LIS SOUZA ROCHA

ANÁLISES *IN SILICO* DAS PROTEÍNAS DE SUPERFÍCIE E SECRETADAS DE *Staphylococcus aureus* PARA DISCRIMINAR ENTRE AS MASTITES BOVINA CLÍNICA E SUBCLÍNICA

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Aplicada, para obtenção do título de *Magister Scientiae.*

VIÇOSA
MINAS GERAIS - BRASIL
2018

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
R672a
2018

Rocha, Lis Souza, 1993-
Análises *in silico* das proteínas de superfície e secretadas de *Staphylococcus aureus* para discriminar entre as mastites bovina clínica e subclínica / Lis Souza Rocha. – Viçosa, MG, 2018.
x, 62f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.
Orientador: Andréa de Oliveira Barros Ribon.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 49-57.

1. Mastite. 2. Bovinos - Doenças. 3. *Staphylococcus aureus*. I. Universidade Federal de Viçosa. Departamento de Bioquímica e Biologia Molecular. Programa de Pós-Graduação em Bioquímica Aplicada. II. Título.

CDD 22. ed. 636.2089819

LIS SOUZA ROCHA

ANÁLISES *IN SILICO* DAS PROTEÍNAS DE SUPERFÍCIE E SECRETADAS DE *Staphylococcus aureus* PARA DISCRIMINAR ENTRE AS MASTITES BOVINA CLÍNICA E SUBCLÍNICA

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Aplicada, para obtenção do título de *Magister Scientiae*.

APROVADA: 26 de fevereiro de 2018.

Denise Mara Soares Bazzolli

Tiago Antônio de Oliveira Mendes

Andréa de Oliveira Barros Ribon
(Orientadora)

AGRADECIMENTOS

A Deus pela vida;

À Universidade Federal de Viçosa e ao Departamento de Bioquímica e Biologia Molecular, todos os seus professores e funcionários;

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa concedida;

À professora e orientadora Andréa de Oliveira Barros Ribon pela orientação a atenção incondicionais, sempre conduzindo todos os meus passos em todos os momentos;

Ao professor e coorientador Tiago Antônio de Oliveira Mendes por toda a orientação e paciência na condução das análises *in silico*;

Aos professores Denise Mara Soares Bazzolli e Tiago Antônio de Oliveira Mendes pelas sugestões e participação na banca examinadora;

Ao bioinformata Pedro Marcus Pereira Vidigal, por todas as valiosas instruções nas análises *in silico*;

Aos meus pais Lourdes e José por todo amor e encorajamento;

À minha avó Adelícia e minha tia Cristina pelo amor incondicional e total dedicação para a realização deste sonho;

Aos meus irmãos e melhores amigos, Blenda, Karak e Marina, pela cumplicidade verdadeira e amor infinitos;

Às minhas avós Leontina e Inez *(in memorian)* por todo amor, carinho e orações;

Aos colegas do LBM Ananda, Amanda, Fernanda, Camila, Géssica, Gilza, Higor, Lílian, Lucas, Mônica, Patrícia, Renato, Sérgio, Silvana e Valquíria, pela convivência, pela paciência, pelas ajudas constantes, paciência, e por terem se tornado meus amigos queridos;

Ao Eduardo Pereira Monteiro, secretário do Programa de Pós-graduação em Bioquímica Aplicada pela competência indiscutível e pela companhia;

Aos meus familiares e amigos que se fizeram presentes na transmissão de boas energias e orações.

SUMÁRIO

LISTA DE ABREVIATURAS

BHI – *Brain heart infusion* (infusão de cérebro e coração)

BLAST – *Basic Local Alignment Search Tool (*ferramenta de busca de alinhamento local)

bp – *Base pair/* par de bases (pb)

CCS - Contagem de células somáticas

CDS – *Coding DNA sequence* (sequência de DNA codificante)

Cepea - Centro de Estudos Avançados em Economia Aplicada

CMT - *California Mastitis Test*

CNA - Confederação da Agricultura e Pecuária do Brasil

CWA *(proteins) - cell wall anchored (*proteínas covalentemente ancorada à proteína parede celular)

dN/dS - *ratio of the number of nonsynonymous substitutions per non-synonymous site (pN) to the number of synonymous substitutions per synonymous site (pS) (*taxa do número de substituições não sinônimas por sítio não sinônimo em relação ao número de substituições sinônimas por sítio sinônimo)

DNA – *Deoxyribonucleic acid (*ácido desoxirrobonucleico)

dNTP – *Deoxynucleotide triphosphate (*desoxirribonucleotídeos trifosfatados)

IBGE - Instituto Brasileiro de Geografia e Estatística

IL-1β - *Interleukin*-1beta (Interleucina 1-beta)

IWG-SCC - *International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements*

Kb – Kilo bases

LBM - Laboratório de Biotecnologia Molecular

L-PG - lisil-fosfatidilglicerol

LPL – *Lipoproteins* (lipoproteínas)

MGEs – *Mobile genetic elements (*elementos genéticos móveis)

NCBI – *National Center for Biotechnology Information*

ORF – *Open reading frames* (quadros de leitura aberta)

PBS - *Phosphate-buffered saline* (tampão fosfato salino)

PCR – *Polymerase chain reaction* (reação em cadeia da polimerase)

RAST - *Rapid Annotation using Subsystem Technology*

S. aureus – Staphylococcus aureus

SaPIs - *Staphylococcal pathogenicity islands* (ilhas de patogenicidade estafilocócicas)

SCC - *Staphylococcal cassette chromosomes* (cassete cromossômico estafilocócico)

SCCmec - Staphylococcal cassette chromosome mec (cassete cromossômico estafilocócico mec)

SNP – *Single nucleotide polymorphism* (polimorfismo de nucleotídeo único)

SSL - *Staphylococcal superantigen-like* (superantígenos estafilocócicos)

ST – S*equence type* (tipo de sequência)

TNF-α - *Tumor necrosis factor alfa* (fator de necrose tumoral alfa)

WGS - *Whole genome sequencing* (sequenciamento de genomas completos)

LISTAS DE FIGURAS E TABELAS

RESUMO

ROCHA, Lis Souza, M.Sc., Universidade Federal de Viçosa, fevereiro de 2018. **Análises *in silico* das proteínas de superfície e secretadas de *Staphylococcus aureus* para discriminar entre as mastites bovina clínica e subclínica.** Orientadora: Andréa de Oliveira Barros Ribon.

Staphylococcus aureus é uma das principais bactérias causadoras da mastite bovina, uma doença predominante nos rebanhos leiteiros. A mastite bovina pode se manifestar de forma clínica, onde os sinais de inflamação são observados, ou de forma subclínica, sem sintomas aparentes, e que frequentemente evolui para a cronicidade. Os genomas de quatro isolados de *S. aureus* causadores de mastite bovina subclínica (170, 302, 1269, 1364) foram sequenciados por nosso grupo e foram contrastados com os genomas de duas cepas isoladas de mastite clínica (*S. aureus* RF122 e *S. aureus* N305) para identificar diferenças que pudessem ser ligadas à mastite subclínica. Não foi possível associar a presença de fatores de virulência ao tipo de manifestação. Mais genes relacionados à produção de enterotoxinas foram encontrados no genoma de *S. aureus* RF122. Os genes que codificam a toxina 1 da síndrome do choque tóxico (*tsst-1*), a variante bovina da enterotoxina C (*secbov*), a enterotoxina t e dois homólogos da streptolisina S de *Streptococcus pyogenes* foram exclusivos da cepa RF122. A presença de polimorfismos de nucleotídeo único (SNPs) nos genomas subclínicos foi averiguada comparando 33 genes que codificam fatores de virulência com seus ortólogos em *S. aureus* RF122. SNPs não sinônimos foram encontrados nos genes *clfa* (fator de aglutinação A), *srta* (sortase A) e *sspa* (protease). Uma proteína transportadora (cl3316) foi identificada *in silico* como exclusiva das cepas subclínicas. Diferentes programas de predição foram utilizados para a identificação das proteínas de superfície e secretadas das seis cepas, as quais foram usadas na construção de um plot de escala multidimensional (MDS). Os resultados mostraram alta similaridade entre as proteínas das cepas estudadas. Pela inspeção visual, identificaram-se dois grupos de ortólogos, cl3309 e cl3700, correspondentes a uma proteína hipotética e a uma lipoproteína, respectivamente, que possuem regiões de maior similaridade entre as clínicas ou subclínicas. Os dados encontrados *in silico* foram validados pela reação em cadeia da polimerase em isolados bacterianos coletados em fazendas leiteiras. Todos os isolados

subclínicos foram corretamente identificados com os *primers* desenhados para as cepas subclínicas, porém existe a necessidade de redesenhar novos *primers* para a correta discriminação dos isolados clínicos.

ABSTRACT

ROCHA, Lis Souza, M.Sc., Universidade Federal de Viçosa, February, 2018**. *In silico* analysis of surface and secreted proteins of *Staphylococcus aureus* to discriminate between subclinical and clinical bovine mastitis.** Adviser: Andréa de Oliveira Barros Ribon.

Staphylococcus aureus is one of the major representative pathogenic bacteria causing bovine mastitis, a predominant disease in dairy cattle worldwide. Bovine mastitis may manifest in a clinical form, where the sings of inflammation are observable, or in a subclinical form, without apparent symptoms, and that frequently evolves to chronicity. The genomes of four *S. aureus* isolates causing subclinical bovine mastitis (170, 302, 1269, 1364) were sequenced by our group and were contrasted to the genomes of two strains isolated from clinical mastitis (*S. aureus* RF122 e *S. aureus* N305) in order to identify differences that could be linked to subclinical mastitis. It was not possible to associate virulence factors to the type of manifestation. More genes related to the production of enterotoxins were found in the genome of *S. aureus* RF122. The genes that code the toxic shock syndrome toxin 1 (*tsst-1*), bovine variant of the staphylococcal enterotoxin C (*secbov*), enterotoxin T, and two homologues of streptolysin S from *Streptococcus pyogenes* were exclusively present in the strain *S. aureus* RF122. The presence of single nucleotide polymorphisms (SNPs) in the subclinical genomas was investigated comparing 33 genes that codify virulence factors against their orthologous in *S. aureus* RF122. Nonsynonymous SNPs were found in the genes of *clfa* (clumping factor A), *srta* (sortase A) and *sspa* (protease). A transporter protein (cl3316) was identified *in silico* as exclusive to the subclinical strains. Different prediction programs were used for the identification of surface and secreted proteins from the six strains and their sequences were used for the construction of a multidimensional scaling (MDS) plot. The results showed high similarity among the proteins. Upon visual inspection, two orthologus groups were identified, cl3309 and cl3700, corresponding to a hypothetical protein and to a lipoprotein, respectively, which possess greatest similarity among clinical or subclinical strains. The data found *in silico* were validated through polymerase chain reaction using DNA from bacterial isolates collected from dairy herds. All subclinical isolates were correctly identified with the primers designed for the

subclinical strains. However, there is still the necessity to redesign primers for the proper discrimination of the clinical isolates.

REVISÃO DE LITERATURA

O leite ocupa a sexta posição entre os produtos mais importantes da agropecuária brasileira (Embrapa Gado de Leite, 2016). De acordo com o Centro de Estudos Avançados em Economia Aplicada (Cepea) e a Confederação da Agricultura e Pecuária do Brasil (CNA), o setor movimentou mais de R$ 63,6 milhões em 2016, empregando diretamente 3,6 milhões de pessoas em mais de um milhão e cem mil propriedades rurais. O Brasil é o quinto maior produtor de leite do mundo, atrás apenas da Índia, Estados Unidos, China e Paquistão. Minas Gerais é o líder da produção nacional, responsável por 26,7% do montante de leite produzido no país (IBGE, 2016).

A produção leiteira é fortemente afetada por doenças que acometem o rebanho, como a mastite, tristeza parasitária bovina, brucelose, tuberculose bovina e febre aftosa. Dentre essas, a mais comum é a mastite, que além de reduzir a qualidade do leite, diminui consideravelmente o lucro do produtor, devido à diminuição na produção, ao alto custo do tratamento, à eliminação de animais e ao descarte do produto pelo risco de contaminação com antibióticos (Geary *et al.,* 2012).

A mastite bovina é uma infecção que se inicia quando o agente etiológico, geralmente bacteriano, invade a glândula mamária causando uma resposta inflamatória. A doença é classificada epidemiologicamente como contagiosa se o reservatório primário é o úbere da vaca infectada ou ambiental quando o reservatório é o ambiente onde ela vive (Cervinkova *et al.,* 2013). Infecções causadas por patógenos ambientais, como *Escherichia coli*, *Klebsiella* spp., *Streptococcus dysgalactiae* e *Streptococcus uberis*, são geralmente clínicas e de duração curta (Smith *et al.,* 2005), e podem ser prevenidas por melhorias na higiene das instalações. Patógenos contagiosos, como *Staphylococcus aureus, Streptococcus agalactiae, Mycoplasma* spp. e *Corynebacterium*, espalham-se de animal para animal e sobrevivem no úbere, podendo estabelecer infecções subclínicas e de longa duração (Radostits *et al.,* 2007, Nickerson, 2011). A mastite contagiosa pode ser prevenida por meio de melhorias da higiene durante o processo de ordenha, além da assepsia pós-ordenha e o tratamento da vaca seca. Frequentemente a infecção torna-se persistente e não responsiva ao tratamento com antibióticos e a segregação ou a eliminação do animal passa a ser a única alternativa de evitar a disseminação da doença pelo rebanho.

Quanto ao grau de severidade, a mastite pode ser classificada como clínica ou subclínica. A mastite clínica aguda é a forma mais severa da doença, sendo caracterizada pela manifestação de vários sintomas repentinos e anormalidades visíveis a olho nu, como inchaço no úbere, endurecimento e dor na glândula mamária, além de redução da produção de leite, que apresenta grumos ou pus. As vacas também podem ser afetadas de forma sistêmica, apresentando febre, desidratação e redução de apetite (Le Maréchal *et al.,* 2011). A forma mais comum de detectar a mastite clínica é por meio do teste da caneca de fundo preto, onde grumos ou sangue são observados no fundo da caneca após a ordenha dos três primeiros jatos de cada teto (Embrapa Gado de Leite, 2008)

A mastite subclínica é caracterizada por uma inflamação sem alterações visíveis no leite ou no úbere. Os animais produzem menos leite e de qualidade inferior e são uma fonte de infecção para as outras vacas no rebanho. A prevalência da mastite subclínica no Brasil chega a 48,64% (Acosta *et al.,* 2016), e o impacto econômico total causado pela mastite chega até a 87,37% (Lopes *et al.,* 2012). Devido à ausência de anormalidades visíveis, a mastite subclínica requer testes diagnósticos específicos, como a contagem de células somáticas (CCS), onde quanto maior a contagem, maior o grau de inflamação. Isso pode ser feito por meio do "California Mastitis Test" (CMT), que se baseia no uso de um surfactante aniônico que rompe a membrana celular das células somáticas presentes no leite e libera o conteúdo intracelular. De acordo com a inspeção visual, o resultado pode ser classificado numa escala de 0 (mistura permanecendo inalterada) a 3 (formação de um gel semi-sólido pela reação do DNA bacteriano com o reagente surfactante), sendo 2 e 3 considerados resultados positivos. Entretanto, o teste apresenta algumas desvantagens, como um elevado limite de detecção, que é acima de 300.000 células/mL e a incapacidade de identificar a espécie bacteriana causadora da mastite (Vinodkumar *et al.,* 2017).

S. aureus é um dos principais causadores da mastite subclínica e está mundialmente distribuído nos rebanhos leiteiros. Embora esse patógeno possa causar a mastite bovina clínica, com prevalência de 0,73% a 2,6% no Brasil, a prevalência da manifestação subclínica é muito maior, podendo chegar até a 48,64% (Acosta *et al.,* 2016). Estima-se que a prevalência da doença no Brasil seja de 20 a 38%, resultando em perdas de 15 a 24% da produção (Milk Point - 2015). A infecção causada por *S. aureus* geralmente é transmitida de úbere a úbere durante o processo de ordenha, seja atráves das máquinas ou pelas mãos

do fazendeiro responsável. Embora esse microorganismo possa sobreviver no ambiente durante algum tempo, ele depende da colonização de animais para garantir sua sobrevivência e multiplicação, o que pode explicar o fato de alguns rebanhos serem livres de Staphylococci (Peton e Le Loir, 2014).

Os mecanismos moleculares envolvidos na patogênese de *S. aureus* têm sido investigados (Ben Zakour *et al.,* 2008, Zadoks *et al.,* 2011), para se buscar uma associação entre cepas bacterianas e as diferentes manifestações clínicas observadas na mastite bovina (Guinane *et al.,* 2010). A comparação entre bactérias isoladas de infecções subclínicas mostrou que aquelas relacionadas a infecções persistentes possuem maior capacidade de produção de biofilme *in vitro* do que as relacionadas a infecções não persistentes (Veh *et al.,* 2015). Esses estudos também revelaram a presença do gene codificador da enterotoxina SEG, uma maior expressão do RNA regulador RNAIII, uma redução da síntese de proteínas de colonização e o aumento das proteínas secretadas nas cepas isoladas de mastite subclínica. Por fim, também foi demonstrado que as cepas com maior expressão do gene *hld*, que codifica a delta-hemolisina, eram mais propensas a serem não persistentes durante ou fora do período de lactação.

O avanço das tecnologias de sequenciamento de genomas completos (*whole genome shotgun sequencing*) tem aumentado o conhecimento sobre as diversidades genética e fisiológica bacterianas (Punina *et al.,* 2015). Análises comparativas revelaram que os genomas de *S. aureus* têm um tamanho de aproximadamente 2,8 Mbp, com conteúdo GC relativamente baixo e regiões genômicas bem conservadas, embora existam diferentes elementos genéticos móveis (MGEs) que conferem alta variabiliade entre as cepas (Chua *et al.,* 2014). Essas regiões variáveis são classificadas como profagos, ilhas de patogenicidade, ilhas genômicas e cassetes cromossômicos estafilocócicos (SCC - Staphylococcal cassette chromosomes) e carregam genes que codificam fatores de virulência e resistência a antibióticos (Baba *et al.,* 2007).

Ilhas de patogenicidade *estafilocócicas (*Staphylococcal pathogenicity islands - SaPIs) são MGEs de 14–17 kb (Malachowa e DeLeo, 2010). Elas formam uma família de genes altamente conservados do core, incluindo fases de leitura aberta (*ORF – open reading frames)* codificantes de proteínas de regulação transcricional, uma região codificante de uma integrase e de uma terminase (Novick e Subedi, 2007, Ubeda *et al.,*

2003). Além dos genes do core, praticamente todas as SAPIs codificam enterotoxinas e toxinas *choque tóxico* estafilocócico (TSST) (Yarwood *et al.,* 2002). SaPIbov2 é uma exceção a essa regra e codifica a proteína de adesão Bap, envolvida nas infecções crônicas de mastite bovina (Carles *et al.,* 2003, Tormo *et al.,* 2005). Estruturalmente, elas estão integradas em um de seis sítios específicos no genoma, sempre na mesma orientação (Novick e Subedi,2007), podendo ser mobilizadas após a infecção por certos bacterióficos estafilocócicos ou por indução por prófagos endógenos (Tormo *et al.,* 2008, Ubeda *et al.,* 2005).

As ilhas genômicas são tipicamente reconhecidas como segmentes discretos de DNA entre cepas relacionadas, e sabe-se que a sua formação contribui para a diversificação e adaptação de microorganismos, tendo assim impactos significativos na plasticidade e evolução do genoma, na disseminação de resistência a antibióticos e na formação de vias metabólicas. (Juhas *et al.,* 2009). Existem três tipos de ilhas genômicas (νSaα, νSaβ, νSaγ) em *S. aureus* (Malachowa e DeLeo, 2010), cada uma polimórfica em seu conteúdo gênico, mas conservada dentro de cada cepa (Hallin *et al.,* 2010, Baba *et al.,* 2007, Holmes *et al.,* 2014). Estruturalmente, elas são flanqueadas por um gene de transposase *upstream* e um tipo de sistema de restrição modificação parcial (RM) *downstream* (Malachowa e DeLeo, 2010). Em νSaα, encontra-se um *cluster* de genes que codificam lipoproteínas (lpl) e genes de *superantígenos estafilocócicos (ssl)* (Lina *et al.,* 2004). νSaβ codifica bacteriocinas, enterotoxinas, hialuronato liase e um cluster de genes que codificam serino-proteases (Baba *et al.,* 2002, Holden *et al.,* Tsuru e Kobayashi, 2008). Por fim, νSaγ contém genes codificantes de modulinas do tipo β solúveis em fenol, e um cluster de genes que codificam *ssl* similares àqueles presentes em νSaα (Gill *et al.,* 2005). Devido à ausência de elementos genéticos típicos necessários ou indicativos de mobilidade, como integrases, excisionases, terminases e sequências repetitivas associadas, as ilhas genômicas têm menor mobilidade que os demais MGEs (Dobrindt *et al.,* 2004, Feng *et al.,* 2008).

Os *SCCs* funcionam como maquinarias de captura de segmentos exógenos de DNA a fim de promover a sobrevivência bacteriana em ambientes tóxico e hostis. O mais representativo de todos os *SCCs, SCCmec* (*staphylococcal cassette chromosome mec*) codifica a resistência ao antibiótico meticilina. A emergência de cepas de estafilococos resistentes a meticilina se deve à aquisição e inserção do elemento *SCCmec* nos

cromossomos de cepas susceptíveis (International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC), 2009). Os *SCCmec* são classificados em seis tipos (de I a VI), com base na sequência de nucleotídeos de duas regiões específicas. Uma região é o complexo *mec*, o qual inclui o gene *mecA*, e a outra é a região do complexo *ccr*, incluindo genes codificantes para duas recombinases (Hiramatsu *et al.,* 2001). Os tipos de *SCCmec* são definidos pela combinação do complexo *ccr,* o qual codifica recombinases, e a classe do complexo *mec, o* qual codifica genes reguladores e sequências de inserção. Esses são os elementos chave do cassete responsáveis pela integração e excisão do *SCCmec* e pelo fenótipo de resistência aos beta-lactâmicos (International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC), 2009).

A primeira cepa de *S. aureus* de origem bovina a ter seu genoma sequenciado foi *S. aureus* RF122, causador da mastite clínica severa (Herron-Olson *et al.,* 2007). Por meio de genômica comparativa foram identificados MGEs únicos, como a pequena ilha genômica vSaBov, de 6,351 bp, que codifica 14 genes inéditos em *S. aureus*, inlcuindo dois homólogos de estreptolisinas, as leucotoxinas potentes do locus *Sag*, originalmente identificadas em Streptococcus *pyogenes*. Dois fagos únicos a essa cepa também foram encontrados: os fagos φSaBov-νSaβ e φ12Bov. O φSaBov-νSaβ, carrega oito genes únicos, além de transposases e diversos genes transportadores de oligopeptídeos, localizados em uma região previamente identificada como variável em diversos isolados de *S. aureus.* Já o fago φ12Bov, menor em tamanho, também carrega genes únicos além de elementos homólogos a outros patógenos gram-positivos.

Também foi observada grande variação alélica entre os genes de virulência presentes em cepas isoladas de infecções humanas e bovinas. Fatores de virulência bem conhecidos, como *spa*, *clfA*, *sdrC* e *ebh*, códons de parada prematuros sendo considerados pseudogenes em *S. aureus* RF122. Os autores sugeriram que a degeneração gênica representa um processo atual na evolução de isolados bovinos bem sucedidos, e que os clones bovinos possam estar em processo de transição para um estilo de vida intracelular. Outro estudo comparativo entre isolados bovinos e humanos encontrou 16 ORFs exclusivas da cepa bovina *S. aureus* RF122, embora nenhuma tenha sido encontrada em todas as 51

estirpes bovinas analisadas. Esses fatores pertencem a categorias diversas, como proteínas hipotéticas, ilhas de patogenicidade e reguladores transcricionais (Kozytska *et al.,* 2010).

O sequenciamento do genoma da cepa *S. aureus* Newbould 305 (N305), isolada de mastite clínica com manifestação crônica e branda, revelou menos genes codificadores de toxinas, porém mais fatores de invasão e colonização quando comparado ao genoma de *S. aureus* RF122, uma possível explicação para a maior dificuldade de detecção e cura da infecção causada por N305 (Peton *et al.,* 2014), visto que esses fatores podem culminar na melhor apatação do patógeno organismo do hospedeiro. O maior número de genes de toxinas em *S. aureus* RF122 explica sua maior propensão à indução de respostas inflamatórias mais graves. Alguns genes foram encontrados exclusivamente em *S. aureus* N305, como aqueles que codificam a proteína de efeito citotóxico delta-hemolisina (Hld), e proteínas de colonização como a variante bovina da proteína ligante de von Willebrand (vWbSbo5), proteína B de ligação a fibronectina (FnbB), proteína de superfície G. de *S. aureus* (SasG) e a 1-acil-sn-glicerol-3-fosfato aciltransferase (PlsC).

A análise comparativa entre os proteomas de *S. aures* N305 e RF122 (Peton *et al.,* 2014) revelou que existem exoproteases comuns às duas cepas, embora algumas delas, como a Spl protease, sejam encontradas exclusivamente na primeira. Também foram encontrados fatores de virulência exclusivos de *S. aureus* N305, como hemolisinas e leucotoxinas (alfa e gama-hemolisina, LukS/F), além de proteínas envolvidas na adesão a tecidos hospedeiros (Newbould305_1324, codificante de uma proteína da matriz extracelular e proteína de ligação ao plasma) e na evasão da resposta imune do hospedeiro (Spa, vWbpSbo5).

Proteínas imunogênicas e fatores de virulência produzidos por duas cepas de *S. aureus* isoladas de ovelhas com infecção clínica gangrenosa (O11) e de subclínica (046) foram contrastadas (Le Marechal *et al.,* 2011). Identificaram-se 89 proteínas imunorreativas, das quais 74 eram proteínas do soroproteoma principal, capazes de induzir uma resposta imune no hospedeiro em infecções causadas por ambas as cepas. As demais 15 proteínas eram do soropretoma acessório, sendo 11 imunogênicas apenas nas infecções causadas pela cepa gangrenosa, sendo 7 fatores de virulência (Sbi, SspB, SspA, Aur, IsdH, Opp1A e VWbp) e 2 fatores de resposta ao stress (AhpF, TrxB). As 3 demais proteínas do soroproteoma acessório eram imunogênicas apenas nas infecções causadas pela cepa

subclínica, sendo elas a transglicosilase IsaA e duas proteínas hipotéticas. No mesmo estudo, as cepas isoladas de mastite gangrenosa ou subclínica foram usadas para induzir mastite experimental em ovelhas. Os resultados mostraram que a cepa subclínica induziu maiores níveis de Interleucina 1-beta (IL-1β) e de Fatores de Necrose Tumoral Alfa (TNF-α) nos lisados da glândula mamária de ovelhas que aumentam a capacidade de eliminação de *S. aureus* intracelular no endotélio de células bovinas (Oviedo-Boyso *et al.,* 2008), o que possivelmente aumentaria a eficiência da resposta imune e eliminação da cepa subclínica (Le Marechal *et al.,* 2011).

Análises comparativas entre grupos de cepas de *S. aureus* isoladas da corrente sanguínea (bacteremia) não-persistente e persistente mostram que há características associadas especificamente ao grupo de cepas persistentes que incluem maior crescimento sob condições de escassez de nutrientes, maior tolerância à toxicidade com ferro e maior expressão de proteínas de superfície celular envolvidas na evasão do sistema imune e na resposta ao stress (Richards *et al.,* 2015). Além disso, análises de genomas completos sequenciados mostram diferentes polimorfismos de nucleotídeos simples no gene *mprF* de todos os isolados com aspectos de persistência adaptativa. Esses genes estão relacionados à sintese e translocação de lisil-fosfatidilglicerol (L-PG) para a membrana celula o que confere carga negativa à superfície celular bacteriana. Propõe-se que o aumento da atividade de *mprF* resulte em maior abundância de L-PG na membrana celular e consequentemente na alteração da sua carga, levando ao aumento da resistência a agentes antimicrobianos catiônicos, como peptídeos de defesa do hospedeiro e daptomicina acoplada a cálcio (Mishra *et al.,* 2013, Peleg *et al.,* 2012, Rubio *et al.,* 2012, Yang *et al.,* 2013, Mishra *et al.,* 2014).

As proteínas de superfície e as proteínas secretadas desempenham um papel crucial na patogenicidade bacteriana, por desempenharem o contato inicial com o sistema imune do hospedeiro (Springer, 1990). Em *S. aureus*, identificam-se quatro grupos de proteínas secretadas, a saber, peptideoglicano hidrolases, enzimas intracelulares (proteases, lipases, etc.), toxinas de membrana (hemolisinas e leucocidinas) e superantígenos (Priatkin e Kuz'menko, 2010). As proteínas de superfície estão em contato direto com o hospedeiro, sendo assim mais susceptíveis à pressão evolutiva para o sucesso nas funções relacionadas

à colonização de tecidos e evasão do sistema imune. As proteínas CWA, covalentemente ancoradas à proteína parede celular (cell wall anchored proteins), estão presentes na superfície de *S. aureus* e são fundamentais ao seu sucesso como bactéria comensal e como patógeno. O repertório de proteínas CWA em *S. aureus* é limitado, o que faz com que muitas delas tenham evoluído de forma a desempenhar diversas funções. Ademais, essas proteínas apresentam redundância fucional (Foster *et al.,* 2014). Ademais, a expressão das proteínas pode ser alterada por condições de crescimento (Mazmanian, *et al.,* 2003, Hammer e Skaar, 2011).

Análises estruturais e funcionais identificaram quatro classes distintas de proteínas CWA em *S. aureus*. Proteínas contendo o motivo transportador de Ferro proximal (NEAT) estão envolvidas na captura do grupo heme da hemoglobina, auxiliando a sobrevivência da bactéria no hospedeiro, onde o mineral ferro é escasso. A proteína A, que consiste em agrupamentos de repetições em tandem de três hélices, é uma proteína CWA multifuncional, ubíqua em *S. aureus*, comumente usada na tipagem de cepas baseada na variação da sequência de DNA codificante da região Xr, que por sua vez é composta por repetições de octapepetídeos altamente variável em tamanho. A família de proteínas contendo repetições G5-E adota uma forma de hélice-β tripla sem função de ligação conhecida. Um exemplo é a proteína de superfície G de *S. aureus* (SasG), similar à proteína de acumulação (Aap), necessária à formação de biofilme em *S. epidermidis*. Por último, a maior família de proteínas CWA, conhecidas como moléculas adesivas da matriz (MSCRAMMs), são caracterizadas por pelo menos dois domínios adjacentes dobrados na forma de IgG na porção amino-terminal A. Essas proteínas desempenham um papel crucial na adesão e invasão de células e tecidos de hospedeiros, formação de biofilme, além de interferirem nas respostas imunes inata e adaptativa (Foster *et al.,* 2014). Como exemplo, as proteínas de superfície ClfA e SdrE são capazes de recrutar reguladores do sistema do complemento e destruírem o componente C3b (Hair *et al.,* 2008, Hair *et al.,* 2010, Sharp, J. A. *et al.,* 2012).

A genômica comparativa tem deixado de focar em meras montagens de genomas por referência e evoluído para análises mais complexas, como a comparação de montagens de genomas individuais (Krzywinski *et al.,* 2009). Enquanto análises de espécies distintas fornecem informações sobre uma espécie de forma a permitir o entendimento de

mecanismos biológicos de uma segunda espécie, os métodos comparativos têm sido agora usados para desvendar diferenças entre indivíduos da mesma espécie, e como essas diferenças afetam as respostas ao ambiente, como susceptibilidade a doenças e resposta a terapias (Krzywinski *et al.,* 2009).

Em trabalho previamente desenvolvido por nosso grupo de pesquisa (Silva *et al.,* 2016), foram anunciadas as sequências de quatro genomas isolados a partir de animais infectados pela mastite subclínica. Definiram-se SNPs em genes de virulência e genes do metabolismo que poderiam ter relação com a persistência da doença. Porém, não foram realizadas análises comparativas das proteínas secretadas e de superfície das bactérias a fim de se investigar sua relação com os tipos de manifestações apresentados pelos animais.

No presente trabalho, a hipótese testada é se variações nas sequências de proteínas de diferentes cepas de *S. aureus* podem ser importantes para a determinação do tipo de manifestação de mastite bovina. Assim, a partir da análise das regiões codificantes dessas proteínas com padrões distintos de conservação entre os genomas clínicos e entre os genomas subclínicos, foram desenhados potenciais marcadores para a mastite clínica e para a mastite subclínica.

REFERÊNCIAS BIBLIOGRÁFICAS

Acosta, A.C., Silva, L.B.G.D., Medeiros, E.S., Pinheiro-Júnior, J.W. and Mota, R.A., 2016. Mastitis in ruminants in Brazil. Pesquisa Veterinária Brasileira, 36:565-573.

Baba T, Bae T, Schneewind O, Takeuchi F, Hiramatsu K. 2008. Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. Journal of Bacteriology, 190:300-310.

Carles, U., Ma Angeles, T., Carme, C., Pilar, T., Timothy, J.F., Inigo, L. and Jose, R.P. 2003. Sip, an integrase protein with excision, circularization and integration activities, defines a new family of mobile *Staphylococcus aureus* pathogenicity islands. Molecular Microbiology, 49:193–210.

Castelani, L., Santos, A.F., Miranda, M., dos S, Zafalon, L.F., Pozzi, C.R. and Arcaro, J.R. 2013. Molecular typing of mastitis causing *Staphylococcus aureus* isolated from heifers and cows. International Journal of Molecular Sciences, 14:4326-4333.

Cervinkova, D., Vlkova, H., Borodacova, I., Makovcova, J., Babak, V., Lorencova, A., Vrtkova, I., Marosevic, D. and Jaglic, Z. 2013. Prevalence of mastitis pathogens in milk from clinically healthy cows. Veterinary Medicine, 58:567–75.

Cheng, A.G., Kim, H.K., Burts, M.L., Krausz, T., Schneewind, O. and Missiakas, D.M. 2009. Genetic requirements for *Staphylococcus aureus* abscess formation and persistence in host tissues. FASEB Journal, 23:3393–3404.

Chua, K.Y.L., Howden, B.P., Jiang J.-H., Stinear, T. and Peleg, A.Y. 2014. Population genetics and the evolution of virulence in *Staphylococcus aureus*. Infection, Genetics and Evolution, 21:554-562.

Collins, F.S., Green, E.D., Guttmacher, A.E. and Guyer, M.S. 2003. A vision for the future of genomics research. Nature, 422:835–847.

Dobrindt, U., Hochhut, B., Hentschel, U. and Hacker, J. 2004. Genomic islands in

pathogenic and environmental microorganisms. Nature Reviews Microbiology, 2:414–424.

Feng, Y., Chen, C.J., Su, L.H., Hu, S., Yu, J. and Chiu, C.H. 2008. Evolution and pathogenesis of *Staphylococcus aureus*: lessons learned from genotyping and comparative genomics. FEMS Microbiology Reviews, 32:23–37.

Geary, U., Lopez-Villalobos, N., Begley, N., McCoy, F., O'Brien, B. and O'Grady, L. 2012. Estimating the effect of mastitis on the profitability of Irish dairy farms. Journal of Dairy Science, 95:3662–3673.

Gill, S.R., Fouts, D.E., Archer, G.L., Mongodin, E.F., DeBoy, R.T. and Ravel, J. 2005. Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilmproducing methicillin-resistant *Staphylococcus epidermidis* strain. Journal of Bacteriology, 187:2426–2438.

González, R.N. and Wilson, D.J. 2003. Mycoplasmal mastitis in dairy herds. Veterinary Clinics of North America: Food animal practice, 19:199-221.

Hair, P. S.,Ward, M. D., Semmes, O. J., Foster, T. J. and Cunnion, K. M. 2010. *Staphylococcus aureus* clumping factor A binds to complement regulator factor I and increases factor I cleavage of C3b. Journal of Infectious Diseases, 198:125–133.

Hair, P.S., Echague, C.G., Sholl, A.M., Watkins, J.A. and Geoghegan, J.A. 2010. Clumping factor: A interaction with complement factor I increases C3b cleavage on the bacterial surface of *Staphylococcus aureus* and decreases complement-mediated phagocytosis. Infection and Immunity, 78:1717–1727.

Hallin, M., De Mendonca, R., Denis, O., Lefort, A., El Garch, F. and Butaye, P. 2011. Diversity of accessory genome of human and livestock-associated ST398 methicillin resistant *Staphylococcus aureus* strains. Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases, 11:290–299.

Hammer, N.D. and Skaar, E.P. 2011. Molecular mechanisms of *Staphylococcus aureus* iron

acquisition. Annual Review of Microbiology, 65:129–147.

Hanssen, A.-M., Ericson, S. and Johanna, U. 2006. SCCmec in staphylococci: genes on the move. FEMS Immunology & Medical Microbiology, 46:8–20.

Hardison, R. C. 2003. Comparative Genomics. PLoS Biology, 1:e58.

Herron-Olson, L., Fitzgerald, J.R., Musser, J.M. and Kapur, V. 2007. Molecular correlates of host specialization in *Staphylococcus aureus*. PLoS One, 2:e1120.

Hiramatsu, K., Cui, L., Kuroda, M. and Ito, T. 2001. The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. Trends in Microbiology, 9:486–493.

Holmes, A., McAllister, G., McAdam, P.R., Hsien Choi, S., Girvan, K. and Robb, A. 2014. Genome-wide single nucleotide polymorphism-based assay for high-resolution epidemiological analysis of the methicillin-resistant *Staphylococcus aureus* hospital clone EMRSA-15. Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 20:124–131.

International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (IWG-SCC). 2009. Classification of staphylococcal cassette chromosome mec (SCCmec): guidelines for reporting novel SCCmec elements. Antimicrobiology Agents Chemotherapy, 53:4961–4967.

Josefsson, E., Hartford, O., O'Brien, L., Patti, J.M. and Foster, T. 2001. Protection against experimental *Staphylococcus aureus* arthritis by vaccination with clumping factor A, a novel virulence determinant. Journal of Infectious Diseases, 184:1572–1580.

Juhas, M., Van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W. and Crook, D.W. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiology Reviews, 33:376–93.

Käll, L., Krogh, A. and Sonnhammer, E.L. 2007. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Research, 35:429-32.

Kozytska, S., Stauß, D., Pawlik, M.C., Hensen, S., Eckart, M., Ziebuhr, W., Ohlsen, K.

2010. Identification of specific genes in *Staphylococcus aureus* strains associated with bovine mastitis. Veterinary Microbiology, 145:360-365.

Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology, 305:567-580.

Krzywinski, M., Schein, J., Birol, I. and Connors, J. 2010. Circos: An information aesthetic for comparative genomics. Genome Research, 19:1639-1645.

Kuzma, K., Malinowski, E., Lassa, H. and Kłossowska, A. 2003. Specific detection of *Staphylococcus aureus* by PCR in intramammary infection. Bulletin of the Veterinary Institute in Pulawy, 47:183-190.

Law, C. J., Maloney, P. C. and Wang D.-N. 2008. Ins and outs of major facilitator superfamily antiporters. Annual Review of Microbiology, 62:289–305.

Le Maréchal, C., Jardin, J., Jan, G., Even, S., Pulido, C., Guibert, J.-M., Hernandez, D., François, P., Schrenzel, J., Demon, D., Meyer, E., Berkova, N., Thiéry, R., Vautor, E. and Le Loir, Y. 2011. *Staphylococcus aureus* seroproteomes discriminate ruminant isolates causing mild or severe mastitis. Veterinary Research, 42:35-10.

Lina, G., Bohach, G. A., Nair, S.P., Hiramatsu, K., JouvinMarche, E. and Mariuzza, R. 2004. Standard nomenclature for the superantigens expressed by *Staphylococcus*. Journal of Infectious Diseases, 189:2334–2336.

Lopes, M.A., Demeu, F.A., Da Rocha, C.M.B.M., Da Costa, G.M., Neto, A.F. and Dos Santos, G., 2012. Avaliação do impacto econômico da mastite em rebanhos bovinos leiteiros. Arquivos do Instituto Biológico, 79:477-483.

Malachowa, N. and DeLeo, F.R. 2010. Mobile genetic elements of *Staphylococcus aureus*. Cellular and Molecular Life Sciences, 67:3057–71.

Mazmanian, S. K., Skaar, E. P., Gasper, A. H., Humayun, M., Gornicki, P., Jelenska, J., Joachimiak, A., Missiakas, D. M. and Schneewind O. 2003. Passage of heme–iron across the envelope of *Staphylococcus aureus*. Science, 299:906–909.

McCarthy, A. J., and Lindsay, J. A. 2010. Genetic variation in *Staphylococcus aureus* surface and immune evasion genes is lineage associated: implications for vaccine design and host–pathogen interactions. BMC Microbiology, 10:173.

Mishra, N.N., Bayer, A.S., Weidenmaier, C., Grau, T., Wanner, S., Stefani, S., Cafiso, V., Bertuccio, T., Yeaman, M.R., Nast, C.C. and Yang, .S-J. 2014. Phenotypic and genotypic characterization of daptomycin-resistant methicillin-resistant *Staphylococcus aureus* strains: relative roles of mprF and dlt operons. PLoS One, 9:e107426.

Mishra, N.N., Yang, S.-J., Chen, L., Muller, C., Saleh-Mghir, A., Kuhn, S., Peschel, A., Yeaman, M.R., Nast, C.C., Kreiswirth, B.N., Crémieux, A.-C. and Bayer, AS. 2013. Emergence of daptomycin resistance in daptomycin-naïve rabbits with methicillin-resistant *Staphylococcus aureus* prosthetic joint infection is associated with resistance to host defense cationic peptides and mprF polymorphisms. PLoS One, 8:e71151

Moon, B.Y., Park, J.Y., Robinson, D.A., Thoma,s J.C., Park, Y.H., Thornton, J.A. and Seo, K.S. 2016. Mobilization of genomic islands of *Staphylococcus aureus* by temperate bacteriophage. PLoS One, 11:e0151409.

Nakai, K. and Kanehisa, M. 1991. Expert system for predicting protein localization sites in Gram-negative bacteria. Proteins: Structure, Function, and Genetics, 11:95-110.

Nakai, K. and Horton, P. 1999. PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization. Trends in Biochemical Sciences, 24:34-35.

Newbould, F. H. S. 1968. Epizootiology of mastitis due to *Staphylococcus aureus*. Journal of the American Veterinary Medical Association, 153:1683-1687.

Nickerson, E.K., Wuthiekanun, V. and Kumar, V. 2011. Emergence of community-associated methicillin-resistant *Staphylococcus aureus* carriage in children in Cambodia. American Journal of Tropical Medicine and Hygiene, 84:313-317.

Novick, R. and Subedi, A. 2007. The SaPIs: mobile pathogenicity islands of *Staphylococcus*. Chemical Immunology and Allergy 93:42–57.

Oviedo-Boyso, J., Cardoso-Correa, B.I., Cajero-Juarez, M., Bravo-Patino, A., ValdezAlarcon, J.J. and Baizabal-Aguirre, V.M. 2008. The capacity of bovine endothelial cells to eliminate intracellular *Staphylococcus aureus* and *Staphylococcus epidermidis* is increased by the proinflammatory cytokines TNF-alpha and IL-1beta. FEMS Immunology and Medical Microbiology, 54:53-59.

Peleg, A.Y., Miyakis, S., Ward, D.V., Earl, A.M., Rubi, O. A., Cameron, D.R., Pillai, S., Moellering, R.C. and Eliopoulos G.M. 2012. Whole genome characterization of the mechanisms of daptomycin resistance in clinical and laboratory derived isolates of *Staphylococcus aureus*. PLoS One, 7:e28316.

Petersen, T., Brunak, S., Heijne, G. and Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods, 8:785-786.

Peton, V. and Le Loir Y. 2014. *Staphylococcus aureus* in veterinary medicine. Infection, Genetics and Evolution, 21: 602–615.

Peton, V., Bouchard, D.S., Almeida, S., Rault, L., Falentin ,H., Jardin, J., Jan G., Hernandez, D., François, P., Schrenzel, J., Azevedo, V., Miyoshi, A., Berkova, N., Even, S. and Le Loir, Y. 2014. Fine-tuned characterization of *Staphylococcus aureus* Newbould 305, a strain associated with mild and chronic mastitis in bovines. Veterinary Research, 45:106-120.

Priatkin R.G. and Kuz'menko O.M. 2010. Secreted proteins of *Staphylococcus aureus*. Zhurnal Mikrobiologii, Epidemiologii, i Immunobiologii 4:118-24.

Radostits, O.M., Gay, C.C., Hinchcliff, K.W. and Constable, P.D. 2007. Veterinary Medicine: A text book of the disease of cattle, horses, sheep, pigs and goats. The Canadian Veterinary Journal, 2065p.

Rahman, O., Cummings, S.P. and Harrington, D.J., Sutcliffe I.C. 2008. Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. World Journal of Microbiology and Biotechnology, 24:2377-2382.

Renier, S., Micheau, P., Talon, R., Hebraud, M. and Desvaux, M. 2012. Subcellular

localization of extracytoplasmic proteins in monoderm bacteria: rational secretomics-based strategy for genomic and proteomic analyses. PLoS ONE, 7:e42982.

Richards, R.L., Haigh, R.D., Pascoe, B., Sheppard, S. K., Priced, F., Jenkinsd, D., Rajakumara,K. and Morrisseya, J. A. 2015. Persistent *Staphylococcus aureus* isolates from two independent cases of bacteremia display increased bacterial fitness and novel immune evasion phenotypes. Infection and Immunity, 83:3311–3324.

Rubio, A., Moore, J., Varoglu, M., Conrad, M., Chu, M., Shaw, W. and Silverman, J.A. 2012. LC-MS/MS characterization of phospholipid content in daptomycin-susceptible and -resistant isolates of *Staphylococcus aureus* with mutations in mprF. Molecular Membrane Biology, 29:1–8.

Sasaki, T., Tsubakishita, S., Tanaka, Y., Sakusabe, A., Ohtsuka, M., Hirotaki, S. and Hiramatsu, K. 2010. Multiplex-PCR Method for Species Identification of Coagulase-Positive Staphylococci. Journal of Clinical Microbiology, 48:765–769.

Seegers, H., Fouricho, C. and Beaudeau, F. 2003. Production effects related to mastitis and mastitis economics in dairy cattle herds. Veterinary record, 34:475–91.

Sharp, J.A., Echague ,C.G., Hair, P.S., Ward, M.D., Nyalwidhe, J.O., Geoghegan, J.A, Foster, T.J. and Cunnion, K.M. 2012. *Staphylococcus aureus* surface protein SdrE binds complement regulator factor H as an immune evasion tactic. PLoS ONE, 7:e38407.

Shim, E.H., Shanks, R.D. and Morin, D.E. 2004. Milk loss and treatment costs associated with two treatment protocols for clinical mastitis in dairy cows. Journal of Dairy Science, 87: 2702-2708.

Silva, D.M., da Silva, M.P., Vidigal, P.M.P., Barcelos, R.M., Klein, R.C., Aguilar, A.P., Fabres-Klein, M.H., Oliveira, G. and Ribon, A.O.B. 2016. Draft genome sequences of *Staphylococcus aureus* strains isolated from subclinical bovine mastitis in Brazil. Genome Announcements, 4: e00350-16.

Smith, K.L., Todhunter, D.A. and Schoenberger, P.S. 1985. Environmental mastitis: cause, prevalence, prevention. Journal of Dairy Science, 68:1531–53.

Springer, T.A., 1990. Adhesion receptors of the immune system. Nature, 346:425-434.

Swofford, D., Olsen, G., Waddell, P. and Hillis, D. 1996. Phylogenetic Inference. Molecular Systematics, 2:407-511.

Tormo, M.A., Ferrer, M.D., Maiques, E., Ubeda, C., Selva, L,. Lasa, I., Calvete, J.J., Novick, R.P., and Penades J.R. 2008. *Staphylococcus aureus* pathogenicity island DNA is packaged in particles composed of phage proteins. Journal of Bacteriology, 190:2434–2440.

Tormo, M.A., Knecht, E., Gotz, F., Lasa, I. and Penades, J.R. 2005. Bapdependent biofilm formation by pathogenic species of *Staphylococcus*: evidence of horizontal gene transfer? Microbiology, 151:2465–2475.

Tsuru, T. and Kobayashi, I. 2008. Multiple genome comparison within a bacterial species reveals a unit of evolution spanning two adjacent genes in a tandem paralog cluster. Molecular Biology and Evolution, 25:2457–2473.

Ubeda, C. 2005. Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in Staphylococci. Molecular Microbiology, 56:836–844.

Ubeda, C., Tormo, M.A., Cucarella, C., Trotonda, P., Foster, T.J., Lasa, I. And Penades, J.R. 2003. Sip, an integrase protein with excision, circularization and integration activities, defines a new family of mobile *Staphylococcus aureus* pathogenicity islands. Molecular Microbiology, 49:193–210.

Vinodkumar, K., Tresamol, P.V., Deepa, P.M. and Latha, C., 2017. Comparison of Efficacy of Three Cow Side Tests for the Diagnosis of Subclinical Mastitis in Dairy Cattle. Imperial Journal of Interdisciplinary Research, 3:417-418.

Yang, S.-J., Mishra, N.N., Rubio, A. and Bayer and A.S. 2013. Causal role of single nucleotide polymorphisms within the mprF gene of *Staphylococcus aureus* in daptomycin resistance. Antimicrob Agents Chemotherapy, 57:5658–5664.

Yarwood, J.M., McCormick, J.K., Paustian, M.L., Orwin, P.M., Kapur, V. and Schlievert

P.M. 2002. Characterization and expression analysis of *Staphylococcus aureus* pathogenicity island 3. The Journal of Biological Chemistry, 277:13138–13147.

OBJETIVOS

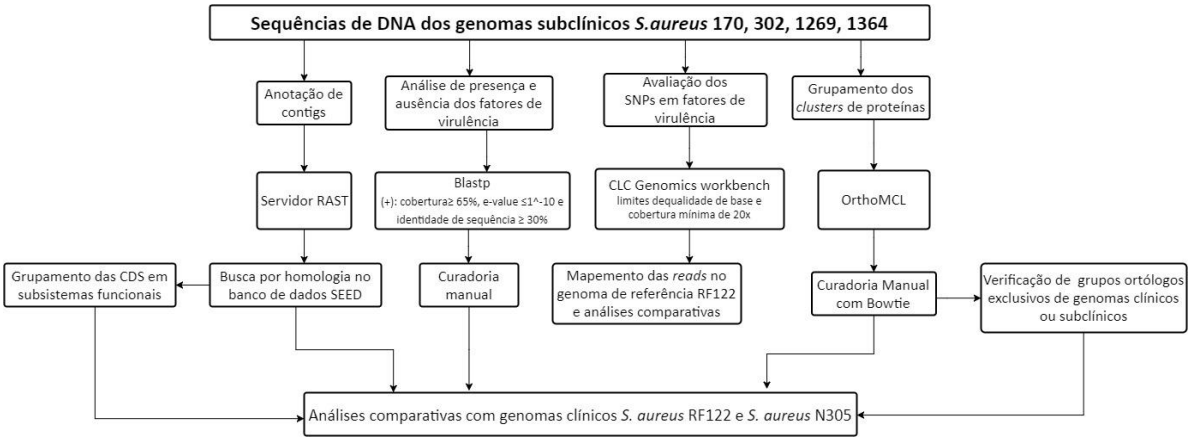
Objetivo Geral

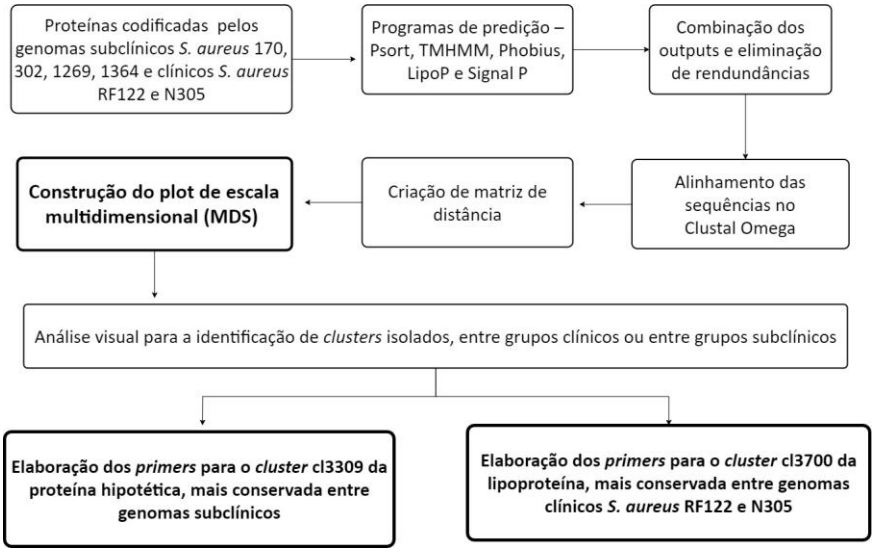
O objetivo geral deste trabalho foi contrastar o repertório de proteínas de superfície e secretadas produzido por cepas de *Staphylococcus aureus* associadas à mastite clínica ou subclínica.

Objetivos Específicos

- Identificar fatores de virulência presentes nos genomas das cepas *S. aureus* 302, 170, 1269 e 1364;
- Contrastar o repertório de proteínas de superfície e secretadas codificado pelos isolados de *S. aureus* relacionados às manifestações subclínica e clínica;
- Identificar diferenças no conteúdo gênico ou na sequência de nucleotídeos dos genes que codificam as proteínas;
- Avaliar por meio de reações da polimerase em cadeia as diferenças encontradas *in silico*.

DELINEAMENTO EXPERIMENTAL





In silico analysis of surface and secreted proteins of *Staphylococcus aureus* to discriminate between subclinical and clinical bovine mastitis

INTRODUCTION

Bovine mastitis is the most common and expensive disease in dairy cattle worldwide. It results in milk contamination by bacteria and their toxins, besides reduction of milk production and its reproof due to the presence of antibiotic residues (Vrieling et al., 2016). In addition, mastitis may also cause cow's death or require the slaughtering of chronically infected cows (Rollin *et al.,* 2015).

Clinical and subclinical mastitis occur in the dairy herd although higher incidence of the later is the normal scenario (Argaw *et al*., 2016). *Staphylococcus aureus* is one of the most representative pathogenic bacteria causing bovine mastitis and widely distributed in dairy cattle herds in several countries. The subclinical mastitis is characterized by lack of symptoms with no apparent changes in milk except for a drop in milk quality and quantity (Shim *et al.,* 2004) that is not always perceived by farmers. Due to its silent development, subclinical manifestations frequently evolve to a chronic infection. It has commonly been assumed that *S. aureus* strains associated to chronic infections are different from those that cause transient infections (Schukken *et al.,* 2011), and are more likely to be transmitted and to prevail through the herd due to the absence of symptoms (Sakwinska *et al.,* 2011; Whelehan *et al.,* 2011). Mastitis caused by *S. aureus* is usually persistent, antibiotic resistant and can easily reoccur. Therefore, the only way to avoid the dissemination to the entire herd is to segregate or to eliminate the animal. Hence, culling of the animals at the subclinical stage is imperative to disease control.

Previous studies have tried to associate virulence factors present in *S. aureus* to the clinical outcome of bovine mastitis (Zecconi *et al.,* 2006, Haveri *et al.,* 2007, Ote *et al.,* 2011, Piccinini *et al.,* 2012) aiming to distinguish virulent isolates from the less harmful ones. The gene *seg*, for example, has been associated with a decreased likelihood of bacteria causing intramammary infections during lactation in dairy cows (Veh *et al.,* 2015).

Surface and secreted proteins comprise very important virulence factors expressed by *S. aureus* that play a pivotal role in bacterial adhesion and immune system evasion. Cell

wall anchored (CWA) proteins are attached to the peptidoglycan by covalent bonds, and participate in the process of commensalism and pathogenicity (Foster *et al.,* 2014). Toxins and exoenzymes are secreted virulence factors that also disarm host immunity and aid in bacterial dissemination and disease progression (Lin and Peterson, 2010). Previous reports have looked for differences in the surface subproteomes of *S. aureus* strains of bovine and human to define therapeutic targets and vaccine antigens or to map differences among strains (Vytvyska *et al.,* 2002; Taverna *et al.,* 2007). Serological proteome analysis was able to identify two secreted proteins differentially expressed in strains of *S. aureus* isolated from ewes with clinical or subclinical mastitis that still has to be tested as marker for acuteness of the disease (Le Marechal *et al.,* 2011). Although there are undeniable advantages in the use of these tools, the current costs of analysis are still a major limitation for many laboratories.

Previously, we have sequenced four *S. aureus* genomes associated with subclinical mastitis, three of which belonged to ST 126 (Silva *et al.,* 2013), a prevalent clone found in herds in southern Brazil, also reported as significantly associated with milk in a study that analyzed intercontinental bovine *S. aureus* isolates (Smith *et al.,* 2005). The general features of the genomes were described, and no further analyses were done. Here we explored the surface and secreted proteins present in the bovine genomes sequenced so far. Given their role in pathogenicity and considering the fate of the infection depends on their interaction with the host organism, we hypothesized that different strains of *S. aureus* presented proteins that could be important to determinate the outcome of bovine mastitis.

MATERIALS AND METHODS

Microorganisms and media used

The isolation and identification of the four isolates of *S. aureus* (Sau 170, Sau 302, Sau 1269, and Sau 1364) causing subclinical mastitis was performed according to the methodology described by Brito *et al.*, (1999), with modifications. A 10 µL aliquot of the milk sample was plated with on a Petri dish containing TSA agar (Tryptic Soy Agar, HiMedia, Mumbai, India), enriched wih 5% defibrinated sheep blood. The plates were kept at 37°C and records were taken after 24h and 48h of incubation. The samples were

considered positive if it were detected the growth of ≥ 2 morphologically identical colonies per plate, even though samples with predominance of three or more morphologically distinct species were considered contaminated. The colonies grown in blood agar plates were analyzed under the parameters of morphology, size, pigmentation and presence of hemolysis. Isolated colonies suspected to be *S. aureus* or *S. agalactiae* were transferred to plates containing BHI agar (Brain Heart Infusion, BHI HiMedia, Mumbai, India) and incubated at 37°C for 24 h. After incubation, the bacteria were identified according to the differential Gram-staining reaction and to the catalase test. The identity of *S. aureus* colonies was confirmed according to the tube coagulase test, acetoin production and amplification of *nuc* gene (Sasaki et al., 2010).

Functional classification, comparative analysis and identification of orthologous proteins

The genomes of the isolates Sau 302, Sau 1364, Sau 170, and Sau 1269 were sequenced on an Ion Torrent Personal Genome Machine (PGM) and contigs were assembled and deposited at NCBI under the accession numbers LNOQ00000000, LNOR00000000, LNOO000000000, and LNOP00000000, respectively, as described previously (Silva *et al.,* 2016).

Prodigal version 2.50 (Hyatt *et al.,* 2010) was used for gene prediction. The protein sets were functionally annotated using BLAST searches (http://Blast.ncbi.nlm.nih.gov/), allowing the proteins to be grouped into Cluster of Orthologous Groups (COG) families (Tatusov *et al.,* 2001). The contigs were also submitted to automatic annotation in the RAST server (Rapid Annotation using Subsystem Technology), through searches for homology in the SEED databank (Overbeek *et al.,* 2005), and contrasted to the already annotated contigs of the clinical strains *S. aureus* RF122 (RF122) (Herron-Olson *et al.,* 2007) and *S. aureus* Newbould 305 (N305) (Peton *et al.,* 2014).

The genomes of the four subclinical strains and of the clinical strains RF122 and N305, deposited at NCBI under the accession numbers NC_007622 and AKYW00000000, respectively, were used for comparative analysis. The genome-scale algorithm OrthoMCL (Li *et al.,* 2003) was used to cluster the protein sequences into orthologous clusters. The

program Bowtie2/2.2.8 (Langmead and Salzberg, 2012) was used to align short sequencing reads to the RF122 and N305 genomes in order to check authenticity of the sequenced genomes and assembly errors. The Bowtie-build tool was used to convert the FASTA sequences to Bowtie indexes. Analyses were run on a Windows 10 Server 64 bits using the terminal emulator PuTTY and submitted to the cluster computer system of the Universidade Federal de Viçosa.

Virulence factors of bovine *Staphylococcus aureus*

The repertoire of virulence factors present in the subclinical strains were defined based on the virulence factors previously listed in the *S. aureus* RF122 genome (Herron-Olson *et al.,* 2007), complemented by other virulence-associated genes (Holtfreter *et al.,* 2010; McCarthy and Lindsay, 2013; Peton *et al.,* 2014) and Virulence Factors Database (http://www.mgc.ac.cn/VFs/).

The FASTA sequences of the proteins were used in a protein BLAST search against the genomes N305, RF122, Sau 170, Sau 302, Sau 1269, and Sau 1364. The protein was considered present (+) in the genome if the match presented a query cover ≥65%, e-value ≤1^-10 and sequence identity ≥ 30%. Manual curation and comparison to the literature were performed in order to verify if the absence or presence of proteins in the genomes was authentic or caused by the thresholds applied in the BLAST search.

Evaluation of SNPs in virulence factors

Total and nonsynonymous single nucleotide polymorphisms (SNPs) were detected with CLC Genomics Workbench, by mapping the reads of the sequenced strains *S. aureus* 170, 302, 1269, and 1364 in the genome of the reference strain *S. aureus* RF122, applying the base quality limits and minimum 20X coverage. SNPs in 33 virulence factors were closely analyzed in order to infer if amino acid changes could possibly alter the virulence of the sequenced strains in comparison to RF122.

Prediction of surface and secreted proteins and search of potential molecular markers for mastitis outcome

The protein sequences of *S. aureus* 170, 302, 1269, 1364, N305, and RF122, were used as the input for the identification of surface and secreted proteins. The programs PSORTb v3.0 (Yu *et al.,* 2010), TMHMM 2.0c (Krogh *et al.,* 2001), Phobius 1.01 (Käll *et al.,* 2004), LipoP 1.0a (Juncker *et al.,* 2003), and Signal P 4.1 (Petersen *et al.,* 2011) were run with default parameters on a Windows 10 Server 64 bits using the terminal emulator PuTTY. The outputs of the five programs were combined, and the redundant protein sequences predicted (with the same FASTA ID) were eliminated. The remaining sequences were aligned with Clustal-Omega 1.2.3 Sievers *et al.,* 2011), and then used to create a distance matrix in which distances were expressed as the number of substitutions per 100 amino acids (Swofford *et al.,* 1996). The distances of the matrix were used to build a Multidimensional Scaling (MDS) scattered plot, using the RStudio Version 1.0.136 and the package bios2mds (from BIOlogical Sequences to MultiDimensional Scaling) (Pelé *et al.,* 2012), where each plot corresponding to a protein sequence was colored in either red (clinical strains) or blue (subclinical strains). To define potential markers associated to mastitis outcome, the plot was visually analyzed to select protein clusters of overlapped sequences originated from the genomes of the subclinical or the clinical strains. The FASTA sequences of the selected proteins were retrieved and used in a BLASTp search to obtain the orthologous protein sequences in the remaining genomes. Additionally, a search was done in the orthologous groups defined by OrthoMCL to find proteins exclusive to the genomes of the subclinical strains.

DNA extraction and validation of the *in silico* analysis

The isolates related to subclinical mastitis, *S. aureus* 170, 302, 1269, 1364 and 322, were used for total DNA extraction with the PureLink® Genomic DNA kit (Invitrogen), adding lysozyme (20 $\mu g.ml^{-1}$) (Sigma-Aldrich) in the initial step. *S. aureus* isolates associated to clinical mastitis (*S. aureus* 76, 216, 1439, 2555, and 3909) were kindly provided by Embrapa Dairy Cattle, Juiz de Fora, Minas Gerais. The bacteria were grown on LB agar plates at 37°C and stocks were made in LB with 20% glycerol and stored at -80°C.

The polymerase chain reaction (PCR) was used to test the specificity of the regions defined *in silico*. A total of 10 *S. aureus* isolates were tested, splitted up in two groups, the first related to the subclinical mastitis (*S. aureus* 170, 302, 1269, 1464 and 322) and the later related to the clinicalmastitis *(S. aureus* 76, 216, 1439, 3909 and 2555). The primer sequences, expected amplicon sizes and amplification conditions are summarized in Table 1. The set of primers au-F3 (5'-TCGCTTGCTATGATTGTGG-3') au-nucR (5'-GCCAATGTTCTACCATAGC-3') (Sasaki *et al.,* 2010) was used as a positive control for all *S. aureus* genomes. Primers for cl3316 were designed based on a sequence found only in the genomes related to subclinical mastitis (Supplementary Figure 1). The sequences of the hypothetical protein cl3309 for each of the six genomes were aligned and used to design two sets of primers. Cl3309subF/cl3309subR were designed over a conserved region among the subclinical genomes, and the primers cl3309cliF/cl3309cliR, overlapped a conserved region of the clinical genomes (Supplementary Figure 2). A single forward primer (LipoP-F-CS) planned over a conserved region among clinical and subclinical strains was used with two reverse primers (LipoP-R-C or LipoP-R-CS). The first one annealed on a conserved region in the clinical genomes and the second one (LipoP-R-CS) targeted a region in clinical and subclinical strains (Supplementary Figure 3).

Table 1 List of primers used in this study.

Primer	Sequence (5'-3')	Description	Target	Expected amplicon size	Amplification conditions*,**
cl3309subF	TGTTGTAGGAGGAACAATCC	Hypothetical protein	Subclinical strains	295bp	Anealing:55°C for 45s Extension: 72°C for 45s
cl3309subR	TTCTAATGTCAGCAACATGC				
cl3309cliF	GCTATTCCTAAATGCTCTATTA	Hypothetical protein	Clinical strains	137bp	Annealing:45°C for 45s Extension: 72°C for 30 s
cl3309cliR	TTTTAAGTATGACATGAATG				
cl3316F	ACGCAAAACCCTTTACTAGT	Transporter protein	Subclinical strains	548bp	Annealing:55°C for 45s Extension: 72°C for 45s
cl3316R	GCAACAACTAGTAGGAGTGA				
LipoP-F-CS	GYTTTGCGAAAACGTTAGAYATGTA	Lipoprotein	Clinical strains	582bp	Annealing:54c for 45s Extension: 72°C for 45s
LipoP-R-C	TGCCTTCATCATTAATTGGACCAATC				
LipoP-F-CS	GYTTTGCGAAAACGTTAGAYATGTA	Lipoprotein	Subclinical and clinical strains	331bp	Annealing:50°C for 45s Extension: 72°C for 1min
LipoP-R-CS	GGTAAAYTCAATGTYCTTATRTCC				
Au-F3	TCGCTTGCTATGATTGTGG	Thermonuclease	All *S. aureus* strains	359bp	**
Au-nucR	GCCAATGTTCTACCATAGC				

*For all sets of primers except Au-F3 and Au-nucR: Initial denaturation: 95.0°C for 5min; 35 cycles of denaturation at 95.0°C for 45s, annealing and extension; final extension at 72.0°C for 10min

**Modification of the amplification conditions described by Sasaki et al (2010): 95.0°C for 5min; 35 cycles of denaturation at 95.0°C for 45s, annealing at 48.8°C for 45s, extension 72°C for 45s; final extension at 72.0°C for 10min.

The reaction mixtures consisted of 100 ng of DNA extract, 1U of Taq DNA polymerase Cellco Biotec, 0.2 µM of each primer, 0.2 mM deoxynucleotide triphosphate mixture, 1.0 mM $MgCl_2$ and 1X reaction buffer and Milli-Q water to increase the reaction volume to the final volume of 50 µL. DNA fragments were analyzed by electrophoresis in 1X Tris-acetate-EDTA on a 1.2% agarose gel stained with 1.0% ethidium bromide.

RESULTS

Functional classification and comparative genomic analysis

General information on the draft genomes of the four *S. aureus* strains related to subclinical mastitis is displayed in Table 2, including their size, the number of coding genes and the number of virulence factors detected in this study. All the strains except Sau1364 belonged to ST 126. Approximately 77% of the proteins were classified into COG families and a similar distribution of the amount of proteins within the categories was seen when the four genomes were compared. On average, 10% of the proteins of the sequenced genomes have unknown function. The categories that comprise the proteins related to transport and amino acid metabolism, protein translation, and carbohydrates metabolism presented the greatest number of constituents, respectively. In the subsystem distribution carried out by SEED, 55% of the CDSs (coding DNA sequences) could be categorized, and 95% out of them were classified as non-hypothetical and 5% as hypothetical. The categories that comprised the greatest number of CDSs were also those related to amino acid metabolism and derived compounds (16%), protein metabolism (10%), and carbohydrates (13%). In the virulence, disease and defense subsystem, there were 68 CDSs for Sau 170, Sau 302, and Sau 1364 and 76 for Sau 1269. In Sau 1364, only one sequence belonged to the phages, prophages, transposable elements, and plasmids subsystem compared to 19 or more genes in the other genomes. The clinical strain RF122 had more proteins related to Cell Wall and Capsule, Cofactors, Vitamins, Prosthetic Groups, Pigments, Phages, Prophages, Transposable elements, Plasmids, RNA metabolism, Regulation and Cell signaling, stress response and carbohydrates compared to the other strains (Figure 1).

Table 2 General features of *Staphylococcus aureus* genomes related to subclinical mastitis and comparison to *S. aureus* N305 and *S. aureus* RF122

Strain number	Draft genome (Mbp)	Coding gene number	SNP[a]	Virulence factors[b]	MLST type
S. aureus 170	2.6	2,599	6,273	105	ST126
S. aureus 302	2.5	2,666	5,481	97	ST126
S. aureus 1269	2.7	2,545	7,420	125	ST126
S. aureus 1364	2.5	2,547	2,491	98	ST1
S. aureus N305[c]	2.7	2752	-	112	ST115
S. aureus RF122[c]	2.7	2664	-	126	ST151

[a] – SNPs were defined based on *S. aureus* RF122 genome.
[b] – Virulence factors described in this study, displayed in Table 3
[c] – Data from Peton *et al (*2014)

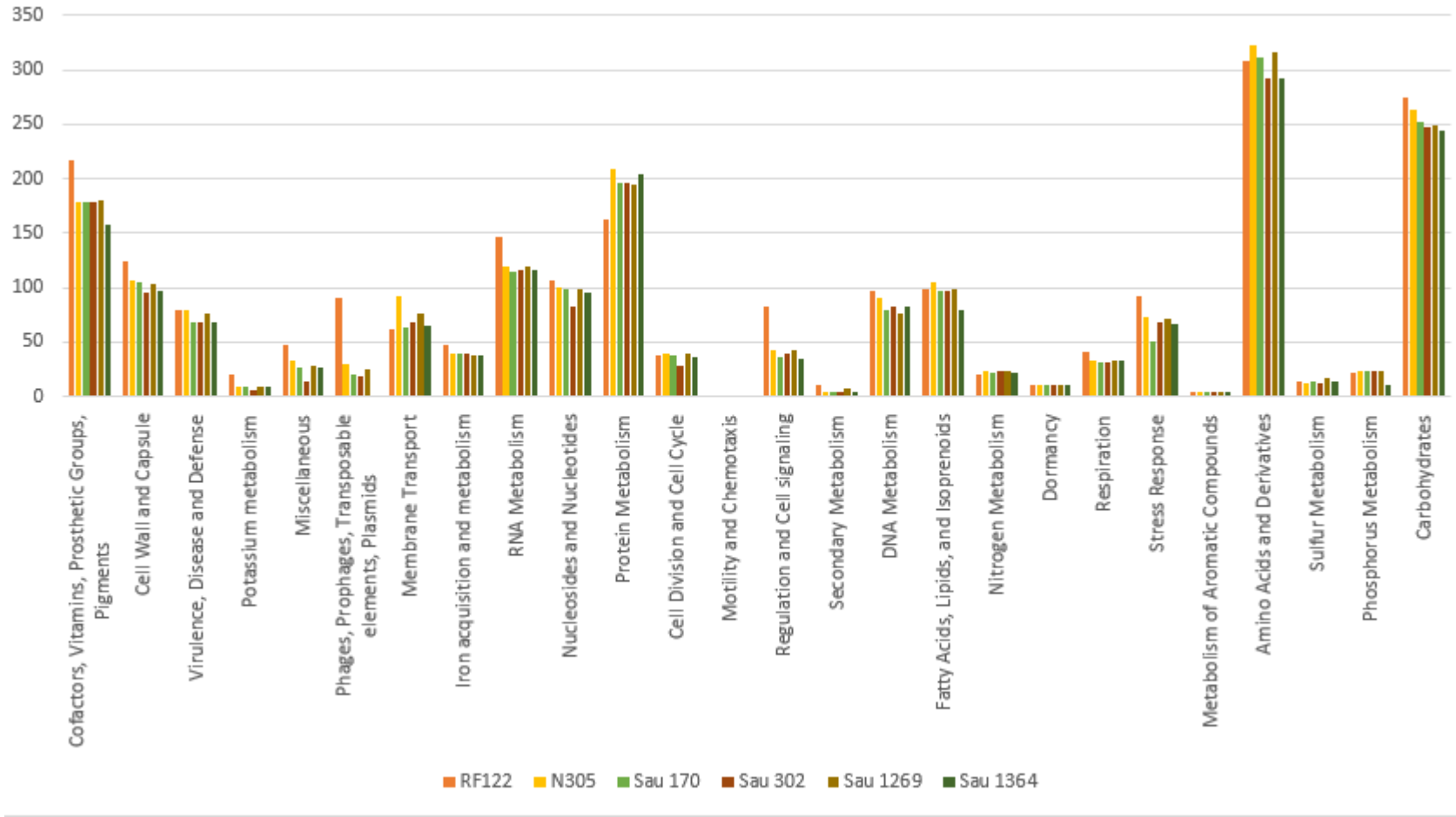


Figure 1 - Functional classification of bovine *Staphylococcus aureus* sequenced genomes CDSs carried out by SEED database. The graph represents an average of CDSs distributions for the genomes of the four subclinical strains *Staphylococcus aureus* 170, 302, 1269, 1364 and the clinical strains *S. aureus* N305 and *S. aureus* RF122.

Repertoire of virulence factors of bovine *Staphylococcus aureus*

Table 3 shows the virulence-associated genes present or absent in Sau 170, Sau 302, Sau 1269, Sau 1364, RF122, and N305, according to the BLAST thresholds used in this study. A total of 149 virulence factors was found, including toxins, exoenzymes, adhesion and cell wall anchored surface proteins, proteins related to host immune evasion, biofilm production, besides regulatory, and miscellaneous proteins. Regulatory enzymes category was the most represented (93%) within all genomes, contrasted to toxins which presence varied within them. A total of 72 virulence factors were present in all the clinical and subclinical genomes, while 4 were absent in all of them: chemotaxis-inhibiting protein *(chp*), sollagen adhesin (*can*), staphylokinase (*sak*), and *S. aureus* surface protein G (*sasG*). Exfoliative toxin A (*eta*), known for causing staphylococcal scalded skin syndrome was found to be present in the genomes related do clinical mastitis and absent in all genomes related to subclinical mastitis.

Six genes were absent in all genomes related to subclinical mastitis, but were present in at least one of the genomes related to clinical mastitis (RF122 and/or N305): bovine variant of enterotoxin C (*sec-bov),* streptolysin S-associated protein SagB homolog, streptolysin-associated protein SagD homolog, toxic shock syndrome toxin 1 (*tst*), transcriptional Repressor SaPI, Serine-rich adhesin for platelet (*sasA*). Comparatively, six proteins were present in all genomes related to subclinical mastitis, but were absent in at least one of the genomes related to clinical mastitis (RF122 and/or N305): Leukocidin chain lukM precursor, Panton-Valentine leukocidin (*PVL*), cell-wall-anchored protein *SasC*, cell-wall-anchored protein *SasK*, fibronectin binding protein B (*fnbB*), and iron-regulated surface determinant protein H (*isdH*).

The strain Sau 1269 was the one that presented the greatest number of virulence factors (125), followed by Sau 170 (105), Sau 1364 (98), and Sau 302 (97). Also, the strain Sau 1269 presented more toxins and genes related to biofilm production when compared to the other strains. On the other hand, the strain Sau 170 presented more exoenzymes and adhesion and cell wall anchored surface proteins, whereas Sau 170 and Sau 1269 together presented more genes related to host immune evasion and regulatory proteins.

Table 3 Presence (+) or absence (-) of virulence-associated genes in sequenced genomes of *Staphylococcus aureus* of bovine origin.

Function	NCBI Accession Number	Gene name	RF122	N305	Sau 170	Sau 302	Sau 1269	Sau 1364
TOXINS								
Bovine variant of Enterotoxin C	ANJ16440.1	*sec-bov*	+	-	-	-	-	-
Enterotoxin B	KLN20762.1	*seb*	+	-	-	-	+	-
Enterotoxin A	AAA26681.1	*sea*	+	+	-	-	+	-
Enterotoxin family protein	ARR25974.1	N/A	+	-	-	-	+	-
Enterotoxin G	BAU35401.1	*seg*	+	-	-	-	+	-
Enterotoxin I	BAU35405.1	*sei*	+	-	-	-	+	-
Enterotoxin J	EEV75450.1	*sej*	+	-	-	-	+	-
Enterotoxin K	AAC28968.1	*sek*	+	-	-	-	+	-
Enterotoxin L	CAI80052.1	*sel*	+	-	-	-	+	-
Enterotoxin M	EFB61033.1	*sem*	+	-	-	-	+	-
Enterotoxin N	BAU35402.1	*sen*	+	-	-	-	+	-
Enterotoxin O	CEF80722.1	*seo*	+	-	-	-	+	-
Enterotoxin T	CRI18413.1	*set*	+	-	-	-	-	-
Enterotoxin, phage associated	BAX06416.1	N/A	+	-	-	-	+	-
Enterotoxin-like protein U	ACI04643.1	N/A	+	-	-	-	+	-
Enterotoxin-like toxin X	AEI60188.1	*selx*	+	+	+	+	+	+
Enterotoxin-like V	ABK27165.1	N/A	+	-	-	-	+	-
Exfoliative toxin A	KPE24689.1	*eta*	+	+	-	-	-	-
Exfoliative toxin B	WP_010994026.1	*etb*	+	-	+	+	-	-
Exotoxin 1	CAI80074.1	*set1*	+	+	+	+	+	+
Exotoxin 10	CAI80065.1	*set10*	+	+	+	+	+	+
Exotoxin 11	CAI80064.1	*set11*	+	+	+	+	+	+
Exotoxin 12	AMV84185.1	*set12*	+	+	-	+	+	-
Exotoxin 13	BBA23069.1	*set13*	+	+	+	+	+	+
Exotoxin 14	BBA23070.1	*set14*	+	+	+	+	+	+

Exotoxin 15	BBA23073.1	*set15*	+	+	+	+	+	+
Exotoxin 2	CAI80071.1	*set2*	+	+	+	+	+	+
Exotoxin 3	CAI80070.1	*set3*	+	+	+	+	+	+
Exotoxin 4	CAI80069.1	*set4*	+	+	-	+	+	-
Exotoxin 5	CAI80068.1	*set5*	+	+	+	-	+	+
Exotoxin 6	BBA23063.1	*set6*	+	+	+	+	+	+
Exotoxin 7	CAI80067.1	*set7*	+	+	+	+	+	+
Exotoxin 8	BAX74372.1	*set8*	+	+	+	+	+	+
Exotoxin 9	CAI80066.1	*set9*	+	+	+	+	+	+
Streptococcal pyrogenic Exotoxin G 1	BAH98136.1	*speG*	+	-	-	-	+	-
Streptolysin S-associated protein SagB homolog	CAI81065.1	SAB1376	+	-	-	-	-	-
Streptolysin-associated protein SagD homolog	CAI81062.1	SAB1373	+	-	-	-	-	-
Toxic shock syndrome toxin 1	BAB43099.1	*tst*	+	-	-	-	-	-
HOST IMMUNE EVASION								
Alpha-hemolysin precursor	OMK04330.1	*hla*	+	+	+	+	+	+
Beta-hemolysin precursor	Q5HEI1.1	*hlb*	+	+	+	+	+	+
Capsular polysaccharide biosynthesis protein Cap5H	BAX04681.1	*cap5H*	-	+	+	+	-	+
Capsular polysaccharide biosynthesis protein Cap5I	BAX04682.1	*cap5I*	-	+	-	-	-	-
Capsular polysaccharide biosynthesis protein Cap5J	BAX04683.1	*cap5J*	-	+	+	+	-	+
Capsular polysaccharide biosynthesis protein Cap5K	BAX04684.1	*cap5K*	-	+	+	+	-	+
Capsular polysaccharide synthesis enzyme cap8H	BAX74115.1	*cap8H*	+	-	-	-	+	-
Capsular polysaccharide synthesis enzyme Cap8I	AMV78743.1	*cap8I*	+	-	-	-	+	-
Capsular polysaccharide synthesis enzyme cap8J	BAX74117.1	*cap8J*	+	+	-	+	+	+
Capsular polysaccharide synthesis enzyme cap8K	BAX74118.1	*cap8K*	+	-	-	-	+	-
Chemotaxis-inhibiting protein CHIPS	BBA24011.1	*chp*	-	-	-	-	-	-
Clumping factor A	CAI80432.1	*clfA*	+	+	+	+	+	+
Clumping factor B	CAA12115.1	*clfB*	+	+	+	+	+	+
Delta-hemolysin	BAH22640.1	*hld*	-	+	+	-	+	-
Fibrinogen-binding protein	CAA50892.1	*efb*	+	+	-	-	+	+
Gamma-hemolysin component A	P0A074.1	*hlgA*	+	+	+	+	+	+

Gamma-hemolysin component B	AAA26639.1	*hlgB*	+	+	+	+	+	+
Gamma-hemolysin component C	AAA26638.1	*hlgC*	+	+	+	-	+	+
Immunoglobulin G binding protein A precursor	ABO39035.1	*spa*	+	+	+	-	+	+
Immunoglobulin-binding protein sbi	BBA24967.1	*setsbi*	+	+	+	+	+	+
Leukocidin chain lukM precursor	SCT10101.1	*lukM*	+	-	+	+	+	+
Leukocidin F subunit	CZQ67463.1	*lukF*	+	+	+	+	+	+
Leukocidin LukD precursor	BAF67990.1	*lukD*	+	+	+	+	+	+
Leukocidin LukE precursor	BAF67991.1	*lukE*	+	+	+	+	+	+
Leukocidin/hemolysin toxin subunit S	EFB60060.1	*lukS*	+	+	-	-	+	+
Panton-Valentine leukocidin	CAI80471.1	*PVL*	+	-	+	+	+	+
Phosphatidylglycerol lysyltransferase	BBA23933.1	*mprF*	+	+	+	+	+	+
Staphylococcal complement inhibitor	BAX06410.1	*scin*	+	+	+	+	+	-
EXOENZYMES								
Catalase	KXA37580.1	*cat*	+	+	+	+	+	+
Cysteine protease inhibitor B	Q9EYW6.1	*sspC*	+	+	+	+	+	+
Cysteine protease precursor B	WP_072468235.1	*sspB*	+	+	+	+	+	-
Glyceraldehyde 3-phosphate dehydrogenase	SAY63819.1	*GAPDH*	+	+	+	+	+	+
Heme transporter IsdDEF, membrane component IsdD	SHD92471.1	*isdD*	+	+	+	+	+	+
High-affinity heme uptake system protein isdE	BBA23683.1	*isdE*	+	+	+	+	+	+
Hyaluronate lyase	KII20513.1	*setlyase*	+	+	+	+	+	+
LPXTG specific sortase A	ASI56539.1	*srtA*	+	+	+	-	-	-
Probable heme-iron transport system permease protein	BBA23684.1	*isdF*	+	+	+	+	+	+
Probable heme-iron transport system permease protein isdF	BBA23684.1	*isdF*	+	+	+	+	+	+
Sortase B	KFL08714.1	*srtB*	+	+	+	+	+	+
Staphylocoagulase	BAH66222.1	*coa*	+	+	+	+	+	+
Staphylokinase	CAA24957.1	*sak*	-	-	-	-	-	-
Transcriptional Repressor SaPI	CDP56461.1	*stl*	-	+	-	-	-	-
V8 protease	KXA36839.1	*sspA*	+	+	+	+	+	+
Von Willebrand factor-binding protein	ADN93319.1	*vwb*	+	+	+	+	+	+
ADHESION AND CELL WALL ANCHORED SURFACE PROTEINS								

Bone sialoprotein-binding protein	CAG39588.1	*bbp*	+	+	+	+	-	+
Cell-wall-anchored protein SasC	ADL65760.1	*sasC*	-	+	+	+	+	+
Cell-wall-anchored protein SasD	ATF40316.1	*sasD*	+	+	+	+	+	+
Cell-wall-anchored protein SasK	EIK30180.1	*sasK*	-	+	+	+	+	+
Collagen adhesin	AAA20874.1	*cna*	-	-	-	-	-	-
Elastin binding protein ebpS	SBE86677.1	*ebpS*	+	+	+	+	+	-
Enolase	AAC17130.1	*eno*	+	+	+	+	+	+
Extracellular adherence protein	ALY21517.1	*eap*	-	+	-	+	+	+
Extracellular matrix-binding protein	EFB49541.1	*ebh*	-	+	-	-	+	-
Fibronectin binding protein A	AEQ55002.1	*fnbA*	+	+	+	+	+	+
Fibronectin binding protein B	CFF40144.1	*fnbB*	-	+	+	+	+	+
Fibronectin/fibrinogen-binding protein	CKH07129.1	*fbe*	+	+	+	+	+	+
Iron-regulated surface determinant protein A	BBA23680.1	*isdA*	+	+	+	+	+	+
Iron-regulated surface determinant protein B	OYP89100.1	*isdB*	+	+	+	+	+	+
Iron-regulated surface determinant protein C	BBA23681.1	*isdC*	+	+	+	+	+	+
Iron-regulated surface determinant protein H	BBA24362.1	*isdH*	+	+	+	+	+	+
Predicted cell-wall-anchored protein SasF (LPXAG motif)	ADC38796.1	*sasF*	+	+	+	+	+	+
S. aureus surface protein G	ONH21632.1	*sasG*	-	-	-	-	-	-
Serine-aspartate repeat containing protein C	CAA06650.1	*sdrC*	-	+	-	-	+	-
Serine-aspartate repeat containing protein D	CAA06651.1	*sdrD*	+	+	+	+	-	+
Serine-aspartate repeat containing protein E	ATV03366.1	*sdrE*	+	+	+	+	-	+
Serine-aspartate repeat containing protein F	SBE93024.1	*sdrF*	-	+	+	-	+	+
Serine-aspartate repeat containing protein G	EZI01607.1	*sdrG*	+	+	+	+	+	+
Serine-rich adhesin for platelet	AMV86383.1	*sasA*	-	+	-	-	-	-
Surface protein SasB	AAR15206.1	*sasB*	-	+	+	+	+	+
Surface protein SasE	AAR15233.1	*sasE*	+	+	+	+	+	+
Surface protein SasH	AAR15278.1	*sasH*	+	+	+	+	+	+
PROTEINS RELATED TO BIOFILM PRODUCTION								
Accumulation-associated protein (associated to biofilm formation)	PKF42030.1	*aap*	-	+	-	-	+	-

Biofilm associated protein	AFS33108.1	*bap*	-	-	+	-	-	+
Polysaccharide intercellular adhesin protein A	ANP43616.1	*icaA*	+	+	+	+	+	+
Polysaccharide intercellular adhesin protein B	AAD52057.1	*icaB*	+	+	+	+	+	-
Polysaccharide intercellular adhesin protein C	AAF60242.1	*icaC*	+	+	+	-	+	+
Polysaccharide intercellular adhesin protein D	SMF97944.1	*icaD*	+	+	+	+	+	+
Polysaccharide intercellular adhesin protein R	AAD52054.1	*icaR*	+	+	+	+	+	+
Thermonuclease	KFB80761.1	*nuc*	+	+	+	+	+	+
REGULATORY PROTEINS								
Accessory gene regulator protein A	ALY21623.1	*agrA*	+	+	+	+	+	+
Accessory gene regulator protein B	ABB17466.1	*agrB*	+	+	+	+	+	+
Accessory gene regulator protein C	AAF23170.1	*agrC*	+	+	+	+	+	+
Accessory gene regulator protein D	BAB95826.1	*agrD*	-	+	-	-	+	-
Histidine protein kinase SaeS	BBA23356.1	*saeS*	+	+	+	+	+	+
HTH-type transcriptional regulator mgrA	BBA23334.1	*mgrA*	+	+	+	+	+	+
HTH-type transcriptional regulator R	AAG35715.1	*sarR*	+	+	+	+	+	+
HTH-type transcriptional regulator S	AFD54312.1	*sarS*	+	+	+	+	+	+
HTH-type transcriptional regulator T	BAB96282.1	*sarT*	+	+	+	+	+	+
HTH-type transcriptional regulator U	AFD54311.1	*sarU*	+	+	+	-	+	-
HTH-type transcriptional regulator V	Q2FVY9.1	*sarV*	+	+	+	+	+	+
HTH-type transcriptional regulator Z	Q2FVN3.1	*sarZ*	+	+	+	+	+	+
Repressor of toxins rot	BBA24393.1	*rot*	+	+	+	+	+	+
Response regulator SaeR	AFD54317.1	*saeR*	+	+	+	+	+	+
Staphylococcal accessory regulator SarY	BAX06750.1	*sarY*	+	+	+	+	+	+
Staphylococcal accessory regulator X	Q2G0D1.2	*sarX*	+	+	+	+	+	+
Transcriptional regulator SarA	KFB79964.1	*sarA*	+	+	+	+	+	-
MISCELLANEOUS PROTEINS								
Bifunctional autolysin precursor	BAF67194.1	*atl*	+	+	+	+	+	+
Glycerol phosphate lipoteichoic acid synthase	OBY01672.1	*itaS*	+	+	+	+	+	+
Immunodominant antigen B IsaB	AGU56232.1	*isaB*	+	+	+	-	+	+
Lipoteichoic acid opeon - ypfP	CAA74741.1	*ypfP*	+	+	+	+	+	+

Lipoteichoic acid protein A	Q2FZP8.2	*ltaA*	+	+	+	+	+	+
Protein translocase subunit A	AAB54024.1	*secA*	+	+	-	-	+	-
Protein translocase subunit E	AAB54017.1	*secE*	+	+	+	-	+	+
Protein translocase subunit F	EFG58204.1	*secF*	+	+	+	+	+	+
Protein translocase subunit G	AFH69089.1	*secG*	+	+	+	+	+	+
Protein translocase subunit Y	AAB54022.1	*secY*	+	+	+	+	+	+
Putative exoprotein DltD	AAD21960.1	*dlt*	+	+	+	-	+	+
Teichoic acid glycerol-phosphate primase	Q2G2X4.1	*tag*	+	+	+	+	+	-
Transglycosylase IsaA	Q2FV52.1	*isaA*	+	+	+	+	+	+
Zinc metalloproteinase aureolysin	BBA25172.1	*aur*	+	+	+	+	+	-

*The protein Spa in *S. aureus* RF122 is considered a pseudogene for having a premature stop codon.

Identification of orthologous proteins

The alignment of short sequencing reads with OrthoMCL predicted several clusters of CDSs possibly exclusive of the genomes of subclinical strains (Supplementary Table 1). However, the alignment of these reads to the RF122 and N305 genomes showed that the lack of correspondence of CDSs between the subclinical and clinical genomes was actually caused by sequencing assembly errors. Only one cluster was confirmed to be exclusively present in the genomes of the strains associated to subclinical mastitis. This cluster, named cl3316, corresponds to a protein member of the Major Facilitator Superfamily, the largest group of secondary active membrane transporters, whose members transport a diverse range of substrates (Law *et al.,* 2008). Therefore, the coding sequence of this protein was used to build the primers cl3316F and cl3316R (Supplementary Figure 1) which were used in PCR reactions with DNA extracted from clinical and subclinical strains (shown below).

SNPs in coding DNA regions and in virulence factors

The number of SNPs found in the CDSs of the genomes of the subclinical strains Sau 170, Sau 302, Sau 1269, and Sau 1364, compared to the reference genome of RF122, is displayed in Figure 2. Insertions, deletions, and nonsense mutations were also considered amino acid changes. 24% to 26% of the SNPs were nonsynonymous, while the majority of them were synonymous. The majority of nonsynonymous mutations occurred among amino acids of the same class. Although the total number of SNPs was different for each *S. aureus* isolate, there was a similar distribution of types of SNPs among them. However, when the SNPs in virulence factors were analyzed, a greater proportion of nonsynonymous SNPs was observed such as 29.0% in Sau 170, 29.7% in Sau 302, 26.8% in Sau 1269, and 33.3% in Sau 1364. Regarding the 33 virulence factors analyzed here, the greatest number of SNPs (307, 212, 188, and 187) were seen in the adhesin genes *fnbA*, *clfA*, *clfB* and in the cell wall-anchored protein *isdA* (307, 212, 188, and 187 SNPs, respectively). The greatest number of nonsynonymous SNPs was also seen in *fnbA*, *clfA*, *isdA*, enterotoxin 9, and *clfB* (82, 73, 72, 60, and 50, respectively). No SNPs were found in serine-aspartate repeat containing proteins D and E (*sdrD* and *sdrE*), and in the translocase proteins *secE* and *secG* (Figure 3).

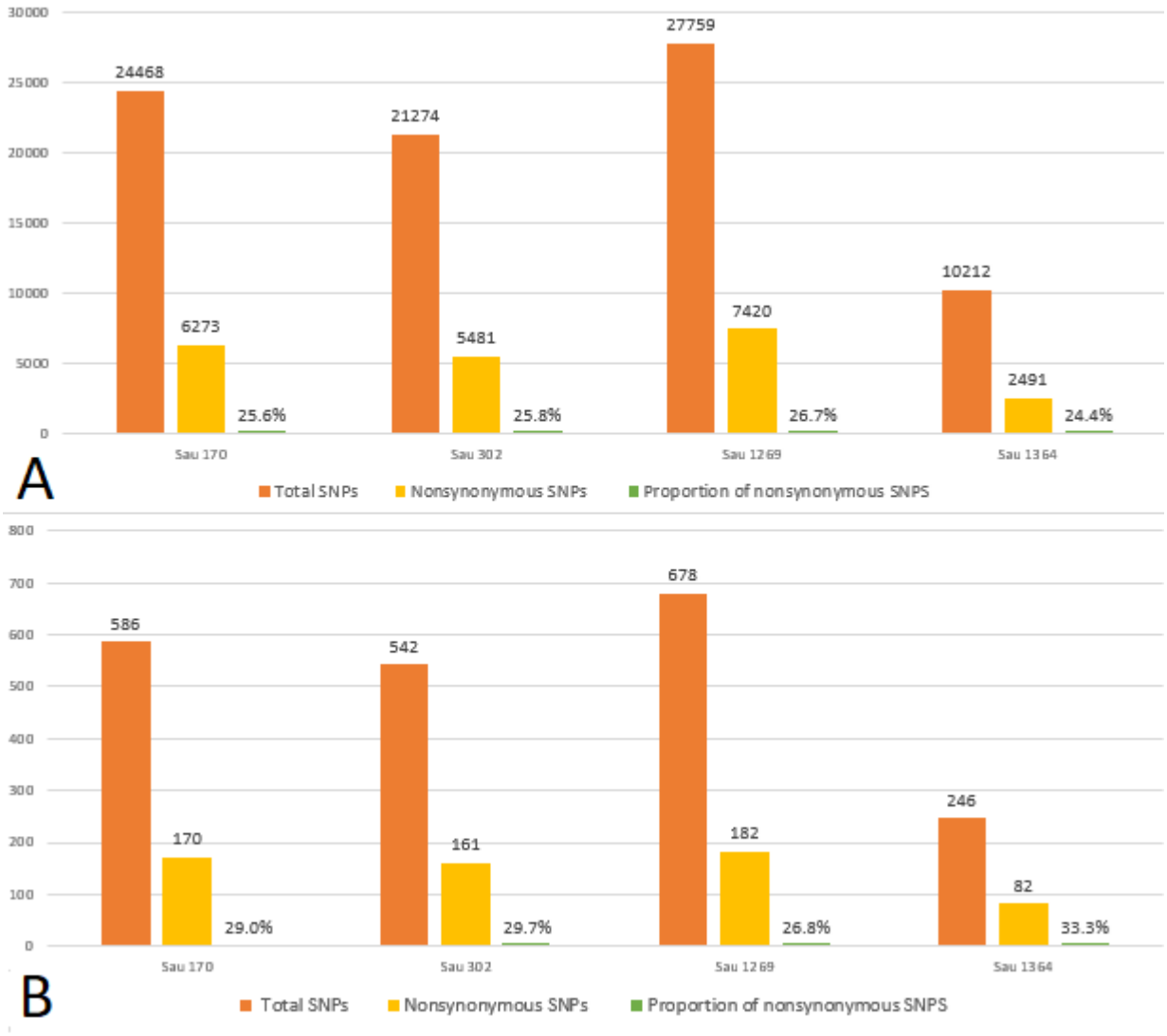


Figure 2 -Analysis of single nucleotide polymorphisms in genes of *Staphylococcus aureus* associated with subclinical of compared to the genome of *S. aureus* RF122, associated with clinical mastitis. In A, the SNPs in all Coding DNA Sequences (CDS) and in B, the SNPs in the 33 virulence factors tested. SNPs were investigated in all coding DNA regions and in main virulence factors described for *S. aureus*.

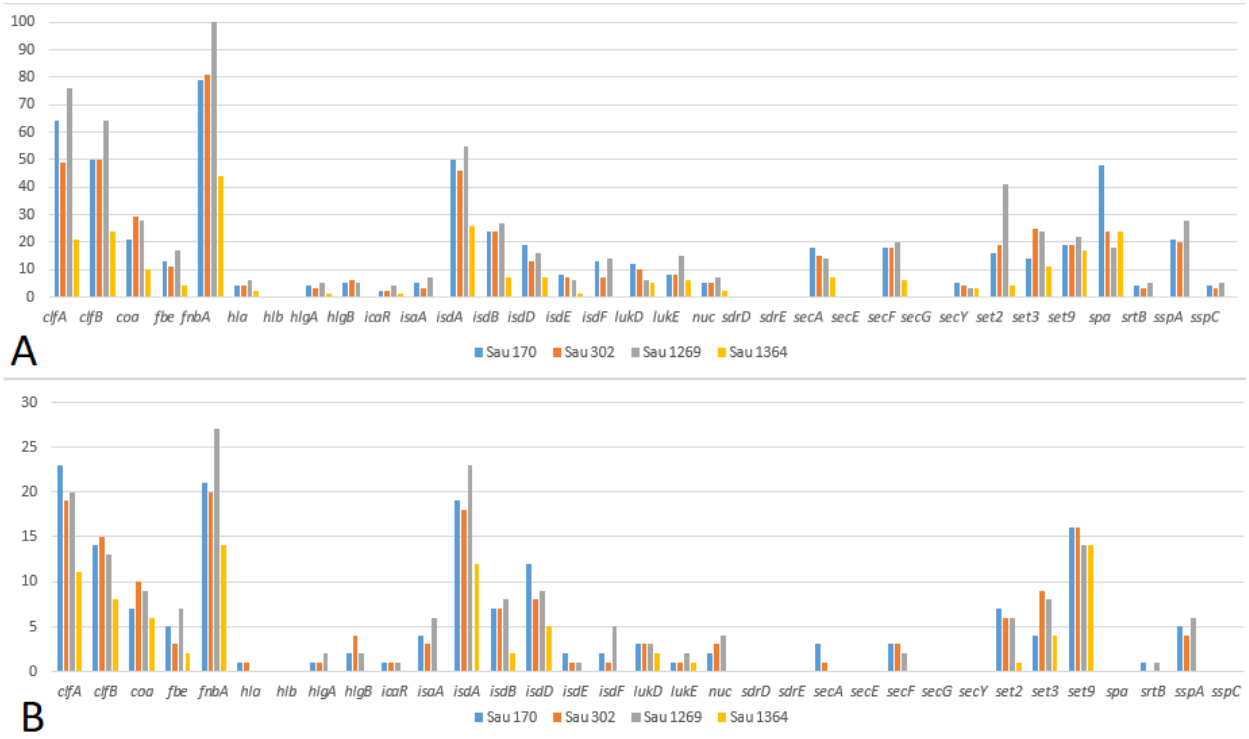


Figure 3 – Distribution of total (A) and nonsynonymous SNPs (B) in main virulence factors present in the genomes of the subclinical strains of *Staphylococcus aureus* compared to the *S. aureus* RF122 genome.

Prediction of surface proteins present in the genomes of the bovine *S. aureus* strains and validation of the *in silico* analysis

The programs PSORTb, TMHMM, Phobius, LipoP, and Signal P predicted a different number of transmembrane and surface proteins for each one of the genomes analyzed. After the combination of the outputs and elimination of the predicted redundant proteins, a total of 922 proteins was predicted for N305, 971 for RF122, 947 for Sau 170, 962 for Sau 302, 984 for Sau 1269, and 954 for Sau 1364 (Table 4). TMHMM and Phobius predicted the greatest number of proteins when compared to the other programs. Most of the protein sequences was highly conserved among the genomes.

Table 4 Output of the programs used for the prediction of surface and secreted proteins in the genomes of bovine *Staphylococcus aureus* strains.

	Lipop	Phobius	PSORTb	Signalp	TMHMM	Total	Total[1]
S. aureus RF122	338	653	125	157	698	1971	971
S. aureus N305	321	666	160	151	689	1987	922
S. aureus 170	325	624	137	146	671	1903	947
S. aureus 302	323	637	145	136	673	1914	962
S. aureus 1269	356	654	130	161	688	1989	984
S. aureus 1364	317	633	146	139	670	1905	954

[1]– total number of proteins without redundancy

Following the alignment of the FASTA sequences of the proteins and the creation of a distance matrix, a MDS plot was built (Figure 4), where the blue dots originated from the genomes related to subclinical mastitis (Sau 170, Sau 302, Sau 1269, and Sau 1364) and the red dots refers to the genomes of *S. aureus* strains associated to clinical mastitis (RF122 and N305).

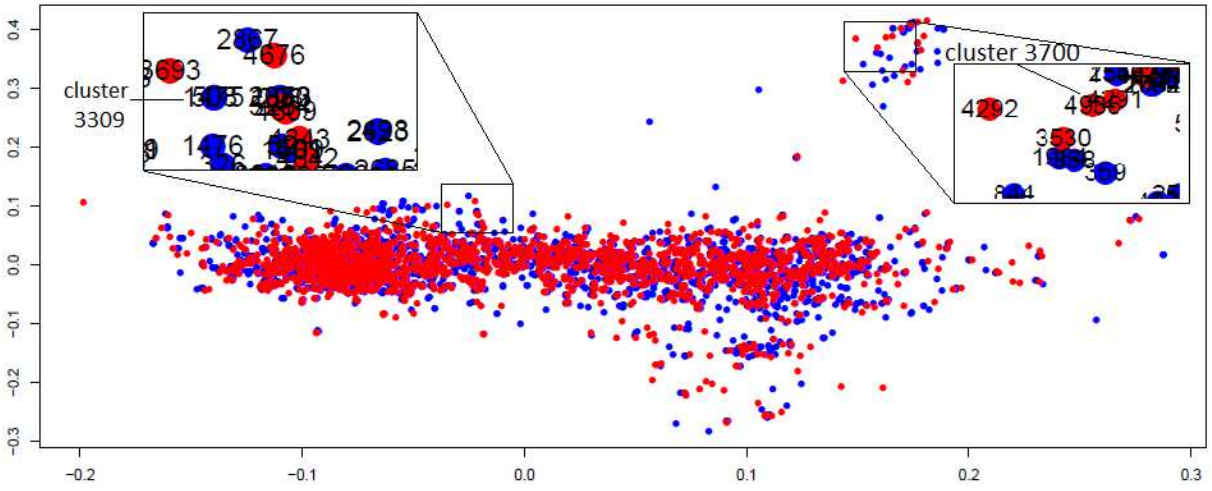


Figure 4 MDS plot of surface and secreted proteins generated from surface proteins of bovine *Staphylococcus aureus* strains predicted with the combined results of the programs PSORTb, TMHMM, Phobius, LipoP, and Signal P. The genomes of *S. aureus* 170, 302, 1269, and 1364 are represented in blue and the genomes of *S. aureus* N305 and RF122, in red. The closely located dots represent more identical protein sequences. In the upper left and in the upper right, the clusters 3009 and cl3700, respectely, were zoomed-in. It contains overlapped blue dots, corresponding to a protein highly conserved among the genomes of the subclinical *S. aureus* strains.

The visual inspection of the MDS plot allowed the identification of two interesting isolated protein clusters (Figure 4, inset). The first cluster (cl3309) corresponded to a hypothetical protein that showed a region 100% identical among the genomes related to subclinical mastitis with smaller identity (58%) to genomes related to clinical mastitis (Supplementary Figure 2). Contrary to expectations the primers cl3309subF/R amplified a 295 bp amplicon in all isolates tested regardless of mastitis outcome (Figure 5). When cl3309 was amplified with cl3309cliF/ cl3309cliR, that should amplify the most conserved region of the protein between the RF122 and N305 genomes, the expected amplicon of 137bp was seen in three out five isolates of clinical manifestation and in three out of five isolates of subclinical manifestation (Figure 5).

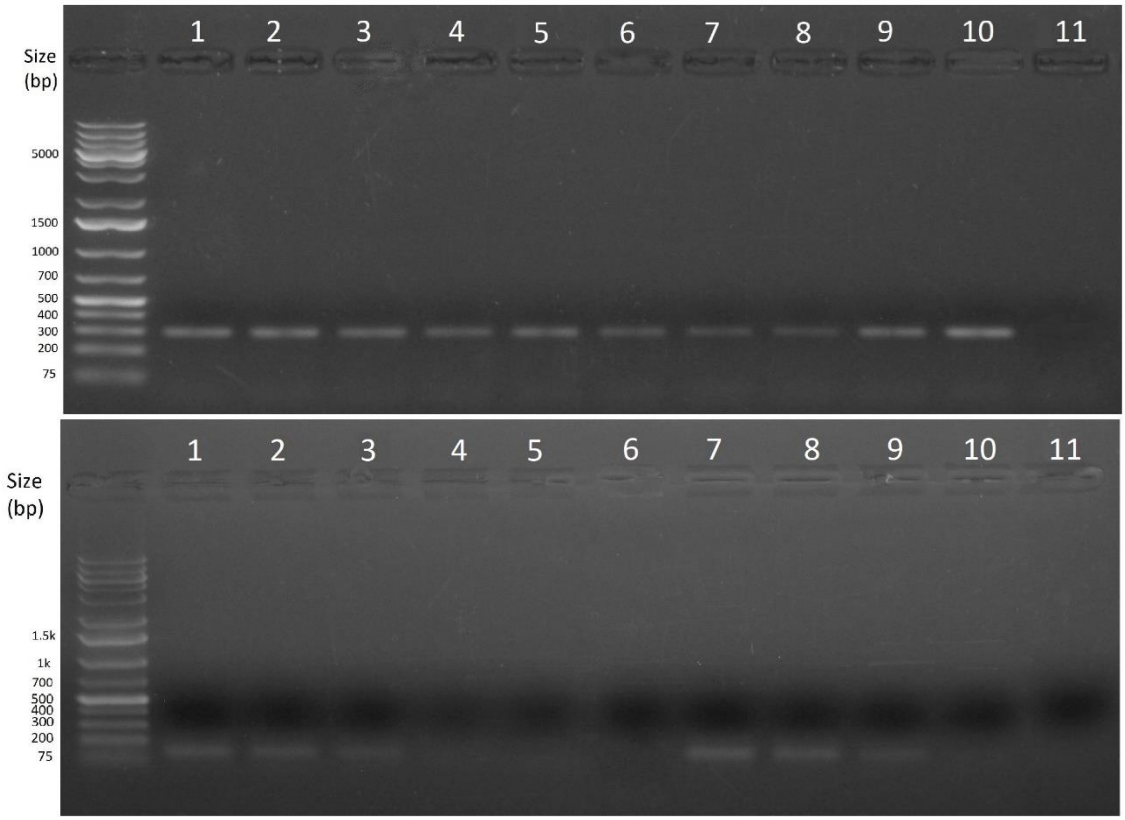


Figure 5 - Amplification of the hypothetical protein cl3309 gene of *Staphylococcus aureus* of bovine origin. At the top, the primers cl3309subF/cl3309subR, targeting subclinical strains, were used in the PCR reaction with DNA extracted from the subclinical isolates *S. aureus* 170 (1), 302 (2), 1269(3), 1364 (4), and the clinical isolates *S. aureus* 322 (5), 216 (6), 3909 (7), 2555 (8), 1439 (9), 76 (10). At the bottom, the primers cl3309cliF/cl3309cliR, targeting clinical strains, were used to amplify the same genomes. Water was used as the negative control (11). Cellco Biotec 1 kb Plus DNA Ladder was used as a molecular marker.

The second group of proteins that showed proximity in the MDS plot corresponded to a tandem-type lipoprotein, named LipoP (cluster cl3700), which possesses a DUF576 domain and has 91.79% identity between the strains related to clinical mastitis and that is less conserved (68%) among the genomes related to subclinical mastitis (Supplementary Figure 3). Three primers were designed to differentially amplify the lipoprotein: a forward (LipoP-F-CS) and reverse primer (Lipo-R-CS), in a region conserved for clinical and subclinical strains, and a second reverse primer (LipoP-R-C) in a region conserved only in the genomes RF122 and N305 (Supplementary Figure 4). Therefore, these sets of primers should harbor a 582bp fragment in all isolates regardless of the manifestation, and a 331bp fragment only in the isolates of clinical manifestation. As expected, the set of primers LipoP-F-CS/ LipoP-R-CS resulted in a successful amplification for all five isolates of subclinical manifestation and all five isolates of clinical manifestation tested (Figure 6). Next, the combination of the primers LipoP-F-CS and LipoP-R-C, which should amplify all the clinical isolates, resulted in the amplification of a single one; no amplicons were seen for the subclinical isolates, as predicted.

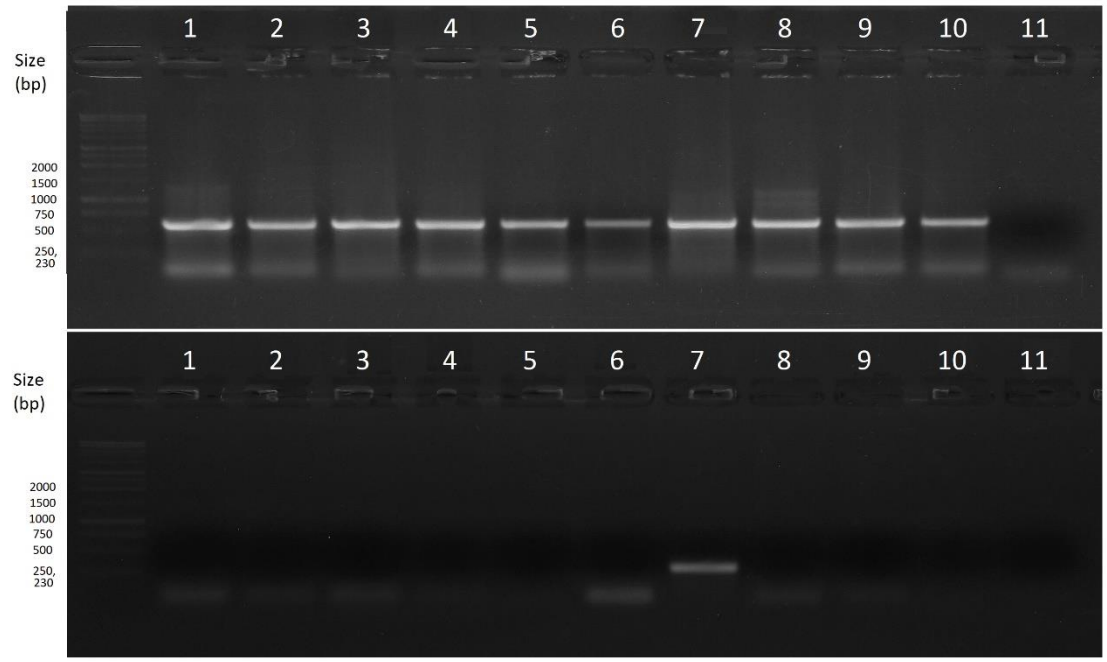


Figure 6 – Amplification of a lipoprotein gene of *Staphylococcus aureus* isolates of bovine origin. At the top, the primers LipoP-F-CS and LipoP-R-CS, targeting subclinical and clinical strains, were used in PCR reaction with the subclinical isolates *S. aureus* 170 (1), 302 (2), 1269(3), 1364 (4), 322 (5), and the clinical isolates *S. aureus* 76 (6) 216 (7), 1439 (8), 2555 (9), and 3909 (10) were used. At the bottom, the primers LipoP-F-CS and LipoP-R-C, targeting clinical strains, were used to amplify the same genomes. Water was used as the negative control (11). Promega 1kb DNA ladder was used as a molecular marker.

To confirm if the sequences detected *in silico* were able to discriminate among subclinical and clinical strains, we performed PCR with the primers listed in Table 1. Amplicon of expected size was seen when the DNA was amplified with primers for the *nuc* gene confirming the isolates as *S. aureus* (Supplementary Figure 4). When the coding sequence of the transporter protein identified as cl3316 by orthoMCL was amplified with the cl3316F and cl3316R, positive results were predicted only in the isolates of subclinical manifestation, since no orthologous proteins were found in the genomes of clinical manifestation (Supplementary Figure 1). However, the expected 548bp-amplicon was seen in all subclinical samples and four out of five clinical samples (Figure 7).

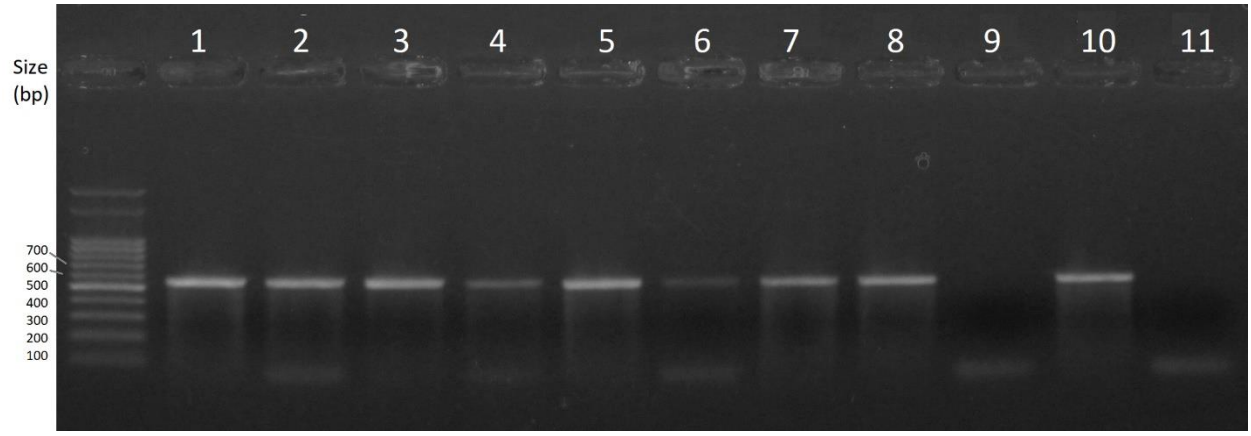


Figure 7 - Amplification of the gene coding the transporter protein cl3316 of *Staphylococcus aureus* isolates of bovine origin. At the top, the primers cl3309subF/cl3309subR were used in the PCR reaction with DNA extracted from the subclinical isolates *S. aureus* 170 (1), 302 (2), 1269(3), 1364 (4), and the clinical isolates *S. aureus* 322 (5), 216 (6), 3909 (7), 2555 (8), 1439 (9), 76 (10). Water was used as the negative control (11). Ludwig Biotec 100bp DNA ladder was used as molecular marker.

DISCUSSION

S. aureus is the major cause of persistent bovine mastitis and a prevalent species in dairy herds worldwide. The control of staphylococcal mastitis significantly impacts milk quantity and quality and has a positive effect on the dairy industry and on public health. Several efforts have been made to characterize isolates circulating in herds in order to find bacterial factors that could be linked to disease outcome. So far, only six *S. aureus* genomes of bovine origin had been sequenced and *S. aureus* RF122, for which the complete genome has been described, is considered a reference genome for comparative analysis.

Our group has sequenced four genomes of *S. aureus* isolated from subclinical mastitis (Silva *et al.,* 2016). The number of contigs produced by the Ion PGM sequencing varied from 93 in Sau 1269 to 568 in Sau 302. Although this could hamper the correct assembly of the genome resulting in the overlapping of noncontiguous reads we were able to map all reads of the four genomes in the reference *S. aureus* RF122, confirming that the amount of contigs did not impair the quality of the results. The total length and GC content are in good agreement with other sequenced genomes of *S. aureus*, which have an average size of 2.8Mb, 32.8% GC, and 2810

CDS. In a comparison to the reference strain *S. aureus* RF122 and to *S. aureus* N305, both isolated from animals with bovine mastitis, the similarity is even greater, since both genomes have 2.7Mb, slightly smaller than the average seen for human isolates (Peton *et al.,* 2014). This is the first report of genomes of bovine *S. aureus* assigned to ST126, a prevalent type in Brazilian herds (Rabello *et al.,* 2007; Silva *et al.,* 2013) that had been claimed more significant targets in the design of methods for controlling *S. aureus* infections, due to their site specificity (Smith et al, 2005).

More than 70% of the genes could be grouped in specific COG categories, similarly as described for *S. aureus* RF122 (Herron-Olson *et al.,* 2007) and N305 (Peton *et al.,* 2014) confirming again good sequencing quality. Regardless of the program used, the categorization of the genes and their distribution in subsystems were similar. Some discrepancies were seen. *S. aureus* RF122 had more sequences related to stress response, regulation and cell signaling, potassium metabolism, cofactors and vitamins, and plasmids, TE, prophages that could confer the ability of the bacteria to cause the severe mastitis caused by RF122. Interestingly, previous studies had already reported an almost 10-fold greater average dn:ds ratio in the RF122 genome relative to human associated sequenced *S. aureus* isolates in several genes involved in host colonization, toxin production, gene regulation, iron metabolism and antibiotic resistance (Herron-Olson *et al.,* 2007). dN/dS is the ratio of the number of nonsynonymous substitutions per non-synonymous site (pN) to the number of synonymous substitutions per synonymous site (pS) (Bioinformatics I/O – 2018). It suggests that bovine strains related to clinical manifestation of mastitis may rely on other alternative mechanisms for specialized pathogenicity.

The phages, prophages, transposable elements, and plasmids subsystem of all subclinical strains but Sau1364 had a similar number of genes. This may happen when there is a high density of reads assembling together, resulting in their annotation as a single CDS rather than as multiple CDSs. Genetic variability should also be considered. Sau1364 belongs to ST1 that has been isolated from humans and bovines (Smith *et al.,* 2005) in opposition to the other strains from ST126 and strongly associated to bovine infections. In *S. aureus* of human origin, a great variability between strains has been reported in the mobile genomic elements and this could also be also true for bovine strains. As an example, four integrated prophages have been identified in *S. aureus* Newman and only one in *S. aureus* COL. Tn*554* have copies in *S. aureus* N315 and none in USA300 (Baba *et al.,* 2008). Here, we expanded the repertoire of virulence factors

previously described for bovine strains of *S. aureus* (Herron-Olson *et al.,* 2007, Peton *et al.,* 2014) and looked for other determinants reported for human strains (Holtfreter *et al.,* 2010, McCarthy and Lindsay, 2013). The content of proteins was quite similar among strains and associations between mastitis outcome and virulence factors were difficult to make. However, there were some exceptions. Enterotoxins were mostly present in RF122 and Sau 1269. These proteins are considered superantigens due to their stimulation of the T-cell proliferation (Ortega *et al.,* 2010) and may impact disease severity in a rabbit model (Stach *et al.,* 2016). However, Sau 1269 was isolated from an animal diagnosed with a subclinical infection suggesting that host factors and not only toxin presence could affect mastitis outcome. Additionally, a special concern exists on bovine isolates that produce enterotoxins since they could pose risk to consumers' health upon ingestion of contaminated milk (Akindolire *et al.,* 2015).

We found that the exfoliative toxin A gene (*eta*) is present in RF122 and N305 that cause clinical mastitis, although the later is reported to cause milder symptoms. Other toxins such as Tsst-1 and streptolysin S-associated protein SagB/D homologs were exclusive to RF122. *Eta* is a superantigen, which has mitogenic properties over T cells, and has proteolytic activity directly responsible for skin exfoliation (Bukowski *et al.,* 2010). Tsst-1 is also a superantigen produced by some human strains that has been implicated in disease progression. Streptolysins are potent cytolytic toxins produced by *Streptococcus pyogenes* that had been first described in *S. aureus* RF122 present in the genomic island *νSaBov* (Herron-Olson *et al.,* 2007).

A great number of SNPs was found in the genes responsible for adhesion (*clfA*, *clfB* and *fnbpA*), which are important for promoting the adhesion of bacterial cells to a myriad of molecules and surfaces (Tsompanidou *et al.,* 2012). Genes like *spA, secA/E/Y* and *sdrD/E* had only synonymous SNPs while for others (*isdA*, *set9*) almost 50% was represented by nonsynonymous SNPs may change amino acid sequence and therefore impact protein structure and/or function. Set 9 is a member of the exoproteins family, known as the staphylococcal exotoxin-like (Set) proteins, further renamed staphylococcal superantigen-like (Ssl) proteins. Studies suggest that allelic variations within the *ssl* cluster between strains of animal origin may result in adaptive advantage to various conditions during the infection of diverse animal tissues (Smyth *et al.,* 2006).

ClfA has been shown to impede macrophage phagocytosis (Palmqvist *et al.,* 2004) and to participate in the inactivation of C3b, the central complement component (Hair *et al.,* 2008). Additionally, *S. aureus* ClfA and -B and FnbA and -B proteins are capable of inhibiting the deposition of opsonins to the pathogen due to its ability of binding fibrinogen (Higgins *et al.,* 2006). Studies have detected key genetic polymorphisms in ClfA and FnBPA, which result in a conformation change in the tridimensional protein, significantly lowering fibrinogen-binding affinities (Stutz *et al.,* 2010). Additionally, ClfA in RF122 show a deletion resulting in a translational frameshift that leads to several premature stop codons in the translated protein (Stutz *et al.,* 2010).

All subclinical strains investigated displayed the whole SpA coding sequence, whereas RF122 has a premature codon leading to a truncated protein. Studies suggest that the untranslated sortase motif (LPXTG) in the truncated SpA protein may result in its absence from the bacteria surface (Stutz *et al.,* 2010). We suggest the great rates of polymorphism detected in these genes, in addition to the full SpA, might confer the subclinical strains several advantages to successfully evade the immune system, without triggering immune responses and the manifestation of clinical symptoms, therefore facilitating the establishment of chronic and silent infections.

An aim of this work was to find a putative association between surface and secreted proteins in mastitis outcome Therefore, these proteins were predicted from the genomes by five tools in order to increase the accuracy of the retrieved data. The programs TMHMM and Phobius have a similar prediction output, since both rely on the prediction of transmembrane topologies using a Markov model (HMM). On the other hand, PSORTb and SignalP, with the smaller number of predictions, are more specific, being the first designed to predict subcellular localization and the second to discriminate between signal peptides and transmembrane regions. Finally, with an intermediate number of predictions, LipoP seems to have intermediate stringency, since it was designed to detect lipoprotein signal peptides. At the beginning we expected the proteins to separate in two groups, one clinical and the other subclinical, but instead, no distinguishable group was seen revealing the high similarity of the proteins among the strains. So we decided to look for sequence variability in some clusters composed of overlapped proteins that belonged to the subclinical or the clinical strains. Visual inspection allowed us to find two clusters, cl3309 and the lipoprotein and significant differences could be

found in the nucleotide sequence, allowing the design of primers that were expected to amplify only clinical or subclinical genomes. If we consider the primers designed for cl3309, cl3309subF/R amplified all the subclinical isolates as expected but the clinical isolates were tested positive as well. *S. aureus* 170, 302, 1269 and 1364 were isolated by this research group from cows showing no visible signs of clinical disease or mastitis and tested positive by CMT followed by bacteriological culture of milk samples. However, the isolates *S. aureus* 75, 216, 1439, 2555, and 3909 were provided by a research institution in dairy cattle retrieved from an ancient cryobank with incomplete farm records. There is a possibility that the isolates were misidentified as clinical instead of subclinical. Other explanation for the unexpected result is the small number of bovine genomes sequenced so far that turns out to be a limitation for comparative analysis. In addition, we were not able to test our potential markers in the clinical strains RF122 and N305, originally isolated in Ireland and Canada, respectively. Therefore, it is important to consider that geographic variability is not being taking into account, since all the tested samples are from isolates coming from southern Brazil.

S. aureus 170, 302, 1269, and 1364 were isolated from subclinical mastitis by this research group but were amplified by the primers cl3309cliF/R designed for clinical isolates. When we take a close look at the nucleotide sequences shown in Supplementary Figure 1 there are complementary sequences in these isolates genomes that could anneal to the 4 last bases of cl3309cliF/R. So, these primers should be redesigned to be more specific. Again, the results were as expected for the amplification of the subclinical genomes with LipoP-F/R-CS and LipoP-F-CS/LipoP-R-C. Additionally, the comparative analysis of orthologous proteins conducted here indicated a high similarity among the genomes analyzed except for a protein member of the Major Facilitator Superfamily found to be exclusive in strains related to subclinical mastitis in the i*n silico* analysis. However, the results did not quite reproduce when isolates other than those used in the comparison were tested by PCR. Again, the small number of *S. aureus* of bovine origin sequenced so far could be a reason for this.

Studies usually explore presence or absence of virulence factors and/or their expression to find correlations with mastitis outcome (Haveri *et al.,* 2007, Le Marechal *et al.,* 2011, Veh *et al.,* 2015). Strains producing higher levls of delta-hemolysin that are more likely to be nonpersistent during either lactation or through the dry period were reported (Veh *et al.,* 2015). Also, strains expressing α-hemolysin are strongly associated with gangrenous mastitis, and α-

hemolysin mutants are less virulent in mouse and rabbit models (Zhao and Lacasse, 2008). *S. aureus* strains isolated from gangrenous ovine mastitis are commonly capable of inducing the production of high levels of α-hemolysin and LukM/LukF′-PV in the host (Rainard *et al.,* 2003; Le Marechal *et al.,* 2011a,b). Nonetheless, the pathogenesis of *S. aureus* is complex since purified leukotoxin LukM/F′ was not able to induce inflammation in the bovine mammary gland (Fromageau *et al.,* 2011), suggesting the need for other virulence factors to trigger the symptoms shown by the animals.

In this work, sequence variation was found in the hypothetical protein cl3309 and in the lipoprotein cl3700 that may possibly determine the different outcomes of mastitis. Both are transmembrane proteins, with a portion exposed on surface of the bacterial cell that is in direct contact to the host environment. These differences can affect folding or function that could ultimately cause less activation or deceive of the host the immune system explaining the absence of clinical symptoms presented by the animals presenting subclinical mastitis. Therefore, more attention should be given to sequence variation and host-pathogen interactions.

There are also cow factors that might modulate the behavior of the immune system and disease outcome. In a study with dairy herds in South Ethiopia it was identified the cow-and herd-level potential and estimated the prevalence of mastitis caused by *S. aureus* (Abebe *et al.,* 2016). They found several cow-level factors to be significantly ($p < 0.05$) associated with the presence of mastitis, such as management type, herd size, use of bedding material, breed, age, parity, stage of lactation, milk yield, udder and leg hygiene, udder position, and teat end morphology.

CONCLUSION

Adhesin genes presented a high degree of polymorphism between strains isolated from subclinical or clinical manifestation. The genes that codify the toxic shock syndrome toxin 1(*tsst-1*), bovine variant of the staphylococcal enterotoxin C (*secbov*), enterotoxin t and two homologues of streptolysin S from *Streptococcus pyogenes* were exclusively present in the clinical strain *S. aureus* RF122. We were able to find three primers that could discriminate the subclinical isolates from the clinical ones. The number of isolates will have to be expanded to account for genetic and geographic diversity. Also, these potential molecular markers can

possibly be used to discriminate *S. aureus* from other bacteria species causing bovine mastitis. This result may help in choosing the best treatment strategy for the disease in sick animals.

REFERENCES

Abebe, R., Hatiya, H., Abera, M., Megersa, B. and Asmare, K. 2016. Bovine mastitis: Prevalence, risk factors and isolation of *Staphylococcus aureus* in dairy herds at Hawassa milk shed, South Ethiopia. BMC Veterinary Research, 12:270-271.

Akindolire, M.A., Babalola, O.O. and Ateba, C.N., 2015. Detection of antibiotic resistant *Staphylococcus aureus* from milk: a public health implication. International Journal of Environmental Research and Public Health, 12:10254-10275.

Argaw, A., 2016. Review on epidemiology of clinical and subclinical mastitis on dairy cows. Food Science and Quality Management, 52:56-65.

Bukowski, M., Wladyka, B., Dubin, G. 2010. Exfoliative toxins of *Staphylococcus aureus* Toxins. Basel, 2:1148-1165.

Chua, K.Y.L., Howden, B.P., Jiang J.H., Stinear, T. and Peleg, A.Y. 2014. Population genetics and the evolution of virulence in *Staphylococcus aureus*. Infection, Genetics and Evolution, 21:554–562.

Clarke, S.R., Wiltshire, M.D., Foster and S.J. 2004. IsdA of *Staphylococcus aureus* is a broad spectrum, iron-regulated adhesin. Molecular Microbiology, 51:1509-1519.

Crowe, J. D., I. K. Sievwright, G. C. Auld and N. R. Moore, N. A. Gow, and N. A. Booth. 2003. Candida albicans binds human plasminogen: identification of eight plasminogen-binding proteins. Molecular Microbiology, 47:1637–1651.

Dinges, M.M., Orwin, P.M., Schlievert and P.M. 2000. Exotoxins of *Staphylococcus aureus*. Clinical Microbiology Reviews, 13:16–34.

Dufour, S., Dohoo, I.R., Barkema, H.W., DesCôteaux, L., DeVries, T.J., Reyher, K.K., Roy, J.-

P. and Scholl, D.T. 2012. Manageable risk factors associated with the lactational incidence, elimination, and prevalence of *Staphylococcus aureus* intramammary infections in dairy cows. Journal of Dairy Science, 95:1283–1300.

Foster, T.J., Geoghegan, J.A., Ganesh, V.K. and Höök, M. 2014. Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. Nature Reviews Microbiology, 12:49–62.

Fromageau, A., Cunha, P., Gilbert, F.B. and Rainard, P. 2011. Purified *Staphylococcus aureus* leukotoxin LukM/F′ does not trigger inflammation in the bovine mammary gland. Microbial Pathogenesis, 51:396-401.

González, R.N., Wilson and D.J. 2003. Mycoplasmal mastitis in dairy herds. Veterinary clinics of North America: Food Animal Practice, 19:199–221.

Guinane, C. M., Sturdevant, D. E., Herron-Olson, L., Otto, M., Smyth, D. S., Villaruz, A. E., Kapur, V., Hartigan, P. J., Smyth, C. J. and Fitzgerald, J.R. 2008. Pathogenomic analysis of the common bovine *Staphylococcus aureus* clone (ET3): emergence of a virulent subtype with potential risk to public health. The Journal of Infectious Diseases, 197:205–213.

Hair, P. S., M. D. Ward, O. J. Semmes, T. J. Foster, and K. M. Cunnion. 2008. *Staphylococcus aureus* clumping factor A binds to complement regulator factor I and increases factor I cleavage of C3b. Journal of Infectious Diseases, 198:125– 133.

Hammer, N.D. and Skaar, E.P. 2011. Molecular mechanisms of *Staphylococcus aureus* iron acquisition. Annual Review of Microbiology, 65:129–147.

Haveri, M., Roslöf, A., Rantala and L. Pyörälä, S. 2007. Virulence genes of bovine Staphyloccus aureus from persistent and nonpersistent intrammary infections with different clinical characteristics. Journal of Applied Microbiology, 103:993-1000.

Herron-Olson, L., Fitzgerald, J.R., Musser, J.M. and Kapur, V. 2007. Molecular correlates of host specialization in *S. aureus*. PLoS One, 2:e1120.

Higgins, J., A. Loughman, K. P. M. van Kessel, J. A. G. van Strijp, and T. J. Foster. 2006.

Clumping factor A of *Staphylococcus aureus* inhibits phagocytosis by human polymorphonuclear leucocytes. FEMS Microbiology Letters, 258: 290–296.

Holtfreter, S., Kolata, J., Broker and B. M. 2010. Towards the immune proteome of *Staphylococcus aureus* — the anti-*Staphylococcus aureus* antibody response. International Journal of Medical Microbiology, 300:176–192.

Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. and Krogh, A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Science, 12:1652–1662.

Käll, L., Krogh, A. and Sonnhammer, E. L. L. 2004. A Combined Transmembrane Topology and Signal Peptide Prediction Method. Journal of Molecular Biology, 338:1027-1036.

Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology, 305:567-580.

Kuhn, G., Francioli, P. and Blanc, D.S. 2007. Double-locus sequence typing using *clfB* and *spa*, a fast and simple method for epidemiological typing of methicillin-resistant *Staphylococcus aureus*. Journal of Clinical Microbiology, 45:54-62.

Langmead, B. and Salzberg, S. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods, 9:357-359.

Law, C.J., Maloney, P.C. and Wang, D.-N. 2008. Ins and outs of major facilitator superfamily antiporters. Annual Review of Microbiology, 62:289–305.

Le Marechal et al., 2011a: Le Marechal, C., Hernandez, D., Schrenzel, J., Even, S., Berkova, N., Thiery, R., Vautor, E., Fitzgerald, J. R., Francois, P., and Le Loir, Y. 2011. Genome sequences of two *Staphylococcus aureus* ovine strains that induce severe (strain O11) and mild (strain O46) mastitis. Journal of Bacteriology, 193:2353–2354.

Le Marechal et al., 2011b: Le Maréchal, C., Jardin, J., Jan, G., Even, S., Pulido, C., Guibert, J.-M., Hernandez ,D., François, P., Schrenzel, J., Demon, D., Meyer, E., Berkova, N., Thiéry, R., Vautor, E. and Le Loir, Y. 2011. *Staphylococcus aureus* seroproteomes discriminate

ruminant isolates causing mild or severe mastitis. Veterinary Research, 42:35-10.

Li, L., Stoeckert, C.J. and Roos, D.S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research, 13:2178-2189.

Mazmanian, S.K., Liu, G., Jensen, E.R., Lenoy, E. and Schneewind, O. 2000. *Staphylococcus aureus* sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. Proceedings of the National Academy of Sciences of the United States of America, 97:5510–5515.

Mazmanian, S.K., Skaar, E.P., Gasper, A.H., Humayun, M., Gornicki, P., Jelenska, J., Joachimiak, A., Missiakas, D.M. and Schneewind, O. 2003. Passage of heme–iron across the envelope of *Staphylococcus aureus*. Science, 299:906–909.

McCarthy, A.J. and Lindsay, J.A. 2013. *Staphylococcus aureus* innate immune evasion is lineage-specific: a bioinfomatics study. Infection, Genetics and Evolution, 19:7–14.

McCarthy, A.J. and Lindsay, J.A. 2010. Genetic variation in *Staphylococcus aureus* surface and immune evasion genes is lineage associated: implications for vaccine design and host–pathogen interactions. BMC Microbiology, 10:173-177.

Ní Eidhin, D., Perkins, S., Francois, P., Vaudaux, P., Höök, M. and Foster, T.J. 1998. Clumping factor B (ClfB), a new surface- located fibrinogen- binding adhesin of *Staphylococcus aureus*. Molecular Microbiology, 30(2):245-257.

Ote, I., Taminiau, B., Duprez, J.N., Dizier, I. and Mainil, J.G. 2011. Genotypic characterization by polymerase chain reaction of *Staphylococcus aureus* isolates associated with bovine mastitis. Veterinary Microbiology, 153:285–292.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crécy-Lagard, V., Diazm N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., et al. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Research,

33:5691-5702.

Oviedo-Boyso, J., Cardoso-Correa, B.I., Cajero-Juarez, M., Bravo-Patino, A., ValdezAlarcon, J. J. and Baizabal-Aguirre, V. M. 2008. The capacity of bovine endothelial cells to eliminate intracellular *Staphylococcus aureus* and *Staphylococcus epidermidis* is increased by the proinflammatory cytokines TNF-alpha and IL-1beta. FEMS Immunology and Medical Microbiology, 54:53-59.

Palmqvist, N., J. M. Patti, A. Tarkowski, and E. Josefsson. 2004. Expression of staphylococcal clumping factor A impedes macrophage phagocytosis. Microbes and Infection, 6:188–195.

Petersen, T.N., Brunak, S., Von, H.G. and Nielsen, H. 2011 SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods, 8:785–786.

Peton, V. and Le Loir, Y. 2014. *Staphylococcus aureus* in veterinary medicine. Infection, Genetics and Evolution, 21:602–615.

Peton, V., Bouchard D.S., Almeida S., Rault L., Falentin H., Jardin J., Jan G., Hernandez D., François P., Schrenzel J., Azevedo V., Miyoshi A., Berkova N., Even S. and Le Loir Y. 2014. Fine-tuned characterization of *Staphylococcus aureus* Newbould 305, a strain associated with mild and chronic mastitis in bovines. Veterinary Research, 45:106-120.

Piccinini, R., Tassi, R., Daprà, V., Pilla, R., Fenner, J., Carter, B. and Anjum, M.F. 2012. Study of *Staphylococcus aureus* collected at slaughter from dairy cows with chronic mastitis. Journal of Dairy Research, 79:249-255.

Proft, T. 2009. In Microbial Toxins: Current Research and Future Trends. Caister Academic Press. 192p.

Rabello, R.F., Moreira, B.M., Lopes, R.M., Teixeira, L.M., Riley, L.W. and Castro, A.C. 2007. Multilocus sequence typing of *Staphylococcus aureus* isolates recovered from cows with mastitis in Brazilian dairy herds. Journal of Medical Microbiology, 56:1505-1511.

Rainard, P., Corrales, J.C., Barrio, M.B., Cochard, T. and Poutrel, B. 2003. Leucotoxic activities of *Staphylococcus aureus* strains isolated from cows, ewes, and goats with mastitis: importance of LukM/LukF′-PV leukotoxin. Clinical and Diagnostic Laboratory

Immunology, 10:272-277.

Rhee, S.G., Chae, H.Z. and Kim, K. 2005. Peroxiredoxins: a historical overview and speculative preview of novel mechanisms and emerging concepts in cell signaling. Free Radical Biology & Medicine, 38:1543-1552.

Rollin, E., Dhuyvetter, K.C. and Overton, M.W., 2015. The cost of clinical mastitis in the first 30 days of lactation: An economic modeling tool. Preventive Veterinary Medicine, 122:257-264.

Sakwinska, O., Giddey, M., Moreillon, M., Morisset, D., Waldvogel, A. and Moreillon, P. 2011. *Staphylococcus aureus* host range and human-bovine host shift. Applied and Environmental Microbiology, 77:5908-5915.

Sasaki, T., Tsubakishita, S., Tanaka, Y., Sakusabe, A., Ohtsuka, M., Hirotaki, S. and Hiramatsu, K. 2010. Multiplex-PCR Method for Species Identification of Coagulase-Positive Staphylococci. Journal of Clinical Microbiology, 48:765–769.

Schukken, Y. H., K. E. Leslie, D. A. Barnum, B. A. Mallard, J. H. Lumsden, P. C. Dick, G. H. Vessie and M. E. Kehrli. 1999. Experimental *Staphylococcus aureus* intramammary challenge in late lactation dairy cows: quarter and cow effects determining the probability of infection. Journal of Dairy Science, 82:2393-2401.

Shim, E.H., Shanks, R.D. and Morin, D.E. 2004. Milk loss and treatment costs associated with two treatment protocols for clinical mastitis in dairy cows. Journal of Dairy Science, 87:2702-2708.

Sievers, F., Wilm, A., Dineen, D.G., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology, 7:539.

Silva, D.M., da Silva, M.P., Vidigal, P.M.P., Barcelos, R.M., Klein, R.C., Aguilar, A.P., Fabres-Klein, M.H., Oliveira, G. and Ribon, A.O.B. 2016. Draft genome sequences of *Staphylococcus aureus* strains isolated from subclinical bovine mastitis in Brazil.

Genome Announcements, 4: e00350-16.

Silva, N.C.C., Guimarães, F.F., Manzi, M.P., Budri, P.E., Gómez-Sanz, E., Benito, D., Langoni, H., Rall, V.L.M. and Torres, C. 2013 Detection, molecular characterization, and clonal diversity of methicillin- susceptible *Staphylococcus aureus* in milk of cows with mastitis in Brazil. Journal of Dairy Science, 96:6856–6862.

Silva, N.C.C., Guimarães, F.F., Manzim M.P., Budri, P.E., Gomez-Sanz, E., Benito, D., Langoni, H. and Torres, C. 2013. Molecular characterization and clonal diversity of methicillin-susceptible *Staphylococcus aureus* in milk of cows with mastitis in Brazil. Journal of Dairy Science, 96:6856-6862.

Smith, E. M., Green, L. E., Medley, G. F., Bird, H. E., Fox, L. K., Schukken, Y. H., Kruze, J. V., Bradley, A. J., Zadoks, R. N. and Dowson, C. G. 2005. Multilocus sequence typing of intercontinental bovine *Staphylococcus aureus* isolates. Journal of Clinical Microbiology, 43:4737–4743.

Smyth, D.S., Meaney, W.J., Hartigan, P.J. and Smyth, C.J. 2007. Occurrence of ssl genes in isolates of *Staphylococcus aureus* from animal infection. Journal of Medical Microbiology, 56:418-425.

Stach, C. S., Vu, B. G., Merriman, J. A., Herrera, A., Cahill, M. P., Schlievert, P. M. and Salgado-Pabón, W. 2016. Novel tissue level effects of the *Staphylococcus aureus* enterotoxin gene cluster are essential for infective endocarditis. PloS One, 11:e0154762.

Swofford, D., Olsen, G., Waddell, P. and Hillis, D. 1996. Phylogenetic Inference. Molecular Systematics, 2:407-511.

Tatusov, R. L., Natale, D. A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D. and Koonin, E.V. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Research, 29:22–28.

Taverna, F., Negri, A., Piccinini., R., Zecconi, A., Nonnis, S., Ronchi, S. and Tedeschi, G. 2007. Characterization of cell wall associated proteins of a *Staphylococcus aureus* isolated from

bovine mastitis case by a proteomic approach. Veterinary Microbiology, 119: 240–247.

Tsompanidou E., Denham E.L., Sibbald M.J., Yang X.M., Seinen J., Friedrich A.W. and Van Dijl J.M. 2012. The sortase A substrates FnbpA, FnbpB, ClfA and ClfB antagonize colony spreading of *Staphylococcus aureus*. PLoS One, 7:e44646.

Veh, A.K., Klein, R.C., Ster, C., Keefe, G., Lacasse, P., Scholl, D., Roy, J.P., Haine, D., Dufour, S., Talbot, B.G., Ribon A.O. and Malouin, F. Genotypic and phenotypic characterization of *Staphylococcus aureus* causing persistent and nonpersistent subclinical bovine intramammary-quarter infections during lactation and at dry-off. Journal of Dairy Science, 98:155–168.

Visai, L., Yanagisawa, N., Josefsson, E., Tarkowski, A., Pezzali, I., Rooijakkers, S.H., Foster, T.J. and Speziale, P. 2009. Immune evasion by *Staphylococcus aureus* conferred by iron-regulated surface determinant protein IsdH. Microbiology, 155:667-79.

Vrieling, M., Boerhout, E.M., Van Wigcheren, G.F., Koymans, K.J., Mols-Vorstermans, T.G., De Haas, C.J., Aerts, P.C., Daemen, I.J., Van Kessel, K.P., Koets, A.P. and Rutten, V.P., 2016. LukMF′ is the major secreted leukocidin of bovine *Staphylococcus aureus* and is produced in vivo during bovine mastitis. Scientific Reports, 6:37759- 37770.

Vytvytska, O., Nagy, E., Blüggel, M., Meyer, H.E., Kurzbauer, R., Huber, L.A. and Klade, C.S. 2002. Identification of vaccine candidate antigens of *Staphylococcus aureus* by serological proteome analysis. Proteomics, 2:80-590.

Whelehan, C.J., Meade, K.G., Eckersall, P.D., Young, F.J. and O'Farrelly, C. 2011. Experimental *Staphylococcus aureus* infection of the mammary gland induces region-specific changes in innate immune gene expression. Veterinary Immunology and Immunopathology, 140:181–189.

Yamasaki, O., Tristan, A., Yamaguchi, T., Sugai, M., Lina, G., Bes, M., Vandenesch, F. and Etienne, J. 2006. Distribution of the exfoliative toxin D gene in clinical *Staphylococcus aureus* isolates in France. Clinical Microbiology and Infection, 12:585-588.

Yu, N.Y., Wagner, J. R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester,

M., Foster, L. J., et al. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics, 26:1608-1615.

Zecconi, A., L. Cesaris, E. Liandris, V. Dapra, and R. Piccinini. 2006. Role of several *Staphylococcus aureus* virulence factors on the inflammatory response in bovine mammary gland. Microbial Pathogenesis, 40:177–183.

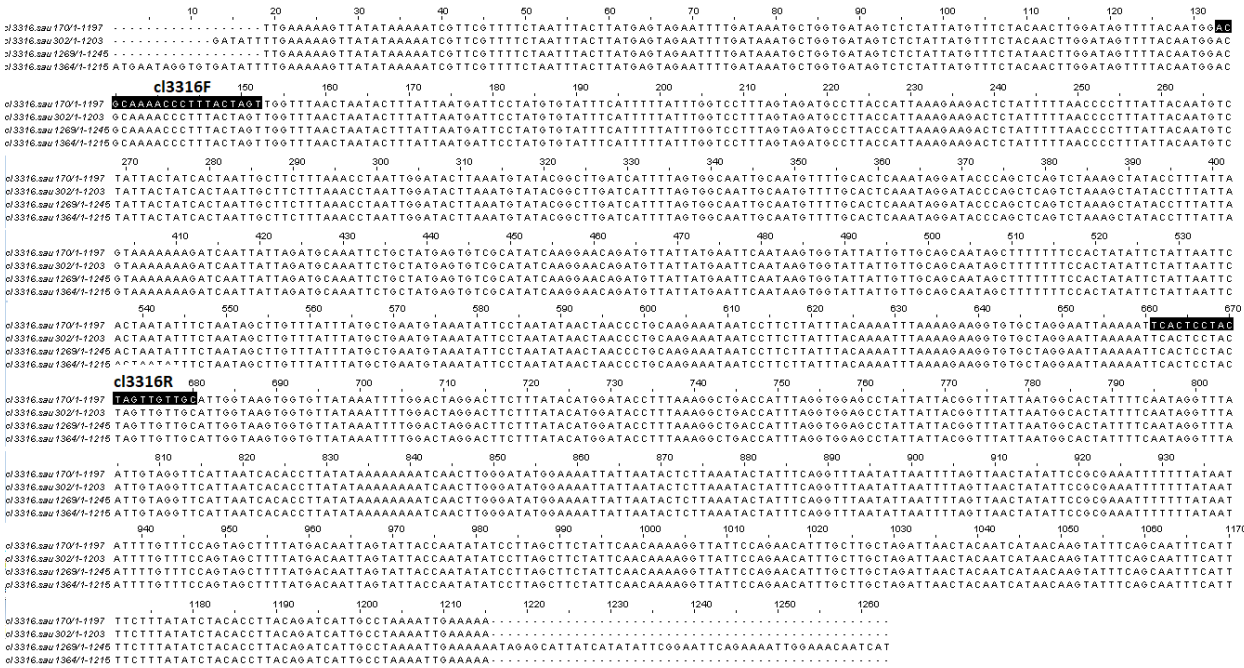
Zhao, X. and Lacasse, P. 2008. Mammary tissue damage during bovine mastitis: causes and control. Journal of Animal Science, 86:57-65.

SUPPLEMENTARY MATERIAL

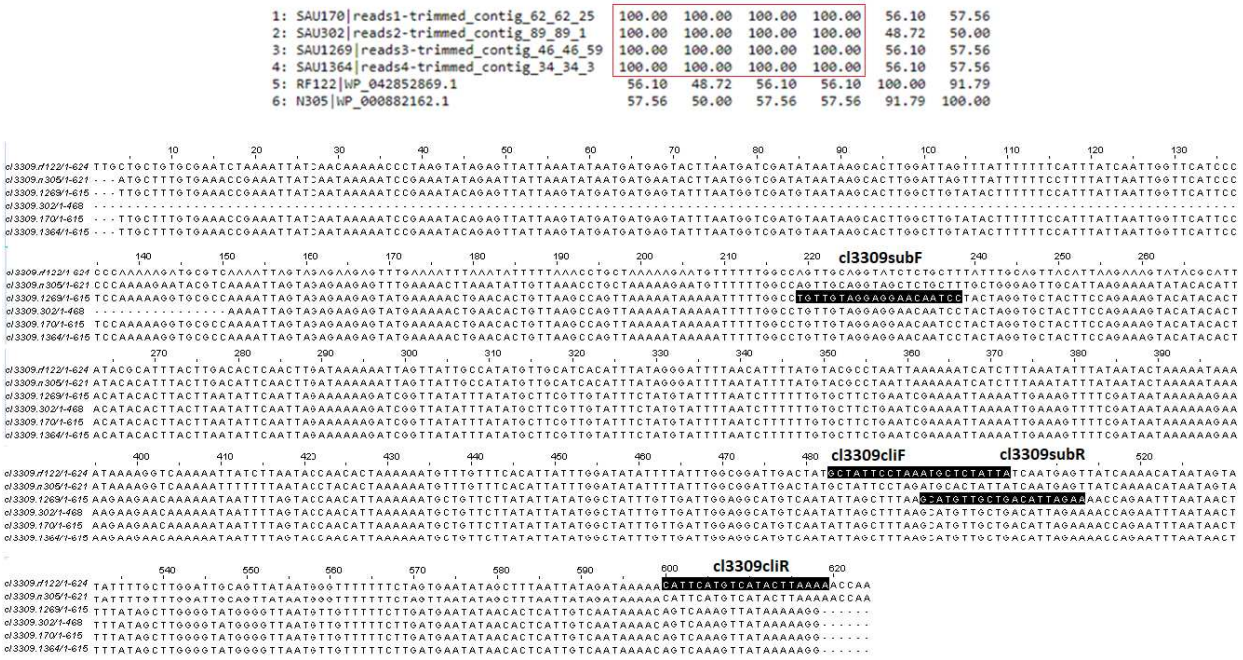
Supplementary Table 1 – Manual curation of the orthoMCL clusters possibly exclusive of the genomes of subclinical strains.

OrthoMCL Cluster number	Presence in the genomes of the subclinical strains	Annotation	Confirmation by Bowtie[a]
cluster3268:	Sau 170, Sau 302, Sau 1269, Sau 1364	2,3-bisphosphoglycerate-independent phosphoglycerate mutase [*Staphylococcus aureus* M0535]	+
cluster3269:	Sau 170, Sau 302, Sau 1269, Sau 1364	Hypothetical protein [*S. aureus*]	-
cluster3270:	Sau 170, Sau 302, Sau 1269, Sau 1364	DNA primase [*S. aureus*]	-
cluster3271:	Sau 170, Sau 302, Sau 1269, Sau 1364	Hypothetical protein SACIGC340D_2457 [*S. aureus* subsp. Aureus CIGC340D]	-
cluster3281:	Sau 170, Sau 302, Sau 1269, Sau 1364	Lactose phosphotransferase system repressor [*S. aureus* M0936]	+
cluster3282:	Sau 170, Sau 302, Sau 1269, Sau 1364	Hypothetical protein O860_00717 [*S. aureus* M0854]	-
cluster3290:	Sau 170, Sau 302, Sau 1269, Sau 1364	Smooth muscle caldesmon [*S. aureus*]	+
cluster3291:	Sau 170, Sau 302, Sau 1269, Sau 1364	Integrase [*S. aureus*]	-
cluster3297:	Sau 170, Sau 302, Sau 1269, Sau 1364	Hypothetical protein V248_02412 [*S. aureus* M32669]	-
cluster3302:	Sau 170, Sau 302, Sau 1269, Sau 1364	CAAX amino protease [*S. aureus*]	-
cluster3305:	Sau 170, Sau 302, Sau 1269, Sau 1364	ATP-dependent DNA helicase recg [*S. aureus* VET1245S]	+
cluster3309:	Sau 170, Sau 302, Sau 1269, Sau 1364	Hypothetical protein [*S. aureus*]	+
cluster3310:	Sau 170, Sau 302, Sau 1269, Sau 1364	Clumping factor B [*S. aureus* T66622]	+
cluster3316:	Sau 170, Sau 302, Sau 1269, Sau 1364	MFS transporter [*S. aureus*]	-
cluster3318:	Sau 170, Sau 302, Sau 1269, Sau 1364	Putative aminopeptidase II [*S. aureus* subsp. Aureus 21318]	+
cluster3322:	Sau 170, Sau 302, Sau 1269, Sau 1364	Hypothetical protein [*S. aureus* subsp. Aureus MW2]	-
cluster3323:	Sau 170, Sau 302, Sau 1269, Sau 1364	Hypothetical protein SAOUHSC_00600 [*S. aureus* subsp. Aureus NCTC 8325]	-

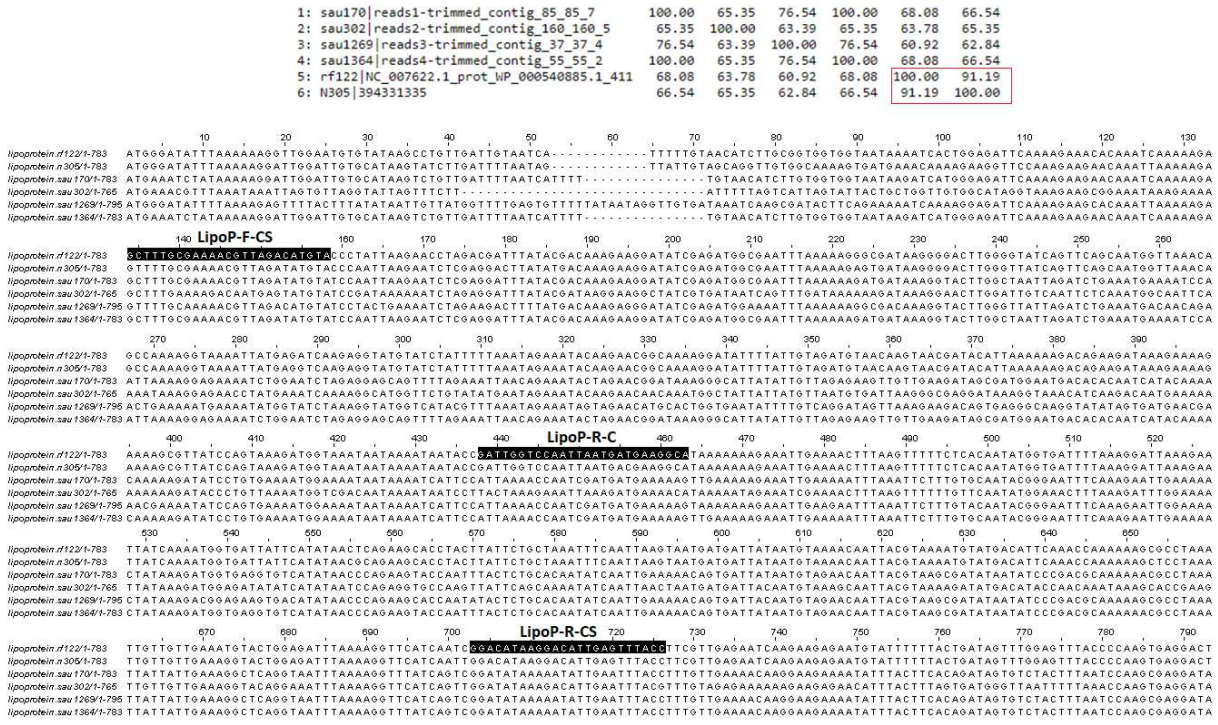
[a] -present (+) or absent (-) in the genomes of clinical strains



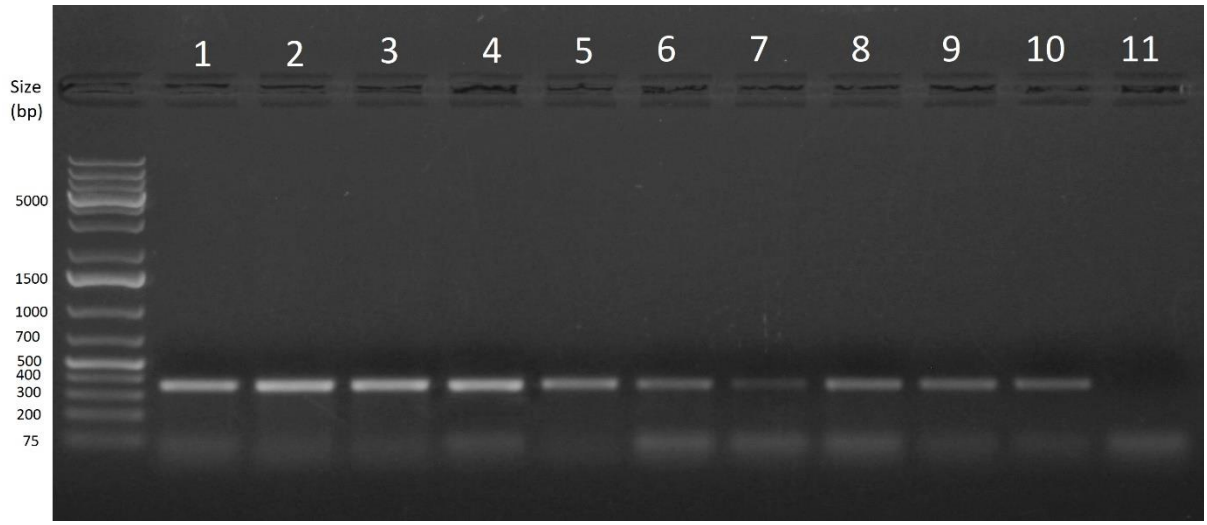
Supplementary Figure 1 – Multiple sequence alignment created with Clustal Omega of the transporter protein cl3316 present only in the genomes of the subclinical strains of *Staplylococcus aureus*. Regions used for primers design are shown.



Supplementary Figure 2 – *In silico* analyses of the hypothetical protein cl3309. At the top, the percent identity matrix created with Clustal 2.1 of cl3309, built with the amino acid sequence of *S. aureus* 170, 302, 1269, 1364, RF122, and N305. The region conserved in the subclinical strains is boxed. At the bottom, the multiple sequence alignment of cl3309, also created with Clustal Omega, and the sets of primers designed over the conserved regions among the subclinical (cl3309subF/ cl3309subR) or clinical (cl3309cliF/ cl3309cliR) strains, respectively.



Supplementary Figure 3 – *In silico* analyses of the lipoprotein cl3700. At the top, the percent identity matrix created with Clustal 2.1 of the lipoprotein, built with its the amino acid sequence for the genomes of *S. aureus* strains Sau 170, Sau 302, Sau 1269, Sau 1364, RF122, and N305. The region conserved in the clinical strains is boxed. At the bottom, the multiple sequence alignment of the lipoprotein, also created with Clustal Omega, and the designing of the sets of primers LipoP-F-CS/LipoP-R-C and LipoP-F-CS/LipoP-R-CS over the regions conserved among the clinical and subclinical strains (LipoP-F-CS and LipoP-R-CS) and over the region conserved between the clinical strains only (LipoP-R-C).



Supplementary Figure 4 – Amplification of the nuclease gene of *Staphylococcus aureus* of bovine origin. The PCR reation was perfomed with DNA extracted from the subclinical isolates *S. aureus* 170 (1), 302 (2), 1269(3), 1364 (4), 322 (5), and the clinical isolates *S. aureus* 76 (6) Sau 216 (7), Sau 1439 (8), Sau 2555 (9), Sau 3909 (10). Negative control (water) (11). Thermo Scientific O'GeneRuler 1 kb Plus DNA Ladder was used as molecular marker.