method has extensively been used for characterization of *S.* aureus isolates, which are commonly associated to host-specific CCs, such as CC133 for ovine isolates (Aires-de-Sousa et al., 2007; Merz et al., 2016; Smith et al., 2005, 2014; Smyth et al., 2009; Zakour and Loir, 2007). Also, previous studies in *S. aureus* have shown variation in genome content (mobile genetic elements and hypothetical proteins) and differences in regulatory genes of ovine strains, leading to host-association (Ben Zakour et al., 2008; Alves et al., 2009).

Comparative genomic analyses can be conducted to understanding the basis of intraspecific evolution and pathogenesis, since bacteria may have strain-specific genetic traits acquired through horizontal gene transfer (e.g. pathogenicity islands and bacteriophages sequences) (Ben Zakour et al., 2008; Dobrindt and Hacker, 2001). Although the number of *S. aureus* complete genomes deposited in the NCBI database has vastly increased in the last years, most of them are related to human isolates and only one (the strain ED133, Accession number CP001996) is from a strain isolated from an ovine (Guinane et al., 2010). Little is known about the genome peculiarities of ruminant isolates and there is no knowledge of the genomic features that might be

involved in the onset of mastitis in sheep (Smith et al., 2005; Binnewies et al., 2006; Ben Zakour et al., 2008; Alves et al., 2009; Peton et al., 2014; Bosi et al., 2016).

In order to contribute to the knowledge of genomic specialization and the genes' role in the establishment of infection in ovine mastitis-associated *S. aureus*, this study presented comparative genomic analyses to infer the genomic characteristics and phylogenetic relation of these genomes.

Material and Methods

Ovine *S. aureus* strains and genomic DNA obtention

Twelve French *S. aureus* strains, isolated from different forms of mastitis in flocks of sheep, were considered for genomic DNA sequencing in this study (Table 1). In order to extract the genomic DNA, these strains were first cultivated on LB (Luria Bertani) agar and, next, on 30mL of LB broth, at 37°C, for 12h. Every culture was centrifuged for bacterial pellet formation and the supernatant was discarded. The pellet was suspended in 600µL of solution (Tris-HCI pH 7.0, 0.5M EDTA pH 8.0, NaCl 5M, and distilled H_2O sufficient to obtain 50 mL) and transferred to a tube of 2mL containing glass beads (VK01) (Bertin Technologies). 3µL of RNase A solution (20mg/mL) was added prior to subjecting bacteria to mechanical lysis. Two homogenization cycles of 15sec each, at 6,500rpm, were performed using Precellys 24 (Bertin Technologies). Subsequently, 1mL of phenol:chloroform:isoamyl alcohol (25:24:1) solution was added to the tube and the mixture was homogenized and centrifuged at 13,000rpm, for 7min. Next, the upper aqueous phase of the mixture was transferred to a new tube, and a second round of phenol:chloroform:isoamyl alcohol purification was performed. Next, the upper aqueous phase was recovered and mixed with 1mL of chloroform. Following centrifugation at 13,000rpm, for 7min, the upper aqueous phase was transferred to a new tube. 1mL of ethyl alcohol, 40µL of 3M NaAc, and 4µL of 20mg/mL glycogen were added. Following gentle inversion, the mixture was placed at -20°C, overnight, for DNA precipitation. Following centrifugation at 13,000rpm, for 15min, the supernatant was discarded. 1mL of 70% ethyl alcohol was added to rehydrate the DNA pellet. A second round of 70% ethyl alcohol wash was performed using centrifugation at 13,000rpm, for 15min. The DNA pellet was placed at

60°C to dry. Finally, the DNA precipitate was suspended in 50 µL of sterilized ultra-pure water. DNA quantity and quality assessments were conducted using NanoDrop™ 2000 (Thermo Scientific™), Qubit Fluorometer (Thermo Scientific™) and 1% agarose gel electrophoresis.

Table 1: *S. aureus* strains isolated from ovine mastitis used in 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	Unespecified
O408	Gangrenous	2010	Alpes de Haute-Provence
O46*	Subclinical	2002	Unespecified
O82	Subclinical	2002	Unespecified
O267	Subclinical	1998	Pyrénées-Atlantiques
O331	Subclinical	2008	Alpes de Haute-Provence
O55	Subclinical	2002	Unespecified

* The strains O11 and O46 were not considered in this DNA extraction because their genomes were previously sequenced.

Sequencing, assembly and annotation of the ovine *S. aureus* genomes

Chromosomal sequencing of every ovine *S. aureus* strain was performed using Hiseq technology (Illumina, San Diego, CA, USA), except for O46 and O11, whose draft genome sequences were previously obtained using Illumina Genome Analyzer GAII (Fasteris, Geneva, Switzerland) and submitted to the NCBI database by Le Maréchal et al., (2011). SPAdes, version 3.9.1 (Bankevich et al., 2012), was used for the *ab initio* assembly of all ovine genomes, including O46 and O11. Contigs were submitted to CONTIGuator, version 2.7 (Galardini et al., 2011), for scaffolding. The gaps resulting from assemblies were manually filled using the CLC Genomics Workbench software, version 7.0 (Qiagen, USA), in which reads were mapped against the reference genome to generate the consensus sequences used for gap filling. To choose the reference genomes, the resulting contigs for each strain were submitted to the Basic Local Alignment Search Tool for nucleotides (BLASTn) (Altschul et al., 1990). The genome that presented the highest hit score and lowest e-value was chosen

as reference for each strain. All genomes were automatically annotated using the RAST server, version 2.0 (Aziz et al., 2008).

Methicillin resistance predicition and comparative genomics of ovine *S. aureus*

A search for the SCCmec IV, a staphylococcal cassette chromosome carrying the *mecA* or *mecC* gene, was performed by submitting all ovine *S. aureus* genomes to SCCmecfinder, version 1.2 (Kaya et al., 2018) (available at cge.cbs.dtu.dk/services/SCCmecFinder). For this analysis, the thresholds of 90% and 60% for minimum identity and minimum length, respectively, were used to compare the genome against the reference database (Kaya et al., 2018). Two genome groups were considered for comparative genomic analyses. The first group comprised 12 ovine strains, whose genomes were sequenced and assembled in this study, and the French ovine strain ED133, whose complete genome was retrieved from the NCBI (National Center for Biotechnology Information) database. An external, second group was formed with 10 *S. aureus* strains isolated from different hosts, including two clinical human isolates. These strains were selected based on the host-diversity and availability of complete genomes on the NCBI database (available at https://www.ncbi.nlm.nih.gov/).

Table 2: *S. aureus* strains, isolated from different hosts, whose genomes were retrieved from the NCBI database.

Strain	Host	Infection/Isolation source	Isolation locality	Accession Number
RF122	Cattle	Mastitis	Ireland	AJ938182.1
K17	Bovine	Milk	India	CP020713.1
K18	Buffalo	Milk	India	CP020714.1
LGA251	Bovine	Bulk milk	England	FR821779.1
ED133	Ovine	Mastitis	France	CP001996
ATCC25923	Human	Clinical isolated	USA	CP009361.1
NCTC13435	Human	Skin abscess	United Kingdom	LN831036.1
ILRI_Emoyle 1/1	Camel	Nasal swab	Kenya	LN626917.1
ISU935	Pig	Nasal swab	USA	CP017090.1
ISU926	Pig	Nasal swab	USA	CP017091.1
ED98	Broiler chicken	Bacterial chondronecrosis with osteomyelitis	Ireland	CP001781.1

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

To determine the Sequence Type (ST) of each selected *S. aureus* strain, an *in-silico* method was used. All genomes were submitted to the MLST (Multilocus sequence typing) server, version 1.8 (available at cge.cbs.dtu.dk//services/MLST) (Larsen et al., 2012). Grouping of STs was performed using eBURST, version 3 (Spratt et al., 2004), to evaluate the clonal complex (CC) formation.

Phylogenetic analyses were conducted, using Mega X, version 10.1 (Kumar et al., 2018), based on alleles of the seven housekeeping genes considered for the MLST analysis. All sequences were aligned using ClustalW (Thompson et al., 1994) and a phylogenetic tree was designed using the Maximum Likelihood method, with 1000 iterations bootstrap, based on the Hasegawa-Kishino-Yano model (Hasegawa et al., 1985). The strain ATCC12228 of *Staphylococcus epidermidis* was used as an outgroup in the generation of the tree.

Phylogenomic analyses of the selected *S. aureus* strains

A phylogenomic tree was generated using the Phylogenomic Tree Tool in PATRIC, version 3.5.17 (Wattam et al., 2017). For this, the variation of amino acid residues in the protein sequences predicted by RAST server was considered. The Maximum Likelihood method was used, with the Automated Progressive Refinement option selected. The strains DAR1907 and ATCC12228 of *S. epidermidis* were used as an outgroup.

A complimentary phylogenomic analysis was conducted using Gegenees, version 2.2.1 (Ågren et al., 2012). For this, fragments of the DNA sequence were obtained, and the fragments content shared by all selected strains was subtracted to generate comparisons among the variable genomic regions. The variable content similarities between strains were calculated and the resulting percentages were used to generate an all-versus-all heatmap (Fig. 6) (Ågren et al., 2012). A distance matrix file in the nexus format was exported to SplitsTree, version 4.14.2, to generate a phylogenomic tree using the UPGMA (Unweighted Pair Group Method using Arithmetic Averages) method (Kloepper and Huson, 2008).

Genome Average Nucleotide Identity (ANI) analysis in the selected *S. aureus* strains

The ANI analysis was used to determine the genome relatedness among the ovine strains and the strains isolated from various hosts. For this, the JSpecies Web Server (available at jspecies.ribohost.com/jspeciesws/#analyse) (Richter et al., 2016) was used with BLASTn (Altschul et al., 1990), selecting the ANIb option. This method of analysis performs pairwise comparisons of 1020bp fragments between a query and a reference genome. Results are obtained using the mean identities of all BLASTn matches with identities equal to or greater than 30% and coverage equal to or greater than 70% (Goris et al., 2007; Richter and Rosselló-Móra, 2009). The strain ED133 genome, retrieved from the NCBI database, was used as the reference in this analysis.

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

To identify conserved genomic regions, including DNA rearrangements and inversions, a synteny analysis was conducted using Mauve (Darling et al., 2004), version 2.4.0, with the most accurate option for alignment selected, Progressive Mauve. The genome of every strain was submitted to PHASTER (Arndt et al., 2016) for the identification and annotation of prophages. Genomic island (GEI) and pathogenicity island (PAI) predictions were performed using GIPSy (Genomic Island prediction Software), version 1.1.2 (Soares et al., 2016), with the genome of *Staphylococcus warneri* strain 231, retrieved from the NCBI database, used as the reference. ACT (Artemis Comparison Tool), version 13.0.0 (Carver et al., 2008), and Double ACT, version 2 (available at http://www.hpa-bioinfotools.org.uk/pise/double_act.html), were used for the visualization and curation of PAIs and GEIs predicted in the aligned genomes. BLAST Ring Image Generator (BRIG), version 0.95 (Alikhan et al., 2011), was used to generate a circular image of all genomes with the predicted prophages, GEIs and PAIs.

Prediction of ovine mastitis-associated *S. aureus* genes and proteins

The software BPGA (Bacterial Pan Genome Analysis Tool), version 1.3 (Chaudhari et al., 2016), was used to determine the core, accessory and singleton genes for the ovine and other 11 *S. aureus* strains isolated from various hosts (Table 2). The complete genomes of these strains were retrieved from the NCBI database and reannotated using RAST (Aziz et al., 2008), in order to homogenize gene annotation throughout all selected strains analyzed. Clustering of the orthologous genes with identities equal to or greater than 80% allowed the identification of accessory or exclusive genes present in the selected strains.

The proteins predicted to exist exclusively in the proteomes of ovine *S. aureus*, according to the BPGA analysis, were assessed using the Basic Local Alignment Search Tool for proteins (BLASTp) tool (Altschul et al., 1990) against the VFDB (Virulence Factors of Pathogenic Bacteria) database (Chen et al., 2005). This analysis was conducted to predict the virulence factors that might be specifically implicated in ovine mastitis. For this purpose, similarities were determined when the e-value was lower than or equal to $1e^{-5}$, and the identities and coverages of queries, in relation to their subjects, were equal to or greater than 30% (Lindahl and Elofsson, 2000; Yang and Honig, 2000) and 70%, respectively. In order to classify into categories the proteins shared exclusively by the strains associated with a single form of mastitis (Table 1), the Protein Family Sorter Tool, version 3.5.17 [available in PATRIC (Pathosystems Resource Integration Center)] (Wattam et al., 2017), was used. Protein functions were assigned using the RAST (Rapid Annotation using Subsystem Technology) server (Aziz et al., 2008). Categorization of proteins into families was conducted using the Markov Cluster (MCL) algorithm (Davis et al., 2016).

Results

Genomic statistics and MLST of the ovine *S. aureus* strains

The chromosomal genomes of the twelve ovine *S. aureus* strains sequenced in this study ranged from 2.7 to 2.8 Mb in size, with approximately 33% of GC content. The RAST annotation server predicted between 2,519 and 2,754 gene coding

sequences. As expected, all of the ovine strains were predicted to be Methicillin Sensitive *S. aureus* (MSSA). Although they do not carry the SCCmec chromosomal cassette, other genes that can contribute to the methicillin resistance phenotype, such as the *fmtB* gene, are found in all ovine strains.

In silico MLST analysis revealed that the majority of the ovine strains considered in this study is grouped into two major clonal complexes, CC130 and CC133. The exceptions were the strain O55, which is grouped into CC425, together with the bovine strain LGA251, and the strain O217, which comprises CC30 together with the strains ATCC25923 and ILRI_Emoyle1/1, isolated from human and camel, respectively. The strain O331 did not share sufficient alleles to belong to a pre-defined clonal complex, therefore it is considered a singleton. All alleles predicted in these analyses are shown in Table S1.

Table 3: General genomic features of twelve ovine *S. aureus* strains analyzed.

Strain	Genomes' length (pb)	GC content (%)	CDS Predicted
O268	2842531	32.84	2754
O17	2772344	32.83	2645
O322	2768115	32.79	2638
O326	2778005	32.69	2658
O217	2764231	32.75	2623
O11	2784926	32.70	2668
O408	2778443	32.68	2672
O46	2815614	32.76	2701
O82	2761422	32.83	2643
O267	2864635	32.85	2803
O331	2695455	32.75	2519
O55	2794102	32.79	2673

Table 4: Sequence types and clonal complexes of the ovine *S. aureus* strains used in this study.

Strain	ST	Definition
O82	2011	CC130
O46	2490	
O326	700	
O11	700	
O408	700	
O55	425	CC425
LGA251	425	
O268	133	CC133
O17	133	
O322	133	
O267	133	
ED133	133	
O331	59	Singleton
ATCC25923	243	CC 30
O217	30	
ILRI_Emoyle1/1	30	

MLST-based phylogenetic analysis of the ovine *S. aureus* strains

A phylogenetic analysis based on the MLST data was conducted (Fig. 1) and showed that all strains belonging to the CC133, which were categorized into the same ST (133), comprise a single clade, suggesting a strong genetic relationship between them. This clade likely shares a close ancestral with the clade formed by the strains belonging to CC425, with a bootstrap value of 98. Another single clade, formed by the strains comprising the CC30, is located apart from the clades formed by all other ovine *S. aureus* strains evaluated in this study. However, a bootstrap value of 30 between these two groups suggests possibility of a stronger MLST relatedness among the CC30 strains and some of the other ovine strains considered in this analysis.

On the other hand, the strains belonging to the CC130 were divided into two distinct phylogenetic clades, one of them comprising the strains O46 and O82 (STs 2490 and 2011, respectively), and the other the strains O326, O11, and O408 (ST 700). The low bootstrap values separating one clade from the other and these two clades from the remainder ovine *S. aureus* strains suggest greater genetic diversity in the group of CC130 strains. This suggestion is corroborated by the fact that the CC130 group, differently form the remainder clonal complexes identified in this study, is formed by a

greater variety of sequence types (STs 2011, 2490 and 700) (Table 4). Noteworthy, O46 and O82 share a combination of 6 out of 7 possible alleles (*aroE*_57, *glpF*_45, *gmk*_2, *pta*_7, *tpi*_14, and *yqiL*_52). Also, O46 and O82 share 6 alleles (with the exception of *tpi*_14) and 5 alleles (with the exception of *arcC*_200 and *tpi*_14), respectively, with the strains O326, O11, and O408 (Table S1).

Interestingly, the bovine strain RF122 likely shares greater MLST relatedness with the strains of CC130, CC425 and CC133, in comparison with the strains of CC30. The MLST-based phylogenetic analysis also revealed that the strain RF122, isolated from bovine, likely shares one common ancestral with the ovine strain O331, which was predicted to form a clonal complex not yet characterized. It is also evident the existence of a close phylogenetic relationship between the clade formed by RF122 and O331 and the clade of O326, O11, and O408.

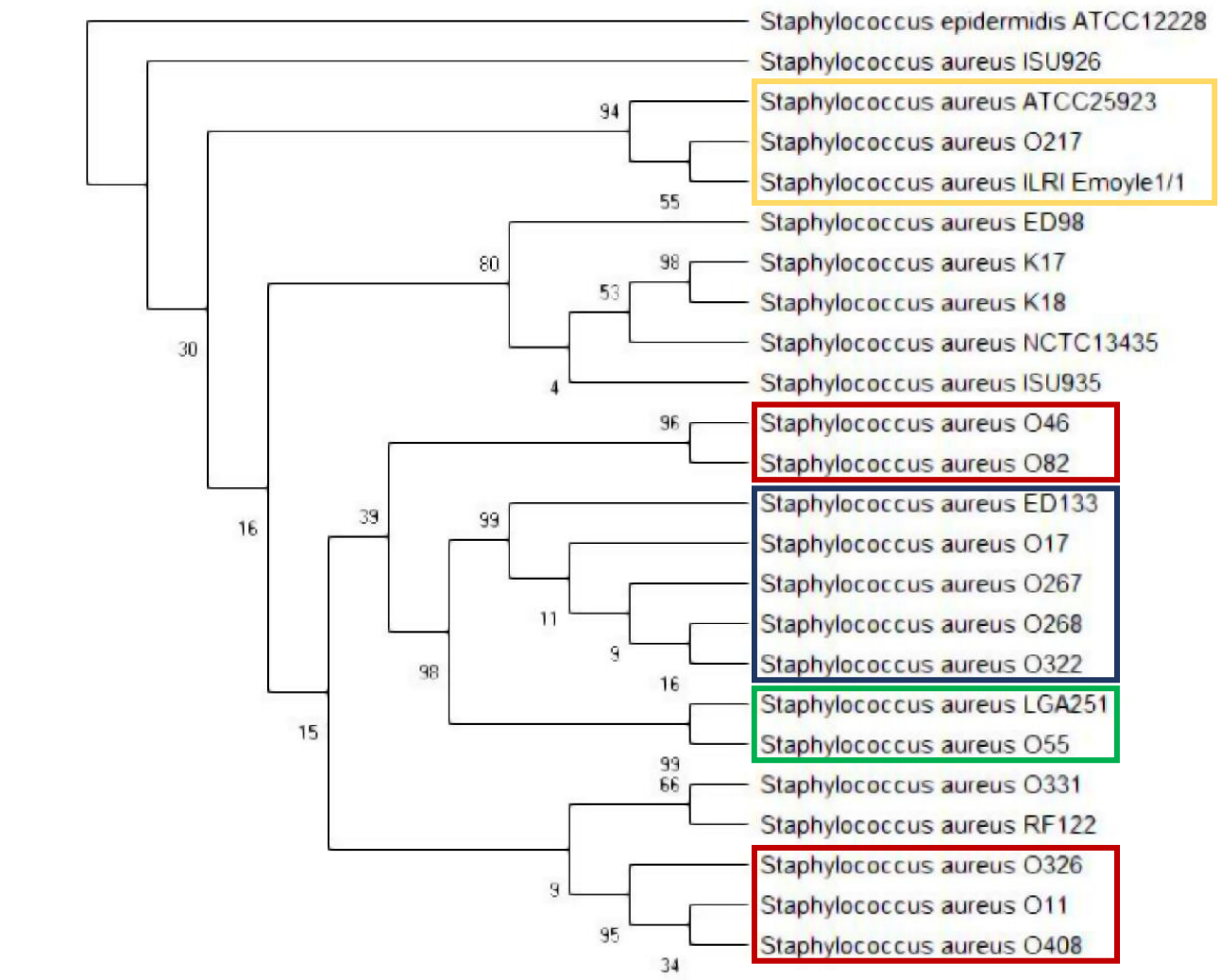

Fig. 1: MLST phylogenetic analysis of all selected *S. aureus* strains, based on alleles of the genes *arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, *yqiL*. Each CC is highlighted as a coloured box: CC133 (blue), CC130 (red), CC425 (green), CC30 (yellow). *Staphylococcus epidermidis* ATCC12228 was used as an outgroup.

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

Phylogenomic analyses were conducted to assess genome-scale phylogenies among the ovine *S. aureus* strains and the strains isolated from the other hosts. First, a phylogenomic tree was generated, using the Phylogenomic Tree Tool and considering the presence/absence of amino acids in the shared proteomes (Fig. 2). In contrast to the previous MLST-based phylogeny findings, the strains belonging to CC130 (O82, O46, O326, O11, and O408) formed a single clade. This result clearly indicates that the

determination of phylogeny benefitted from a proteome scale analysis, when compared to the limited number of gene *loci* considered in the MLST-based phylogenetic analysis. Similarly to the MLST-based results, the strains belonging to CC130 and CC425 form two distinct clades, whilst the O55 strain comprises a third clade together with the bovine strain LGA251. Also, O217 forms a single clade together with the strains ATCC25923 (from human) and IRLI_Emoyle1/1 (from camel), and the strain O331 forms another clade with the bovine strain RF122.

Another phylogenomic approach was considered, using Gegenees, for the comparison among all selected *S. aureus* strains. The variable DNA sequences present in the *S. aureus* genomes were pairwise-compared to generate a phylogenomic tree and plot a heatmap, demonstrating the phylogenomic relatedness among all strains (Fig. 3). The strains belonging to CC130 form a single clade, as well as the CC133 strains form another clade. Once again, the strain O55 is grouped with the strain LGA251 and the strain O217 is grouped with the strains ATCC25923 and IRLI_Emoyle1/1.

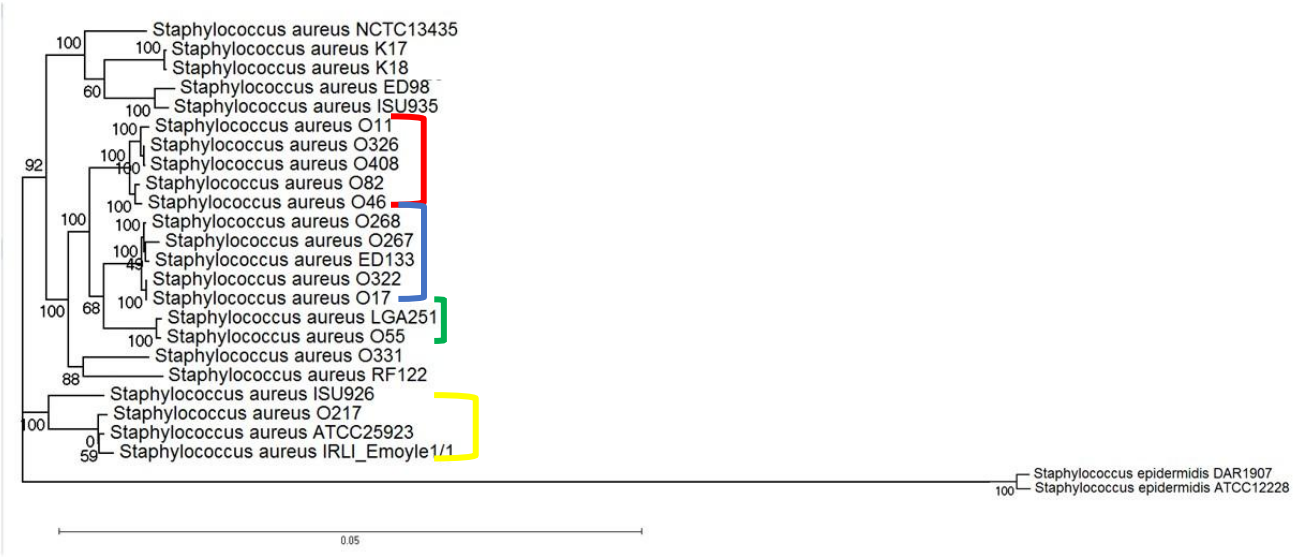

Fig. 2.: Molecular phylogenomic analysis based on presence/absence of amino acids from *S. aureus* genomes. Each CC is highlighted with a coloured bracket: CC133 (blue), CC130 (red), CC425 (green), CC30 (yellow). *Staphylococcus epidermidis* ATCC12228 and DAR1907 were used as outgroup.

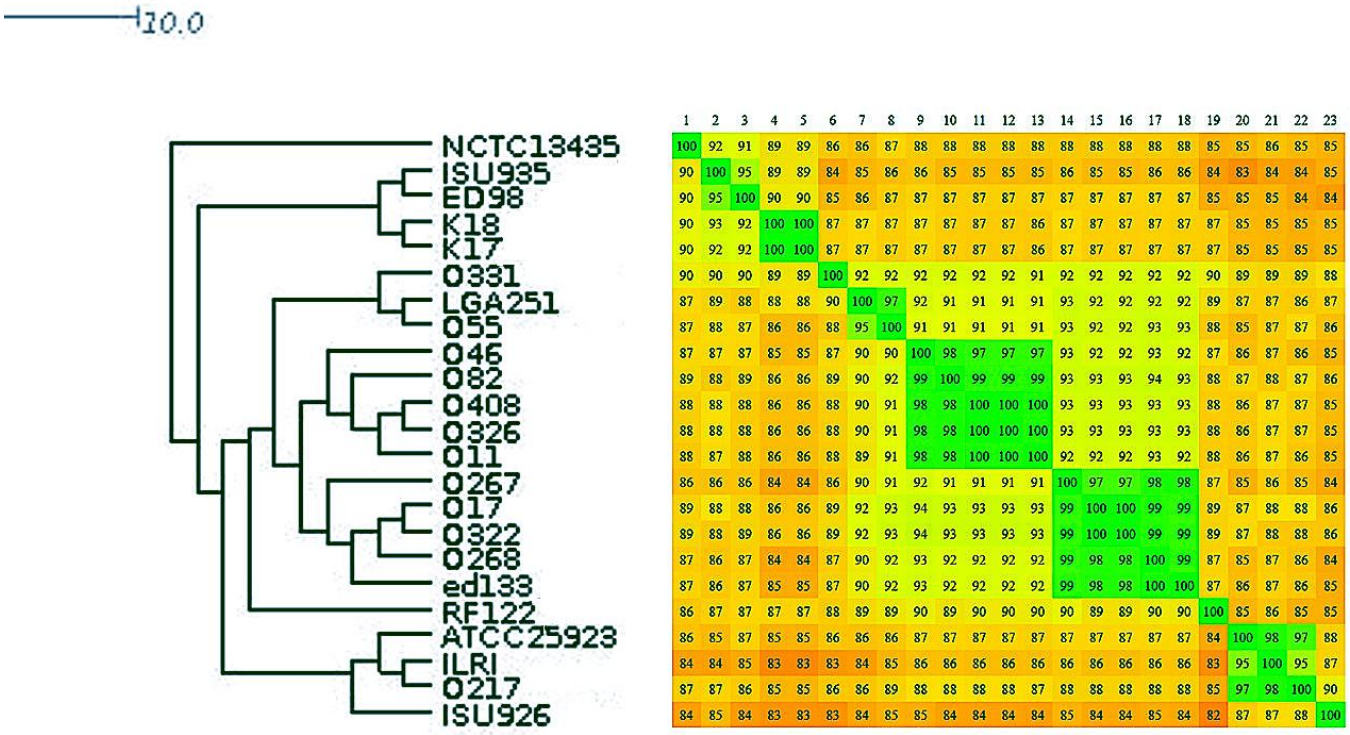

Fig. 3: Phylogenomic tree and heatmap of the 23 *S. aureus* strains, generated using Genenees. The similarity percentages are presented in an all-versus-all fashion, ranging from red (lower similarities) to green (higher similarities). The numerals above the heatmap correspond, in the same order, to the strains listed vertically in the tree.

ANI analysis evidences ovine *S. aureus* genomic relatedness

Average Nucleotide Identity predictions were conducted, using all selected *S. aureus* strains, to achieve a better understanding of how the ovine strains genetically correlate with each other and with the strains from the other hosts. All genomes presented a high degree of similarity with the genome of the reference strain, ED133 (ANI > 97%). The majority of the ovine *S. aureus* genomes were clustered in the highest similarities group (Fig. 4), with the exception of strain O217 (ANI < 98%). The genomes from O17, O267, O268, and O322 presented the greatest similarities (nearly 100%) with ED133. Interestingly, LGA251 and RF122 (ANI > 98%) were the non-ovine strains with the greatest similarity with the ED133 reference, corroborating the phylogenetic and phylogenomic clustering previously observed in this work (Fig. 2).

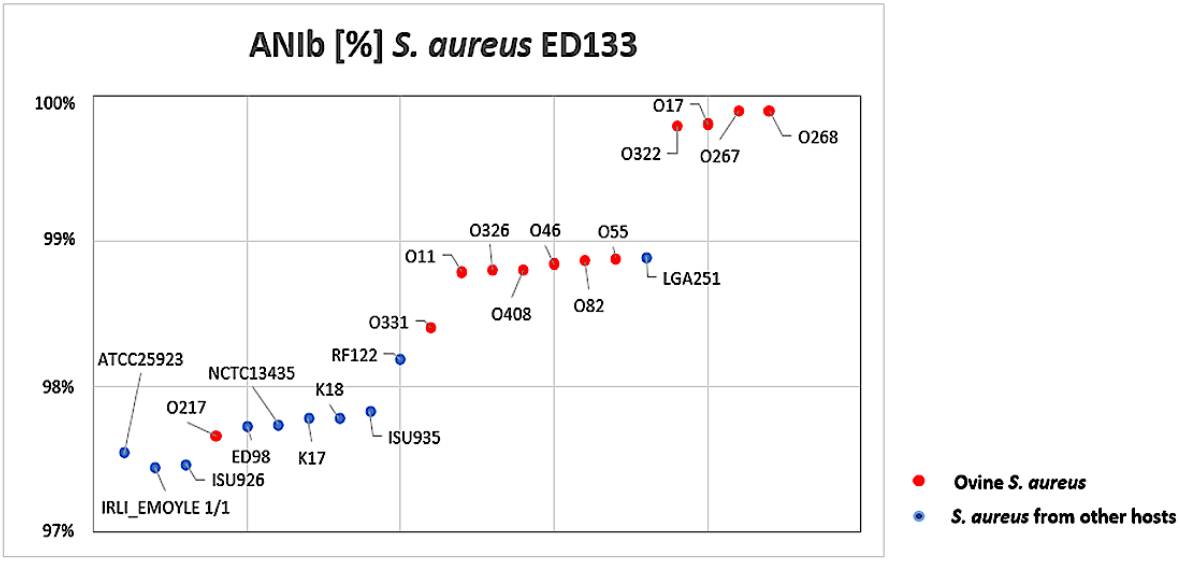

Fig. 4: ANI similarities (%) of the 23 selected *S. aureus* genomes with the reference ovine strain ED133.

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

The 23 *S. aureus* genomes were separated into two groups for the comparative genomics analysis. The first group consists of genomes isolated from ovines (Table 1) and the second group is composed of genomes isolated from other hosts (Table 2). The comparative genomics yielded 3,964 different genes found in both groups, representing ~1,47 fold of the average of total genes in the genomes studied. Using a threshold identity of 80% among the coding sequences, 281 genes were found to belong exclusively to the ovine strains, while the group of other hosts has 714 exclusive genes. 2,969 genes are shared between the groups; however, 859 of these genes are accessory genes (i.e., genes that are present in at least two genomes but not in all of the genomes) (Fig. 5- A). Comparing the core genome between the groups, 2,110 genes are present in all strains of both groups. 2,149 genes comprise the core genome of ovines, while the core genome of the other hosts' group present 2,119 genes (Fig. 5- B).

The genes belonging to the core, accessory and exclusive genome of each group were categorized into cellular processes and functions, in accordance with the terms of COG (Fig. 6). Most of the core genes are responsible for cellular metabolism processes, mainly transport and metabolism of amino acids. For the accessory genes and genes that

are unique to a specific strain, the most represented COG functions are replication (information storage) and recombination and repair (processing activities). Additionally, the functions of the proteins exclusively found in the group of ovine *S. aureus* were assessed (Fig. 7)**.** A very high percentage (41%) of the 281 exclusive ovine *S. aureus* genes encode proteins with unknown functions, evidencing the need of further studies on them. Also, 27% of the 281 genes encode proteins involved in different cellular functions, such as iron acquisition proteins, transferases, DNA helicases, and coagulases. Interestingly, 25% of the exclusive ovine genes are found in prophages and mobile genetic elements (25%), suggesting that many of the genetic traits related to the ovine host specificity might be horizontally disseminated.

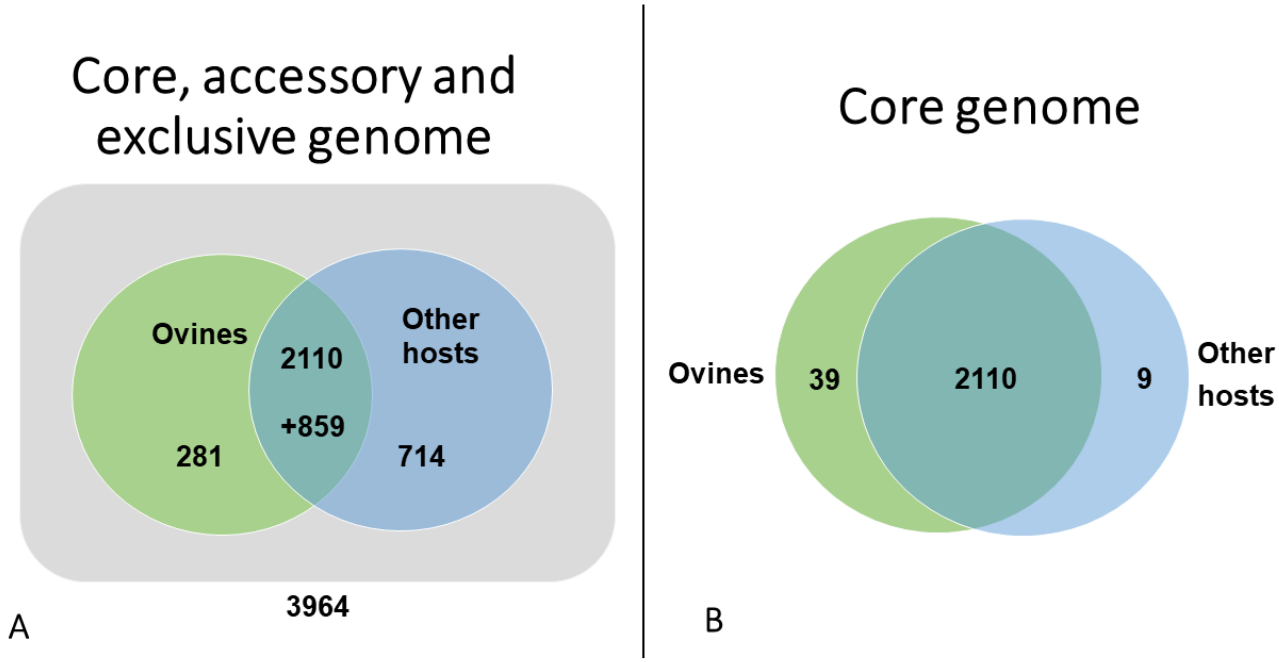

Fig. 5: Venn diagrams presenting the numbers of genes grouped according to the comparative genomic analysis between the ovine *S. aureus* strains (green) and the strains isolated from other hosts (blue). **A)** Number of genes found in both groups, exclusive genes in each group, and the total of non-redundant genes found in the comparative analysis (represented in the grey circle). **B)** Numbers of genes comprising the core genomes of each group and both of the groups combined.

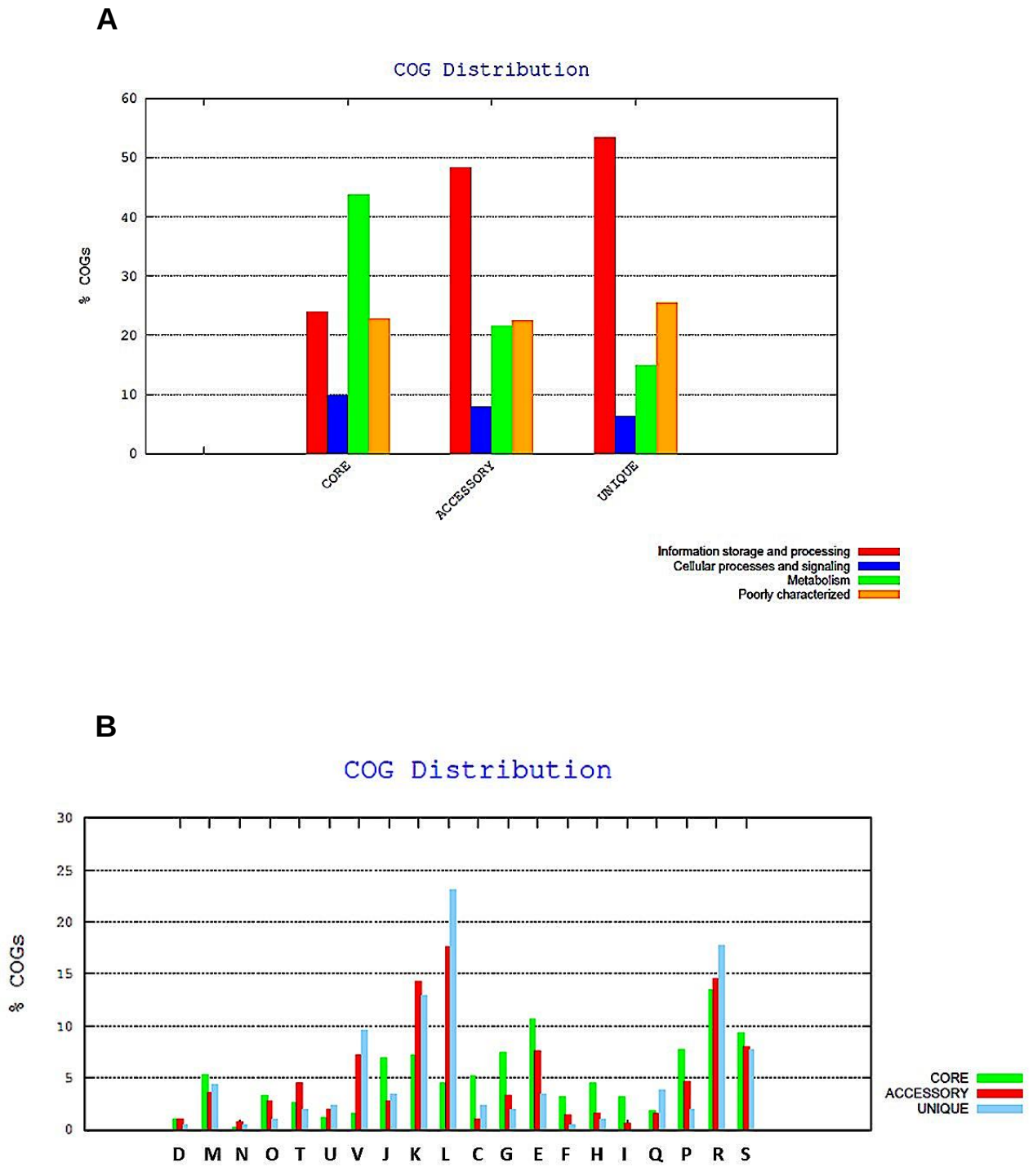

Fig. 6: Graphical representations of COG distribution profiles of the genes assessed in the BPGA comparative analysis of all 23 selected *S. aureus* strains. **A)** Information storage and processing category is in red bars, cellular processes and signaling in blue bars, metabolism in green bars, and poorly characterized genes in orange bars. **B)** Detailed COG functional categorization of core (green bars), accessory (red bars), and unique (blue bars) genes from 23 *S. aureus* genomes. D: Cell cycle control, cell division, chromosome partitioning; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; O: Posttranslational modification, protein turnover, chaperones; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms; J: Translation, ribosomal structure and biogenesis; K:

Transcription; L: Replication, recombination and repair; C: Energy production and conversion; G: Carbohydrate transport and metabolism; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; P: Inorganic ion transport and metabolism; R: General function prediction only; S: Function unknown.

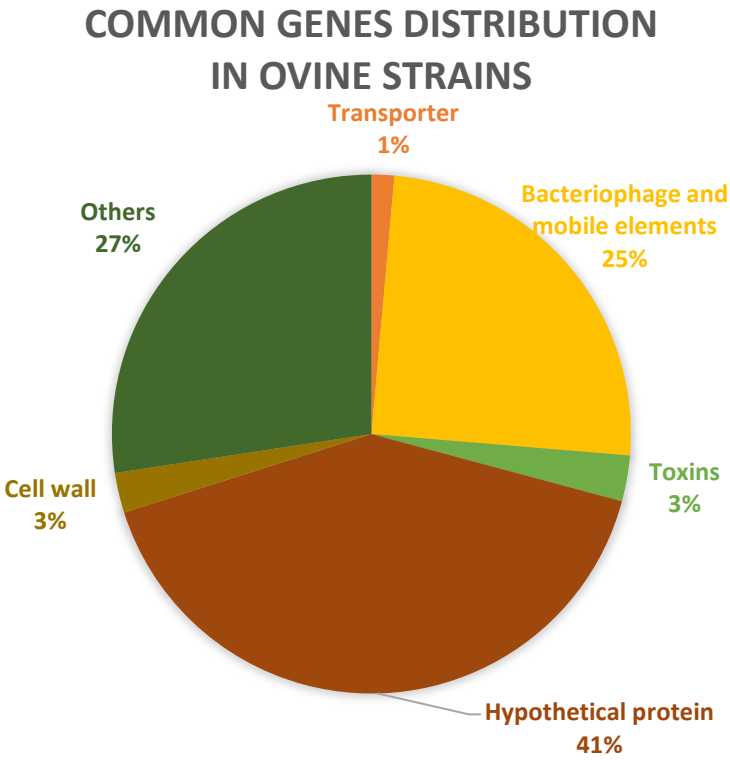

Fig. 7: Categorization of the 281 genes exclusively found in the 13 ovine *S. aureus* strains according to the BPGA prediction and to the RAST annotation.

Virulence genes in the ovine *S. aureus* strains and mastitis-specific proteins

Out of the 281 genes exclusive to the ovine *S. aureus* strains, 14 codify virulence factor proteins, including the prophage-carried *seb* (staphylococcal enterotoxin B), *selk* (staphylococcal enterotoxin K) and *vWpb* (secreted von Willebrand factor-binding protein precursor). Virulence genes encoding transporters, membrane proteins, iron acquisition proteins, and exotoxins are also present in ovine *S. aureus*. All virulence proteins predicted using the VFDB database are listed in Table S2.

The analysis of proteins that are exclusively shared among genomes of *S. aureus* involved in a single form of mastitis resulted in the prediction of hypothetical and

phage-related proteins. However, although these genes might be exclusively related to a specific form of mastitis, they are not shared by all of the strains grouped in the same mastitis condition (Table S3). 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. Also, specific traits of the ovines affected by *S. aureus*, as well as the interference of the environment, should be taken into consideration in order to explain the development of a determined form of mastitis.

Multi-genome synteny and plasticity analyses in the ovine *S. aureus* strains

A very high synteny level is observed for the genomic regions present in the ovine *S. aureus* strains, with only a few regions of DNA inversion. Some differences caused by gene acquisition from phages and DNA recombination events likely explain the variation in size among the genomes of the ovine strains. One example involves the prophage region 5, which is absent only in the O217, O408, and O331 genomes (Fig. S1). Using the strain O268 as the reference, a circular genome representation was generated for the visualization of most of the prophages found in the ovine genomes (Fig. 8). The proteins predicted to be encoded by genes present in these prophages are listed in Table S4. The *S. aureus* genomes from the other hosts were included in this analysis in order to evaluate the sequence conservation across all of the genomes.

All of the prophages predicted in the ovine genomes are listed in Table S5. The prophage region PHAGE_Staphy_StB27_NC_019914 is found in the strains O267, O268, O331, and ED133, while the prophage region PHAGE_Acinet_vB_AbaM_phiAbaA1_NC_031280 is found in the strains O322, O17, O55, O217, and O267. These two prophages are not present in any of the genomes belonging to the non-ovine hosts considered in this study. The first one carries genes for hypothetical proteins and transposases. In contrast, the second one carries genes for the chaperone GroEL, co-chaperone GroES, the von Willebrand factor-binding protein VWbp, in addition to membrane proteins and transposases.

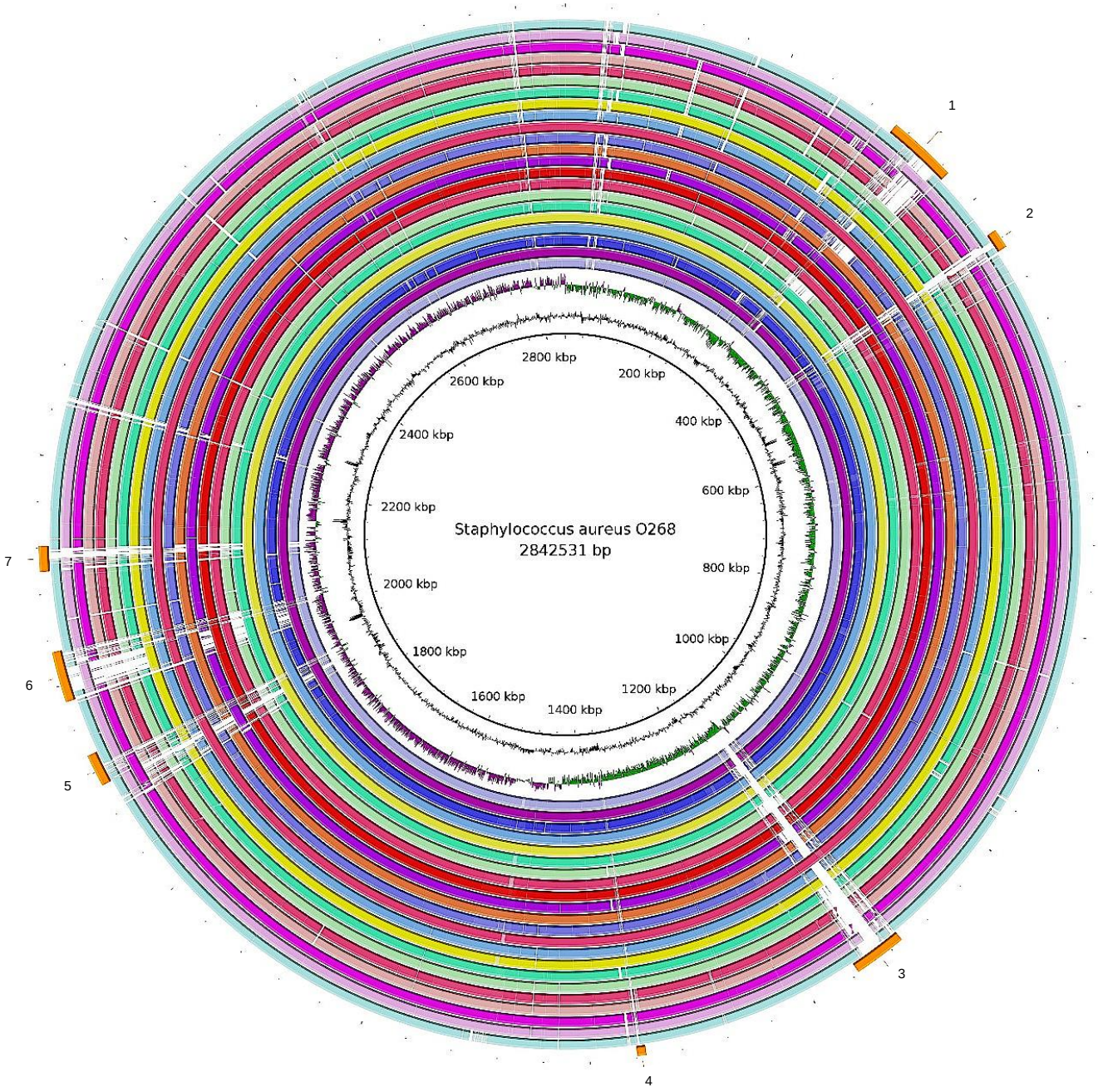

Fig. 8: Circular genome map presenting all prophages found in the strain O268. Representation, from the inner to the outer circle, is: O268, GC Content, GC Skew, O11, O17, O217, O267, O322, O326, O331, O408, O46, O55, O82, ATCC 25923, ED133, ED98, ILRI_Emoyle1/1, ISU926, ISU935, K17, K18, LGA251, NCTC13435, RF122, and prophages (orange color). The numbers represent the prophages: PHAGE_Staphy_StauST398_2_NC_021323 (1); PHAGE_Staphy_PT1028_NC_007045 (2); PHAGE_Staphy_77_NC_005356 (3); PHAGE_Staphy_StB27_NC_019914 (4); PHAGE_Staphy_StauST398_2_NC_021323 (5); PHAGE_Staphy_phiPV83_NC_002486 (6); PHAGE_Acinet_vB_AbaM_phiAbaA1_NC_031280 (7).

The number of PAIs predicted in the ovine *S. aureus* genomes ranges from 8 to 13. In turn, the number of GEIs predicted ranges from 8 to 15 (Table 5). The PAIs and GEIs found in O82 are represented, as a circular image, in Fig. 9. The largest pathogenicity island, PAI1 is found in all of the analyzed genomes; however, part of its sequence is common only to the strains O11, O326, O408, and O46. GEI12 is the genomic island presenting the largest sequence heterogeneity among all of the genomes.

In strain O331, the enterotoxin genes *sei* (Staphylococcal enterotoxin type I) and *selK* (enterotoxin K-like protein) are localized in the PAI5, while the gene encoding the Enterotoxin 15 is found in the GEI2. Also, the iron acquisition genes *mbtB_2* (Iron acquisition yersiniabactin synthesis enzime) and *IgrE* (Iron acquisition yersiniabactin synthesis enzime) are localized in the PAI1 of strains O11 and O326, respectively.

Table 5: Number of PAIs and GEIs predicted in every *S. aureus* genome considered in this study, using GIPSy.

Strain	PAI	GEI
O11	10	12
O17	10	12
O46	8	10
O55	11	8
O82	10	15
O217	9	11
O267	8	10
O268	9	11
O322	12	12
O326	11	10
O331	10	10
O408	9	11
ED133	12	6
K17	11	6
LGA251	9	9
RF122	12	6
K18	12	6
IRLI_Emoyle1/1	13	8
ED98	8	10
ATCC25923	11	-
NCTC13435	11	-
ISU926	12	9
ISU935	8	8

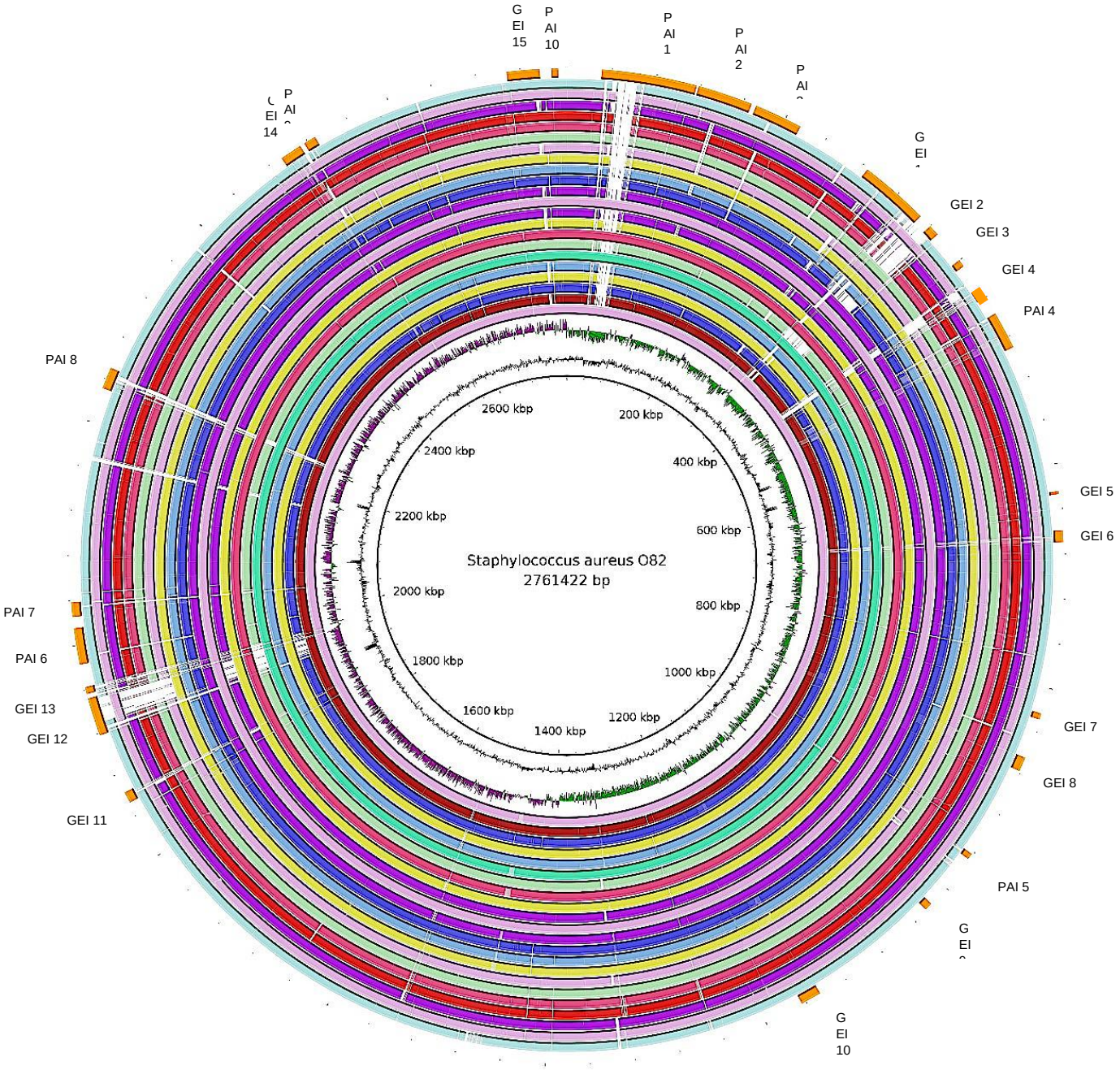

Fig. 9: Circular genome map representing all genomic and pathogenic islands found in the O82 strain. Representation, from the inner to the outer circle, is: O82, GC Content, GC Skew, O11, O17, O217, O322, O267, O326, O331, O408, O268, O55, O46, O55, O46, ED133, ATCC25923, ED98, ILRI_Emoyle1/1, ISU926, ISU935, K17, K18, LGA251, NCTC13435, and RF122.

Discussion

S. aureus is an important pathogen of ovine mastitis (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 the treatment with antibiotics (Oviedo-Boyso et al. 2007). The number of ovine S. *aureus* genomes available in the NCBI database is minimal and genomic studies involving these strains, are scarce. Therefore, more information about the *S. aureus* genomic characteristics are necessary for the development of new strategies to efficiently control infection (Peton et al., 2014).

In this study, new genomic information related to 12 *S. aureus* strains, isolated from ovine herds in France, was acquired. Together with the ovine strain ED133, whose complete genome sequence was previously available on the NCBI database, these strains were predicted to be sensitive to methicillin (MSSA). However, zoonotic transmission of *S. aureus* is of great interest and the horizontal acquisition of resistance genes could increase the relevance of ovine mastitis in human infection with methicillin resistant *S. aureus* (MRSA). 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 has an indirect effect on methicillin resistance, but further biochemical studies are necessary to elucidate the role of *fmtB* in this mechanism of resistance (Komatsuzawa et al., 2000).

To evaluate the clonal relationship of the ovine *S. aureus* strains, MLST analysis was conducted. The sequence typing of *S. aureus* is routinely done with allelic variants of 7 housekeeping genes (*arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, *yqiL*) (Enright et al., 2000), and the combination of these variants defines a specific sequence type. If at least 5 out of the 7 alleles in the different *loci* are shared between two *S. aureus* isolates, these isolates form a single clonal complex (Enright et al., 2000; Maiden et al., 1998; Smith et al., 2005). The CC130, which had already been associated with ovine mastitis (Guinane et al., 2010; Smith et al., 2014), encompassed 5 of the ovine strains considered in this study (STs 2411, 2490, and 700). Also, three of the selected ovine strains were found to belong to the CC30 (STs 30 and 243). Studies have associated these STs with human, bovine and ovine infections in Europe and Asia (Monecke et al., 2008; Rabello et al.,

2007; Smith et al., 2014). The strains ED133, O268, O17, O322, and O267 present the same ST and belong to the CC133, which is associated with intramammary infections in ovines (Smith, et al., 2013; Smyth et al., 2009). Interestingly, only the strain O331 presents the ST59. *S. aureus* ST59 has already been isolated from bovine milk (Hata et al., 2010) and dairy goat mastitis (Chu et al., 2012). The strain O331 is likely an atypical clone among the ovine strains considered in this study, since studies showed that the ST59 isolates originated in the United States and have spread to Asia. Also, ST59 *S. aureus* have also been isolated in Oceania and the North of Europe (Mediavilla et al., 2012; Smith et al., 2014; Tristan et al., 2007). 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 has potential to provide a much higher strain typing resolution compared to the MLST analysis, since a larger number of genes is considered (Hall et al., 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 by the ST700 (Fig. 1). In turn, the phylogenomic analysis based on the amino acid variation in the proteomes of the *S. aureus* strains (Fig. 7) resulted in the grouping of all strains belonging to the CC130, with a bootstrap value of 100. However, the phylogenomic analysis also reinforced the fact that the STs 2490 and 2011 are more closely related to each other than to ST700.

It is also noteworthy the strengthening of confidence observed in the formation of the ovine *S. aureus* clades, following the amino acid-based phylogenomic analyses. For example, the strains RF122 and O331 are grouped, in the MLST analysis, with a bootstrap value of 66 (Fig. 1), while the same clade in the phylogenomic tree is supported by a bootstrap value of 88 (Fig. 7). Considering that RF122 was isolated from cattle and O331 from ovine mastitis, this result provides strong evidence that these two strains are closely related to each other. Also, LGA251 and O55 are once more grouped, reinforcing the close phylogenetic relationship between these two strains that were isolated from different hosts and countries. Interestingly, 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, in which a bootstrap value of 94 was observed. This result suggests that these three strains, although belonging to the same clonal complex (CC30), do not share a very high level of similarities in their proteomes. In the phylogenomic analysis with Gegenees, two ovine clusters were formed with similarities ranging from 97 to 100%. It is possible to observe a clonal-behavior among them, compared with strains from the others hosts. However, three strains from ovines (O217, O55 and O331) are not grouped in these clusters. O217 is grouped with ATCC25923 and IRLI_Emoyle1/1, in agreement with the CC formation. The same clustering agreement is observed for O55, which belongs to the same clonal complex of LGA251.

Another mean to measure evolutionary relatedness among closely related bacterial strains, through identity and similarity values of the total genome sequence is using the ANI method (Kim et al., 2014; Konstantinidis et al., 2006). In this analysis, all ovine genomes presented more than 98% of similarity regarding ED133 genome, except for O217, which presented the minor percentage of similarity among sheep strains. In the same way, this strain has not been grouped with any ovine genomes in MLST analysis. This study showed that it belongs to CC30, important in human and animal infections, which suggests no host- specialization by this strain. On the other hand, genomes isolated from milk or mastitis in cows (LGA251 and RF122) present a similarity level major than 98%, regarding to ED133 (isolated from ovine). Previous studies reported that some CCs were originated from human and acquired genetic adaptation to infect ruminants (Guinane et al., 2010; Sakwinska et al., 2011). Furthermore, in the present study, the evolutionary separation of some ovine and bovine strains is not totally clear.

The genomic diversity of multiple strains can be uncover determining the core, accessory and exclusive genomes (Bosi et al., 2016; Chaudhari et al., 2016). All strains used in this comparative analysis have the average of 2679 genes predicted by RAST server. The comparative genomics performed using BPGA software yielded 3964 genes. In its turn, the core genome is formed by 2110 genes among all strains. It represents 79% 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., 2016). Additionally, the core genome analysis was

performed among ovine strains and among other hosts strains, resulting in 2149 and 2119 genes, respectively. It represents 79% (ovine strains) and 81% (other hosts) of the average genes predicted by RAST server. The number of core genes in this study suggests low diversification in core genome of *S. aureus* strains used for comparison. About the functions core genes, they were classified, by COG categories, playing role in amino acid transport and metabolism. This data is according of the expected, because core genome belongs to the group of housekeeping functions (Tettelin et al., 2005). The accessory genes predicted in this present study (n= 859), are in their majority, involved in replication, recombination and repair (R). These genes play role in lateral gene transfer events from mobile genetics elements, such as transposons and bacteriophages (Bosi et al., 2016; Tettelin et al., 2005). Of the exclusive genes found in ovine genomes (n=281), the major part consists in hypothetical proteins (41%), which are proteins of unknown function, and can contribute with many activities in the genome. For instance, a functional assignment of hypothetical proteins in *S. aureus* study, predicted these as binding proteins, helicases, transporters and virulence factors (Prava et al., 2018). Mobile genetic elements represent 25% of the exclusive ovine genes. They are very important in bacterial diversity due to capacity to transduce host genes and confers novel genetic information for bacteria, such as genes involved in virulence factors (Kwan et al., 2005). Some genes predicted playing 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). Furthermore, two prophages were found in some ovine strains and they were not found in strains of other hosts group. The PHAGE_Acinet_vB_AbaM_phiAbaA1_NC_031280, present in O17, O268, O322, O55, O217 and O267, was predicted as an incomplete phage, with score <70 by PHASTER, and produces 15 proteins. This phage is commonly found in *Actinobacter baumanii* (Turner et al., 2017). Noteworthy, Kwan *et al.,* 2005 have been showed that gene transfer among *S. aureus* phages are more predominant than between of *S. aureus* and other species. The other prophage found was the PHAGE_Staphy_StB27_NC_019914, present in O268, O331, ED133 and O267, which carries principally hypothetical proteins, that may play role in virulence factors (Lima-Mendez et al., 2011). Both prophages were classified as PAIs or GEIs by GIPSy (Soares et al., 2016). Genes that encode cell wall proteins represent 3% of the genes found in all ovine genomes. These proteins are important in adherence (Silhavy et al., 2010) and antibiotic resistance (Assis et al., 2017). Likewise, 3% of the genes were predicted as

toxins precursor. Toxins have an important role in pathogenicity, in this context, *S. aureus* exotoxins are a leading cause of gastroenteritis in human from consumption of contaminated food, principally of raw milk and raw milk cheese, which can have contaminations from animal origins due to infections, such as mastitis (Balaban and Rasooly, 2000; Le Loir et al., 2003). Additionally, exfoliative toxins were found in *S. aureus* isolated from cows, and these proteins are agent of scalded- skin syndrome in humans (Vautor et al., 2009). In addition, the ovine exclusive genes were compared to the VFDB database through BLASTp, and resulted 14 proteins predicted, such as the coagulase (*coa*), which is secreted by *S. aureus* and causes clotting in the host's plasma, as a result of mechanisms of escape from immune system (Javid et al., 2018), hyaluronate lyase precursor (*hysA*), which plays role in subcutaneous infection (Ibberson et al., 2014; Makris et al., 2004) and toxins.

Finally, the comparative genomics among the isolates of the same mastitis form, resulted in hypothetical and phage-associated proteins present in some but not in all genomes of the same mastitis isolates. Therefore, further studies are needed to define the bacterial virulence mechanisms of the mastitis spectrum. It was previously reported that different levels of iron metabolism, transcriptional regulators and exoprotein production can contribute to ovine mastitis severity, as well as, different levels of toxin expression were related to pathogenic potential of genomes isolated from bovine mastitis (Ben Zakour et al., 2008; Maréchal et al., 2011).

Conclusion

In this study, comparative genomic analyses were performed with *S. aureus* genomes isolated from different forms of ovine mastitis, against genomes retrieved from the NCBI database and isolated from different hosts. The MLST and phylogenomic analyses revealed two major clonal complexes formed by the ovine genomes, each one comprised of 5 strains. However, three of the ovine strains studied do not belong to these groups and have probably evolved from distinct ancestors. These atypical ovine strains have indispensable genic regions shared with isolates from other hosts. Also, some of the ovine and bovine strains seem to be closely related. The comparative genomic analyses revealed a high gene conservation level among the *S. aureus* strains considered for study. However, accessory genes encoding virulence factors and unknown proteins that might play an important role in the establishment of infection are exclusively found in the ovine genomes. This work brings to light new evidences of genomic specialization in ovine mastitis-associated *S. aureus.*

References

Ågren, J., Sundström, A., Håfström, T., and Segerman, B. (2012). Gegenees: Fragmented Alignment of Multiple Genomes for Determining Phylogenomic Distances and Genetic Signatures Unique for Specified Target Groups. *PLoS ONE* 7. doi:10.1371/journal.pone.0039107.

Aires-de-Sousa, M., Parente, C. E. S. R., Vieira-da-Motta, O., Bonna, I. C. F., Silva, D. A., and Lencastre, H. de (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, 3845–3849. doi:10.1128/AEM.00019-07.

Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12, 402. doi:10.1186/1471-2164-12-402.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2.

Alves, P. D. D., McCulloch, J. A., Even, S., Le Maréchal, C., Thierry, A., Grosset, N., 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, 190–195. doi:10.1016/j.vetmic.2008.12.014.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16-21. doi:10.1093/nar/gkw387.

Assis, L. M., Nedeljković, M., and Dessen, A. (2017). New strategies for targeting and treatment of multi-drug resistant Staphylococcus aureus. *Drug Resist. Updat. Rev. Comment. Antimicrob. Anticancer Chemother.* 31, 1–14. doi:10.1016/j.drup.2017.03.001.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9, 75. doi:10.1186/1471-2164-9-75.

Balaban, N., and Rasooly, A. (2000). Staphylococcal enterotoxins. *Int. J. Food Microbiol.* 61, 1–10.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19, 455–477. doi:10.1089/cmb.2012.0021.

Ben Zakour, N. L., Sturdevant, D. E., Even, S., Guinane, C. M., Barbey, C., Alves, P. D., et al. (2008). Genome-Wide Analysis of Ruminant Staphylococcus aureus Reveals Diversification of the Core Genome. *J. Bacteriol.* 190, 6302–6317. doi:10.1128/JB.01984-07.

Bergonier, D., de Crémoux, R., Rupp, R., Lagriffoul, G., and Berthelot, X. (2003). Mastitis of dairy small ruminants. *Vet. Res.* 34, 689–716. doi:10.1051/vetres:2003030.

Binnewies, T. T., Motro, Y., Hallin, P. F., Lund, O., Dunn, D., La, T., et al. (2006). Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Funct. Integr. Genomics* 6, 165–185. doi:10.1007/s10142-006-0027-2.

Bosi, E., Monk, J. M., Aziz, R. K., Fondi, M., Nizet, V., and Palsson, B. Ø. (2016). Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc. Natl. Acad. Sci.* 113, E3801–E3809. doi:10.1073/pnas.1523199113.

Carver, T., Berriman, M., Tivey, A., Patel, C., Böhme, U., Barrell, B. G., et al. (2008). Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24, 2672–2676. doi:10.1093/bioinformatics/btn529.

Chaudhari, N. M., Gupta, V. K., and Dutta, C. (2016). BPGA- an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* 6. doi:10.1038/srep24373.

Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., et al. (2005). VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 33, D325–D328. doi:10.1093/nar/gki008.

Chu, C., Yu, C., Lee, Y., and Su, Y. (2012). Genetically divergent methicillin-resistant Staphylococcus aureus and sec-dependent mastitis of dairy goats in Taiwan. *BMC Vet. Res.* 8, 39. doi:10.1186/1746-6148-8-39.

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res.* 14, 1394–1403. doi:10.1101/gr.2289704.

Davis, J. J., Gerdes, S., Olsen, G. J., Olson, R., Pusch, G. D., Shukla, M., et al. (2016). PATtyFams: Protein Families for the Microbial Genomes in the PATRIC Database. *Front. Microbiol.* 7. doi:10.3389/fmicb.2016.00118.

Dobrindt, U., and Hacker, J. (2001). Whole genome plasticity in pathogenic bacteria. *Curr. Opin. Microbiol.* 4, 550–557.

Enright, M. C., Day, N. P. J., Davies, C. E., Peacock, S. J., and Spratt, B. G. (2000). Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones ofStaphylococcus aureus. *J. Clin. Microbiol.* 38, 1008–1015.

Galardini, M., Biondi, E. G., Bazzicalupo, M., and Mengoni, A. (2011). CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol. Med.* 6, 11. doi:10.1186/1751-0473-6-11.

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi:10.1099/ijs.0.64483-0.

Guinane, C. M., Zakour, B., L, N., Tormo-Mas, M. A., Weinert, L. A., Lowder, B. V., et al. (2010). Evolutionary Genomics of Staphylococcus aureus Reveals Insights into the Origin and Molecular Basis of Ruminant Host Adaptation. *Genome Biol. Evol.* 2, 454–466. doi:10.1093/gbe/evq031.

Halasa, T., Nielen, M., Huirne, R. B. M., and Hogeveen, H. (2009). Stochastic bio-economic model of bovine intramammary infection. *Livest. Sci.* 124, 295–305. doi:10.1016/j.livsci.2009.02.019.

Hall, B. G., Ehrlich, G. D., and Hu, F. Z. (2010). Pan-genome analysis provides much higher strain typing resolution than multi-locus sequence typing. *Microbiology* 156, 1060–1068. doi:10.1099/mic.0.035188-0.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22, 160–174.

Hata, E., Katsuda, K., Kobayashi, H., Uchida, I., Tanaka, K., and Eguchi, M. (2010). Genetic Variation among Staphylococcus aureus Strains from Bovine Milk and Their Relevance to Methicillin-Resistant Isolates from Humans. *J. Clin. Microbiol.* 48, 2130–2139. doi:10.1128/JCM.01940-09.

Ibberson, C. B., Jones, C. L., Singh, S., Wise, M. C., Hart, M. E., Zurawski, D. V., et al. (2014). Staphylococcus aureus hyaluronidase is a CodY-regulated virulence factor. *Infect. Immun.* 82, 4253–4264. doi:10.1128/IAI.01710-14.

Javid, F., Taku, A., Bhat, M. A., Badroo, G. A., Mudasir, M., and Sofi, T. A. (2018). Molecular typing of Staphylococcus aureus based on coagulase gene. *Vet. World* 11, 423–430. doi:10.14202/vetworld.2018.423-430.

Kaya, H., Hasman, H., Larsen, J., Stegger, M., Johannesen, T. B., Allesøe, R. L., et al. (2018). SCCmecFinder, a Web-Based Tool for Typing of Staphylococcal Cassette Chromosome mec in Staphylococcus aureus Using Whole-Genome Sequence Data. *mSphere* 3, e00612-17. doi:10.1128/mSphere.00612-17.

Kayansamruaj, P., Dong, H. T., Hirono, I., Kondo, H., Senapin, S., and Rodkhum, C. (2017). Comparative genome analysis of fish pathogen Flavobacterium columnare reveals extensive sequence diversity within the species. *Infect. Genet. Evol.* 54, 7–17. doi:10.1016/j.meegid.2017.06.012.

Kim, M., Oh, H.-S., Park, S.-C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351. doi:10.1099/ijs.0.059774-0.

Kloepper, T. H., and Huson, D. H. (2008). Drawing explicit phylogenetic networks and their integration into SplitsTree. *BMC Evol. Biol.* 8, 22. doi:10.1186/1471-2148-8-22.

Komatsuzawa, H., Ohta, K., Sugai, M., Fujiwara, T., Glanzmann, P., Berger-Bächi, B., et al. (2000). Tn551-mediated insertional inactivation of the fmtB gene encoding a cell wall-associated protein abolishes methicillin resistance in Staphylococcus aureus. *J. Antimicrob. Chemother.* 45, 421–431. doi:10.1093/jac/45.4.421.

Komatsuzawa, H., Sugai, M., Ohta, K., Fujiwara, T., Nakashima, S., Suzuki, J., 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. *Antimicrob. Agents Chemother.* 41, 2355–2361.

Konstantinidis, K. T., Ramette, A., and Tiedje, J. M. (2006). The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 1929–1940. doi:10.1098/rstb.2006.1920.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096.

Kwan, T., Liu, J., DuBow, M., Gros, P., and Pelletier, J. (2005). The complete genomes and proteomes of 27 Staphylococcus aureus bacteriophages. *Proc. Natl. Acad. Sci.* 102, 5174–5179. doi:10.1073/pnas.0501140102.

Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., et al. (2012). Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. *J. Clin. Microbiol.* 50, 1355–1361. doi:10.1128/JCM.06094-11.

Le Loir, Y., Baron, F., and Gautier, M. (2003). Staphylococcus aureus and food poisoning. *Genet. Mol. Res. GMR* 2, 63–76.

Lima-Mendez, G., Toussaint, A., and Leplae, R. (2011). A modular view of the bacteriophage genomic space: identification of host and lifestyle marker modules. *Res. Microbiol.* 162, 737–746. doi:10.1016/j.resmic.2011.06.006.

Lindahl, E., and Elofsson, A. (2000). Identification of related proteins on family, superfamily and fold level11Edited by F. C. Cohen. *J. Mol. Biol.* 295, 613–625. doi:10.1006/jmbi.1999.3377.

Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. (1998). Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci.* 95, 3140–3145. doi:10.1073/pnas.95.6.3140.

Makris, G., Wright, J. D., Ingham, E., and Holland, K. T. (2004). The hyaluronate lyase of Staphylococcus aureus - a virulence factor? *Microbiol. Read. Engl.* 150, 2005–2013. doi:10.1099/mic.0.26942-0.

Maréchal, C. L., Seyffert, N., Jardin, J., Hernandez, D., Jan, G., Rault, L., et al. (2011). Molecular Basis of Virulence in Staphylococcus aureus Mastitis. *PLOS ONE* 6, e27354. doi:10.1371/journal.pone.0027354.

Mediavilla, J. R., Chen, L., Mathema, B., and Kreiswirth, B. N. (2012). Global epidemiology of community-associated methicillin resistant Staphylococcus aureus (CA-MRSA). *Curr. Opin. Microbiol.* 15, 588–595. doi:10.1016/j.mib.2012.08.003.

Monecke, S., Slickers, P., and Ehricht, R. (2008). Assignment of Staphylococcus aureus isolates to clonal complexes based on microarray analysis and pattern recognition. *FEMS Immunol. Med. Microbiol.* 53, 237–251. doi:10.1111/j.1574-695X.2008.00426.x.

Oviedo-Boyso, J., Valdez-Alarcón, J. J., Cajero-Juárez, M., Ochoa-Zarzosa, A., López-Meza, J. E., Bravo-Patiño, A., et al. (2007). Innate immune response of bovine mammary gland to pathogenic bacteria responsible for mastitis. *J. Infect.* 54, 399–409. doi:10.1016/j.jinf.2006.06.010.

Peton, V., Bouchard, D. S., Almeida, S., Rault, L., Falentin, H., Jardin, J., et al. (2014). Fine-tuned characterization of Staphylococcus aureus Newbould 305, a strain associated with mild and chronic mastitis in bovines. *Vet. Res.* 45. doi: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. doi:10.1016/j.ijbiomac.2017.10.169.

Rabello, R. F., Moreira, B. M., Lopes, R. M. M., Teixeira, L. M., Riley, L. W., and Castro, A. C. D. (2007). Multilocus sequence typing of Staphylococcus aureus isolates recovered from cows with mastitis in Brazilian dairy herds. *J. Med. Microbiol.* 56, 1505–1511. doi:10.1099/jmm.0.47357-0.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci.* 106, 19126–19131. doi:10.1073/pnas.0906412106.

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., and Peplies, J. (2016). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931. doi:10.1093/bioinformatics/btv681.

Sakwinska, O., Giddey, M., Moreillon, M., Morisset, D., Waldvogel, A., and Moreillon, P. (2011). Staphylococcus aureus Host Range and Human-Bovine Host Shift. *Appl. Environ. Microbiol.* 77, 5908–5915. doi:10.1128/AEM.00238-11.

Silhavy, T. J., Kahne, D., and Walker, S. (2010). The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* 2, a000414. doi:10.1101/cshperspect.a000414.

Smith, E. M., Green, L. E., Medley, G. F., Bird, H. E., Fox, L. K., Schukken, Y. H., et al. (2005). Multilocus Sequence Typing of Intercontinental Bovine Staphylococcus aureus Isolates. *J. Clin. Microbiol.* 43, 4737–4743. doi:10.1128/JCM.43.9.4737-4743.2005.

Smith, E. M., Needs, P. F., Manley, G., and Green, L. E. (2014). Global distribution and diversity of ovine-associated Staphylococcus aureus. *Infect. Genet. Evol.* 22, 208–215. doi:10.1016/j.meegid.2013.09.008.

Smyth, D. S., Feil, E. J., Meaney, W. J., Hartigan, P. J., Tollersrud, T., Fitzgerald, J. R., et al. (2009). Molecular genetic typing reveals further insights into the diversity of animal-associated Staphylococcus aureus. *J. Med. Microbiol.* 58, 1343–1353. doi:10.1099/jmm.0.009837-0.

Soares, S. C., Geyik, H., Ramos, R. T. J., de Sá, P. H. C. G., Barbosa, E. G. V., Baumbach, J., et al. (2016). GIPSy: Genomic island prediction software. *J. Biotechnol.* 232, 2–11. doi:10.1016/j.jbiotec.2015.09.008.

Soares, S. C., Silva, A., Trost, E., Blom, J., Ramos, R., Carneiro, A., et al. (2013). The Pan-Genome of the Animal Pathogen Corynebacterium pseudotuberculosis Reveals Differences in Genome Plasticity between the Biovar ovis and equi Strains. *PLOS ONE* 8, e53818. doi:10.1371/journal.pone.0053818.

Spratt, B. G., Hanage, W. P., Li, B., Aanensen, D. M., and Feil, E. J. (2004). Displaying the relatedness among isolates of bacterial species -- the eBURST approach. *FEMS Microbiol. Lett.* 241, 129–134. doi:10.1016/j.femsle.2004.11.015.

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." *Proc. Natl. Acad. Sci.* 102, 13950–13955. doi:10.1073/pnas.0506758102.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi:10.1093/nar/22.22.4673.

Tristan, A., Bes, M., Meugnier, H., Lina, G., Bozdogan, B., Courvalin, P., et al. (2007). Global Distribution of Panton-Valentine Leukocidin–positive Methicillin-resistant Staphylococcus aureus, 2006. *Emerg. Infect. Dis.* 13, 594–600. doi:10.3201/eid1304.061316.

Turner, D., Ackermann, H.-W., Kropinski, A., Lavigne, R., Sutton, J., and Reynolds, D. (2017). Comparative Analysis of 37 Acinetobacter Bacteriophages. *Viruses* 10, 5. doi:10.3390/v10010005.

Vautor, E., Cockfield, J., Le Marechal, C., Le Loir, Y., Chevalier, M., Robinson, D. A., et al. (2009). Difference in virulence between Staphylococcus aureus isolates causing gangrenous mastitis versus subclinical mastitis in a dairy sheep flock. *Vet. Res.* 40. doi:10.1051/vetres/2009039.

Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., et al. (2017). Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* 45, D535–D542. doi:10.1093/nar/gkw1017.

Yang, A.-S., and Honig, B. (2000). An integrated approach to the analysis and modeling of protein sequences and structures. II. On the relationship between sequence and structural similarity for proteins that are not obviously related in sequence11Edited by F. Cohen. *J. Mol. Biol.* 301, 679–689. doi:10.1006/jmbi.2000.3974.

Zakour, N. B., and Loir, Y. (2007). “Designing Primers for Whole Genome PCR Scanning Using the Software Package GenoFrag,” in *PCR Primer Design*, ed. A. Yuryev (Totowa, NJ: Humana Press), 348–368. doi:10.1007/978-1-59745-528-2_18.

Supplementary Material

Table S1: Alleles prediction of the seven housekeeping genes considered in MLST analyses.

	Alleles						
Strains	***arcC***	***aroE***	***glpF***	***gmk***	***pta***	***tpi***	***yqiL***
O11 O326 O408	6	57	45	2	7	95	52
O46	6	57	45	2	7	14	52
O82	200	57	45	2	7	14	52
O268 O267 ED133 O322 O17	6	66	46	2	7	50	18
O55 LGA251	18	33	6	20	7	50	48
O217 IRLI_Emoyle1/1	2	2	2	2	6	3	2
ATCC25923	2	2	5	2	6	3	2

Table S2: Virulence proteins of exclusive genes from ovine *S. aureus* found in VFDB database.

VFDB prediction	Query ID	Subject ID	%ID	%Query coverage
VFG000365(gb|NP_405474) (ybtP) lipoprotein inner membrane ABC-transporter [Yersiniabactin (VF0136)] [Yersinia pestis CO92]	sa_o82_0051	VFG000365(gb|NP_405474)	31.034	83
VFG000365(gb|NP_405474) (ybtP) lipoprotein inner membrane ABC-transporter [Yersiniabactin (VF0136)] [Yersinia pestis CO92]	sa_o82_0050	VFG000365(gb|NP_405474)	32.869	85
VFG000359(gb|NP_405468) (ybtE) yersiniabactin siderophore biosynthetic protein [Yersiniabactin (VF0136)] [Yersinia pestis CO92]	sa_o408_0042	VFG000359(gb|NP_405468)	38.856	94
VFG000368(gb|NP_405477) (ybtS) salicylate synthase Irp9 [Yersiniabactin (VF0136)] [Yersinia pestis CO92]	sa_o11_0052	VFG000368(gb|NP_405477)	41.210	76
VFG002423(gb|YP_001332910) (scn) complement inhibitor SCIN [SCIN (VF0425)] [Staphylococcus aureus subsp. aureus str. Newman]	sa_o46_0410	VFG002423(gb|YP_001332910)	50.427	100
VFG001801(gb|AAA26628) (etb) exfoliative toxin B [Exfoliative toxin (VF0009)] [Staphylococcus aureus]	sa_o82_2066	VFG001801(gb|AAA26628)	57.762	99
VFG002418(gb|YP_001331791) (vWbp) secreted von Willebrand factor-binding protein precursor [vWbp (VF0420)] [Staphylococcus aureus subsp. aureus str. Newman]	sa_o268_1995	VFG002418(gb|YP_001331791)	60.000	100
VFG002419(gb|NP_645021) (coa)	sa_o267_0167	VFG002419(gb|NP_645021)	78.199	89

staphylocoagulase precursor [Staphylocoagulase (VF0421)] [Staphylococcus aureus subsp. aureus MW2]				
VFG001280(gb|NP_645334) (sdrD) Ser-Asp rich fibrinogen-binding bone sialoprotein-binding protein [SDr (VF0019)] [Staphylococcus aureus subsp. aureus MW2]	sa_o408_0574	VFG001280(gb|NP_645334)	79.409	85
VFG001315(gb|NP_646946) (hysA) hyaluronate lyase precursor [Hyaluronate lyase (VF0013)] [Staphylococcus aureus subsp. aureus MW2]	sa_o331_1632	VFG001315(gb|NP_646946)	88.022	99
VFG001275(gb|NP_647161) (hlgB) gamma-hemolysin component B [<gamma>-hemolysin (VF0011)] [Staphylococcus aureus subsp. aureus MW2]	sa_o82_2277	VFG001275(gb|NP_647161)	90.476	69
VFG001327(gb|NP_646755) (selk) staphylococcal enterotoxin K precursor [SE (VF0020)] [Staphylococcus aureus subsp. aureus MW2]	sa_o331_0728	VFG001327(gb|NP_646755)	96.680	99
VFG001326(gb|NP_646754) (selq) staphylococcal enterotoxin G precursor [SE (VF0020)] [Staphylococcus aureus subsp. aureus MW2]	sa_o331_0729	VFG001326(gb|NP_646754)	97.934	100
VFG001802(gb|AAA88550) (seb) staphylococcal enterotoxin B [SE (VF0020)] [Staphylococcus aureus S6]	sa_o331_0749	VFG001802(gb|AAA88550)	100.000	100

Table S3: Shared proteins among *S. aureus* ovine genomes isolated of the same mastitis form.

ID Protein	Product	Chronic mastitis				
		O217				
PLF_1279_00004852	Hypothetical protein	x				
PLF_1279_00005354	Hypothetical protein	x				
PLF_1279_00125358	Hypothetical protein	x				

ID Protein	Product	Gangrenous mastitis	
		O11	O408
PLF_1279_00005879	Hypothetical protein	x	
PLF_1279_00028431	Hypothetical protein		x

ID Protein	Product	Subclinical mastitis				
		O46	O82	O267	O331	O55
PLF_1279_00002691	Hypothetical protein	x	x			
PLF_1279_00002997	Hypothetical protein			x		x
PLF_1279_00016586	Hypothetical protein	x	x			
PLF_1279_00123752	Hypothetical protein		x	x		

ID Protein	Product	Clinical mastitis				
		O268	O17	O322	O326	ED133
PLF_1279_00002376	Phage associated protein	x				x
PLF_1279_00002379	Phage associated protein	x				x
PLF_1279_00002380	Phage terminase, large subunit	x				x
PLF_1279_00002381	Phage major capsid protein	x				x
PLF_1279_00002382	Phage protein	x				x
PLF_1279_00002383	Phage protein	x				x
PLF_1279_00002384	Prophage Clp protease-like protein	x				x
PLF_1279_00002385	Phage protein	x				x
PLF_1279_00002386	Phage protein	x				x
PLF_1279_00002387	Phage protein	x				x
PLF_1279_00002388	Phage-associated homing endonuclease	x				x
PLF_1279_00002390	Phage protein	x				x
PLF_1279_00002391	Phage transcriptional terminator	x				x
PLF_1279_00002861	Hypothetical protein	x	x	x		
PLF_1279_00010931	Hypothetical protein	x		x		
PLF_1279_00018094	Hypothetical protein	x				x
PLF_1279_00026606	Hypothetical protein		x	x		
PLF_1279_00029130	Hypothetical protein		x			x
PLF_1279_00030260	Cl-like repressor, phage associated	x				x

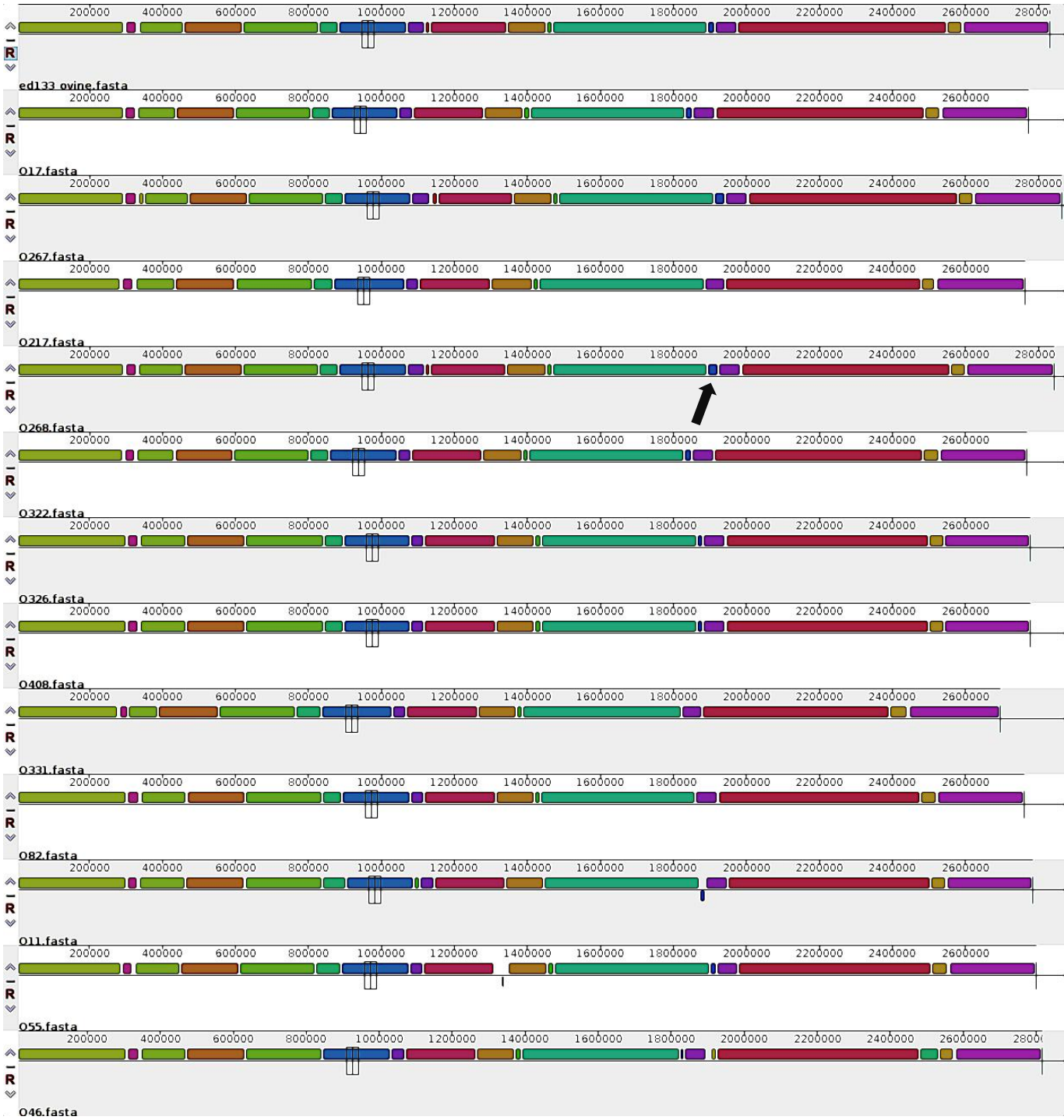

Fig. S1: Synteny analysis of *S. aureus* strains isolated from ovines. This alignment was performed using Progressive Mauve option of the software Mauve. Each block represents sequences shared by genomes (locally collinear blocks). The black arrow, represents a prophage, which is absent in O217, O408 and O331 genomes.

Table S4: Features of prophages and their proteins predicted in the genome *S. aureus* O268.

Intact phage	Genes	Product
Region 1 PHAGE_Staphy_StauST398_2_NC_021323	psuG, sgIT, nanA, nanE, int, hel	Perfringolysin O regulator protein PfoR, Pseudouridine kinase, Pseudouridine 5' -phosphate glycosidase, Predicted sialic acid transporter, N- acetylneuraminate lyase, N- acetylmannosamine kinase, Sialic acid utilization regulator- RpiR Family, N-acetylmannosamine-6-phosphate 2-epimerase, Predicted histidine uptake transporter, Triacylglycerol lipase, Phage integrasse, CAAX amino terminal protease Family, Phage repressor protein, Phage antirepressor protein, Single- stranded DNA- binding protein, Phage replication initiation protein, DNA helicase, Phage associated, Transcriptional regulator, Cro/CI Family, Dimeric dUTPase, Transcriptional activator rinB- phage associated, DNA primase- phage associated, DNA helicase- phage associated
Incomplete phage	**Genes**	**Product**
Region 2 PHAGE_Staphy_PT1028_NC_007045	guaB, guaA, pri, tst, sel, entC1	Inosine- 5' monophosphate dehydrogenase / CBS domain, GMP synthase [glutamine- hydrolyzing], aminotransferase subunit/ GMP synthase [glutamine - hydrolyzing], ATP pyrophosphatase subunit, Integrase, Superantigen - encoding pathogenicity islands SaPI, Putative primase, Superantigen-encoding pathogenicity islands SaPI, Phage terminase small subunit, Toxic shock syndrome toxin 1, Superantigen enterotoxin, Superantigen enterotoxin SEL
Region 3 PHAGE_Staphy_77_NC_005356	ssb_1, isdA_1, isdA_2, isdC, isdD, isdE	LSU ribosomal protein, Phage DNA invertase, Cro, Phage antirepressor protein, Single-stranded DNA binding protein, Phage replication initiation protein, DNA helicase- phage associated, Deoxyuridine 5'- triphosphate nucleotidohydrolase (EC 3.6.1.23) [SA bacteriophages 11, Mu50B], Transcriptional activator rinB - phage associated, Phage terminase- large subunit, Phage portal (connector) protein, Prophage Clp protease- like protein, Phage major capsid protein, Phage transcriptional terminator, Phage tail length tape- measure protein, Structural protein, phage associated, Enterotoxin- phage associated, Phage holin, Phage lysin, N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28), HNH homing endonuclease, Phage lysin, N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28), Cell surface receptor IsdB for hemoglobin and hemoglobin complexes, Cell surface protein IsdA, Transfers heme from hemoglobin to apo-IsdC, NPQTN cell wall anchored protein IsdC, Heme transporter IsdDEF, membrane component IsdD, Heme transporter IsdDEF, lipoprotein IsdE
Region 4 PHAGE_Staphy_StB27_NC_019914		Hypothetical proteins, Transposase

Region 5 PHAGE_Staphy_StauST398_2_NC_021323	lukD, lukE_1, lukE_2, hemY, hemH, hemE, EcsB, ecsA_4	Mobile element protein, Leukotoxin LukD, Leukotoxin LukE, Leukotoxin LukE, Phage integrase, Phage lysin, N-acetylmuramoyl-L-alanine amidase (EC 3.5.28), Phage antirepressor protein, Phage repressor, Integrase, tRNA-Ser-GGA, tRNA-Glu-TTC, tRNA-Asn-GTT, tRNA-Gly-TCC, tRNA-His-GTG, tRNA-Phe-GAA, tRNA-Asp-GTC, tRNA-Met-CAT, Protoporphyrinogen IX oxidase - aerobic - HemY (EC 1.3.3.4), Ferrochelatase, protoheme ferro-lyase (EC 4.99.1.1), Uroporphyrinogen III decarboxylase (EC 4.1.1.37), ABC transporter- permease protein EscB, ABC transporter- permease protein EscA, Histidine triad nucleotide-binding protein- similarity with At5g48545 and yeast YDL125C
Region 7 PHAGE_Acinet_vB_AbaM_phiAbaA1_NC_031280	vWbp, Int-Tn, groL, groS	Permease of the drug/metabolite transporter (DMT) superfamily, 5-aminolevulinate synthase, Potassium uptake protein- integral membrane component, KtrB, Secreted von Willebrand factor-binding protein VWbp, Involved in expression of fibrinogen binding protein- phage associated, Putative primase- superantigen-encoding pathogenicity islands SaPI, Phage antirepressor protein, Integrase- superantigen-enconding pathogenicity islands SaPI, Heat shock protein 60 family chaperone GroEL, Heat shock protein 60 family co-chaperone GroES
Questionable phage	**Genes**	**Product**
Region 6 PHAGE_Staphy_phiPV83_NC_002486	bcglA, lukF-PV, hlgA_1, yqbO, ssbA,	Type I restriction- modification system- DNA-methyltransferase subunit M, Panton-Valentine leukocidin chain F precursor, Panton-Valentine leukocidin chain S precursor, Phage lysin, N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28), Phage tail lenght tape-measure protein, Phage DNA packaging, Phage capsid protein, Phage head maturation protease, Phage portal protein, Phage terminase, large subunit, Phage terminase- small subunit, Transcriptional activator rinB- phage associated, Dimeric dUTPase (EC 3.6.1.23), DNA helicase- phage associated, Phage replication initiation, Pathogenesis-related transcriptional factor and ERF, Single-stranded DNA- binding protein, Phage repressor, Phage integrase

Table S5: Region of prophages found in ovine *S. aureus* genomes.

Strains	Regions		Status
	Beginning	End	
O11	326K	371K	Intact
	410K	437K	Intact
	884K	914K	Questionable
	1,95M	1.99M	Questionable
O17	320K	361K	Intact
	1.91M	1.96M	Questionable
	2.03M	2.05M	Incomplete
O46	334K	380K	Intact
	419K	446K	Intact
	1.89M	1.97M	Intact
O55	313K	359K	Intact
	872K	889K	Questionable
	1.31M	1.35M	Intact
	2.06M	2.08M	Incomplete
O82	329K	375K	Intact
	413K	441K	Intact
	1.92M	1.96M	Questionable
O217	313K	359K	Intact
	1.31K	1.32K	Incomplete
	2.03M	2.04M	Incomplete
O267	320K	384K	Intact
	1.12M	1.17M	Intact
	1.37M	1.38M	Incomplete
	1.93M	1.95M	Incomplete
	1.99M	2.05M	Intact
	2.12M	2.14M	Incomplete

Strains	Regions		Status
	Beginning	End	
O268	307K	368K	Intact
	431K	448K	Incomplete
	1.10M	1.15M	Incomplete
	1.35M	1.35M	Incomplete
	1.90M	1.93M	Incomplete
	1.98M	2.02M	Questionable
	2.09M	2.12M	Incomplete
O322	279K	318K	Intact
	1.87M	1.92M	Questionable
	1.99M	2.01M	Incomplete
O326	328K	373K	Intact
	412K	439K	Intact
	1.94M	1.99M	Questionable
O331	818K	836K	Questionable
	1.27M	1.28M	Incomplete
O408	328K	373K	Intact
	412K	439K	Intact
	1.94M	1.99M	Questionable
ED133	307K	369K	Intact
	432K	449K	Incomplete
	1.11M	1.15M	Incomplete
	1.35M	1.36M	Incomplete
	1.97M	2.02M	Intact
	2.09M	2.11M	Incomplete

5. Perspectivas

Este trabalho tem como perspectivas:

- Realizar a anotação funcional das proteínas hipotéticas encontradas exclusivamente em genomas isolados de ovinos, a fim de elucidar os possíveis papeis dessas proteínas na infecção e associação ao hospedeiro;

- Detalhar os resultados obtidos na análise de predição de ilhas genômicas e correlacioná-los com os genomas isolados de diferentes formas da mastite;

- Avaliar os genomas de *S. aureus* isolados de ovinos através de polimorfismos de nucleotídeo único, com o intuito de se estabelecer taxas de mutação entre as linhagens;

6. Referências Bibliográficas

Alves, P. D. D.; *et al*. Molecular characterisation of Staphylococcus aureus strains isolated from small and large ruminants reveals a host rather than tissue specificity. **Veterinary Microbiology**, v. 137, n.1-2, p. 190-195, may 2009.

Barkema, H. W.; *et al*. Invited Review: The role of cow, pathogen, and treatment regimen in the therapeutic success of bovine Staphylococcus aureus mastitis. **Journal of Dairy Science**, v. 89, n. 6, p. 1877–1895, jun. 2006.

Berglund, E. C.; *et al*. Next-generation sequencing technologies and applications for human genetic history and forensics. **Investigative Genetics**, v. 2, n. 1, p. 23, 2011.

Bradley, A. J. Bovine Mastitis: An Evolving Disease. **The Veterinary Journal**, v. 164, n. 2, p. 116–128, set. 2002.

Contreras, A. *et al*. Mastitis in small ruminants. **Small Ruminant Research**, v. 68, n. 1, p. 145–153, 1 mar. 2007.

Da Silva, E. R. *et al*. Identification and in vitro antimicrobial susceptibility of Staphylococcus species isolated from goat mastitis in the Northeast of Brazil. **Small Ruminant Research**, v. 55, n. 1–3, p. 45–49, out. 2004.

Deleo, F. R. & Chambers, H. F. Reemergence of antibiotic-resistant Staphylococcus aureus in the genomics era. **Journal of Clinical Investigation**, v. 119, n. 9, p. 2464–2474, 1 set. 2009.

Delsuc, F. Army Ants Trapped by Their Evolutionary History. **PLoS Biology**, v. 1, n. 2, p. e37, 17 nov. 2003.

EFSA PANEL ON ANIMAL HEALTH AND WELFARE (AHAW). Scientific Opinion on the welfare risks related to the farming of sheep for wool, meat and milk production. **EFSA Journal**, v. 12, n. 12, dez. 2014.

Feil, E. J. *et al*. How Clonal Is Staphylococcus aureus? **Journal of Bacteriology**, v. 185, n. 11, p. 3307–3316, 1 jun. 2003.

Fleischmann, R. D. *et al*. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. **Science (New York, N.Y.)**, v. 269, n. 5223, p. 496–512, 28 jul. 1995.

Gordon, R. J. & Lowy, F. D. Pathogenesis of Methicillin-Resistant Staphylococcus aureus Infection. **Clinical infectious diseases : an official publication of the Infectious Diseases Society of America**, v. 46, n. Suppl 5, p. S350–S359, 1 jun. 2008.

Harris, L. G.; *et al*. An introduction to Staphylococcus aureus, and techniques for identifying and quantifying S. aureus adhesins in relation to adhesion to biomaterials: review. **European Cells & Materials**, v. 4, p. 39–60, 31 dez. 2002.

Harris, S. R.; *et al*. Evolution of MRSA during hospital transmission and intercontinental spread. **Science (New York, N.Y.)**, v. 327, n. 5964, p. 469–474, 22 jan. 2010.

Holden, M. T. G.; *et al*. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. **Genome Research**, v. 23, n. 4, p. 653–664, 1 abr. 2013.

Jain, M.; *et al*. Improved data analysis for the MinION nanopore sequencer. **Nature Methods**, v. 12, n. 4, p. 351–356, abr. 2015.

Kuroda, M.; *et al*. Whole genome sequencing of meticillin-resistant Staphylococcus aureus. **The Lancet**, v. 357, n. 9264, p. 1225–1240, 21 abr. 2001.

Le Loir, Y.; *et al*. Staphylococcus aureus and food poisoning. **Genetics and molecular research: GMR**, v. 2, n. 1, p. 63–76, 31 mar. 2003.

Le Maréchal, C.; *et al*. Mastitis impact on technological properties of milk and quality of milk products–a review. **Dairy Science & Technology**, v. 91, n. 3, p. 247–282, 2011.

Lents, C. A.; *et al*. Effects of dry cow treatment of beef cows on pathogenic organisms, milk somatic cell counts, and calf growth during the subsequent lactation1. **Journal of Animal Science; Champaign**, v. 86, n. 3, p. 748–55, mar. 2008.

Loman, N. J.; *et al*. Twenty years of bacterial genome sequencing. **Nature Reviews Microbiology**, v. 13, n. 12, p. 787–794, dez. 2015.

Lowy, F. D. Staphylococcus aureus Infections. **New England Journal of Medicine**, v. 339, n. 8, p. 520–532, 20 ago. 1998.

Kaur, R & Malik, C. P. Next generation sequencing: a revolution in gene sequencing. p. 1–20, 2013.

Mardis, E. R. Next-Generation Sequencing Platforms. **Annual Review of Analytical Chemistry**, v. 6, n. 1, p. 287–303, 12 jun. 2013.

Maréchal, C. L. *et al*. Molecular Basis of Virulence in Staphylococcus aureus Mastitis. **PLOS ONE**, v. 6, n. 11, p. e27354, 11 nov. 2011.

Mavrogianni, V. S. *et al*. Teat disorders predispose ewes to clinical mastitis after challenge with *Mannheimia haemolytica*. **Veterinary Research**, v. 37, n. 1, p. 89–105, jan. 2006.

Metzker, M. L. Sequencing technologies - the next generation. **Nature Reviews. Genetics**, v. 11, n. 1, p. 31–46, jan. 2010.

Murray, P. R.; AMERICAN SOCIETY FOR MICROBIOLOGY (EDS.). **Manual of clinical microbiology**. 6th ed ed. Washington, D.C: ASM Press, 1995.

Myllys, V. *et al*. Effect of Abrasion of Teat Orifice Epithelium on Development of Bovine Staphylococcal Mastitis. **Journal of Dairy Science**, v. 77, n. 2, p. 446–452, 1 fev. 1994.

Novick, R. P. **Molecular Biology of the Staphylococci**. [s.l.] VCH, 1990.

Oviedo-Boyso, J. *et al*. Innate immune response of bovine mammary gland to pathogenic bacteria responsible for mastitis. **Journal of Infection**, v. 54, n. 4, p. 399–409, abr. 2007.

Peton, V. *et al*. Fine-tuned characterization of Staphylococcus aureus Newbould 305, a strain associated with mild and chronic mastitis in bovines. **Veterinary Research**, v. 45, n. 1, 2014.

Puranik, R. *et al*. A pipeline for completing bacterial genomes using *in silico* and wet lab approaches. **BMC Genomics**, v.16, jun 2014.

Roy, J. P & Keefe, G. Systematic review: what is the best antibiotic treatment for Staphylococcus aureus intramammary infection of lactating cows in North America? **The Veterinary clinics of North America. Food animal practice**, v. 28, n. 1, p. 39–50, viii, mar. 2012.

Smith, E. M. *et al*. Global distribution and diversity of ovine-associated Staphylococcus aureus. **Infection, Genetics and Evolution**, v. 22, n. 100, p. 208–215, mar. 2014.

Vautor, E. *et al*. Genotyping of Staphylococcus aureus isolated from various sites on farms with dairy sheep using pulsed-field gel electrophoresis. **Veterinary Microbiology**, v. 96, n. 1, p. 69–79, 8 out. 2003.

Vautor, E. *et al*. Difference in virulence between Staphylococcus aureus isolates causing gangrenous mastitis versus subclinical mastitis in a dairy sheep flock. **Veterinary Research**, v. 40, n. 6, 2009.

Zakour, B. N. *et al.* Genome-Wide Analysis of Ruminant *Staphylococcus aureus* Reveals Diversification of the Core Genome. **Journal of Bacteriology**, v. 190, n. 19, 2008.