



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA E BIOCÊNCIAS

Bruno Hech Dominski

**Análise genômica de *Staphylococcus aureus* isolados de colonização e de infecções
invasivas**

Florianópolis

2022

Bruno Hech Dominski

**Análise genômica de *Staphylococcus aureus* isolados de colonização e de infecções
invasivas**

Dissertação submetida ao Programa de Pós-Graduação em Biotecnologia e Biociências da Universidade Federal de Santa Catarina para a obtenção do título de mestre em Biotecnologia e Biociências.

Orientador: Prof^a. Dr^a. Fabienne Antunes Ferreira

Coorientadora: Prof^a. Dr^a. Andrea Rita Marrero

Florianópolis

2022

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Dominski, Bruno Hech Dominski

Análise genômica de *Staphylococcus aureus* isolados de colonização e de infecções invasivas / Bruno Hech Dominski Dominski ; orientadora, Fabienne Antunes Ferreira, coorientadora, Andrea Marrero, 2022.

55 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências Biológicas, Programa de Pós Graduação em Biotecnologia e Biociências, Florianópolis, 2022.

Inclui referências.

1. Biotecnologia e Biociências. 2. Genômica. 3. *Staphylococcus aureus*. I. Antunes Ferreira, Fabienne . II. Marrero, Andrea. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Biotecnologia e Biociências. IV. Título.

Bruno Hech Dominski

**Análise genômica de *Staphylococcus aureus* isolados de colonização e de infecções
invasivas**

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca
examinadora composta pelos seguintes membros:

Prof. Dr. Ricardo Ruiz Mazzon
Universidade Federal de Santa Catarina

Profa. Dra. Marília Miotto
Universidade Federal de Santa Catarina

Profa. Dra. Fabienne Antunes Ferreira
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi
julgado adequado para obtenção do título de mestre em Biotecnologia e Biociências

Prof. Dr. Glauber Wagner
Coordenação do Programa de Pós-Graduação

Profa. Dra. Fabienne Antunes Ferreira
Orientadora

Florianópolis, 2022

Esse trabalho é dedicado aos meus pais e amigos.

AGRADECIMENTOS

À minha orientadora Fabienne, sua orientação e ajuda durante todo o percurso foram vitais para a conclusão do presente trabalho. Ela é uma cientista exemplar e amiga muito dedicada, querida e iluminada. Sou eternamente grato e me sinto muito honrado desde que entrei no laboratório para meu trabalho de conclusão de curso.

Aos meus pais e irmão, por todo o apoio e por me darem todas as condições possíveis para que eu seguisse meu caminho na vida. Sem vocês, nada disso teria sido possível.

Aos amigos do Laboratório de Genética Molecular de Bactérias (GeMBac), professor Ricardo, Maria, Amanda, Sayonara e Marcel, por todas as conversas, desabafos e discussões científicas que muito contribuíram para minha formação.

Aos meus amigos Marcelo, Sofia, Vini, Rennan e Andie por estarem comigo durante muitos anos e mantermos uma amizade muito importante para mim.

Aos meus amigos mais distantes fisicamente, Matheus e Juliano, pela companhia diária e todo o suporte durante esses anos pandêmicos.

Aos meus amigos do trabalho, Camila, Dina, Victória, Carime, Giovana, Rafa e em especial ao Madson, Matheus, Vanessa e João, por sempre me incentivarem e tornarem meus dias mais agradáveis.

Aos professores Andrea, Marcos e Jussara pelas importantes contribuições científicas a esse trabalho.

À banca examinadora desta dissertação.

A todos os professores, alunos e laboratórios que de alguma forma contribuíram para a minha formação e realização dessa etapa.

A CAPES pelo suporte financeiro.

RESUMO

Staphylococcus aureus é uma bactéria de grande importância clínica mundial, devido, principalmente, à sua potencial patogenicidade e multirresistência a diversos antimicrobianos na forma de *Staphylococcus aureus* resistentes à meticilina (MRSA). Cerca de 30% da população humana mundial é colonizada por esta bactéria nas fossas nasais, de forma persistente e assintomática. Em cerca de 80% dos casos a mesma cepa que causa o processo infeccioso já colonizava previamente o indivíduo. Porém, os mecanismos dessa transição colonização-infecção não estão completamente elucidados. Sabe-se que se trata de um processo multifatorial, envolvendo fatores da bactéria, do ambiente/microbiota e do hospedeiro. De forma a compreender melhor os fatores bacterianos associados a estes mecanismos, a hipótese do presente estudo foi de que há um ou mais fatores genéticos bacterianos que podem ser determinantes nesta transição, visto que as bactérias podem adquirir ou perder genes de resistência e/ou virulência principalmente via elementos genéticos móveis. Assim, o presente estudo investigou se existem um ou mais potenciais marcadores genéticos de infecção através da análise genômica de 415 genomas completamente sequenciados de *S. aureus*. Os genomas foram agrupados em dois grandes grupos: (1) genomas de cepas isoladas de colonização e (2) genomas de cepas isoladas de infecções invasivas, entre outros subgrupos. Através de análises *in silico*, os dois grandes grupos foram comparados com relação à (i) análise pangênômica, (ii) alinhamento genômico completo, (iii) polimorfismos de único nucleotídeo e (iv) distribuição dos genes de virulência, resistência aos antimicrobianos e elementos genéticos móveis. As análises iniciais (i, ii e iii) indicaram que os genomas apresentam suas particularidades como genomas únicos, sem diferenças que predominam em um dos dois grandes grupos. A partir de análise estatística, a distribuição dos genes de virulência indicou maior prevalência dos genes *sea*, *tst* e genes do *cluster egc* (*seg*, *seu*, *sem*, *sen*, *sei* e *seo*) nos genomas de infecção invasiva. Porém, não foram identificados genes de resistência aos antimicrobianos mais prevalentes em um dos dois grupos. Apenas o plasmídeo rep7C foi encontrado com maior frequência nos genomas de colonização, enquanto a sequência de inserção (IS) ISLgar5 foi encontrado mais comumente nos genomas de infecção. Apesar de alguns dos genes encontrados nos genomas analisados já terem sido correlacionados com amostras de *S. aureus* de infecção, este estudo apresenta limitações e vieses importantes que não permitem estabelecer um marcador genômico de infecção apenas com os dados

analisados. Mais estudos precisam ser realizados para confirmar a relevância destes genes em amostras isoladas de infecções estafilocócicas. Adicionalmente, a literatura científica vem evidenciando que modificações genômicas em *S. aureus* parecem não ser cruciais nos mecanismos de transição colonização-infecção, uma vez que, em geral, não se observada diferenças significativas entre os dois grupos. Outras abordagens metodológicas como transcriptoma e proteoma podem ser sugeridos de para entender melhor os mecanismos da transição.

Palavras-chave: *Staphylococcus aureus*; MRSA; análise genômica; virulência; transição colonização-infecção

ABSTRACT

Staphylococcus aureus is a bacterium of great clinical importance worldwide, mainly due to its potential pathogenicity and multi-resistance to several antimicrobials in the form of methicillin-resistant *Staphylococcus aureus* (MRSA). About 30% of the world's human population is persistently and asymptotically colonized by this bacterium in the nasal cavity. In around 80% of cases, the same strain that causes the infectious process has previously colonized the individual. However, the mechanisms of this colonization-infection transition are not completely elucidated. It is known that it is a multifactorial process, the factors of the bacteria, the environment/microbiota and the host. In order to better understand the bacterial factors associated with these mechanisms, the hypothesis of the present study was that there are one or more bacterial genetic factors that can be decisive in this transition, since bacteria can acquire or lose resistance and/or virulence genes, mainly via mobile genetic elements. Thus, the present study investigated whether there are one or more potential genetic markers of infection through the genomic analysis of 415 completely sequenced *S. aureus* genomes. The genomes were grouped into two large groups: (1) genomes of strains isolated from colonization and (2) genomes of strains isolated from invasive infections, among other subgroups. Through *in silico* analyses, the two large groups were compared with respect to (i) pangenomic analysis, (ii) complete genomic alignment, (iii) single nucleotide polymorphisms (SNPs) and (iv) distribution of virulence genes, antimicrobial resistance and mobile genetics. The initial analyzes (i, ii and iii) indicated that the genomes present their particularities as unique genomes, without differences that predominate in one of the two large groups. Based on statistical analysis, the distribution of virulence genes indicated a higher prevalence of *sea*, *tst* and *egc* cluster genes (*sec*, *seu*, *sem*, *sen*, *sei* and *seo*) in the genomes of invasive infection. However, antimicrobial resistance genes with higher prevalence were not identified in one of the two main groups. Only the rep7C plasmid was found more frequently in colonization genomes, while the insertion sequence (IS) ISLgar5 was found more commonly in infection genomes. Although some of the genes found in the analyzed genomes have already been correlated with *S. aureus* isolated from infection, this study has important limitations and biases that do not allow establishing a genomic marker of infection with the analyzed data alone. More studies need to be carried out to confirm the relevance of these genes in samples isolated from staphylococcal infections. Additionally, the scientific literature has shown that genomic changes in *S. aureus* do not seem to be crucial in the colonization-infection transition mechanisms, since, in general, no significant differences are observed between the two main groups. Other methodological approaches such as transcriptome and proteome can be suggested to better understand the transition mechanisms.

Keywords: *Staphylococcus aureus*; MRSA; genomic analysis; Virulence; colonization-infection transition

LISTA DE FIGURAS

Figura 1-Gráficos representativos da porcentagem de MRSA ou MSSA entre os genomas sequenciados e a porcentagem dos tipos de SCCmec identificados nos genomas associados a amostras MRSA. (A) representa os genomas associados à colonização: 11 genomas (33,3%) são MSSA enquanto 22 genomas (66,7%) são MRSA, com a predominância do SCCmec IV (16 genomas; 48%). (B) representa os genomas associados à infecção invasiva: 25 genomas (42,4%) são MSSA e 34 (57,6%) são MRSA. Os gráficos foram confeccionados no Microsoft Power Point (Microsoft®). MSSA: Methicilin-Susceptible *Staphylococcus aureus*. MRSA: Methicilin-Resistant *Staphylococcus aureus*. SCCmec: staphylococcal cassette chromosome mec..... 29

Figura 2-Gráfico representativos dos ST (*Sequencing typing*; derivados do MLST) de genomas de *S. aureus* sequenciados a partir de amostras de colonização. (A) genomas associados à colonização. Os STs associados as amostras de colonização são: ST8 (18), ST239 (3), ST30(2), ST121 (2), ST228 (2). “Outros” STs são: ST5, ST15, ST50, ST59, ST254 e ST464 cada um com apenas um genoma. (B) genomas associados à infecção invasiva. STs relacionados a amostras de infecção, sendo: ST5 (8), ST8 (7), ST15 (5), ST239 (5), ST1 (4), ST22(3) e ST121 (3). “Outros” nos STs de infecção são: ST25 (2), ST59 (2), ST72 (2), ST105 (2), ST228 (2), ST254 (2), ST2389 (2), ST 30 (1), ST36 (1), ST45 (1), ST93 (1), ST152 (1), ST237 (1), ST240 (1), ST 246 (1), ST291 (1) e ST398 (1). O número entre parênteses representa o número de genomas pertencente a cada ST. Os gráficos foram confeccionados no Microsoft Power Point (Microsoft®). 30

Figura 3- Diagrama de Venn representando a distribuição dos STs (*Sequence Typing*) dos 92 genomas selecionados entre os continentes (América, Ásia, Europa e Oceania). Alguns ST foram encontrados exclusivamente em um dos continentes (Ex.: ST15 no continente Europeu), enquanto outros são encontrados em mais de 1 continente (Ex.: ST5 foi encontrado no continente americano e asiático; indicando a área sobreposta) ou em todos os continentes (Ex.: ST239; sobreposição de todas as áreas do diagrama). O Programa Microsoft Power Point (Microsoft®) foi utilizado para a confecção do diagrama. 31

Figura 4-Análise pan-genômica considerando os 92 genomas de *S. aureus* selecionados. O gráfico representa o número de genes compartilhados (eixo y) relacionado ao número de genomas (eixo x). Cada coluna representa o número de genes compartilhado entre dado número de genomas. O *core* genoma (presente em todas as amostras) é a última coluna com 2415 genes. Gráfico obtido a partir da análise realizada pelo Roary, conforme descrito na metodologia..... 32

Figura 5-Gráfico representando o pangenoma dos genomas ST239 (8 genomas). O *core* genoma está representado em azul escuro, com 2415 genes; o *shell* genoma (genes compartilhados com apenas um genoma ou com até 7 outros genomas) está representado na cor lilás mais escuro e o *cloud* genoma (genes compartilhados apenas com um outro genoma) na cor lilás mais claro. Gráfico obtido a partir da análise realizada pelo Roary, conforme descrito na metodologia. 33

Figura 6-Matriz relacionada a ausência e presença dos genes de cada amostra do ST239 e sua relação filogenética. Gráfico obtido a partir da análise realizada pelo Roary, conforme descrito na metodologia..... 34

Figura 7-Representação do alinhamento dos 8 genomas completos pertencentes ao ST239, destacando os genomas do grupo colonização (em azul; 3 genomas) e do grupo infecção invasiva (em vermelho; 5 genomas). Todos os genomas são de MRSA. Figura obtida diretamente do MAUVE, apenas com a adição do nome do genoma. 35

Figura 8-Representação do alinhamento dos 25 genomas completos pertencentes ao ST239, destacando os genomas do grupo colonização (em azul; 18 genomas) e do grupo infecção invasiva (em vermelho; 7 genomas). Os genomas que são de MRSA estão sublinhados. Figura obtida diretamente do MAUVE, apenas com a adição do nome do genoma. 36

Figura 9-Árvores filogenéticas construída com base em diferenças em SNPs em todos os genomas do estudo. (A) Árvore obtida através do CSI Phylogeny. (B) Árvore obtida através do RealPhy. Genomas em vermelho escuro representam os genomas de infecção invasiva MRSA, vermelho claro genomas de infecção invasiva MSSA, azul escuro genomas de colonização MRSA e por fim em azul claro genomas de colonização MSSA. 38

LISTA DE TABELAS

Tabela 1 - Genomas completos de *S. aureus* sequenciados a partir de amostras de colonização. Os 33 genomas foram selecionados a partir do banco de dados GenBank e agrupados de acordo com o ST (obtido pelo MLST). Além do ST, a tabela inclui o nome do genoma ou amostra, o número de acesso Genbank, ano de isolamento da bactéria, tecnologia de sequenciamento utilizada, sítio biológico de isolamento da bactéria, classificação como MRSA ou MSSA, tipagem SCCmec e região geográfica de isolamento. 26

Tabela 2- Genomas completos de *S. aureus* sequenciados a partir de amostras de infecção invasiva. Os 59 genomas foram selecionados a partir do banco de dados GenBank e agrupados de acordo com o ST (obtido pelo MLST). Além do ST, a tabela inclui o nome do genoma ou amostra, o número de acesso Genbank, ano de isolamento da bactéria, tecnologia de sequenciamento utilizada, sítio biológico de isolamento da bactéria, classificação como MRSA ou MSSA, tipagem SCCmec e região geográfica de isolamento. 28

Tabela 3-Diversidade de genes de virulência identificados nos genomas segundo o programa VirulenceFinder. Os genomas estão agrupados entre o grupo de colonização (azul) e o grupo de infecções invasivas (vermelho); primeira coluna. O ST (*Sequence Typing*) e o perfil de sensibilidade à meticilina (S: sensível; MSSA; R: resistente; MRSA) de cada genoma estão indicados; segunda e terceira colunas, respectivamente. Os quadrados pintados em verde indicam a presença do gene. Grupo de genes foram separados conforme função de acordo com o proposto pelo VirulenceFinder: Imun. Hosp.: Imunidade do hospedeiro; Exoenzima ou toxina.....40

Tabela 4-Diversidade de genes de resistência aos antimicrobianos identificados nos genomas segundo o programa ResistanceFinder. Os genomas estão agrupados entre o grupo de colonização (azul) e o grupo de infecções invasivas (vermelho); primeira coluna. O ST (*Sequence Typing*) e o perfil de sensibilidade à meticilina (S: sensível; MSSA; R: resistente; MRSA) de cada genoma estão indicados; segunda e terceira colunas, respectivamente. Os quadrados pintados em verde indicam a presença do gene. Grupo de genes foram separados conforme classe de antimicrobianos.....42

Tabela 5-Diversidade de plasmídeos presentes nos genomas do estudo segundo o programa PlasmidFinder. Em azul estão os genomas associados à colonização e, em vermelho, os genomas associados à infecções invasivas; primeira coluna. O ST (*Sequence Typing*) e o perfil de sensibilidade à meticilina (S: sensível; MSSA; R: resistente; MRSA) de cada genoma estão indicados na segunda e terceira colunas, respectivamente. Os quadrados pintados em verde indicam a presença do gene. Grupo de plasmídeos foram separados conforme o proposto pelo PlasmidFinder.....43

Tabela 6-Diversidade de sequências de inserção presentes nos genomas do estudo segundo o programa ISFinder. Em azul estão os genomas associados à colonização e, em vermelho, os genomas associados à infecções invasivas; primeira coluna. O ST (*Sequence Typing*) e o perfil de sensibilidade à meticilina (S: sensível; MSSA; R: resistente; MRSA) de cada genoma estão indicados na segunda e terceira colunas, respectivamente. Os quadrados pintados em verde indicam a presença do gene. Grupo de genes foram separados conforme o proposto pelo ISFinder.....45

LISTA DE ABREVIATURAS E SIGLAS

- ACME – elemento genético catabólico de arginina (do inglês: *arginine-catabolic mobile elemento*)
- CC- Complexo clonal (do inglês: *clonal complex*)
- CGE - *Center for Genomic Epidemiology*
- EGM – Elementos Genéticos Móveis
- FnBP- Proteína ligante de fibronectina (do inglês: *fibronectin-binding protein*)
- HLA – alpha-hemolisina
- IS – Sequência de inserção
- MLST- Tipagem de sequências multiloco (do inglês: *Multilocus sequence typing*)
- MRSA- *Staphylococcus aureus* resistentes à Meticilina (do inglês: *Methicilin-resistant Staphylococcus aureus*)
- MSSA- *Staphylococcus aureus* sensíveis à Meticilina (do inglês: *Methicilin-susceptible Staphylococcus aureus*)
- NCBI - *National Center for Biotechnology Information*
- ONT – *Oxford Nanopore Technologies*
- PBP2a- Proteína ligante de Penicilina 2a (do inglês: *Penicilin-binding protein 2a*)
- PFGE- Eletroforese em gel submetido a campos elétricos alternados (do inglês: *Pulsed-Field Gel Electrophoresis*)
- PVL- Leucocidina de Panton-Valentine (do inglês: *Panton-Valentine Leukocidin*)
- SaPI - ilhas de patogenicidade estafilocócicas (do inglês: *staphylococcal pathogenicity islands*)
- SCC*mec*- Cassete cromossômico estafilocócico *mec* (do inglês: *Staphylococcal cassette chromosome mec*)
- SE- Enterotoxinas estafilocócicas (do inglês: *Staphylococcal enterotoxins*)
- SNP - polimorfismos de único nucleotídeo (do inglês: *single nucleotide polymorphism*)
- ST- sequência tipo (do inglês: *sequence typing*)
- TSST-1- Toxina da síndrome do choque tóxico-1 (do inglês: *Toxic shock syndrome toxin-1*)
- VRSA- *Staphylococcus aureus* resistentes à Vancomicina (do inglês: *Vancomycin-resistant S.*)

SUMÁRIO

INTRODUÇÃO	12
1.1 <i>Staphylococcus aureus</i>	12
1.2 Colonização da cavidade nasal.....	12
1.3 Patogenicidade dos <i>S. aureus</i>	14
1.4 <i>Staphylococcus aureus</i> Resistentes à Meticilina (MRSA).....	17
1.5 Utilização do sequenciamento genômico como ferramenta para análises biológicas	18
2 OBJETIVOS	21
2.1 Objetivo Geral	21
2.2 Objetivos Específicos	21
3 METODOLOGIA	22
3.1 Seleção dos genomas de <i>S. aureus</i>	22
3.2 Análises Genômicas	22
3.2.1 Análise dos genes de virulência	23
3.2.2 Análise dos genes de resistência aos antimicrobianos	23
3.2.3 Análise de elementos genéticos móveis (EGM)	24
3.3 Análise estatística.....	24
4 RESULTADOS	25
4.1 Características gerais dos Genomas de Colonização.....	25
4.2 Características gerais dos genomas de Infecção invasiva.....	27
4.3 Análise pan-genômica.....	31
4.4 Alinhamento dos genomas	34
4.5 Análise filogenética com base em alterações de SNPs.....	37
4.6 Distribuição e análise dos genes de virulência.....	38
4.7 Distribuição e análise dos genes de Resistência.....	41
4.8 Distribuição e análise de Elementos genéticos móveis	43
5 DISCUSSÃO	46
REFERÊNCIAS	52

INTRODUÇÃO

1.1 *Staphylococcus aureus*

O gênero *Staphylococcus* é composto por cerca de 65 espécies e subsespécies de bactérias Gram-positivas, com membros de grande importância nas áreas de saúde humana e animal, assim como na produção de alimentos. Estes microrganismos de formato esférico (cocos) são imóveis, anaeróbios facultativos, catalase positivos, capazes de crescer em variações de temperatura de 18° a 40°C e em concentrações de até 10% de cloreto de sódio. São encontrados em diversos ambientes, desde a pele e membranas mucosas de humanos, até solo, água e no ar (BITRUS, 2018). Podem ser divididos bioquimicamente em dois grupos, de acordo com a produção da enzima coagulase: *Staphylococcus* coagulase positiva ou negativa. Entre as espécies não produtoras de coagulase (*Staphylococcus* coagulase negativos; SCoN) de importância médica podemos destacar: *Staphylococcus lugdunensis*, *Staphylococcus saprophyticus*, *Staphylococcus hominis*, *Staphylococcus haemolyticus* e *Staphylococcus epidermidis*. A bactéria coagulase positiva de maior importância médica para humanos é *Staphylococcus aureus* (MURRAY; ROSENTHAL; PFALLER, 2009) sendo um dos principais causadores de infecções hospitalares e comunitárias. Além de doenças em seres humanos, *S. aureus* também pode causar mastites em bovinos, dermatites em cães e septicemia e artrite em aves (BITRUS, 2018). As infecções variam desde lesões na pele, como foliculites e furúnculos até infecções mais invasivas como infecção de corrente sanguínea, endocardite, osteomielite e pneumonia (LI, 2018).

Cerca de 30% da população humana é persistentemente colonizada por *S. aureus* na cavidade nasal de forma assintomática (KRISMER, 2017). Portadores prévios dessa bactéria que estejam hospitalizados ou imunocomprometidos possuem maior risco de desenvolverem infecções de corrente sanguínea do que não portadores (EIFF, 2001). Além disso, a cepa que o indivíduo carrega é a que causa a infecção em cerca de 80% dos casos, fazendo com que a colonização prévia seja considerada um fator de risco para o desenvolvimento de uma infecção (LAKHUNDI; ZHANG, 2018).

1.2 Colonização da cavidade nasal

A colonização do vestíbulo nasal por *S. aureus* é um processo multifatorial, envolvendo a imunidade do hospedeiro, interações com a microbiota local e fatores proteicos e não-proteicos expressos pela bactéria (MULCAHY; MCLOUGHLIN,

2016). A primeira linha de defesa do organismo nas narinas são as células do epitélio, os queratinócitos, que podem expressar receptores, peptídeos antimicrobianos e quimiocinas. O receptor *Toll-like 2* (TLR-2) já foi identificado como importante para a colonização em modelo animal. Isso foi observado em um estudo de Quinn e colaboradores em que cepas persistentes de carreadores promoveram uma ativação mais lenta dos TLR-2 em comparação com cepas não associadas à colonização persistente. No mesmo trabalho, constatou-se que falhas na produção de peptídeos antimicrobianos por alterações na expressão de genes do hospedeiro também estão associadas com colonização nasal persistente, uma vez que cepas colonizadoras regulavam negativamente a expressão de β -defensina-3, um peptídeo potente contra estafilococos (QUINN; COLE, 2007).

A microbiota anfibiônica humana também pode interferir no sucesso da colonização, competindo por nutrientes, receptores do epitélio e por meio da produção de antimicrobianos (bacteriocinas) pelas espécies presentes na cavidade nasal. Já foi relatado que os *Staphylococcus spp.* sintetizam essas moléculas com alta frequência e possuem amplo espectro de ação (JANEK, 2016). *Corynebacterium pseudodiphtheriticum* e *Micrococcus luteus* são exemplos de duas espécies que tem seu crescimento inibido pelas bacteriocinas produzidas pelos *Staphylococcus spp.*, sugerindo que sejam grandes competidores (JANEK, 2016). Outro exemplo que merece destaque é a produção de lugdunina pelo *S. lugdunensis*, capaz de inibir e matar *S. aureus*, de modo que humanos colonizados por essa espécie tenham seis vezes menos chance de carrear *S. aureus* em relação a pessoas que não são colonizadas (KRISMER, 2017). Além das bacteriocinas, outras moléculas também podem interagir com *S. aureus*, como por exemplo serina proteases de *Staphylococcus epidermidis*, que foram correlacionadas negativamente com a colonização de *S. aureus* em humanos e ratos, segundo um estudo por Iwase e colaboradores (2010).

Quanto a fatores produzidos pela bactéria, podemos citar adesinas, como o determinante de superfície regulado por ferro A (IsdA) e o fator *clumping B*, que são importantes na ligação com as proteínas loricrina e citoqueratina 10 dos queratinócitos (MULCAHY; MCLOUGHLIN, 2016). O fator *clumping B* tem grande envolvimento na adesão ao epitélio na colonização de roedores e foi a única adesina de superfície que tem papel comprovado na colonização nasal *in vivo* em humanos (WERTHEIM, 2008). Outras proteínas também podem estar envolvidas na colonização, como a proteína de superfície SasX, que foi identificada como fator crítico para ocupação do sítio nasal,

porém seu ligante não foi descoberto ainda (MULCAHY; MCLOUGHLIN, 2016). Além disso, como outras proteínas de superfície de *S. aureus*, SasX parece ter um papel na formação de biofilme e evasão do sistema imune em algumas cepas, fazendo com que seja considerada um fator de virulência e participe nos quadros clínicos causados pela bactéria (DE BACKER et al., 2018).

1.3 Patogenicidade dos *S. aureus*

Staphylococcus aureus é um patógeno em potenciale o principal responsável pela maioria das infecções de pele e tecidos moles (furunculose, impetigo e celulite infecciosa), bem como está envolvido em diversos casos de osteomielites, artrite séptica, endocardites e infecções de corrente sanguínea, nos quais muitos estão associados a utilização de dispositivos médicos implantáveis, como catetes, tubos respiratórios e próteses cardíacas e ortopédicas (TONG, 2015). Inicialmente confinados ao ambiente hospitalar, na metade dos anos 90 houve um aumento no número de casos comunitários. Esses últimos são as infecções adquiridas por pessoas que não apresentam fatores de risco clássicos (imunocomprometimento, extremos de idade, doenças crônicas, uso de dispositivos médicos implantáveis, etc.), tendendo a afetar indivíduos mais jovens e saudáveis (DAVID; DAUM, 2010). Atualmente é difícil distinguir claramente as origens epidemiológicas das amostras, visto que cepas comunitárias vem sendo encontradas em hospitais e linhagens nosocomiais circulam na comunidade, demonstrando uma alta capacidade de adaptação destes microrganismos (FIGUEIREDO; FERREIRA, 2014).

Assim como a colonização, a patogenicidade é multifatorial e está relacionada, ao menos em parte, a um conjunto de fatores de virulência produzidos pela bactéria. Os principais grupos desses fatores são as proteínas de superfície e as exoproteínas (DEINHARDT-EMMER, 2018). Tal qual o exemplo citado no item anterior, as proteínas de superfície auxiliam no estabelecimento da bactéria no hospedeiro, com destaque para as proteínas ligantes de fibronectina (*fibronectin-binding proteins*; FnBPs), ligadas principalmente à adesão e formação de biofilme (FOSTER, 2016); a adesina de colágeno (Cna), que previne a ativação da via clássica do sistema complemento, além de contribuir com a adesão a tecido rico em colágeno (FOSTER et al., 2014); proteína A que inibe a opsonofagocitose e promove mecanismo de inflamação (FOSTER et al., 2014). Das exoproteínas destacam-se clinicamente a Alpha hemolisina (HLA), que forma poros nas membranas de diversos tipos celulares

ao se ligar com o receptor ADAM10, lisando-as e promovendo edema tecidual (BERUBE; WARDENBURG, 2013); as leucocidinas, com todas as cepas de *S. aureus* possuindo pelo menos três delas (CHEUNG; BAE; OTTO, 2021) e as enterotoxinas estafilocócicas (SE), associadas principalmente à intoxicação alimentar. Algumas enterotoxinas ainda atuam como superantígenos, estimulando grande produção de citocinas e induzindo febre alta (HU; NAKANE, 2014).

Muitos dos genes codificantes destes fatores de virulência e também de resistência são adquiridos horizontalmente via elementos genéticos móveis (EGM), como elementos genéticos associados a bacteriófagos, ilhas gênicas, transposons e plasmídeos. Esses elementos constituem o chamado genoma acessório/ auxiliar e tem sido o foco de muitos estudos para o entendimento do surgimento de padrões de resistência e virulência complexos (JANI, 2017), visto que quase todas as cepas de *S. aureus* possuem pelo menos um desses EGM (COPIN; SHOPSIN; TORRES, 2018). Acredita-se que a alta variabilidade desse genoma acessório seja a razão do sucesso adaptativo da bactéria a diversos ambientes e hospedeiros (HAABER; PENADÉS; INGMER, 2017).

Os fagos ou bacteriófagos são vírus que infectam bactérias e que podem integrar segmentos do genoma viral no genoma da célula hospedeira. O material genético viral pode carrear genes de resistência e/ou virulência, como o fator de virulência denominado leucocidina de Pantone Valentine (PVL), que já foi associado epidemiologicamente a síndromes clínicas específicas, como pneumonia necrotizante e casos graves de infecção de pele e tecidos moles (COPIN; SHOPSIN; TORRES, 2018). Após infectar um hospedeiro, o fago integra seu genoma no genoma bacteriano e passa a ser denominado como profago. A excisão desse profago pode ocorrer espontaneamente ou via ativação após dano no DNA bacteriano ou resposta à exposição a antimicrobianos (HAABER; PENADÉS; INGMER, 2017). Após a excisão, o profago entra em ciclo lítico, produzindo mais fagos e lisando a célula. Durante a montagem dos capsídeos, segmentos do DNA bacteriano podem ser empacotados, passando a ser chamado de partícula transdutora. A transferência de material genético por fagos transdutores é intitulada transdução e ocorre quando um fago, carreando material genético bacteriano de um processo infeccioso anterior, invade uma nova célula bacteriana e entra em ciclo lítico (HAABER; PENADÉS; INGMER, 2017). A partir desse ciclo, o DNA da primeira bactéria infectada pode se recombinar com o DNA da nova bactéria hospedeira, fazendo com que adquira novas

informações genéticas. Além disso, podem mobilizar outros EGMs, como plasmídeos e ilhas de patogenicidade estafilocócicas (SaPIs- do inglês *staphylococcal pathogenicity islands*). Essas últimas não possuem mecanismos para sua transmissão horizontal e se utilizam dos capsídeos de fagos (conhecidos como *helper phages*) para sua transdução (NOVICK; CHRISTIE; PENADÉS, 2010). Fagos são muito presentes nas linhagens de *S. aureus*, com todos os genomas sequenciados até 2018 contendo ao menos um fago (COPIN; SHOPSIN; TORRES, 2018). Já foi demonstrada experimentalmente a aquisição de genes de resistência à tetraciclina, eritromicina e cloranfenicol via bacteriófagos (HAABER, 2016).

Os plasmídeos são segmentos de DNA extracromossômico de tamanho variável (desde kilobases até megabases) com capacidade de auto replicação, podendo ser transmitido verticalmente (na fissão binária), ou horizontalmente, sendo conjugado por outra célula (PARTRIDGE, 2018). Além dos genes necessários para sua transmissão e replicação, os plasmídeos podem carrear genes de resistência. O surgimento de cepas de *S. aureus* resistentes à vancomicina (VRSA; do inglês: *Vancomycin-resistant S. aureus*), está relacionado à aquisição de um plasmídeo de *Enterococcus faecalis* contendo o operon *vanA* de resistência a esse antimicrobiano (TURNER, 2019).

As ilhas gênicas são segmentos de DNA que carregam genes que impactam no *fitness* (aptidão, capacidade de se adaptar a diferentes ambientes) da bactéria (NOVICK; RAM, 2016). Geralmente contém genes codificando para integrases ou transposases e em *S. aureus* 3 famílias já foram descritas, cada uma flanqueada por um gene de transposase truncado (DOBRINDT, 2004; TURNER, 2019). Quanto à virulência, essas ilhas geralmente contém um ou dois genes codificantes de superantígenos, como a toxina da síndrome do choque tóxico 1 (TSST-1) e as enterotoxinas B e C (SEB e SEC), que promovem uma resposta inflamatória exacerbada, por ligação ao complexo principal de histocompatibilidade de classe 2 de células apresentadoras de antígeno, que por sua vez ativam linfócitos T inespecificamente, induzindo a liberação de grandes quantidades de citocinas de forma abrupta, podendo levar ao choque séptico (LEE, 2018). Dois exemplos de fatores de virulência adquiridos via elementos genéticos móveis de maior destaque são o elemento genético catabólico de arginina (*arginine-catabolic mobile element-ACME*, restrito majoritariamente à linhagem de MRSA USA300) e a PVL. Uma das enzimas do ACME, arginina desaminase, inibe resposta imune inata e adaptativa, aumentando a sobrevivência bacteriana no hospedeiro (TURNER, 2019). A leucocidina de Pantou

Valentine possui dois componentes que se ligam na membrana dos leucócitos e outros tipos celulares do hospedeiro, formando poros e causando morte lítica na célula (LEE, 2018). A questão de aquisição de resistência aos antimicrobianos por esses elementos será melhor exemplificada no item a seguir, com o cassete cromossômico estafilocócico *mec* (*SCCmec*).

1.4 *Staphylococcus aureus* Resistentes à Meticilina (MRSA)

Um ano após a introdução da penicilina para uso clínico nas infecções por *S. aureus*, já apareciam cepas resistentes a esse beta-lactâmico (RAMMELKAMP; MAXON, 1942). Mesmo com o desenvolvimento de novas penicilinas semissintéticas na década de 60, como a Meticilina, que atuavam mesmo com a ação das beta-lactamases bacterianas, surgiram cepas com resistência, que foram denominadas *Staphylococcus aureus* resistentes à Meticilina (MRSA; do inglês: *Methicillin-resistant S. aureus*) (O’GARA, 2017). Atualmente 13% a 74% das infecções de *S. aureus* são causadas por MRSA (HASSOUN; LINDEN; FRIEDMAN, 2017). Países como Portugal, Itália e Romênia tem uma incidência maior que 60% das amostras coletadas (HASSOUN; LINDEN; FRIEDMAN, 2017). Na América Latina, já foi registrada uma incidência de 90% de MRSA no Chile (WHO, 2014). No Brasil, não existem dados randomizados sobre a incidência, apenas de alguns estados como o Rio de Janeiro, em que foi observada uma taxa de 46% das amostras de *S. aureus* como MRSA (BRIDE, 2019). Alguns autores estimam que a incidência no país esteja em torno de 60% (ROSSI, 2011).

Essa resistência foi possível pela aquisição do gene *mecA*, que codifica uma proteína ligante de penicilina modificada, PBP2a, com menor afinidade ao anel beta lactâmico dos antimicrobianos dessa classe, permitindo que a síntese da parede celular bacteriana continue. Esse gene faz parte do cassete cromossômico estafilocócico *mec* (*SCCmec*), um elemento genético móvel e ilha de resistência, e sua aquisição por um *S. aureus* sensível à Meticilina (MSSA; do inglês *Methicillin-susceptible S. aureus*) torna-o um MRSA (O’GARA, 2017). Além da resistência aos beta-lactâmicos de modo geral, alguns *SCCmec* ainda podem conter genes de resistência à aminoglicosídeos, tetraciclina e metais, como Cádmio e Mercúrio. É comum também que cepas MRSA apresentem outros mecanismos de resistência além do *SCCmec*, como enzimas que alteram o sítio alvo de antimicrobianos, como a fosfotransferase (*aphA*) na neomicina (FOSTER, 2017). Até o presente momento foram descritos 13

tipos de *SCCmec*, sendo diferenciados por sua estrutura molecular (URUSHIBARA, 2020). Diferentes metodologias podem ser usadas para a classificação de linhagens genéticas de MRSA como a tipagem de fago, padrão de PFGE (do inglês: *Pulsed-Field Gel Electrophoresis*), tipagem de sequências multilocus (MLST), tipagem *spa* (proteína A), tipagem de *SCCmec* e sequenciamento completo (GIULIERI, 2016). A maioria dos estudos utiliza o MLST em conjunto com a tipagem *SCCmec* para determinar uma linhagem, com o *sequence type* (ST) seguido do tipo de cassete (por exemplo: ST239-*SCCmecIII*). O MLST se baseia no sequenciamento de sete genes constitutivos (*housekeeping genes*). A comparação das pequenas variações nesses genes entre as cepas contra um banco de dados (<http://saureus.mlst.net/>), gera um perfil alélico que corresponde a um ST, que posteriormente pode ser agrupado em um complexo clonal (CC) para fins de classificação (ABDELBARY, 2017).

Atualmente, a vancomicina é um dos antimicrobianos mais utilizados para tratamento de pacientes infectados por MRSA em situações clínicas graves, porém isolados resistentes (VRSA) já foram identificados em várias regiões do globo, inclusive no Brasil (ROSSI, 2014), fazendo com que novos antimicrobianos e alternativas terapêuticas sejam necessários para tratamento das infecções.

1.5 Utilização do sequenciamento genômico como ferramenta para análises biológicas

O sequenciamento genômico total (WGS- do inglês *whole genome sequencing*) vem se tornando cada vez mais acessível tanto na pesquisa quanto na prática clínica, podendo ser usado na predição de sensibilidade aos antimicrobianos, investigação de surtos e mecanismos de virulência, identificação de cepas/linhagens e para estudos epidemiológicos (GIULIERI, 2016). Essa metodologia tem potencial de entregar dados genômicos completos dos isolados e permitir um tratamento personalizado, havendo, em geral, boa concordância entre WGS e os métodos clássicos para detecção de resistência em patógenos Gram positivos (BABIKER, 2019a). Contudo, o custo e dedicação/conhecimento para montagem e anotação das sequências geradas nessa metodologia ainda é um fator limitante para o uso rotineiro em muitos países e regiões do mundo (KWONG, 2015). Avanços na velocidade e facilidade dos sequenciamentos fizeram com que fosse possível um monitoramento genômico de surtos de infecção, sendo *S. aureus* uma das bactérias mais frequentemente sequenciadas no mundo (COPIN; SHOPSIN; TORRES, 2018). Porém, com cada vez mais sequenciamentos sendo depositados em bancos de dados, se torna difícil analisar tanta informação em

constante expansão, sendo necessárias mais ferramentas de bioinformática e mais estudos que permitam um melhor aproveitamento das informações genômicas geradas (COPIN; SHOPSIN; TORRES, 2018). O uso de WGS para predição de sensibilidade aos antimicrobianos já foi demonstrada em *Mycobacterium tuberculosis*, fornecendo um resultado semanas antes do que os ensaios fenotípicos de cultura (GIULIERI, 2016). Seu uso para estudo de surtos também já foi publicado: um trabalho em uma unidade neonatal do Reino Unido analisou por WGS um surto envolvendo 17 recém nascidos (HUMPHREYS; COLEMAN, 2019). De acordo com a metodologia de PFGE todos os 17 isolados obtidos pertenciam à linhagem USA300 (ST8-SCC*mecIV*), porém o sequenciamento completo mostrou que 12 (70,5%) deles eram geneticamente distintos e que, na verdade, não estavam relacionados ao surto. Juntamente aos avanços nas tecnologias de sequenciamento, as ferramentas de bioinformática vêm se tornando cada vez mais amigáveis e práticas a usuários sem muito treinamento (BABIKER, 2019b), com plataformas baseadas na web como o CARD (do inglês-*Comprehensive Antibiotic Resistance Database*) (MCARTHUR, 2013) e ResFinder (<https://cge.cbs.dtu.dk/services/ResFinder/>) para a análise de resistência em genomas depositados pelo simples *upload* do genoma desejado.

Em cerca de 80% dos casos de infecção por *S. aureus*, a cepa que já estava colonizando previamente o indivíduo de forma assintomática é a que está causando a infecção. Porém, o gatilho para que uma cepa comensal se torne infecciosa encontra-se pouco elucidado. Sabe-se que é um processo multifatorial, envolvendo fatores da bactéria (agente etiológico), do hospedeiro e da microbiota associada. Com um foco na parte bacteriana desse processo, é possível que existam diversos determinantes genéticos bacterianos participando dessa transição colonização-infecção. Um gene ou grupo de genes pode ser adquirido ou perdido por meio de elementos genéticos móveis e ter um papel crucial nestes mecanismos. Alguns grupos de pesquisa já tentaram estudar potenciais marcadores genéticos bacterianos nesta transição, porém com relativamente poucas amostras ou amostras oriundas de uma mesma localidade/estudo, podendo não representar um marcador global. Assim, a fim de aproveitar o grande volume de informações genômicas disponíveis gratuitamente em bancos de dados *online*, com mais de 400 genomas de *S. aureus* completamente sequenciados, montados e anotados, o presente estudo formulou a hipótese de que existem um ou mais fatores genéticos preditivos em potencial que podem contribuir

de forma significativa na transição bacteriana do estado comensal para o infeccioso. Estudos buscando marcadores genéticos em potencial de infecção podem auxiliar na compreensão dos mecanismos biológicos da infecção bacteriana e também para encontrar possíveis alvos para o controle e monitoramento de infecções.

2 OBJETIVOS

2.1 Objetivo Geral

Encontrar potenciais marcadores genéticos de infecção em *S. aureus* através de análises de genomas completos depositado em um banco de dados.

2.2 Objetivos Específicos

- Selecionar genomas completamente sequenciados associados à amostras de *S. aureus* isoladas de colonização e associados à infecções invasivas em humanos;
- A partir da seleção dos genomas em dois grandes grupos (colonização e infecção invasiva), subgrupar os genomas de acordo o ST (*Sequence Typing*), sensibilidade aos antimicrobianos (MSSA ou MRSA), localização geográfica, ano de isolamento e, se possível, a plataforma de sequenciamento utilizada;
- Realizar (i) análise pan-genômica, (ii) análise de alinhamento genômico e (iii) análises de polimorfismos de único nucleotídeo (SNP; do inglês: *single nucleotide polymorphism*) entre os genomas pertencentes aos dois grandes grupos (colonização e infecção invasiva);
- Analisar nos genomas selecionados a distribuição de genes relacionados à (i) virulência bacteriana, (ii) a resistência aos antimicrobianos e (iii) presença de elementos genéticos móveis (EGM).
- Comparar os genomas associados à amostras bacterianas isoladas de colonização com os genomas associados à amostras isoladas de infecções invasivas, considerando também os subgrupos descritos, com o auxílio de métodos estatísticos.

3 METODOLOGIA

3.1 Seleção dos genomas de *S. aureus*

Foi realizada uma busca por todos os genomas completamente sequenciados e montados (genomas completos, excluindo-se draft) de *S. aureus* (n=432) depositados até 30 de Junho de 2020 no banco de dados da seção Genome do NCBI (*National Center for Biotechnology Information*, <https://www.ncbi.nlm.nih.gov/genome/>). A partir destes genomas, foi realizada uma triagem para selecionar genomas oriundos de (critérios de inclusão):

- Amostras bacterianas isoladas de humanos;
- Amostras bacterianas isoladas a partir de infecções invasivas. Foi considerado como infecção “invasiva” amostras bacterianas isoladas de sítios normalmente estéreis, como a corrente sanguínea, tecido ósseo, líquido cefalorraquidiano e líquido pericárdico. Também foram incluídas amostras bacterianas isoladas de um caso de pneumonia, por ser uma infecção de trato respiratório inferior.
- Amostras bacterianas isoladas de colonização. Foi considerado como “colonização” genomas descritos como “colonização” ou amostras bacterianas isoladas de pele íntegra, vestíbulo nasal, períneo e intestino, sem qualquer descrição de infecção pela bactéria.

A escolha de genomas associados a infecções invasivas em detrimento de quaisquer outras amostras isoladas de infecção se deu como forma de priorização de infecções clinicamente mais graves, supondo que nestas as mudanças genéticas na bactéria poderão ser tão marcantes quando comparadas a infecções com menor gravidade clínica, como infecções de pele superficiais por exemplo. Os genomas foram divididos em dois grandes grupos: colonização e infecção invasiva. Em seguida, foram subgrupados de acordo com: localização geográfica, ano de isolamento, metodologia de sequenciamento (se disponível), sensibilidade a antimicrobianos (MSSA ou MRSA) e a linhagem genética a qual pertenciam pela classificação de ST (*sequence typing*), de acordo com o MLST. Adicionalmente, as amostras MRSA também foram agrupadas pela tipagem *SCCmec*. Foram adicionados também outros tipos de informações genômicas, quando possível, como tamanho e cobertura do genoma.

3.2 Análises Genômicas

Os genomas selecionados foram comparados utilizando o Roary versão 3.11.2 (ANDREW et al 2015; <https://sanger-pathogens.github.io/Roary/>), pipeline de análise de pan-genoma, realizando a análise dos dois grandes grupos de interesse (grupo de colonização e grupo de infecção), além da análise de subgrupos dentro deles: por exemplo, amostras de MRSA de colonização com amostras MRSA de infecção. Também foi realizada a análise dentro de um mesmo ST, a fim de identificar possíveis grupos de genes que possam ser característicos a uma linhagem. Adicionalmente, entre os mesmos grupos, foi realizada uma análise filogenética com base em polimorfismos de único nucleotídeo (SNP; do inglês: *single nucleotide polymorphism*) utilizando 2 softwares disponíveis *online*: REALPHY versão 1.12 (BERTELS, 2014; <https://realphy.unibas.ch/realphy/>) e CSIPhylogeny versão 1.4 (KAAS, 2014; <https://cge.cbs.dtu.dk/services/CSIPhylogeny>). Por fim, ainda foi realizado um alinhamento do genoma completo utilizando o MAUVE versão 2.4.0 (DARLING, 2004; <https://darlinglab.org/mauve/mauve.html>). Todas as plataformas/pipelines são gratuitas e foram utilizadas as configurações padrão (*default*) para as análises.

3.2.1 Análise dos genes de virulência

Os genomas foram comparados quanto a distribuição de genes associados a virulência bacteriana utilizando a plataforma VirulenceFinder versão 1.7, (<https://cge.cbs.dtu.dk/services/VirulenceFinder/>; JOENSEN, 2014), do *Center for Genomic Epidemiology* (CGE), localizado na Universidade Técnica da Dinamarca e curado por um time multidisciplinar e multinacional. Essa plataforma *online* faz um alinhamento do genoma que o usuário insere, nos formatos *fasta* (genoma montado) ou *fastq* (dados brutos), com os genes de virulência depositados em um banco de dados e detecta a porcentagem de identidade entre as sequências, traçando um perfil de virulência de cada genoma. Também foi utilizado o banco de dados do software RidomSeqSphere+ versão 8.3 (JÜNEMANN, 2013; <https://www.ridom.de/seqsphere/>, foi utilizada a versão trial, gratuita), que analisa a distribuição de genes associados a virulência bacteriana de forma similar, para comparação e validação dos dados. Em ambas as plataformas, foram utilizadas as configurações padrão (*default*) para as análises.

3.2.2 Análise dos genes de resistência aos antimicrobianos

Para análise da distribuição dos genes de resistência aos antimicrobianos nos genomas, foi utilizada a plataforma ResFinder versão 3.2 (BORTOLAIA et al., 2020; <https://cge.cbs.dtu.dk/services/ResFinder/>) também do CGE. Configurações padrão (*default*) foram utilizadas para as análises. A partir do *upload* do genoma (arquivo *fasta*), a plataforma faz um alinhamento com os genes no banco de dados e detecta similaridade entre as sequências, traçando um perfil de resistência para o genoma.

3.2.3 Análise de elementos genéticos móveis (EGM)

Foi realizada uma análise para verificar a distribuição de EGM nos genomas utilizando as seguintes plataformas:

- a) PlasmidFinder versão 2.1 (CARATTOLI et al., 2014; <https://cge.cbs.dtu.dk/services/PlasmidFinder/>) também do CGE, em que as sequências do genoma submetido, no formato *fasta* ou *fastq*, são comparados com o banco de dados de plasmídeos, apresentando uma tabela com dados de similaridade.
- b) ISFinder (SIGUIER et al., 2006; <https://www-is.biotoul.fr/index.php>) um banco de dados de sequências de inserção (IS), que são elementos de transposição simples, amplamente distribuídos em genomas procarióticos. IS podem afetar positivamente ou negativamente a patogenicidade de *S. aureus* quando inseridos em genes relacionados a patogênese bacteriana (VARANI et al., 2021).

3.3 Análise estatística

Para análise estatística da frequência dos genes encontrados nos genomas dos diferentes grupos, foi utilizado o teste exato de Fisher (Fisher, 1934), com o auxílio do Software MedCalc Statistical versão 19.2.6 (MedCalc Software bv, Ostend, Belgium; <https://www.medcalc.org>). O Prof Marcos Jose Machado, do Departamento de Análises clínicas da UFSC, auxiliou nesta etapa do estudo.

4 RESULTADOS

Dentre os 432 genomas completos de *S. aureus* disponíveis no banco de dados GenBank, foram selecionados para análise neste estudo 92 genomas completos após a triagem, que cumpriram os requisitos propostos. Entre os 92 genomas, 33 (36%) foram associados a amostras bacterianas isoladas de colonização do hospedeiro humano, e 59 (64%) foram sequenciados a partir de amostras bacterianas isoladas de sítios clínicos de infecções consideradas invasivas.

Quanto ao ano de isolamento das amostras, 29 (32%) foram isoladas entre 1900 e 1999, 19 (21%) nos anos 2000 até 2010, 31 (34%) de 2011 até 2018 e 13 (14%) deles não possuíam informação do período de isolamento. A plataforma mais utilizada para o sequenciamento dos genomas foi Illumina, com 36 amostras (39%), seguido de PacBio SMRT (*Single Molecule Real Time*) com 12 amostras (13%), o sequenciamento por nanoporos da *Oxford Nanopore Technologies* (ONT) com sete amostras (8%) e pirosequenciamento Roche 454 com quatro amostras (4%). Além disso, 12 amostras (13%) foram sequenciadas de forma híbrida, utilizando mais de uma metodologia, sendo oito (9%) com Illumina e PacBio, três (3%) com Illumina e ONT e uma (1%) por Illumina e pirosequenciamento 454. Não foi possível obter as informações sobre qual foi o método utilizado para sequenciar 19 genomas (21%).

4.1 Características gerais dos Genomas de Colonização

A Tabela 1 contém todos os 33 genomas associados a amostras de colonização e suas principais características. Entre os 33 genomas associados à colonização, seis foram sequenciados a partir de amostras isoladas da região do períneo (18%) e 27 foram isoladas da cavidade nasal (82%). Genomas associados a amostras bacterianas isoladas de pele foram excluídos por não haver especificação se eram provenientes de um processo infeccioso ou colonização do hospedeiro. Entre estes genomas, 11 (33%) são MSSA e 22 (67%) MRSA, com o SCC*mec* de tipo IV tendo maior incidência (16 genomas, 73%) seguido dos tipos III e I (três genomas, 14% e dois genomas, 9%, respectivamente) e um genoma que não foi possível tipar o SCC*mec*, mesmo utilizando o SCC*mec*Finder v1.2 (<https://cge.cbs.dtu.dk/services/SCCmecFinder/>; plataforma online e gratuita). Como este genoma continha o gene *mecA*, foi classificado como MRSA (Figura 1A). Cabe ressaltar que 12 entre os 92 genomas foram classificadas como provenientes de amostras MRSA ou MSSA utilizando o SCC*mec*Finder v1.2, por

não haver informação da resistência à metilina disponível no acesso/referência depositado no GenBank. Os ST (*Sequence Typing*) encontrados entre estes 33 genomas foram o ST8 com 18 genomas (55%), ST239 com três (9%), os STs 30, 121 e 228 com dois genomas cada (6% cada), e os STs 464, 254, 5, 50, 15, 59 cada um com uma única amostra (3% cada) (Figura 2A). Quanto à origem geográfica dos isolados relacionados aos genomas, a maioria é proveniente da América do Sul (14 amostras, 42%), seguida da Europa (12 amostras, 36%), América do Norte (cinco amostras, 15%) e Leste Asiático (duas amostras, 6%). A distribuição dos genomas conforme o ST e a origem geográfica continental podem ser observados na Figura 3.

Tabela 1 - Genomas completos de *S. aureus* sequenciados a partir de amostras bacterianas de colonização. Os 33 genomas foram selecionados a partir do banco de dados GenBank e agrupados de acordo com o ST (*Sequence Typing*). Além do ST, a tabela inclui o nome do genoma ou amostra, o número de acesso Genbank, ano de isolamento da bactéria, tecnologia de sequenciamento utilizada, sítio biológico de isolamento da bactéria, classificação como MRSA ou MSSA, tipagem *SCCmec* e região geográfica de isolamento.

Nome do genoma ou amostra	Accession number Genbank	Ano de isolamento	Método/Equipamento	Infecção	MLST	MRSA	SCCmec	Região geográfica
82	GCA_003425885.1	1993	Illumina HiSeq 2500	Cavidade nasal	ST8	MSSA	-	Alemanha
164	GCA_003354885.1	1993	Illumina HiSeq 2500	Cavidade nasal	ST8	MSSA	-	Alemanha
1969.N	GCA_001717705.2	2011	Pacbio RSII Illumina MiSeq	Cavidade nasal	ST8	MRSA	IV	EUA
2148.N	GCA_001717665.2	2011	Pacbio RSII Illumina MiSeq	Cavidade nasal	ST8	MSSA	-	EUA
5118.N	GCA_001717645.3	2010	Pacbio RSII Illumina MiSeq	Cavidade nasal	ST8	MRSA	IV	EUA
MOZ66	GCA_003426145.1	1993	Illumina HiSeq 2500	Cavidade nasal	ST8	MSSA	-	Alemanha
USA300-SUR1	GCA_003595365.1	2013	Illumina MiSeq	Períneo	ST8	MRSA	IVa	República do Suriname
USA300-SUR10	GCA_002000585.1	2013	Illumina MiSeq	Cavidade nasal	ST8	MRSA	IVa	República do Suriname
USA300-SUR11	GCA_002000605.1	2013	Illumina MiSeq	Períneo	ST8	MRSA	IVa	República do Suriname
USA300-SUR15	GCA_002000685.1	2013	Illumina MiSeq	Cavidade nasal	ST8	MRSA	IVa	República do Suriname
USA300-SUR16	GCA_002000705.1	2013	Illumina MiSeq	Períneo	ST8	MRSA	IVa	República do Suriname
USA300-SUR18	GCA_002000745.1	2013	Illumina MiSeq	Cavidade nasal	ST8	MRSA	IVa	República do Suriname
USA300-SUR19	GCA_002000765.1	2013	Illumina MiSeq	Períneo	ST8	MRSA	IVa	República do Suriname
USA300-SUR20	GCA_002000785.1	2013	Illumina MiSeq	Cavidade nasal	ST8	MRSA	IVa	República do Suriname
USA300-SUR23	GCA_002000845.1	2013	Illumina MiSeq	Períneo	ST8	MRSA	IVa	República do Suriname
USA300-SUR24	GCA_002000865.1	2013	Illumina MiSeq	Períneo	ST8	MRSA	IVa	República do Suriname
USA300-SUR8	GCA_003595505.1	2013	Illumina MiSeq	Cavidade nasal	ST8	MRSA	IVa	República do Suriname
USA300-SUR9	GCA_002000565.1	2013	Illumina MiSeq	Cavidade nasal	ST8	MRSA	IVa	República do Suriname
HC1335	CP012012	2001	ND	Cavidade nasal	ST239	MRSA	III	Brasil
HC1340	CP012011	2001	ND	Cavidade nasal	ST239	MRSA	III	Brasil
M92	GCA_002097595.2	1980	PacBio RSII	Cavidade nasal	ST239	MRSA	III	Canadá
78	GCA_003354905.1	1993	Illumina HiSeq 2500	Cavidade nasal	ST121	MSSA	-	Alemanha
422	GCA_003354605.1	1993	Illumina HiSeq 2500	Cavidade nasal	ST121	MSSA	-	Alemanha
277	GCA_003354685.1	1993	Illumina HiSeq 2500	Cavidade nasal	ST30	MSSA	-	Alemanha
466	GCA_003354745.1	1993	Illumina HiSeq 2500	Cavidade nasal	ST30	MSSA	-	Alemanha
191	GCA_003354765.1	1993	Illumina HiSeq 2500	Cavidade nasal	ST228	MRSA	I	Alemanha
628	GCA_003354985.1	1993	Illumina HiSeq 2500	Cavidade nasal	ST228	MRSA	I	Alemanha
135	GCA_003354925.1	1993	ND	Cavidade nasal	ST15	MSSA	-	Alemanha
545	GCA_003354945.1	1993	Illumina HiSeq 2500	Cavidade nasal	ST254	MRSA	IV	Alemanha
ATCC BAA-39	GCA_000146385.1	1993	ND	Cavidade nasal	ST464	MRSA	-	Hungria
502A	GCA_000597965.1	1963	PacBio	Cavidade nasal	ST5	MSSA	-	EUA
GN3	GCA_003609895.1	2005	PacBio RS II	Cavidade nasal	ST50	MSSA	-	Japão
SA40	GCA_000470865.1	2005	DNA microarray	Cavidade nasal	ST59	MRSA	IV	Taiwan

(-) Representa genomas de MSSA, que não carregam o *SCCmec*

EUA: Estados Unidos da América.

ND: não determinado

4.2 Características gerais dos genomas de Infecção invasiva

A Tabela 2 contém todos os 59 genomas associados a amostras de infecção invasiva e suas principais características. Entre estes 59 genomas, 46 (78%) foram sequenciados a partir de amostras isoladas de infecções de corrente sanguínea (ICS; podendo incluir bacteremia ou septicemia), 10 (17%) de osteomielites, duas (3%) de endocardites e uma (1,7%) isolada de pneumonia. Quanto à sensibilidade à meticilina, 25 (42,4%) são MSSA e 34 (57,6%) são MRSA, sendo o SCC*mec* II (12 genomas, 35%) o mais prevalente entre os *S. aureus* resistentes à meticilina, seguido dos tipos IV (9 genomas, 26%), tipo III (sete genomas, 21%) e tipo X (seis genomas; 18%) (Figura 1B). O ST mais prevalente foi o ST5 com oito genomas (14%), sendo o ST8 o segundo mais frequente com sete genomas (12%), seguidos do ST239 e ST15 cada um com 5 genomas (8% cada) e ST1 com quatro genomas (7%). Também foram encontrados o ST22 e ST121 com três genomas (5%), os STs 105, 228, 254, 2389, 72, 59 e 25 com duas amostras cada (3%), e por fim os STs 45, 291, 30, 246, 152, 237, 36, 398, 240 e 93 cada um com uma amostra (2% cada) (Figura 2B). Quanto à região geográfica de coleta das amostras a maioria provém da Europa (28 amostras, 47%), seguido do Leste Asiático (18 amostras, 31%), América do Norte (seis amostras, 10%), América do Sul (quatro amostras, 7%) e Oceania (três amostras, 5,1%). A distribuição dos genomas conforme o ST e a origem geográfica continental podem ser observados na Figura 3.

Tabela 2- Genomas completos de *S. aureus* sequenciados a partir de amostras bacterianas isoladas de infecção invasiva. Os 59 genomas foram selecionados a partir do banco de dados GenBank e agrupados de acordo com o ST (*Sequence Typing*). Além do ST, a tabela inclui o nome do genoma ou amostra, o número de acesso Genbank, ano de isolamento da bactéria, tecnologia de sequenciamento utilizada, sítio biológico de isolamento da bactéria, classificação como MRSA ou MSSA, tipagem *SCCmec* e região geográfica de isolamento.

Nome do genoma ou amostra	Accession number Genbank	Ano de isolamento	Método/Equipamento	Infecção	MLST	MRSA	SCCmec	Região geográfica
FORC_027	GCA_001725965.1	ND	ND	Sangue	ST5	MRSA	II	Coréia do Sul
FORC_062	GCA_003351925.1	ND	PacBio	Sangue	ST5	MRSA	II	Coréia
HOU1444-VR	GCA_001278745.1	2012	Illumina MiSeq e PacBio RS II	Sangue	ST5	MSSA	-	Brasil
Mu3	GCA_000010445.1	1996	ND	Pneumonia	ST5	MRSA	II	Japão
NCCP14558	GCA_001640905.1	ND	ND	Sangue	ST5	MRSA	II	Coréia do Sul
NCCP14562	GCA_001640925.1	ND	PacBio SMRT	Sangue	ST5	MRSA	II	Coréia do Sul
ZJ5499	GCA_001640885.1	ND	ND	Sangue	ST5	MRSA	II	China
CHU15-056	GCA_003194405.1	ND	Illumina MiSeq	Sangue	ST5	MRSA	IV	Brasil
28	GCA_003425835.1	ND	ND	Sangue	ST8	MSSA	-	Alemanha
54	GCA_003354805.1	ND	ND	Sangue	ST8	MSSA	-	Alemanha
199	GCA_003425545.1	ND	ND	Sangue	ST8	MSSA	-	Alemanha
EDCC5458	GCA_002786465.1	ND	Illumina MiSeq e PacBio RSII	Osteomielite	ST8	MSSA	-	Alemanha
NRS137	GCA_001018735.2	1960	ND	Sangue	ST8	MSSA	-	Reino Unido
USA300_TCH1516	GCA_000017085.1	2007	Pirosequenciamento 454	Sangue	ST8	MRSA	IVa	EUA
USA300-ISMMS1	GCA_000568455.1	ND	PacBio RSII	Sangue	ST8	MRSA	IV	EUA
Bmb9393	GCA_000418345.1	1993	Pirosequenciamento 454	Sangue	ST239	MRSA	III	Brasil
JKD6008	GCA_000145595.1	2003	Illumina e 454	Sangue	ST239	MRSA	III	Nova Zelândia
TW20	GCA_000027045.1	2010	ND	Sangue	ST239	MRSA	III	Reino Unido
WCH-SK2	GCA_003394105.1	2015	PacBio RSII	Sangue	ST239	MRSA	II	Austrália
Z172	GCA_000485885.1	2010	Illumina/Solexa e PacBio	Sangue	ST239	MRSA	III	Taiwan
FORC_045	GCA_002209325.1	ND	ND	Sangue	ST1	MSSA	-	Coréia do Sul
JMUB3031	GCA_003573835.1	ND	PacBio RS e Illumina MiSeq	Sangue	ST1	MRSA	IVc	Japão
MSSA476	GCA_000011525.1	2004	ND	Osteomielite	ST1	MSSA	-	Reino Unido
MW2	GCA_000011265.1	2002	Sanger	Sangue	ST1	MRSA	IV	EUA
JH1	GCA_000017125.1	2007	ND	Endocardite	ST105	MRSA	IIa	EUA
JH9	GCA_000016805.1	2007	ND	Endocardite	ST105	MRSA	IIa	EUA
27	GCA_003354545.1	ND	ND	Sangue	ST121	MSSA	-	Alemanha
128	GCA_003354645.1	ND	ND	Sangue	ST121	MSSA	-	Alemanha
XQ	GCA_001444345.1	ND	Ion Torrent	Sangue	ST121	MSSA	-	China
45	GCA_003354865.1	ND	ND	Sangue	ST15	MSSA	-	Alemanha
ST20130938	GCA_001611405.1	2016	Ion Torrent	Osteomielite	ST15	MSSA	X	França
ST20130939	GCA_001611425.1	2016	Ion Torrent	Osteomielite	ST15	MSSA	X	França
ST20130940	GCA_001611325.1	2016	Ion Torrent	Osteomielite	ST15	MSSA	X	França
ST20130941	GCA_001611345.1	2016	Ion Torrent	Osteomielite	ST15	MSSA	X	França
EDCC5464	GCA_002786535.1	ND	Illumina MiSeq e PacBio RSII	Osteomielite	ST22	MSSA	-	Alemanha
IT1-S	GCA_003184985.1	2013	Illumina MiSeq	Endocardite	ST22	MSSA	-	Europa
IT4-R	GCA_003185005.1	2013	Illumina MiSeq	Endocardite	ST22	MSSA	-	Europa
61	GCA_003354705.1	ND	ND	Sangue	ST228	MRSA	I	Alemanha
187	GCA_003354565.1	ND	ND	Sangue	ST228	MRSA	I	Alemanha
TUM9458	GCA_003945405.1	ND	Illumina MiSeq e MinION	Sangue	ST2389	MSSA	-	Japão
TUM9463	GCA_003945425.1	ND	Illumina MiSeq e MinION	Sangue	ST2389	MRSA	II	Japão
ST20130942	GCA_001611365.1	2016	Ion Torrent	Osteomielite	ST25	MSSA	X	França
ST20130943	GCA_001611385.1	2016	Ion Torrent	Osteomielite	ST25	MSSA	X	França
165	GCA_003354665.1	ND	ND	Sangue	ST254	MRSA	IV	Alemanha
Newman	GCA_000010465.1	2008	ND	Osteomielite	ST254	MSSA	-	Reino Unido
E16SA093	GCA_003342735.1	2016	PacBio RSII	Sangue	ST72	MRSA	IV-like	Coréia do Sul
F17SA003	GCA_003342775.1	2017	PacBio RSII	Sangue	ST72	MRSA	IV-like	Coréia do Sul
CA-347	GCA_000412775.1	2005	Illumina e PacBio RS	Sangue	ST45	MRSA	II	EUA
SA957	GCA_000470845.1	2000	DNA microarray	Sangue	ST59	MRSA	V	Taiwan
JP080	GCA_003609855.1	ND	Illumina MiSeq e PacBio RS II	Sangue	ST291	MSSA	-	Japão
143	GCA_003354725.1	ND	ND	Sangue	ST30	MSSA	-	Alemanha
SA268	GCA_000737615.1	2012	Illumina HiSeq2000	Sangue	ST59	MRSA	V	China
V605	GCA_001641045.1	2018	ND	Sangue	ST246	MRSA	III	Coréia do Sul
55-99-44	GCA_002803885.1	2018	Illumina e MinION	Sangue	ST152	MRSA	XIII	Dinamarca
Be62	GCA_001515685.1	1996	ND	Sangue	ST237	MRSA	III	Brasil
MRSA252	GCA_000011505.1	2004	ND	Sangue	ST36	MRSA	II	Reino Unido
ST398	GCA_000009585.1	2010	Pirosequenciamento 454	Sangue	ST398	MRSA	V	Países Baixos
NCTC9944	GCA_900474575.1	1956	ND	Sangue	ST240	MRSA	III	Reino Unido
JKD6159	GCA_000144955.1	2010	Illumina e 454	Sangue	ST93	MRSA	IV	Austrália

(-) Representa genomas de MSSA, que não carregam o *SCCmec*

EUA: Estados Unidos da América.

ND: não determinado

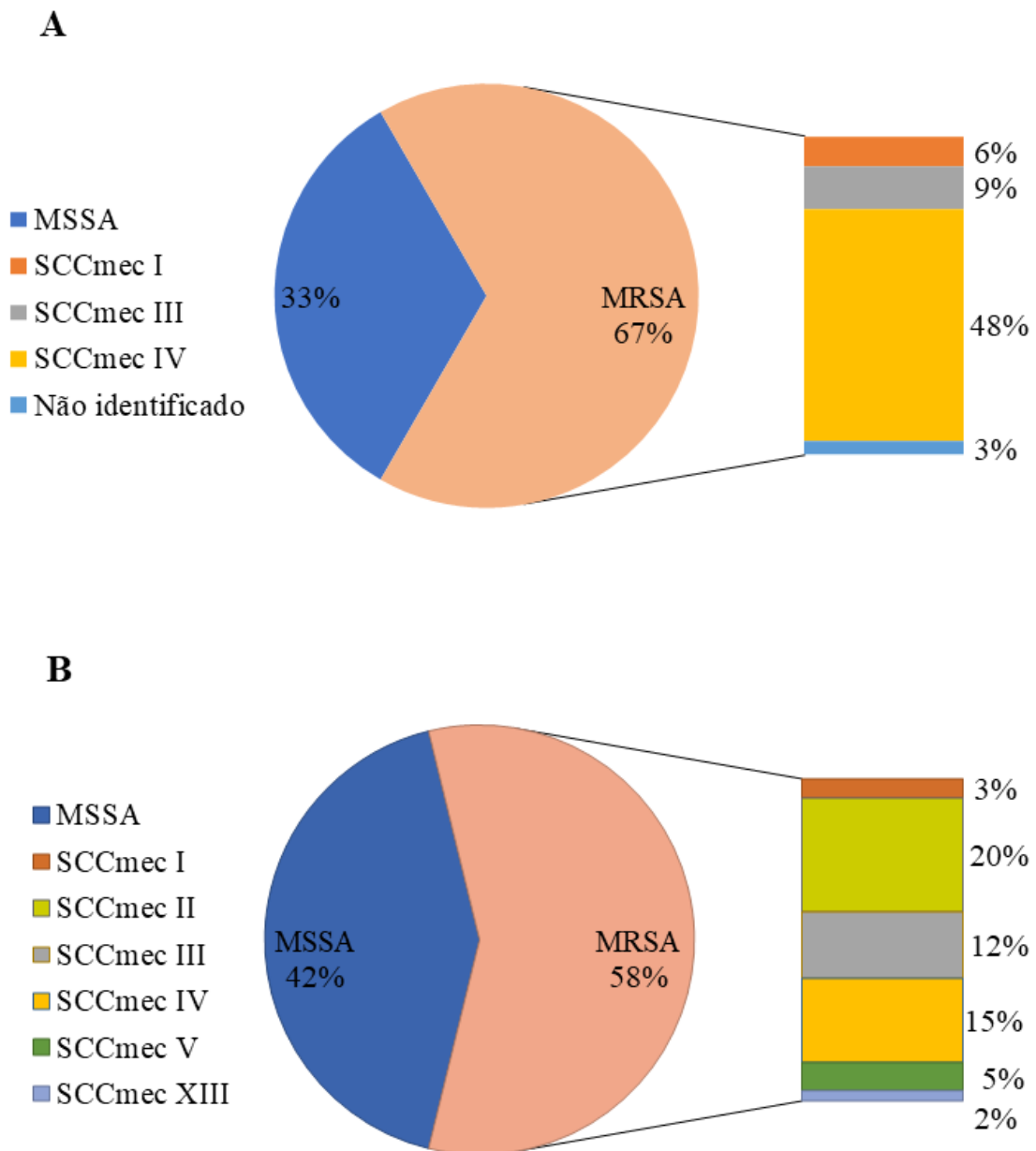
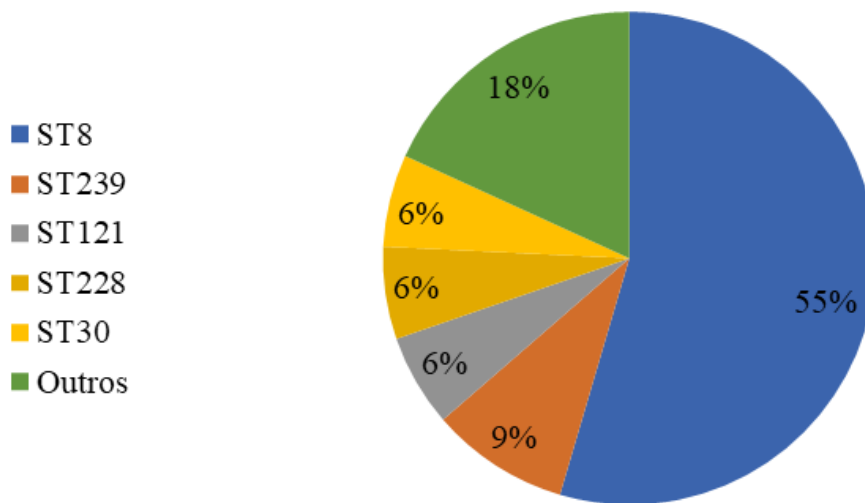


Figura 1-Gráficos representativos da porcentagem de MRSA ou MSSA entre os genomas sequenciados e a porcentagem dos tipos de *SCCmec* identificados nos genomas associados a amostras MRSA. (A) representa os genomas associados à colonização: 11 genomas (33,3%) são MSSA enquanto 22 genomas (66,7%) são MRSA, com a predominância do *SCCmec* IV (16 genomas; 48%). (B) representa os genomas associados à infecção invasiva: 25 genomas (42,4%) são MSSA e 34 (57,6%) são MRSA. Os gráficos foram confeccionados no Microsoft Power Point (Microsoft®). MSSA: Methicilin-Susceptible *Staphylococcus aureus*. MRSA: Methicilin-Resistant *Staphylococcus aureus*. *SCCmec*: staphylococcal cassette chromosome *mec*.

A



B

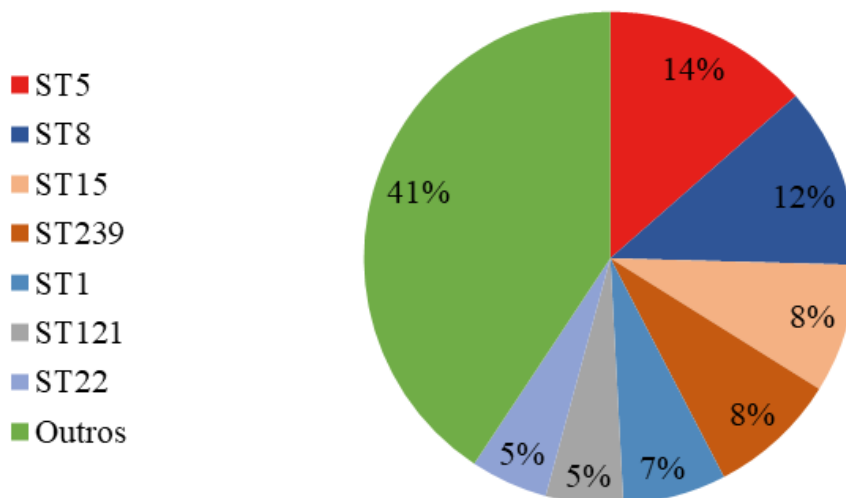


Figura 2-Gráfico representativos dos ST (*Sequencing typing*; derivados do MLST) de genomas de *S. aureus* sequenciados a partir de amostras de colonização. (A) genomas associados à colonização. Os STs associados as amostras de colonização são: ST8 (18), ST239 (3), ST30(2), ST121 (2), ST228 (2). “Outros” STs são: ST5, ST15, ST50, ST59, ST254 e ST464 cada um com apenas um genoma. (B) genomas associados à infecção invasiva. STs relacionados a amostras de infecção, sendo: ST5 (8), ST8 (7), ST15 (5), ST239 (5), ST1 (4), ST22(3) e ST121 (3). “Outros” nos STs de infecção são: ST25 (2), ST59 (2), ST72 (2), ST105 (2), ST228 (2), ST254 (2), ST2389 (2), ST 30 (1), ST36 (1), ST45 (1), ST93 (1), ST152 (1), ST237 (1), ST240 (1), ST 246 (1), ST291 (1) e ST398 (1). O número entre parênteses representa o número de genomas pertencente a cada ST. Os gráficos foram confeccionados no Microsoft Power Point (Microsoft®).

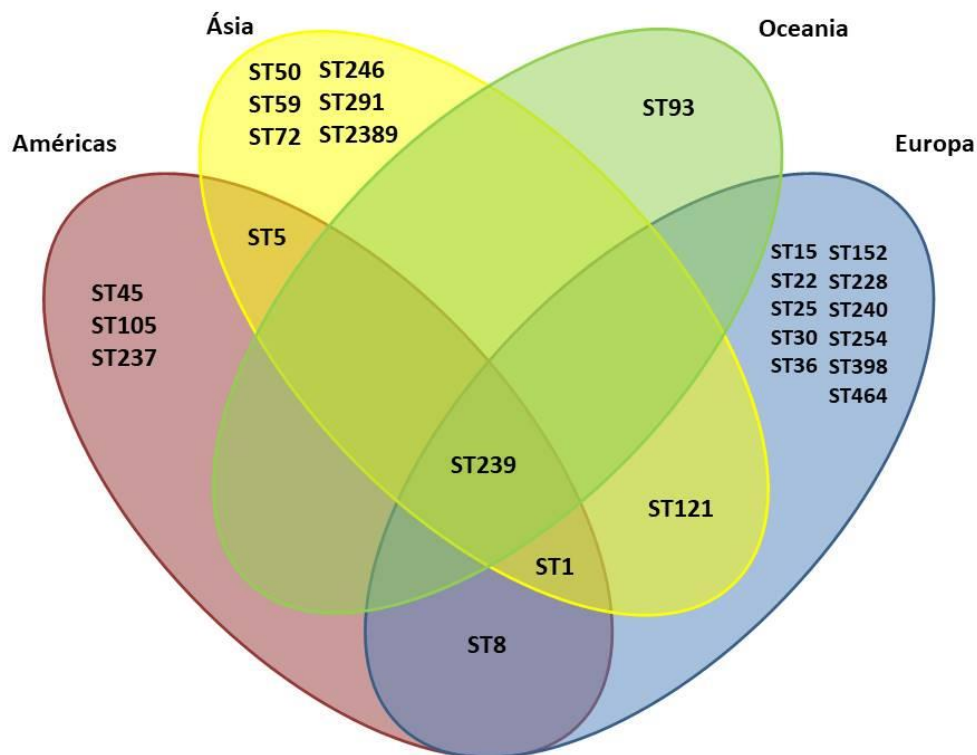


Figura 3- Diagrama de Venn representando a distribuição dos STs (*Sequence Typing*) dos 92 genomas selecionados entre os continentes (América, Ásia, Europa e Oceania). Alguns ST foram encontrados exclusivamente em um dos continentes (Ex.: ST15 no continente Europeu), enquanto outros são encontrados em mais de 1 continente (Ex.: ST5 foi encontrado no continente americano e asiático; indicando a área sobreposta) ou em todos os continentes (Ex.: ST239; sobreposição de todas as áreas do diagrama). O Programa Microsoft Power Point (Microsoft®) foi utilizado para a confecção do diagrama.

4.3 Análise pan-genômica

Inicialmente, a análise pan-genômica foi realizada com todos os 92 genomas de *S. aureus* selecionados. A análise revelou que 2278 genes formam o *core* genoma (genes presentes em 91 a 92 das amostras), 476 formaram o *soft core* (genes compartilhados entre 40 a 91 das amostras), 365 genes no *shell core* (genes compartilhados em 1 a 40 das amostras) e 374 genes são compartilhados apenas entre dois genoma (Figura 4). A partir destes dados, não foi possível estabelecer uma diferenciação entre os genomas de colonização e infecção, considerando o número de genes ou sua localização no genoma. Esta análise indicou que os genomas apresentam suas particularidades como genomas únicos, sem considerar a classificação nos dois grandes grupos.

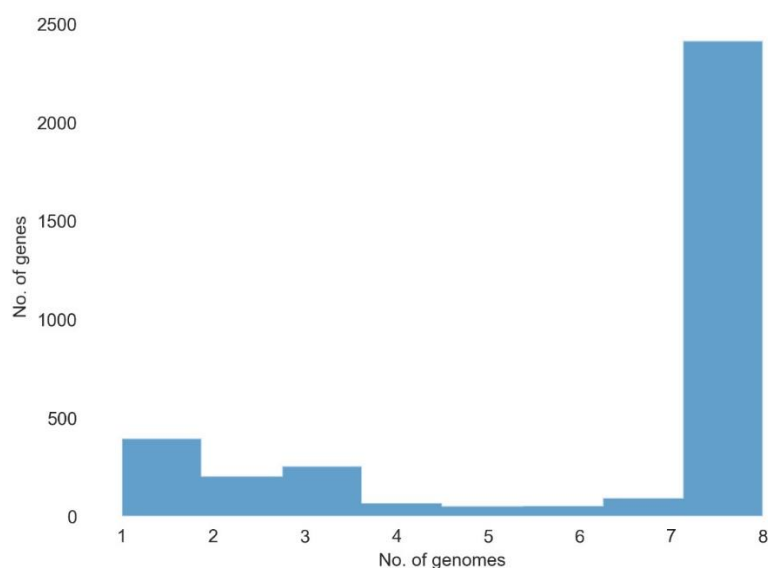


Figura 4-Análise pan-genômica considerando os 92 genomas de *S. aureus* selecionados. O gráfico representa o número de genes compartilhados (eixo y) relacionado ao número de genomas (eixo x). Cada coluna representa o número de genes compartilhado entre dado número de genomas. O *core* genoma (presente em todas as amostras) é a última coluna com 2415 genes. Gráfico obtido a partir da análise realizada pelo Roary, conforme descrito na metodologia.

Posteriormente, a análise pan-genômica foi realizada apenas com os genomas MRSA pertencentes ao ST239 (8 genomas), por ter um número de genomas mais representativo entre os grupos infecção (5 genomas) e colonização (3 genomas) e por ser geneticamente mais homogêneo (mesmo ST e tipo de *SCCmec*) entre todos os genomas selecionados. A análise revelou que 2415 genes formaram o *core* genoma, não houve formação de um *soft core*, 739 genes formaram o *shell* genoma (genes compartilhados entre 1 a 7 amostras) e 397 são compartilhados apenas entre dois genomas (Figuras 5 e 6). A partir destes dados, novamente a análise indicou que os genomas apresentam suas particularidades como genomas únicos, sem diferenças que predominam em um dos dois grandes grupos (colonização e infecção).

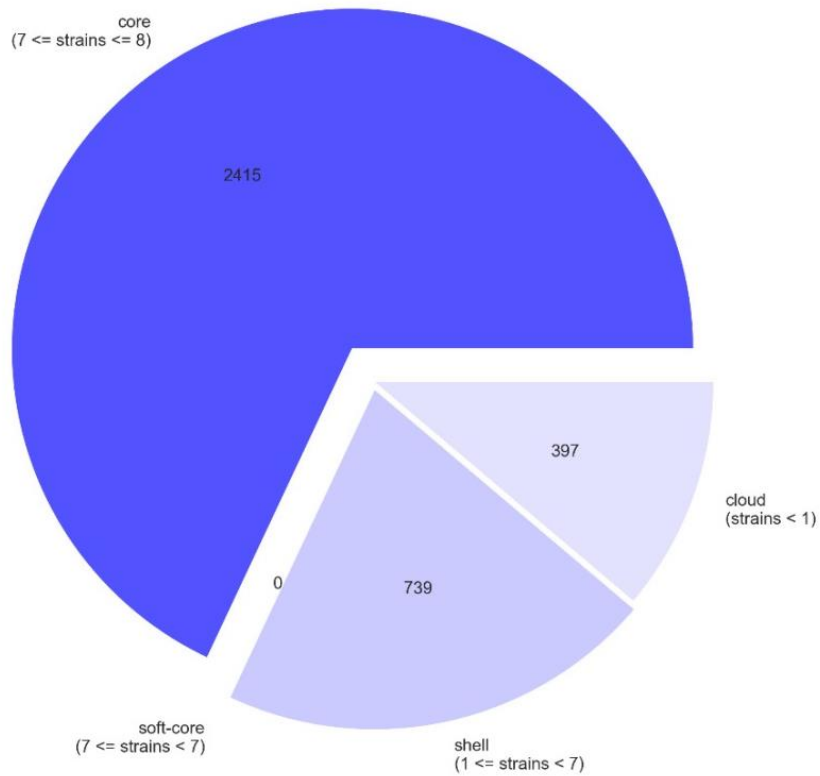


Figura 5-Gráfico representando o pangenoma dos genomas ST239 (8 genomas). O *core* genoma está representado em azul escuro, com 2415 genes; o *shell* genoma (genes compartilhados com apenas um genoma ou com até 7 outros genomas) está representado na cor lilás mais escuro e o *cloud* genoma (genes compartilhados apenas com um outro genoma) na cor lilás mais claro. Gráfico obtido a partir da análise realizada pelo Roary, conforme descrito na metodologia.

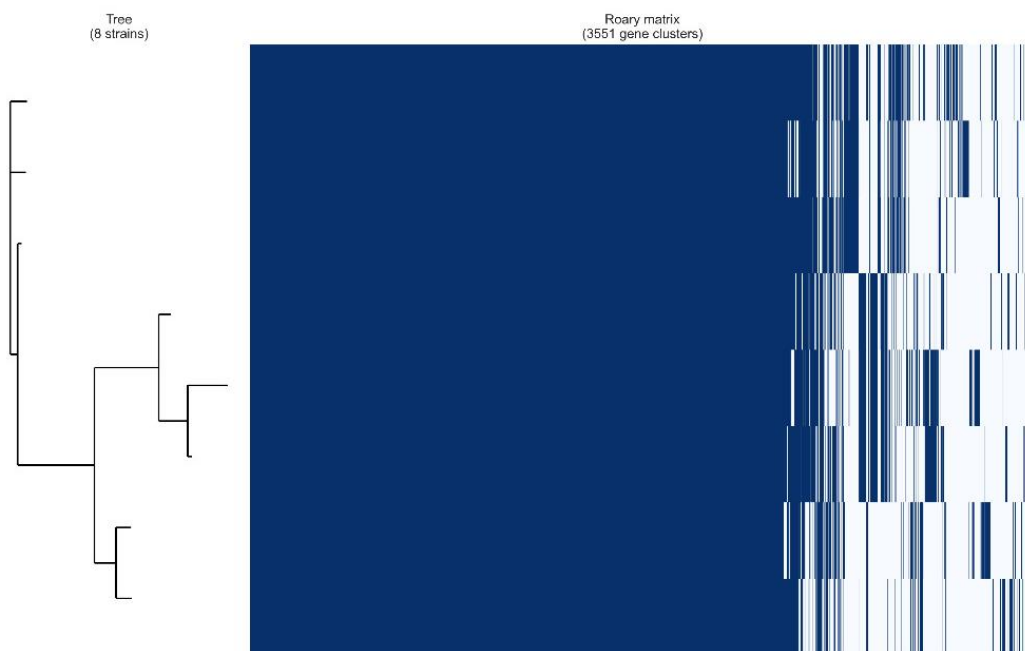


Figura 6-Matriz relacionada a ausência e presença dos genes de cada amostra do ST239 e sua relação filogenética. Gráfico obtido a partir da análise realizada pelo Roary, conforme descrito na metodologia.

4.4 Alinhamento dos genomas

O alinhamento dos genomas completos foi realizado utilizando genomas pertencentes apenas ao ST8 e ao ST239 (em separado), uma vez que foram os únicos STs representativos com relação ao número de amostras nos dois grandes grupos (infecção e colonização). Os genomas ST8 apresentam genomas de MRSA e MSSA, enquanto ST239 só foram selecionados genomas MRSA. A análise visual dos resultados do alinhamento através do MAUVE não revelou nenhuma diferença perceptível entre os grupos de colonização e de infecção invasiva (Figuras 7 e 8). Foi possível observar apenas diferenças individuais entre os genomas, como o genoma ST239 WCH-SK2, que apresenta diversas regiões genômicas em posições diferentes dos demais genomas ST239 (Figura 7), provavelmente relacionado com o fato deste genoma ter sido isolado de uma pequena variante de colônia (do inglês: *Small colony variant*, SCV) (BUI; KIDD, 2015). Entre os genomas ST8 (Figura 8), não foi possível observar diferenças perceptíveis entre os dois grupos (infecção e colonização) com relação à resistência aos betalactâmicos codificada pelo gene *mecA*. Adicionalmente, os genomas ST8 isolados de amostras de MRSA de colonização e que tem como nome o prefixo “SUR” são todas oriundas de um único estudo e localização geográfica, por isso

são tão semelhantes (SABAT et al., 2017). Entre os genomas ST8 em geral, também é possível observar uma heterogenicidade maior entre a posição das regiões genômicas, que pode estar relacionada aos diferentes *SCCmec* presentes (IV e IVa), que podem acompanhar outros genes adicionais, já que as amostras ST239 são todas *SCCmec*III.

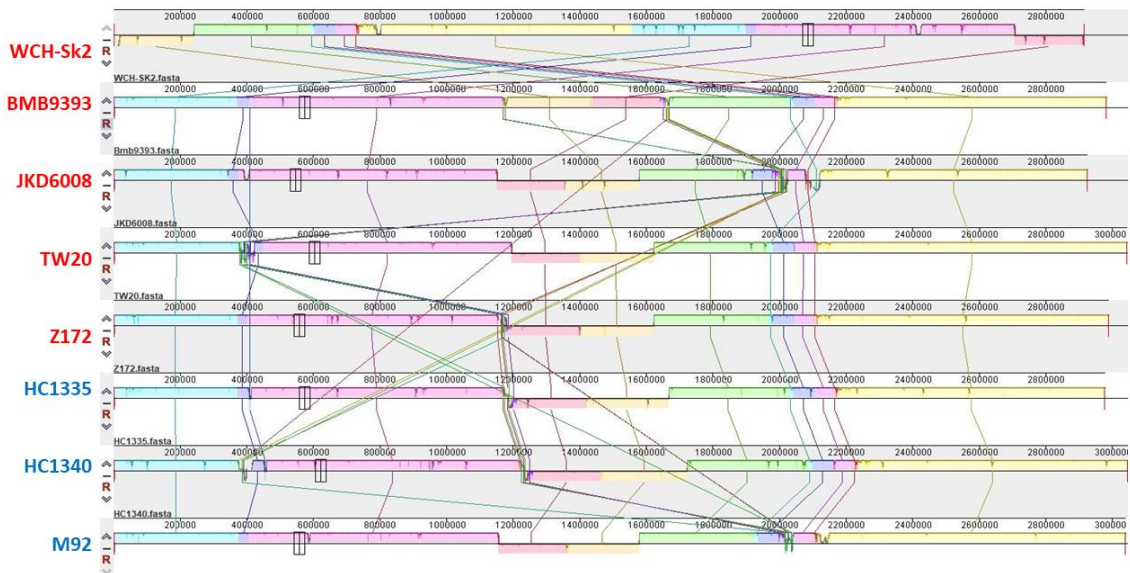


Figura 7-Representação do alinhamento dos 8 genomas completos pertencentes ao ST239, destacando os genomas do grupo colonização (em azul; 3 genomas) e do grupo infecção invasiva (em vermelho; 5 genomas). Todos os genomas são de MRSA. Figura obtida diretamente do MAUVE, apenas com a adição do nome do genoma.

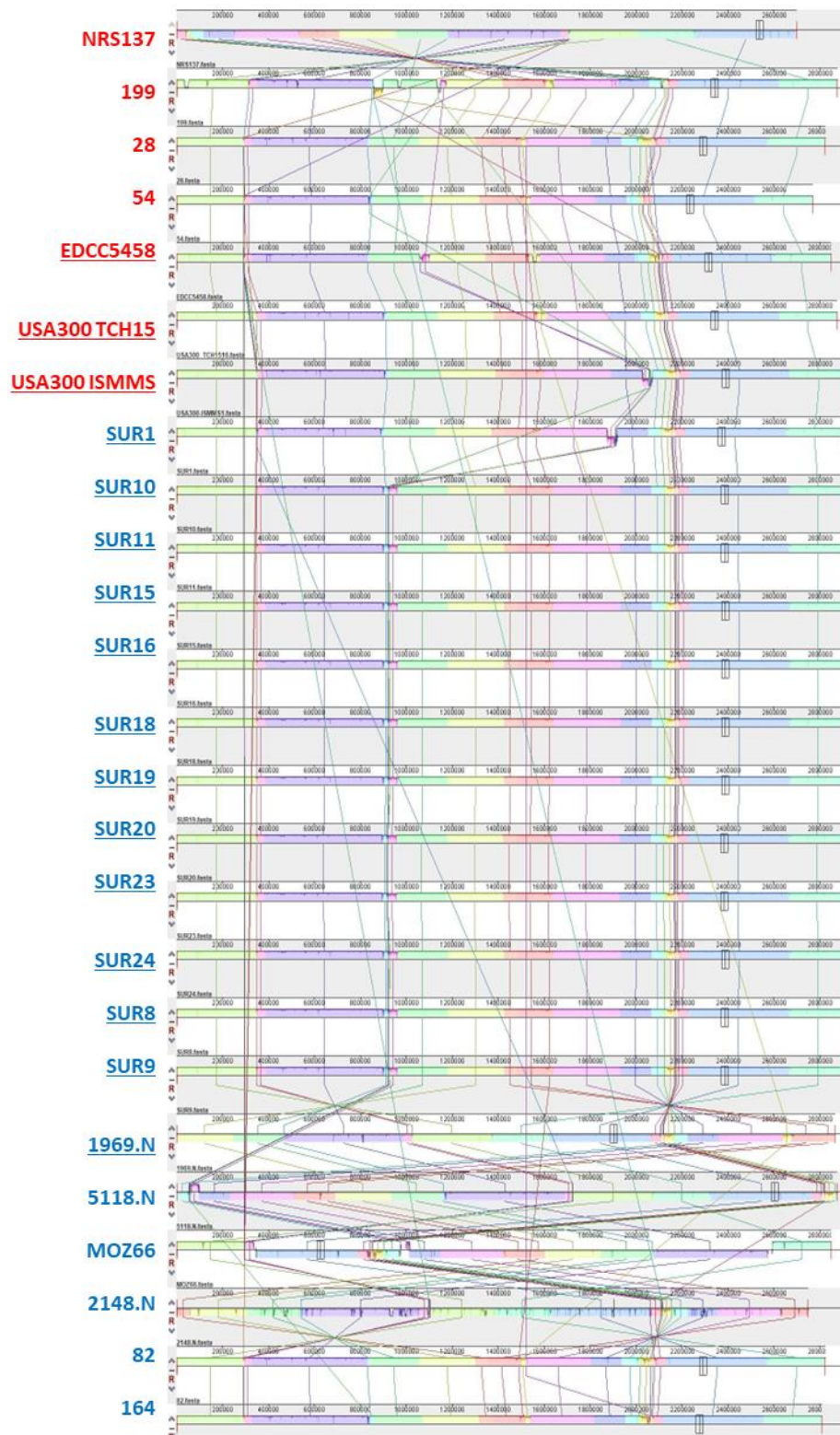
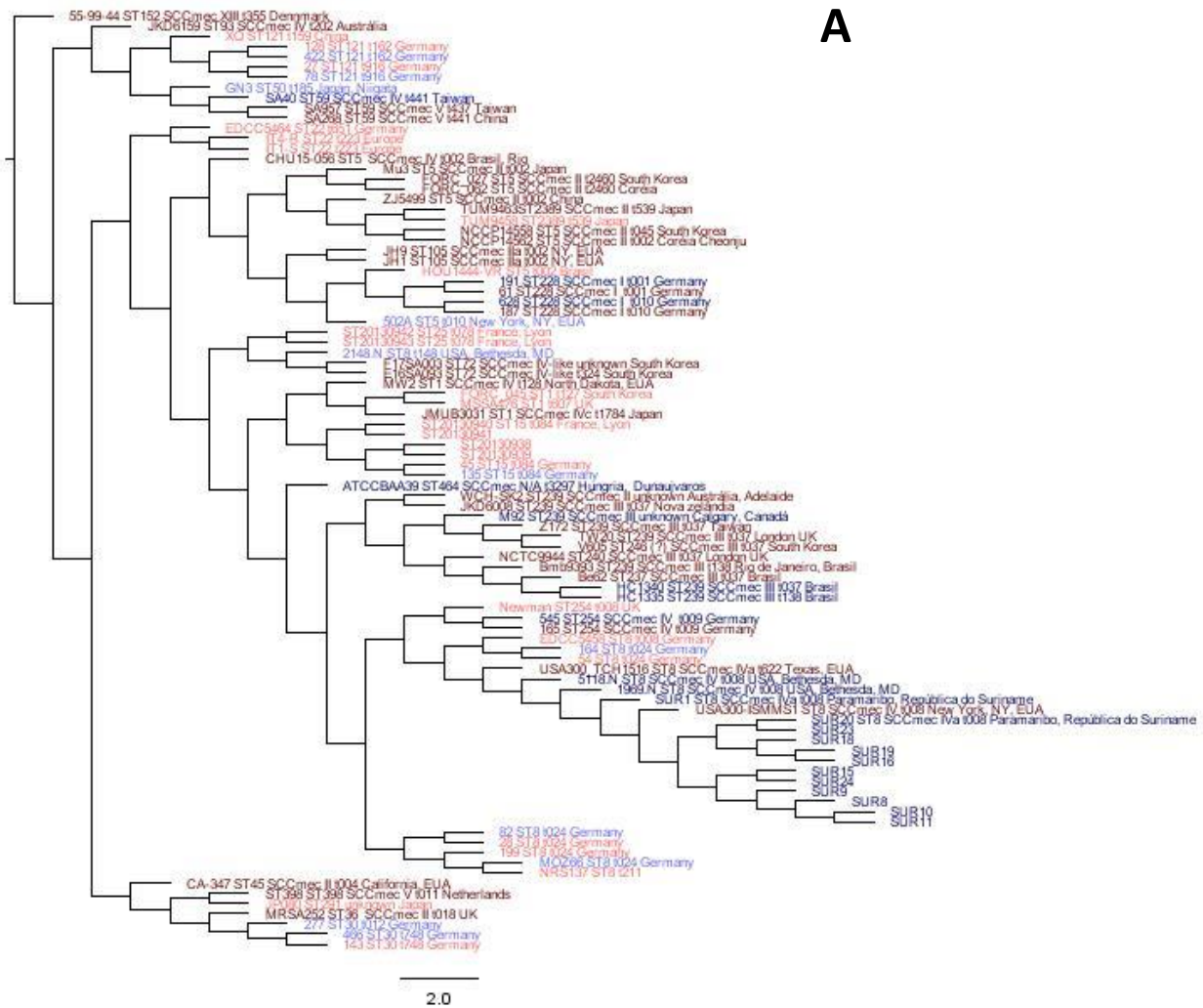


Figura 8-Representação do alinhamento dos 25 genomas completos pertencentes ao ST239, destacando os genomas do grupo colonização (em azul; 18 genomas) e do grupo infecção invasiva (em vermelho; 7 genomas). Os genomas que são de MRSA estão sublinhados. Figura obtida diretamente do MAUVE, apenas com a adição do nome do genoma.

4.5 Análise filogenética com base em alterações de SNPs

De forma a verificar uma possível diferença entre a distribuição de SNPs (*single nucleotide polymorphism*) no *core* genoma entre os dois grandes grupos (colonização e infecção), foram utilizados programas de bioinformática que montam uma árvore filogenética com base nestas diferenças. As árvores filogenéticas construídas com os programas CSIPhylogeny (Figura 9A) e RealPhy (Figura 9B) mostraram uma clusterização (agrupamento) que seguiu primariamente os STs e a tipagem de *spa*. Dentro de um mesmo ST, não foi observada a formação de grupos filogenéticos entre amostras de colonização ou de infecção. O ST8 possui 12 amostras de um mesmo estudo, que formaram um *cluster* (grupo). Não houve agrupamento visível por região geográfica. Com os dados de SNPs do CSIPhylogeny, não foi encontrada nenhuma diferença entre genomas de colonização e infecção de forma geral e nem dentro de um mesmo ST (Figura 9).



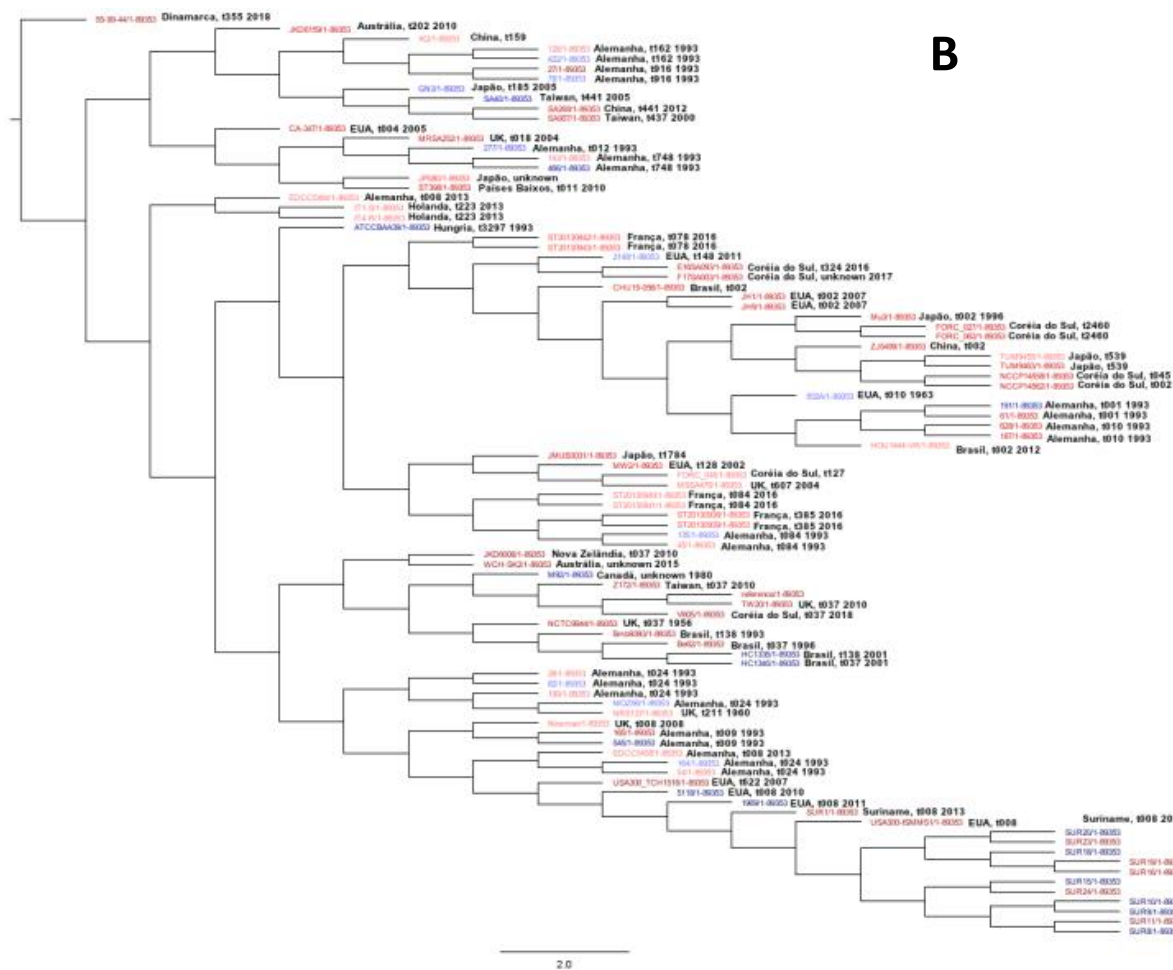


Figura 9-Árvores filogenéticas construída com base em diferenças em SNPs em todos os genomas do estudo. (A) Árvore obtida através do CSI Phylogeny. (B) Árvore obtida através do RealPhy. Genomas em vermelho escuro representam os genomas de infecção invasiva MRSA, vermelho claro genomas de infecção invasiva MSSA, azul escuro genomas de colonização MRSA e por fim em azul claro genomas de colonização MSSA.

4.6 Distribuição e análise dos genes de virulência

A tabela 3 indica todo os genes de virulência dos 92 genomas de *S. aureus* encontrados pelo VirulenceFinder, sendo apresentado genomas marcados em azul os genomas do grupo colonização e marcados em vermelho os genomas do grupo infecção invasiva.

Inicialmente, os genes de virulência que foram identificados e apresentaram maior incidência estatisticamente significativa no grupo dos genomas associados às amostras de colonização foram: (1) o gene codificante de ACME, identificado em 14 genomas associados a amostras de colonização (14/33; 42,4%) e 3 genomas associados a amostras de infecção invasiva (3/59; 5,1%) ($p = <0.0001$); (2) os genes *lukS*-PV/*lukF*-PV, codificantes da Leucocidina de Pantone-Valentine e identificados em 14 genomas de

colonização (14/33; 42,4%) e 8 genomas associados a amostras de infecção invasiva (8/59; 13,5%) ($p=0,0042$); genes *sek* e *seq*, codificantes das enterotoxinas K e Q, respectivamente, cada gene identificado em 18 genomas de colonização (18/33; 54,5%) e 15 genomas de infecção (15/59; 25,4%) ($p=0,0069$). Interessante notar que os genes *seq* e *sek* foram identificados simultaneamente nos mesmos genomas. No entanto, 12 entre 14 genomas positivos para todos estes genes são de um mesmo estudo em um mesmo hospital (SABAT et al., 2017; genomas de colonização com prefixo SUR; tabela 3), caracterizando-se como um resultado com viés. De forma a minimizar este viés, a análise estatística foi refeita considerando apenas 1 genoma “SUR”. Após esta análise, nenhum dos genes permaneceu com diferença significativa entre os genomas associados à colonização e infecção: 3 genomas de colonização (3/22; 13,6%) e 3 genomas de infecção (3/59; 5,1%) carregam ACME ($p=0,3367$), enquanto 3 genomas de colonização (3/22; 13,6%) e 8 genomas de infecção carregam *lukS*-PV/*lukF*-PV (8/59; 13,5%) ($p=0,9999$). Adicionalmente, a análise também foi refeita da mesma forma para os genes *sek* e *seq*, ambos identificados em 7 genomas de colonização (7/22; 31,8%) (excluindo-se 11 dos 12 genomas SUR) e 15 genomas de infecção (15/59; 25,4%) ($p=0,5834$) e sem diferença estatística.

Os genes de virulência que foram identificados e apresentaram maior incidência estatisticamente significativa no grupo dos genomas associados às amostras de infecção foram: *sea*, codificante da enterotoxina A [identificado em 13 genomas de infecção (13/59; 22,0%) e apenas 1 genoma de colonização (1/33; 3,0%); $p=0,0156$], *tst*, codificante da toxina 1 da síndrome do choque tóxico [identificado em 12 genomas de infecção (12/59; 20,3%) e apenas 1 genoma de colonização (1/33; 3,0%); $p=0,0277$]; e os genes *seg*, *seu*, *sem*, *sen*, *sei*, *seo*, codificantes das enterotoxinas G, U, M, N, I e O, respectivamente. Todos estes genes estão presentes no cluster *egc* (NAGARAJ et al 2016). O gene *seg* foi identificado em 8 genomas de colonização (8/33; 23,2%) e 27 genomas de infecção (27/59; 44,0%) ($p=0,0470$); os genes *seu* e *sem* em 9 genomas de colonização (9/33; 27,2%) e 27 genomas de infecção cada (27/59; 45,7%) ($p=0,0044$); o gene *sen* em 9 genomas de colonização (9/33; 27,2%) e 25 genomas de infecção (25/59; 42,3%) ($p=0,0042$); e os genes *sei* e *seo* em 9 genomas de colonização (9/33; 27,2%) e 25 genomas de infecção cada (42,3%) ($p=0,0069$) (Tabela 3). Demais genes com pouca incidência (menos de 10 genomas positivos para o gene considerando os dois grupos; colonização e infecção) não foram avaliados estatisticamente entre os grupos por terem sido considerados sem significado biológico relevante para o objetivo do estudo.

O coeficiente de Kappa de concordância entre os genes das plataformas RidomSeqSphere+ e Virulence Finder teve uma média de 0,624938, indicando uma boa concordância. Os genes *aur*, *hlgA*, *hlgB*, *hlgC*, *sep*, *sed*, *sej*, *ser*, *edinC* e *etb* apresentaram coeficiente de concordância menor que 0,6. Os genes *seb*, *seh* e *eta* tiveram concordância total em todas as amostras pelas duas plataformas analisadas. Os genes *LukSPV*, *sek*, *seq*, *sea*, *tst* e *seg* tiveram um coeficiente maior que 0,84, sendo considerado um nível de concordância muito bom.

4.7 Distribuição e análise dos genes de Resistência

A tabela 4 indica todos os genes de resistência dos 92 genomas de *S. aureus* encontrados pelo ResistanceFinder, sendo apresentados genomas marcados em azul os genomas do grupo colonização e marcados em vermelho os genomas do grupo infecção invasiva. Assim como na análise de distribuição dos genes de virulência, inicialmente os genes de resistência relacionados às seguintes classes apresentaram diferença estatisticamente significativa (teste de Fisher) entre os dois grandes grupos (colonização e infecção invasiva): aminoglicosídeo [genes *ant(6)-Ia* ($p=0,0008$) e *aph(3')-III* ($p=0,0008$); 17/33 (52%)], trimetoprima [gene *dfrG* ($p=0,0104$); 11/33 (33,3%)] e macrolídeos [genes *mph(C)* ($p<0,0001$); 13/33 (39,39%), *msr(A)* ($p<0,0001$); 13/33 (39,39%) e *erm(C)* ($p=0,0232$); 6/33 (18,18%)]. Todos estes genes foram mais encontrados nos genomas relacionados à colonização. No entanto, 12 entre 14 genomas positivos para alguns destes genes são genomas de colonização com prefixo SUR; SABAT et al., 2017), caracterizando-se novamente como um resultado com viés (Tabela 4). De forma a minimizar este viés, a análise estatística foi refeita considerando apenas 1 genoma “SUR”. Após esta nova análise, nenhum dos genes permaneceu com diferença significativa entre os dois grupos.

Nos plasmídeos analisados via PlasmidFinder, apenas o rep7c apresentou significância estatística entre os grupos, estando com maior frequência nos genomas associados à colonização (19/33 genomas; 57,5%) do que nos genomas associados a infecção (16/59; 35,6%) ($p=0,0067$). A tabela 6 indica todas as sequências de inserção (IS) dos 92 genomas de *S. aureus* encontrados na plataforma online ISFinder, sendo apresentado genomas marcados em azul os genomas do grupo colonização e, marcados em vermelho, os genomas do grupo infecção invasiva. Apenas a sequência de inserção ISLgar5 apresentou significância estatística ($p= 0,0033$), aparecendo com maior frequência em genomas associados a infecções invasivas (13/59; 22,03%) (Tabela 6).

Tabela 6-Diversidade de seqüências de inserção presentes nos genomas do estudo segundo o programa ISFinder. Em azul estão os genomas associados à colonização e, em vermelho, os genomas associados à infecções invasivas; primeira coluna. O ST (*Sequence Typing*) e o perfil de sensibilidade à meticilina (S: sensível; MSSA; R: resistente; MRSA) de cada genoma estão indicados na segunda e terceira colunas, respectivamente. Os quadrados pintados em verde indicam a presença da seqüência de inserção. Grupo de genes de plasmídeos foram separados conforme o proposto pelo ISFinder.

28	ST8	S					
54	ST8	S					
199	ST8	S					
EDCC5458	ST8	S					
NRS137	ST8	S					
SA300_TCH15	ST8	R					
SA300-ISMM5	ST8	R					
FORC_027	ST5	R					
FORC_062	ST5	R					
HOU1444-VR	ST5	S					
Mu3	ST5	R					
NCCP14558	ST5	R					
NCCP14562	ST5	R					
ZJ5499	ST5	R					
CHU15-056	ST5	R					
Bmb9393	ST239	R					
KD6008	ST239	R					
TW20	ST239	R					
WCH-SK2	ST239	R					
Z172	ST239	R					
27	ST121	S					
128	ST121	S					
XQ	ST121	S					
45	ST15	S					
ST20130938	ST15	S					
ST20130939	ST15	S					
ST20130940	ST15	S					
ST20130941	ST15	S					
FORC_045	ST1	S					
JMUB3031	ST1	R					
MSSA476	ST1	S					
MW2	ST1	R					
61	ST228	R					
187	ST228	R					
EDCC5464	ST22	S					
IT1-S	ST22	S					
IT4-R	ST22	S					
SA268	ST59	R					
SA957	ST59	R					
143	ST30	S					
165	ST254	R					
Newman	ST254	S					
TUM9458	ST2389	S					
TUM9463	ST2389	R					
ST20130942	ST25	S					
ST20130943	ST25	S					
JH1	ST105	R					
JH9	ST105	R					
MRSA252	ST36	R					
ST398	ST398	R					
CA-347	ST45	R					
E16SA093	ST72	R					
55-99-44	ST152	R					
JP080	ST291	S					
Be62	ST237	R					
F17SA003	ST72	R					
NCTC9944	ST240	R					
V605	ST246	R					
KD6159	ST93	R					
82	ST8	S					
164	ST8	S					
1969_N	ST8	R					
2148_N	ST8	S					
5118_N	ST8	R					
MOZ66	ST8	S					
SUR1	ST8	R					
SUR10	ST8	R					
SUR11	ST8	R					
SUR15	ST8	R					
SUR16	ST8	R					
SUR18	ST8	R					
SUR19	ST8	R					
SUR20	ST8	R					
SUR23	ST8	R					
SUR24	ST8	R					
SUR8	ST8	R					
SUR9	ST8	R					
502A	ST5	S					
HC1335	ST239	R					
HC1340	ST239	R					
M92	ST239	R					
78	ST121	S					
422	ST121	S					
135	ST15	S					
191	ST228	R					
628	ST228	R					
SA40	ST59	R					
277	ST30	S					
466	ST30	S					
545	ST254	R					
ATCCBAA39	ST464	R					
GN3	ST50	S					
Genoma	Sequence Type	Resistência à Meticilina	ISSau3	IS1181	IS1272	IS256	IS1gar5

5 DISCUSSÃO

A análise genômica é um dos recursos que possibilitam estudar sobre a biologia de bactérias tanto no âmbito da pesquisa básica quanto da pesquisa aplicada (CAPUTO; FOURNIER; RAOULT, 2019). Estudar possíveis diferenças entre genomas de *S. aureus* completamente sequenciados a partir de amostras bacterianas isoladas de colonização e infecção invasiva em humanos pode desvendar potenciais novos marcadores de infecção, auxiliando no entendimento da transição do processo de colonização para o processo de infecção, que ocorre comumente em *S. aureus* (CHEUNG; BAE; OTTO, 2021).

O presente estudo analisou 92 genomas completamente sequenciados de *S. aureus* com características bastante heterogêneas, considerando plataforma de sequenciamento, ano de isolamento, linhagem (ST e tipo *SCCmec*), localização geográfica e sítio biológico de isolamento da bactéria. Esperava-se que esta heterogeneidade auxiliasse no estudo de um possível marcador de infecção universal, uma vez que a maioria dos estudos com o mesmo objetivo deste trabalho utilizaram genomas com características mais homogêneas, seja considerando genomas de uma mesma linhagem, seja considerando genomas de uma mesma localização geográfica ou de um mesmo contexto clínico (BENOIT; FRANK; BESSESEN, 2018). No entanto, o presente estudo revelou que em termos de análise pangenômica, alinhamento genômico completo e diferenças em SNPs no *core* genoma, não foi possível visualizar diferenças que permitissem uma separação entre os genomas de colonização e infecção invasiva, impossibilitando a identificação de regiões genômicas com potencial para serem regiões importantes no contexto de infecção.

Com relação a distribuição dos genes de virulência, verificou-se que os genes codificantes da enterotoxina A (*sea*), toxina 1 da síndrome do choque tóxico (*tst*) e enterotoxinas G, U, M, N, I e O (*seg*, *seu*, *sem*, *sem*, *sei*, *seo*, respectivamente) apresentaram-se com maior frequência nos genomas do grupo infecções invasivas. Todas estas toxinas têm a característica em comum de terem atividade de superantígeno, cuja estimulação inespecífica e exacerbada de células T está correlacionada a uma ativação desregulada da resposta imunológica que, se sustentada, promove uma sobrecarga maciça de citocinas, prenunciando características clínicas de início rápido de febre, falência múltipla de órgãos e mortalidade significativa (FISHER, 2018).

As enterotoxinas estafilocócicas são classicamente associadas a intoxicações alimentares em razão do efeito emético associado a algumas delas (como a enterotoxina

G e A). Entretanto, estas enterotoxinas já foram descritas em associação com diversas manifestações clínicas em humanos, como pneumonia, infecções relacionadas a sepse, síndrome do choque tóxico, infecções respiratórias e desenvolvimento de doenças autoimunes (FISHER, 2018). Apesar deste potencial clínico, cerca de 80% dos isolados de *S. aureus*, incluindo cepas comensais e clínicas carregam em média 5-6 genes codificantes de enterotoxinas (NAGARAJ, 2016; FISHER, 2018). A enterotoxina A (codificada por *sea*) é uma enterotoxina clássica descrita desde a década de 1960, com efeito emético bem documentado, sendo principalmente associada a surtos de intoxicações alimentares (CHEN, 2018; HAGHI, 2021). Até o momento, não há evidência científica da prevalência do gene *sea* em amostras de infecção invasiva.

Os genes *seg*, *seu*, *sem*, *sem*, *sei* e *seo* estão localizados no *cluster* *egc*, que corresponde a um operon policistrônico, que carrega, além destes genes, dois pseudogenes (ϕ ent1 e ϕ ent2) (LETERTRE, 2003; SCHWENDIMANN, 2021). Interessantemente, estes genes do operon não foram encontrados com a mesma frequência em todos os genomas positivos (Tabela 3), evidenciando que nem sempre os genes codificantes destas enterotoxinas podem ser detectados simultaneamente. Os genes presentes no operon podem apresentar variações na sequência (NAGARAJ, 2016), que talvez não sejam detectadas com identidade mínima no banco de dados das plataformas analisadas. O *cluster* *egc* já foi associado a doenças como febre escarlatina (enterotoxina G e I), eczema atópico, fascite necrotizante, enteropatia neonatal, endocardite e sepse (NAGARAJ, 2016). No entanto, este cluster também já foi encontrado em cepas consideradas não patogênicas (NAGARAJ, 2016).

A toxina 1 da síndrome do choque tóxico (TSST-1, codificada pelo gene *tst*) está associada a síndrome do choque tóxico, que é caracterizada por febre, hipotensão, *rash* e descamação cutânea e falência múltipla dos órgãos, bem como sintomas gastrointestinais, disfunção renal, falência renal, trombocitopenia e danos ao sistema nervoso central (TAKI, 2022). Esta síndrome pode ou não ser relacionado ao uso de tampões menstruais e é considerada uma infecção grave, com taxa de mortalidade documentada de até 8,5% (TAKI, 2022). O gene *tst* é comumente encontrado entre amostras clínicas de *S. aureus*, uma vez que cepas de *S. aureus* isoladas de pacientes com esta síndrome carregam o gene em 81.8-100% dos casos (TAKI, 2022). No entanto, estima-se que 20% de todas as amostras de *S. aureus* carregam *tst*, incluindo cepas comensais (TAKI, 2022). Cabe ressaltar que a enterotoxina A (gene *sea*) também já foi identificada como agente causador da síndrome do choque tóxico, bem como a

enterotoxina B e C, cujos genes codificantes (*seb* e *sec*, respectivamente) não foram detectados nos genomas de infecção com maior frequência (TAKI, 2022).

A presença e a funcionalidade de genes de resistência aos antimicrobianos em cepas de *S. aureus* pode estar correlacionada com a sobrevivência, multiplicação e persistência da bactéria no hospedeiro (DURÃO; BALBONTÍN; GORDO, 2018; KNIGHT; BUDD; LINDSAY, 2013). Em geral, a resistência aos antimicrobianos é mais frequentemente encontrada em amostras de *S. aureus* isoladas de locais de atendimento à saúde, provavelmente em razão da pressão seletiva ser maior nestes locais pelo uso aumentado de antimicrobianos, em comparação com a comunidade (KATEETE, 2019). Não foi possível obter a informação se as amostras bacterianas de colonização utilizados para a obtenção dos genomas foram ou não obtidas em um contexto de local de atendimento à saúde ou de grande pressão seletiva com uso de antimicrobianos, não sendo possível investigar melhor a distribuição de genes de resistência entre estes genomas. Diversos estudos sugerem que modificações genéticas que conferem resistência à antibióticos podem causar uma diminuição nas taxas de crescimento da bactéria, interferindo assim nos processos fisiológicos normais do microrganismo, pois essas modificações geram um custo de manutenção elevado, devido a expressão simultânea de muitos genes (DURÃO; BALBONTÍN; GORDO, 2018). Dessa maneira, a aquisição de um fenótipo de resistência pode ser um importante contribuinte para um impacto negativo no *fitness* dos isolados, que afeta seu crescimento e, conseqüentemente, sua disseminação (FREY, 2021). Assim, o impacto negativo no *fitness* causado pelas marcas de resistência aos antimicrobianos pode ser um fator limitante para o mecanismo de transição colonização-infecção ocorrer. Porém, mais estudos são necessários para entender esta dinâmica, não tendo sido possível, neste estudo, investigar o impacto da distribuição destes genes neste mecanismo.

Além dos genes centrais (*core* genoma), existem vários elementos genéticos móveis (EGM) adquiridos por transferência horizontal de genes que podem ser benéficos para *S. aureus* sob certas condições ambientais, sendo importantes para plasticidade bacteriana e evolução (ALIBAYOV, 2014). Diante disso, EGMs podem ter papel relevante nos mecanismos de patogenicidade desta bactéria e, por representarem cerca de 15% dos genomas de *S. aureus* (ALIBAYOV, 2014), foram investigados neste estudo. Com relação a distribuição de elementos genéticos móveis entre os genomas, até o momento não foi encontrada nenhuma evidência científica publicada sobre como o plasmídeo *rep7c* e o *ISLgar5* poderiam estar correlacionados a mecanismos de

colonização e patogenicidade de *S. aureus* no hospedeiro humano, respectivamente. rep7c já foi documentado carregando genes relacionados a resistência a metais pesados e genes de virulência (HOLDEN, 2004). Adicionalmente, uma variante deste plasmídeo (rep7) já foi detectado em genomas de MRSA obtidos a partir de amostras de *S. aureus* isoladas de cavidade nasal (colonização) de trabalhadores de fazendas de animais de produção na Grécia (KARAMPATAKIS, 2021), o que pode inferir alguma importância deste no contexto de colonização. Porém, esta potencial importância precisa ser melhor investigada através de mais estudos.

Classicamente, *S. aureus* são distribuídos de forma clonal, sendo frequentemente detectados em surtos ocasionados por linhagens que predominam em determinados contextos e localidades (FIGUEIREDO & FERREIRA, 2014). Cada clone/linhagem pode ser definido a partir do CC/ST e do tipo de SCC mec e comumente cada clone/linhagem tem um repertório de distribuição de genes e outros segmentos de DNA característicos (FIGUEIREDO & FERREIRA, 2014). Por exemplo: o clone epidêmico brasileiro (CEB) é classificado comumente como CC8, ST239 e SCC mec III e geralmente é um clone que apresenta multirresistência, tendo um vasto repertório de genes de resistência (ANDRADE, 2020). As análises genômicas do presente estudo corroboraram com os dados da literatura em mostrar que os genomas de cada clone/linhagem são bastante semelhantes entre si, principalmente quanto a distribuição dos genes analisados (Tabela 3, 4, 5 e 6). No entanto, esta característica também pode representar um importante viés do presente estudo, uma vez que se houver mais genomas de um determinado clone/linhagem sendo analisado em um grupo (conforme ocorreu para os genomas “SUR”), um marcador de infecção por ser erroneamente apontado, pois a prevalência será oriunda da linhagem e não do grupo infecção.

Além das características de cada clone/linhagem de *S. aureus*, a distribuição dos genes de resistência antimicrobiana também pode ser influenciada pela localização geográfica, uma vez o uso de antimicrobianos varia entre países/regiões, podendo assim exercer pressões seletivas diferentes, que contribuem na diferenciação do repertório destes genes (BROWNE, 2021). Um exemplo deste cenário ocorreu no estudo de Guimarães e colaboradores (2015), onde duas cepas pertencentes ao clone USA400 (linhagem ST1-SCC mec IV), sendo uma isolada dos Estados Unidos da América e outra do Brasil (Rio de Janeiro) foram comparadas. A cepa norte americana foi isolada na comunidade (portanto, supõe-se com menor pressão seletiva), enquanto a cepa brasileira foi isolada no hospital. Observou-se que, apesar de muito similares geneticamente, o

perfil de resistência aos antimicrobianos era diferente, sendo a cepa brasileira resistente a maior número de antimicrobianos, sugerindo, portanto, que a pressão seletiva é importante na distribuição do repertório destes genes, conforme já observado em outros estudos (PENG, 2018). Países em desenvolvimento tendem a ter cepas hospitalares mais resistentes a antimicrobianos quando comparadas a cepas oriundas de países desenvolvidos (WHO, 2014).

Outras limitações nas análises genômicas do presente estudo também podem ser apontadas, como: (i) variação na cobertura dos genomas, que pode impactar o alinhamento das sequências contra os bancos de dados; (ii) os próprios bancos de dados das plataformas utilizadas, que podem não incluir variantes gênicas importantes para a identificação de uma sequência; (iii) diferenças entre plataformas de sequenciamento, que também podem influenciar na cobertura e nas variantes; (iv) o conjunto de genomas selecionado não contempla todos os *SSCmec* ou linhagens já descritos, limitando e enviesando as análises; e (v) não ter sido realizada uma análise de proteínas hipotéticas, regiões não codificantes e do transcriptoma dessas amostras, que poderiam auxiliar na elucidação dos mecanismos.

Mesmo o presente estudo tendo encontrado alguns genes de virulência que podem ser potenciais marcadores de infecção, estes dados apresentam limitações metodológicas e vieses importantes, devendo estes marcadores serem melhor investigados quanto sua importância na transição colonização-infecção. Em geral, estudos comparando genomas completos ou sequências de DNA específicas entre cepas de *S. aureus* isolados de colonização ou infecção não encontram diferenças significativas entre os dois grupos (BENOIT; FRANK; BESSESEN, 2018, YOUNG et al., 2012, FITZGERALD, 2014). Além disso, fatores do hospedeiro ou da microbiota podem ser determinantes na transição, Mulcahy e McLoughlin (2016) citam a produção de H₂O₂ por *S. pneumoniae*, que causa a ativação de fagos lisogênicos de *S. aureus*. Uma colonização persistente por *S. aureus* está ligada a uma falha genética na produção de peptídeos antimicrobianos, como no caso de SNPs no gene DEFB-1, um produtor de beta defensinas.

5.1 Conclusão

Com os dados apresentados, não foi possível refutar e nem confirmar a hipótese. O presente trabalho evidenciou que o mecanismo de transição colonização-infecção é complexo, e necessita de outras abordagens metodológicas para ser mais bem

compreendido em *S. aureus*. Análises transcriptômicas/proteômicas/metabolômicas, bem como experimentos *in vivo* são abordagens que podem auxiliar no entendimento deste mecanismo.

REFERÊNCIAS

- ABDELBARY, M. M. H. et al. The Evolution and Dynamics of Methicillin-Resistant *Staphylococcus aureus*. [s.l.] Elsevier Inc., 2017.
- BABIKER, A. et al. The use of Online Tools for Antimicrobial Resistance Prediction by Whole Genome Sequencing in MRSA and VRE. **Journal of Global Antimicrobial Resistance**, 2019a.
- BABIKER, A. et al. The use of Online Tools for Antimicrobial Resistance Prediction by Whole Genome Sequencing in MRSA and VRE. **Journal of Global Antimicrobial Resistance**, 2019b.
- BERUBE, B. J.; WARDENBURG, J. B. *Staphylococcus aureus* α -toxin: Nearly a century of intrigue. **Toxins**, v. 5, n. 6, p. 1140–1166, 2013.
- BITRUS, A. A. et al. *Staphylococcus aureus*: A Review of Antimicrobial Resistance Mechanisms. **Veterinary Sciences: Research and Reviews**, v. 4, n. 2, 2018.
- BRIDE, L. DE L. et al. Differences in resistance profiles and virulence genes among methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* of different lineages at a public tertiary hospital. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 52, p. 1–7, 2019.
- BUI, L. M. G.; KIDD, S. P. A full genomic characterization of the development of a stable Small Colony Variant cell-type by a clinical *Staphylococcus aureus* strain. **Infection, Genetics and Evolution**, v. 36, p. 345–355, 2015.
- CARATTOLI, A. et al. In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. **Antimicrobial Agents and Chemotherapy**, v. 58, n. 7, p. 3895–3903, 2014.
- COPIN, R.; SHOPSIN, B.; TORRES, V. J. After the deluge: mining *Staphylococcus aureus* genomic data for clinical associations and host–pathogen interactions. **Current Opinion in Microbiology**, 2018.
- DAVID, M. Z.; DAUM, R. S. Community-associated methicillin-resistant *Staphylococcus aureus*: Epidemiology and clinical consequences of an emerging epidemic. **Clinical Microbiology Reviews**, v. 23, n. 3, p. 616–687, 2010.
- DE BACKER, S. et al. Remarkable geographical variations between India and Europe in carriage of the staphylococcal surface protein-encoding *sasX/sesI* and in the population structure of methicillin-resistant *Staphylococcus aureus* belonging to clonal complex 8. **Clinical Microbiology and Infection**, p. 1–7, 2018.
- DEINHARDT-EMMER, S. et al. Virulence patterns of *Staphylococcus aureus* strains from nasopharyngeal colonization. **Journal of Hospital Infection**, v. 100, n. 3, p. 309–315, 2018.
- DOBRINDT, U. et al. Genomic islands in pathogenic and environmental microorganisms. **Nature Reviews Microbiology**, v. 2, n. 5, p. 414–424, 2004.
- DURÃO, P.; BALBONTÍN, R.; GORDO, I. Evolutionary Mechanisms Shaping the Maintenance of Antibiotic Resistance. **Trends in Microbiology**, v. 26, n. 8, p. 677–

691, 2018.

EIFF, V. Nasal Carriage As a Source of Staphylococcus Aureus Bacteremia. **English Journal**, v. 344, n. 1, p. 11–16, 2001.

FIGUEIREDO, A. M. S.; FERREIRA, F. A. The multifaceted resources and microevolution of the successful human and animal pathogen methicillin-resistant Staphylococcus aureus. **Memorias do Instituto Oswaldo Cruz**, v. 109, n. 3, p. 265–278, 2014.

FOSTER, T. J. et al. Adhesion, invasion and evasion: the many functions of the surface proteins of Staphylococcus aureus. **Nat.Rev.Microbiol.**, v. 12, n. 1740–1534 (Electronic), p. 49–62, 2014.

FOSTER, T. J. The remarkably multifunctional fibronectin binding proteins of Staphylococcus aureus. **European Journal of Clinical Microbiology and Infectious Diseases**, v. 35, n. 12, p. 1923–1931, 2016.

FOSTER, T. J. Antibiotic resistance in Staphylococcus aureus. Current status and future prospects. **FEMS Microbiology Reviews**, v. 41, n. 3, p. 430–449, 2017.

GIULIERI, S. G. et al. Use of bacterial whole-genome sequencing to understand and improve the management of invasive Staphylococcus aureus infections. **Expert Review of Anti-Infective Therapy**, v. 14, n. 11, p. 1023–1036, 2016.

HAABER, J. et al. Bacterial viruses enable their host to acquire antibiotic resistance genes from neighbouring cells. **Nature Communications**, v. 7, 2016.

HAABER, J.; PENADÉS, J. R.; INGMER, H. Transfer of Antibiotic Resistance in Staphylococcus aureus. **Trends in Microbiology**, v. 25, n. 11, p. 893–905, 2017.

HASSOUN, A.; LINDEN, P. K.; FRIEDMAN, B. Incidence, prevalence, and management of MRSA bacteremia across patient populations-a review of recent developments in MRSA management and treatment. **Critical care** (London, England), v. 21, n. 1, p. 211, 2017.

HU, D. L.; NAKANE, A. Mechanisms of staphylococcal enterotoxin-induced emesis. **European Journal of Pharmacology**, v. 722, n. 1, p. 95–107, 2014.

HUMPHREYS, H.; COLEMAN, D. C. Contribution of whole-genome sequencing to understanding of the epidemiology and control of methicillin-resistant Staphylococcus aureus **Journal of Hospital Infection**, 2019.

IWASE, T. et al. Staphylococcus epidermidis Esp inhibits Staphylococcus aureus biofilm formation and nasal colonization. **Nature**, v. 465, n. 7296, p. 346–349, 2010.

JANEK, D. et al. High Frequency and Diversity of Antimicrobial Activities Produced by Nasal Staphylococcus Strains against Bacterial Competitors. **PLoS Pathogens**, v. 12, n. 8, 2016.

JANI, M. et al. Deciphering pathogenicity and antibiotic resistance islands in methicillin-resistant Staphylococcus aureus genomes. **Open Biology**, v. 7, n. 12, 2017.

JOENSEN, K. G. et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. **Journal of**

- Clinical Microbiology**, v. 52, n. 5, p. 1501–1510, 2014.
- KRISMER, B. et al. The commensal lifestyle of *Staphylococcus aureus* and its interactions with the nasal microbiota. **Nature reviews. Microbiology**, v. 15, n. 11, p. 675–687, 2017.
- KWONG, J. C. et al. Whole genome sequencing in clinical and public health microbiology. **Pathology**, v. 47, n. 3, p. 199–210, 2015.
- LAKHUNDI, S.; ZHANG, K. Methicillin-Resistant *Staphylococcus aureus* : Molecular Characterization, Evolution, and Epidemiology. **Clinical Microbiology Reviews**, v. 31, n. 4, p. 1–103, 12 set. 2018.
- LI, Z. A Review of *Staphylococcus aureus* and the Emergence of Drug-Resistant Problem. **Advances in Microbiology**, v. 08, n. 01, p. 65–76, 2018.
- MCARTHUR, A. G. et al. The comprehensive antibiotic resistance database. **Antimicrobial Agents and Chemotherapy**, v. 57, n. 7, p. 3348–3357, 2013.
- MULCAHY, M. E.; MCLOUGHLIN, R. M. Host – Bacterial Crosstalk Determines *Staphylococcus aureus* Nasal Colonization. **Trends in Microbiology**, v. 24, n. 11, p. 872–886, 2016.
- NOVICK, R. P.; CHRISTIE, G. E.; PENADÉS, J. R. The phage-related chromosomal islands of Gram-positive bacteria. **Nature Reviews Microbiology**, v. 8, n. 8, p. 541–551, ago. 2010.
- NOVICK, R. P.; RAM, G. The Floating (Pathogenicity) Island: A Genomic Dessert. **Trends in Genetics**, v. 32, n. 2, p. 114–126, 2016.
- O’GARA, J. P. Into the storm: Chasing the opportunistic pathogen *Staphylococcus aureus* from skin colonisation to life-threatening infections. **Environmental Microbiology**, v. 19, n. 10, p. 3823–3833, 2017.
- PARTRIDGE, S. R. et al. Mobile genetic elements associated with antimicrobial resistance. **Clinical Microbiology Reviews**, v. 31, n. 4, p. 1–61, 2018.
- QUINN, G. A.; COLE, A. M. Suppression of innate immunity by a nasal carriage strain of *Staphylococcus aureus* increases its colonization on nasal epithelium. **Immunology**, v. 122, n. 1, p. 80–89, 2007.
- RAMMELKAMP, C. H.; MAXON, T. Resistance of *Staphylococcus aureus* to the Action of Penicillin. **Experimental Biology and Medicine**, v. 51, n. 3, p. 386–389, 1942.
- ROSSI, F. The challenges of antimicrobial resistance in Brazil. **Clinical Infectious Diseases**, v. 52, n. 9, p. 1138–1143, 2011.
- ROSSI, F. et al. Transferable Vancomycin Resistance in a Community-Associated MRSA Lineage. **New England Journal of Medicine**, v. 370, n. 16, p. 1524–1531, 2014.
- SABAT, A. J. et al. Complete-genome sequencing elucidates outbreak dynamics of CA-MRSA USA300 (ST8-spa t008) in an academic hospital of Paramaribo, Republic of Suriname. **Scientific Reports**, v. 7, n. October 2016, p. 1–12, 2017.

- SIGUIER, P. et al. ISfinder: the reference centre for bacterial insertion sequences. **Nucleic acids research**, v. 34, n. Database issue, p. 32–36, 2006.
- TONG, S. Y. C. et al. Staphylococcus aureus infections: Epidemiology, pathophysiology, clinical manifestations, and management. **Clinical Microbiology Reviews**, v. 28, n. 3, p. 603–661, 2015.
- TURNER, N. A. et al. Methicillin-resistant Staphylococcus aureus: an overview of basic and clinical research. **Nature Reviews Microbiology**, v. 17, n. 4, p. 203–218, 2019.
- URUSHIBARA, N. et al. Novel staphylococcal cassette chromosome mec (SCCmec) type XIV (5A) and a truncated SCCmec element in SCC composite islands carrying speG in ST5 MRSA in Japan. **Journal of Antimicrobial Chemotherapy**, v. 75, n. 1, p. 46–50, 2020.
- WERTHEIM, H. F. L. et al. Key role for clumping factor B in Staphylococcus aureus nasal colonization of humans. **PLoS Medicine**, v. 5, n. 1, p. 0104–0112, 2008.
- WORLD HEALTH ORGANIZATION. Antimicrobial resistance – global report on surveillance. World Health Organization; 2014.22