



Estudio de la variabilidad genética de genes relacionados con factores de virulencia en aislados clínicos de *Staphylococcus aureus* causantes de bacteriemias.

Herdoiza Montero, Jean Pierre

Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Trabajo de titulación, previo a la obtención del título de Ingeniero en Biotecnología

Grijalva Silva, Rodrigo Marcelo M. D. PhD.

20 de enero del 2022



Herdoiza_Jean_Proyecto de Titulación_SinPortada para Copy...

Scanned on: 19:6 January 27, 2022 UTC



Overall Similarity Score



Results Found



Total Words in Text

Identical Words	741
Words with Minor Changes	24
Paraphrased Words	233
Ommited Words	0



Proyecto de Titulación_SinPortada para Copy...
RODRIGO MARCELO
GRIJALVA SILVA



Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Certificación

Certifico que el trabajo de titulación, “**Estudio de la variabilidad genética de genes relacionados con factores de virulencia en aislados clínicos de *Staphylococcus aureus* causantes de bacteriemias**” fue realizado por el señor **Herdoiza Montero, Jean Pierre** el cual ha sido revisado y analizado en su totalidad por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 27 de Enero, 2022

Firma:



Firmado electrónicamente por:
**RODRIGO MARCELO
GRIJALVA SILVA**

.....

Dr. Marcelo Grijalva M.D., Ph.D.

C.C.: 1706590641



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Autorización de publicación

Yo, **Herdoiza Montero, Jean Pierre**, con C.C. 0401906102, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: "**Estudio de la variabilidad genética de genes relacionados con factores de virulencia en aislados clínicos de *Staphylococcus aureus* causantes de bacteriemias**" en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 27 de enero de 2022

Herdoiza Montero, Jean Pierre

C.C.: 0401906102



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

Responsabilidad de autoría

Yo, **Herdoiza Montero, Jean Pierre**, con C.C. 0401906102, declaro que el contenido, ideas y criterios del trabajo de titulación: **"Estudio de la variabilidad genética de genes relacionados con factores de virulencia en aislados clínicos de *Staphylococcus aureus* causantes de bacteriemias"** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 27 de enero de 2022

Herdoiza Montero, Jean Pierre

C.C.: 0401906102

Índice de Contenidos

Índice de Tablas	10
Índice de Figuras	11
Listado de Abreviaturas.....	13
Resumen	15
Palabras clave:	15
Abstract.....	16
Keywords:	16
Capítulo I: Introducción	17
Formulación del problema.....	17
Justificación del problema	18
Objetivos de la Investigación	20
Objetivo General.....	20
Objetivos Específicos	20
Capítulo II: Revisión Bibliográfica.....	20
<i>Staphylococcus aureus</i>	20
Generalidades	20
Importancia Clínica	22
Factores de Virulencia.....	23
Genómica	26
Fundamento Teórico de los Métodos Empleados	29
Extracción de ADN.	29
Secuenciación de Nueva Generación y Secuenciación Illumina	33

Preparación de Librerías	33
Secuenciación.	36
Análisis de Datos derivados de Secuenciación Illumina.	39
Hipótesis	42
Capítulo III: Metodología.....	43
Zona de estudio	43
Población de estudio.....	43
Técnicas de Laboratorio.....	45
Extracción de ADN	45
Preparación de la Muestra.	45
Lisis Celular.	45
Precipitación de ADN.....	45
Purificación de ADN.	45
Tratamiento con RNAsas.	46
Evaluación de la Integridad, Cuantificación y Conservación del ADN.	46
Preparación de Librerías para Secuenciación NGS Illumina	47
Tagmentación.	47
Limpieza Post-tagmentación.	48
Amplificación del ADN Tagmentado.....	49
Limpieza Post-PCR.....	50
Control de Calidad de Librerías.....	51
Pool de Librerías.	51
Desnaturalización de las Librerías.	52

Análisis Bioinformático	52
Control de Calidad de las Secuencias Obtenidas.	52
Preprocesamiento de las Secuencias Obtenidas.	52
Identificación de Especie y Evaluación de Presencia de Contaminantes.	53
Cobertura y Profundidad de Secuenciación.	53
Ensamblaje <i>de novo</i>	54
Anotación del genoma	54
Tipificación de SCCmec	54
Tipificación MLST	54
Core genome SNPs.....	56
Determinación de genes de resistencia y virulencia.	56
Variabilidad genética de los genes (fnbA, fhuD, sstD, hla, sdrC).	56
Capítulo IV Resultados	58
Extracción de ADN	58
Preparación de librerías para secuenciación NGS Illumina.	61
Análisis Bioinformático	62
Control de Calidad de las Secuencias Obtenidas	62
Preprocesamiento de las Secuencias Obtenidas	68
Identificación de Especie y Evaluación de la Presencia de Contaminantes.....	70
Cobertura y Profundidad de Secuenciación	72
Ensamblaje de novo.....	73
Anotación del Genoma	74
Tipificación de SCCmec.	75

Tipaje MLST.....	76
Core Genome SNPs	77
Búsqueda de genes de resistencia y virulencia	80
Variabilidad genética de los genes (fnbA, fhuD, sstD, hla, sdrC).....	84
Capítulo V: Discusión	109
Capítulo VI: Conclusiones	121
Bibliografía	124

Índice de Tablas

Tabla 1 <i>Reactivos usados comúnmente en la extracción de ADN genómico bacteriano.</i>	30
Tabla 2 <i>Lista de contaminantes y sus posibles efectos en la secuenciación Illumina.</i>	32
Tabla 3 <i>Número de ciclos recomendados de acuerdo a la masa de entrada de DNA.</i>	35
Tabla 4 <i>Valores de Phred Quality Score.</i>	39
Tabla 5 <i>Información disponible sobre los aislados de estudio y los factores de virulencia de interés.</i>	44
Tabla 6 <i>Programa de Tagmentación en el termociclador.</i>	47
Tabla 7 <i>Programa de Post-Tagmentación en el termociclador.</i>	48
Tabla 8 <i>Programa de Amplificación de Ciclos Limitados en el Termociclador.</i>	49
Tabla 9 <i>Concentración de ADN extraído y los valores de calidad 260/280, 260/230</i>	60
Tabla 10 <i>Identificación de especie y contaminantes mediante el software Kraken V.2</i>	71
Tabla 11 <i>Cobertura y Profundidad de Las Lecturas Preprocesadas.</i>	72
Tabla 12 <i>Parámetros de calidad de los ensamblajes obtenidos mediante el software Quast.</i>	74
Tabla 13 <i>Número promedio de clases anotadas con Prokka para todos los aislados de SA.</i>	75
Tabla 14 <i>Tipificación SCCmec en los aislados de SA.</i>	75
Tabla 15 <i>Tipificación MLST de los aislados de SA en estudio.</i>	76
Tabla 17 <i>Genes de resistencia encontrados en los aislados clínicos de SA.</i>	81
Tabla 18 <i>Genes de Virulencia encontrados en los aislados clínicos de SA.</i>	82
Tabla 19 <i>Variantes genéticas observadas entre los aislados con regulación al alza de fnbA.</i>	85
Tabla 20 <i>Variantes genéticas observadas entre los aislados con regulación al alza de sstD.</i>	102
Tabla 21 <i>Variantes genéticas observadas entre los aislados con regulación al alza de hla.</i>	105

Índice de Figuras

Figura 1 <i>Staphylococcus aureus</i> en Tinción Gram.....	21
Figura 2 Modelo de adherencia e invasión de SA a una célula epitelial.	24
Figura 3 Operones de síntesis y captura de hierro en SA.	25
Figura 4 Mecanismo de acción de la alfa hemolisina.	26
Figura 5 Elementos genéticos típicos encontrados en MRSA.....	27
Figura 6 Estructura típica de SCCmec de MRSA	28
Figura 7 Pasos a seguir para realizar el tipaje MLST por el método tradicional.....	29
Figura 8 Preparación de librerías de ADN según el protocolo Illumina DNA Prep.	34
Figura 9 Amplificación en Puente en Secuenciación Illumina.	37
Figura 10 Principio de la Secuenciación por Síntesis en Illumina.	38
Figura 11 Estructura del archivo con formato FASTQ.	40
Figura 12 Estructura del Archivo con Formato Tipo BAM.	41
Figura 13 Diferencia entre grado de cobertura y profundidad de cobertura.....	41
Figura 14 Esquema del ensamblaje de novo.	42
Figura 15 Gel de electroforesis que muestra el ADN genómico	59
Figura 16 Librerías de ADN en gel de electroforesis en distintos puntos del proceso.....	62
Figura 17 Millones lecturas generadas para cada aislado de SA.....	63
Figura 18 Gráficos de calidad visualizados a través de MULTIQC.....	64
Figura 19 Contenido de Base por Secuencia para AJ.	67
Figura 20 Diagrama de calidad resultante del preprocesamiento de las secuencias.	69
Figura 21 Árbol Filogenético de máxima verosimilitud y bayesiano basado en MLST.....	77
Figura 22 Distribución de los genes según su tipo en el pangenoma.	78

Figura 23 <i>Árbol Filogenético de máxima verosimilitud y bayesiano basado en el cgSNPs.</i>	79
Figura 24 <i>Árbol Filogenético de máxima verosimilitud y bayesiano para fnbA.</i>	84
Figura 25 <i>Árbol filogenético de máxima verosimilitud y bayesiano para el gen sstD.</i>	101
Figura 26 <i>Árbol filogenético de máxima verosimilitud y bayesiano para el gen hla.</i>	104
Figura 27 <i>Árbol Filogenético de máxima verosimilitud y bayesiano para sdrC.</i>	108

Listado de Abreviaturas

%GC	Porcentaje Guanina-Citosina
ADN	Ácido Desoxirribonucleico
arcC	Carbamato Quinasa
ARN	Ácido Ribonucleico
aroE	Shikimato Deshidrogenasa
BAM	Mapa de Alineamiento Binario
BLAST	Herramienta Básica de Búsqueda de Alineación Local
BLT	Transposoma Ligado a Cuentas
BWA	Alineador Burrows-Wheeler
cgSNPs	Polimorfismos de un Solo Nucleótido en el Genoma Core
EPM	Mezcla PCR Mejorada
fnbA	Proteína A de Unión a la Fibronectina
fhuD1	Proteína determinante de superficie regulada por el hierro F1
GFF	Formato de Búsqueda de Genes
glpF	Facilitador de la Captación de Glicerol
Gmk	Guanilato Quinasa
hla	Alfa-Toxina
MCMC	Cadena de Markov de Monte Carlo
MEMS	Coincidencias Exactas Máximas
MLST	Tipificación Multilocus de Secuencias
MRSA	<i>Staphylococcus aureus</i> Resistente a la Meticilina
MSCCARMMS	Componentes de la Superficie Microbiana que Reconocen las Moléculas de la Matriz Adhesiva
NCBI	Centro Nacional de Información Biotecnológica
NGS	Secuenciación de Próxima Generación
Pb	Pares de Bases
PBP2a	Proteína de Unión a la Penicilina 2ª
PCR	Reacción en Cadena de la Polimerasa
PE	Extremo Emparejado

Pta	Fosfato Acetiltransferasa
RSB	Tampón de Resuspensión
SA	<i>Staphylococcus aureus</i>
SAM	Mapa de Alineación de Secuencias
SCCmec	Cassette Cromosómico Estafilocócico mec
sdrC	Proteína que Contiene Repeticiones de Serina-Aspartato
SNPs	Polimorfismo de un Solo Nucleótido
SPB	Perlas de Purificación de Muestras
sstD	Proteína de Unión al Sustrato del Transportador ABC de Sideróforos
TB1	Tampón de Tagmentación 1
Tpi	Triosafofosfato Isomerasa
TSB	Tampón de Parada de la Etiqueta
TWB	Tampón de Lavado de Tagmentación
yqiL	Acetil Coenzima A Acetiltransferasa

Resumen

Staphylococcus aureus (SA) es un agente causal de infecciones graves vinculadas a una fuerte presión en el sistema sanitario. La patogenicidad observada está asociada a factores de virulencia. En este estudio se propuso estudiar la variabilidad de los genes de factores de virulencia Proteína de Unión a la Fibronectina A (fnbA), Proteína que Contiene Repeticiones de Serina-Aspartato (sdrC), Alfa-Toxina (hla), Proteína determinante de superficie regulada por el hierro F1 (fhuD1), Proteína de Unión al Sustrato del Transportador ABC de Sideróforos (sstD), mediante secuenciación de siguiente generación (NGS) del genoma completo de aislados clínicos de *Staphylococcus aureus* causantes de bacteriemia y relacionar esta variabilidad con datos de regulación de expresión génica obtenidos en un estudio previo. El análisis de genoma completo de los aislados adicionalmente permitió realizar tipificación mediante Tipificación Multilocus de Secuencias (MLST), tipificar el Cassette Cromosómico Estafilocócico mec (SCCmec) y estudiar las relaciones filogenéticas entre los aislados. No se encontró relación entre el patrón de expresión génica y las variantes genéticas observadas, lo que está en concordancia con observaciones previas que dictan que la regulación de los factores de virulencia es compleja y multifactorial. En esta investigación se identificó por primera vez SA ST2625 SCCmec IVa en Ecuador, clon que en Europa ha sido relacionado con brotes en unidades pediátricas. Además, el estudio de Polimorfismos de un Solo Nucleótido en el Genoma Core (cgSNPs) reveló posibles eventos de transmisión. Toda la información aquí expuesta surge de la secuenciación de nueva generación y recalca su importancia para la toma de decisiones basadas en evidencia.

Palabras clave:

- **STAPHYLOCOCCUS AUREUS**
- **SECUENCIACIÓN NGS**
- **FACTORES DE VIRULENCIA**

Abstract

Staphylococcus aureus (SA) is a causative agent of serious infections linked to severe pressure on the healthcare system. On the other hand, the observed pathogenicity is associated with virulence factors. Virulence factors are genes that allow the bacteria to adapt to hostile environments. Next-generation sequencing has made it possible to identify and understand the role of virulence factors in infection. This study aimed at studying, through next generation sequencing (NGS) the variability of virulence factor genes Fibronectin Binding Protein A (fnbA), Serine-Aspartate Repeat Containing Protein (sdrC), Alpha-Toxin (hla), F1 Iron Regulated Surface Determining Protein (fhuD1), and Siderophore ABC Transporter Substrate Binding Protein (sstD) in bacteremia-causing *Staphylococcus aureus* clinical isolates and draw a possible role of genetics variants in Gene expression regulation data obtained in a previous study. Whole genome NGS allowed to molecularly characterize SA isolates, including Multi Locus Sequence Typing (MLST), Staphylococcal Chromosome Cassette mec (SCCmec) typing along with phylogenetic relationships. No relationship was found between the pattern of gene expression and the genetic variants observed, which is in agreement with previous observations that the regulation of virulence factors is complex and multifactorial. In this research, SA ST2625 SCCmec IVa was reported for the first time in Ecuador, a clone that in Italy has been linked to outbreaks in pediatric units. In addition, possible transmission events within this ST were found. Also, a genetic variant found in the fnbA gene was proposed as a potential prognostic marker for bacteremia. All the information presented here arises from next-generation sequencing and underscores its importance for evidence-based decision making.

Keywords:

- **STAPHYLOCOCCUS AUREUS**
- **NGS SEQUENCING**
- **VIRULENCE FACTORS**

Capítulo I: Introducción

Formulación del problema

Staphylococcus aureus (SA) es una bacteria oportunista que coloniza persistentemente las fosas nasales, piel y perineo de hasta el 30% de la población mundial (Thomer et al., 2016). Esta bacteria puede permanecer en el huésped durante semanas e incluso meses sin causar síntomas. Sin embargo, SA puede ingresar a tejidos y/o sangre causando desde infecciones benignas como foliculitis hasta potencialmente mortales como infección del torrente sanguíneo o bacteriemia (Silva et al., 2020).

SA es uno de los principales agentes causantes de bacteriemia. A la infección asociada a SA antes mencionada se le atribuye una mortalidad en un período de 30 días de hasta el 40% (Silva et al., 2020). En comparación con otros patógenos que causan bacteriemia se ha reportado que SA contribuye a un tiempo de hospitalización, gasto en el tratamiento y riesgo de muerte elevados (Powers & Wardenburg, 2014). Esto se debe en parte a la aparición de SA con un amplio espectro de resistencia a antibióticos que dificultan el tratamiento de la infección (Guo et al., 2020). Por ejemplo, las bacteriemias asociadas a SA resistente a la meticilina (MRSA) se vinculan con el doble de mortalidad que las asociadas a SA susceptible a la meticilina (MSSA) (Kim et al., 2021).

La severidad de la infección causada por SA se debe al arsenal de factores de virulencia que esta bacteria posee (Bitrus et al., 2018). Los factores de virulencia son genes que la bacteria expresa para facilitar la invasión al huésped, la adhesión al tejido y la evasión de la respuesta inmune (Turner et al., 2019). Además, los factores de virulencia contribuyen a la adquisición de nutrientes que le permiten a la bacteria sobrevivir en ambientes hostiles (Lalaouna et al., 2018). Existe una asociación entre el curso de infección por SA y los factores de virulencia presentes en el genoma de la bacteria. Por tanto, la detección temprana de los factores de virulencia permitiría una intervención terapéutica oportuna

(Mason, 2013). Sin embargo, a pesar de la creciente literatura, la base de la virulencia y el cambio entre el estado de comensal a patógeno no ha sido completamente dilucidada (Bonar et al., 2018).

Recientemente una investigación en modelos animales encontró factores de virulencia putativos (sdrC, fnbA, fhuD, sstD y hla) que intervienen en el cambio de la bacteria de estado comensal a patógeno (Jenkins et al., 2015). En otro estudio, se observó que los factores de virulencia antes mencionados se expresaron al alza de forma diferencial en aislados clínicos de SA (Noboa, Kirsty, 2020). En la presente investigación se utilizó la secuenciación de genoma completo para caracterizar aislados clínicos de SA y determinar variantes genéticas en los factores de virulencia (sdrC, fnbA, fhuD, sstD y hla). Al estudiar la variabilidad genética de dichos genes será posible entender su implicación en la patogenicidad y virulencia de los aislados clínicos.

Justificación del problema

La bacteria oportunista SA puede causar intoxicaciones alimentarias, infecciones en heridas e infecciones sistémicas como bacteriemias (Bien et al., 2011) A nivel global el número de muertes atribuidos a bacteriemias por SA es mayor que para el SIDA, la tuberculosis y la hepatitis viral en conjunto (van Hal et al., 2012). Los casos de bacteriemia al año debido a SA ascienden a 100000 en Estados Unidos (Gnanamani et al., 2017). Lo anterior genera una carga significativa sobre el sistema de salud pública en términos de mortalidad, morbilidad y costos asociados. Lo cual es aún más relevante en el contexto de países en desarrollo (Murray et al., 2022).

A lo largo de la historia, la mortalidad atribuida a infección por SA se redujo de un 80% a un 15-20 % debido al uso de antibióticos. (Abdelbary et al., 2017). Sin embargo, una preocupación importante acerca de SA es su creciente espectro de resistencia a antibióticos, el cual abarca incluso a los antibióticos de último recurso (Gould et al., 2010). De especial interés es MRSA, que se ha posicionado como la bacteria multirresistente más común en varios sectores del mundo. MRSA ha desarrollado

resistencia incluso a linezolid, el único antibiótico nuevo introducido al mercado en los últimos 20 años (Turner et al., 2019). Sin embargo, aunque MRSA posee mayores tasas de morbilidad que MSSA, los dos tipos son responsables de infecciones comunitarias y nosocomiales graves (Guo et al., 2020).

La creciente preocupación acerca de la resistencia antimicrobiana ha obligado a los investigadores a idear nuevas estrategias de tratamiento. Como alternativa, las terapias antivirulencia buscan interferir las toxinas bacterianas y así reducir la patogenicidad de la infección. Conocer los perfiles de virulencia y su diversidad es importante para elaborar estas estrategias (Zhang et al., 2018). Además, se ha reconocido que el análisis de los perfiles de genes de virulencia en SA es importante para predecir el curso de la infección (X. Li et al., 2018). Los factores de virulencia pueden servir como marcadores para diferenciar entre patógenos que causan infección y colonizadores. La implementación de esta esta diferenciación en los laboratorios permitiría un tratamiento adecuado y contribuiría a reducir el uso indiscriminado de antibióticos (Shettigar & Murali, 2020).

En un estudio previo realizado por (Noboa, Kirsty, 2020), se encontró diferencias en la expresión de los genes asociados a factores de virulencia: *fnbA*, *hla*, *sstD* y *fhuD* en aislados de SA causantes de bacteriemia. En este trabajo se plantea encontrar posibles variantes genéticas que expliquen la diferencia de expresión de los genes antes mencionados. Además, se indagará en las posibles implicaciones de estas variantes en la patogenicidad y evasión de la respuesta inmune. Para lograrlo se recurrirá a la secuenciación de genoma completo y se caracterizará a cada aislado clínico de SA. Con esta información será posible desarrollar a futuro sistemas diagnósticos rápidos para factores de virulencia específicos y para variantes genéticas relacionadas con alta virulencia y mortalidad. También será posible utilizar los blancos genéticos identificados en este estudio en futuros proyectos de desarrollo de potenciales vacunas y nuevos tratamientos antimicrobianos.

Objetivos de la Investigación

Objetivo General

Estudiar la variabilidad genética de genes relacionados con factores de virulencia en aislados clínicos de *Staphylococcus aureus* causantes de bacteriemias.

Objetivos Específicos

Extraer y purificar ADN de aislados clínicos de *Staphylococcus aureus* con patrones de expresión al alta (up-regulation) o la baja (down-regulation) de los genes (sdrC, fnbA, fhuD, sstD y hla).

Preparar y secuenciar librerías de ADN mediante tecnologías de secuenciación de siguiente generación (NGS)

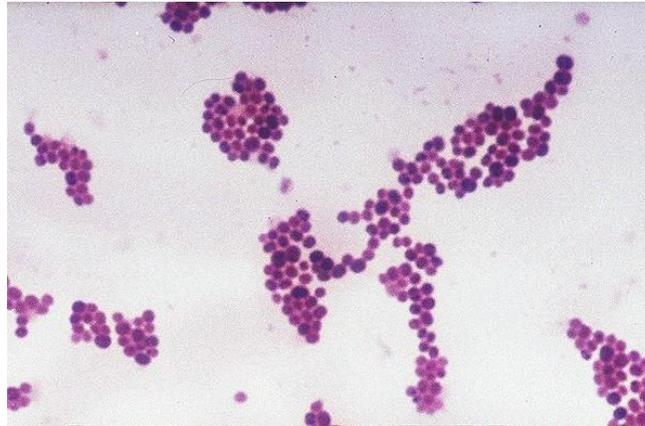
Determinar la variabilidad genética de los genes estudiados mediante el análisis bioinformático de las secuencias obtenidas.

Capítulo II: Revisión Bibliográfica

Staphylococcus aureus

Generalidades

Staphylococcus aureus es una bacteria Gram negativa esférica. El nombre “*Staphylococcus*” proviene del griego staphyle que significa racimo de uvas y kokkos que significa bayas. Lo anterior debido a que la morfología típica se asemeja a un racimo de uvas bajo microscopía óptica (Turner et al., 2019). En la Figura 1 se observa a SA sometido a tinción Gram.

Figura 1*Staphylococcus aureus* en Tinción Gram

Nota: Tomado de (Kamel & Jarjes, 2015)

Cuando crece en medio de cultivo sólido, se destaca el color amarillo de las colonias. Este color se debe a la producción de carotenoides (Turner et al., 2019). El carotenoide producido por SA recibe el nombre de estafiloxantina que tiene como rol evadir las especies reactivas de oxígeno y defender a SA de la fagocitosis (Thomer et al., 2016).

Es frecuente observar hemólisis en un cultivo de SA en agar sangre, lo anterior debido a que dicha bacteria puede contener cuatro hemolisinas (α , β , γ , δ) (Turner et al., 2019). SA es capaz de coagular la sangre a partir de una enzima denominada coagulasa y la proteína de unión al factor von Willebrand (vWbp) (Thomer et al., 2016). De hecho, un método comúnmente usado para identificar a SA se basa en la producción de la enzima coagulasa.

Además, SA es capaz de crecer en altas concentraciones de sal, incluso en un medio que contiene 7.5% de cloruro de sodio. Otras características son que el organismo es catalasa positivo y oxidasa negativo (Turner et al., 2019).

Importancia Clínica

En países desarrollados, la presión de las infecciones por SA en el sistema hospitalario es bien conocida. Por ejemplo, en Estados Unidos se reporta anualmente 490000 hospitalizaciones, 93000 eventos de bacteriemia y 35000 eventos de sepsis acompañada o no de endocarditis. Existe un mayor riesgo de infección para la población en contacto con la asistencia sanitaria: pacientes con catéter, intubación endotraqueal, implantación médica, traumatismo, hemodiálisis, heridas quirúrgicas, diálisis peritoneal, diabetes y que han sido sujetos a tratamiento oncológico o inmunosupresor (Thomer et al., 2016).

La resistencia frente a los antibióticos β -lactámicos se evidenció inmediatamente después de la introducción de la penicilina para tratar infecciones por SA a principios de los 40s. La primera penicilina semisintética dirigida a combatir infecciones por SA se desarrolló en 1960 y tan solo un año después de su uso clínico se identificó ya cepas MRSA. La resistencia observada se debe al gen *mecA* que codifica para una proteína de unión a la penicilina (PBP2a) que se une con menor afinidad a los antibióticos β -lactámicos causando su ineffectividad. El gen *mecA* es adquirido mediante transferencia horizontal de un elemento genético móvil denominado SCCmec (Dadashi et al., 2020; Turner et al., 2019).

Uno de los aspectos más importantes en SA es el espectro de resistencia antimicrobiana en el cual destacan todos los antibióticos empíricamente recetados, lo que resulta en un pronóstico negativo para el paciente. Los MRSA más exitosos no solo poseen resistencia a los antibióticos β -lactámicos sino también, a los macrólidos, quinolonas e incluso los aminoglucósidos. En la actualidad, MRSA es tratado con vancomicina, linezolid o daptomicina, antibióticos que ya están generando una presión selectiva generando resistencia antimicrobiana (Gould et al., 2010).

Otros motivos por los cuales MRSA se ha convertido en una preocupación global, es su presencia en prácticamente todos los países, lo anterior debido en parte a la expansión clonal de cepas resistentes. Además, MRSA es reconocido por su alta virulencia, lo que conlleva a desarrollar una enfermedad severa. Esto es de particular interés en pacientes inmunocomprometidos quienes son susceptibles a padecer complicaciones. Sin embargo, MRSA es capaz de ocasionar enfermedad grave incluso en jóvenes saludables en la comunidad(Gould et al., 2010).

MRSA representa un impacto adicional a MSSA en los sistemas de salud, debido a que MRSA no ha logrado reemplazar a MSSA. Con el paso del tiempo se ha reconocido la relevancia de las infecciones atribuidas a MSSA. Los primeros estudios revelaron que la mortalidad asociada a MSSA era la mitad de la atribuida a MRSA. Sin embargo, las publicaciones recientes sugieren que esta diferencia se ha acortado (Gould et al., 2010).

SA es uno de los agentes principales de bacteriemia. En Estados Unidos se presentan al año de 4,3-38,2 casos por cada 100.000 personas(Tong et al., 2015). Los valores de decesos asociados a bacteriemia por SA no se han visto alterados desde la década de los 90s. Un factor de riesgo para el desarrollo de estas infecciones son los dispositivos prostéticos, catéteres venosos, prótesis ortopédicas, materiales implantados quirúrgicos. También destaca el uso de fármacos intravenosos y comorbilidades como neoplasias, pacientes inmunosuprimidos y diabetes(Ribeiro et al., 2012).

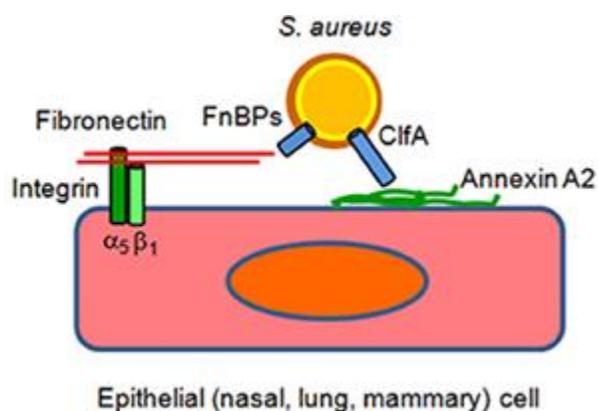
Factores de Virulencia

SA posee una serie de factores de virulencia entre los que destacan la familia MSCRAMMs (Compuestos Microbianos de Superficie que Reconocen Moléculas de Adhesión a la Matriz)(Uribe-García et al., 2021). En ellos destaca fnbA (proteína de unión al fibrinógeno A y sdrC (proteínas de repeticiones serina aspartato C). La función de estas proteínas es adherirse a las células epiteliales del huésped y este paso es considerado un requisito para la infección por SA(Lim et al., 2015). Se ha

probado que SdrC y fnbA se expresan durante la colonización nasal (Pivard et al., 2021). En la Figura 2 se esquematiza el proceso de adherencia de SA en células epiteliales.

Figura 2

Modelo de adherencia e invasión de SA a una célula epitelial.



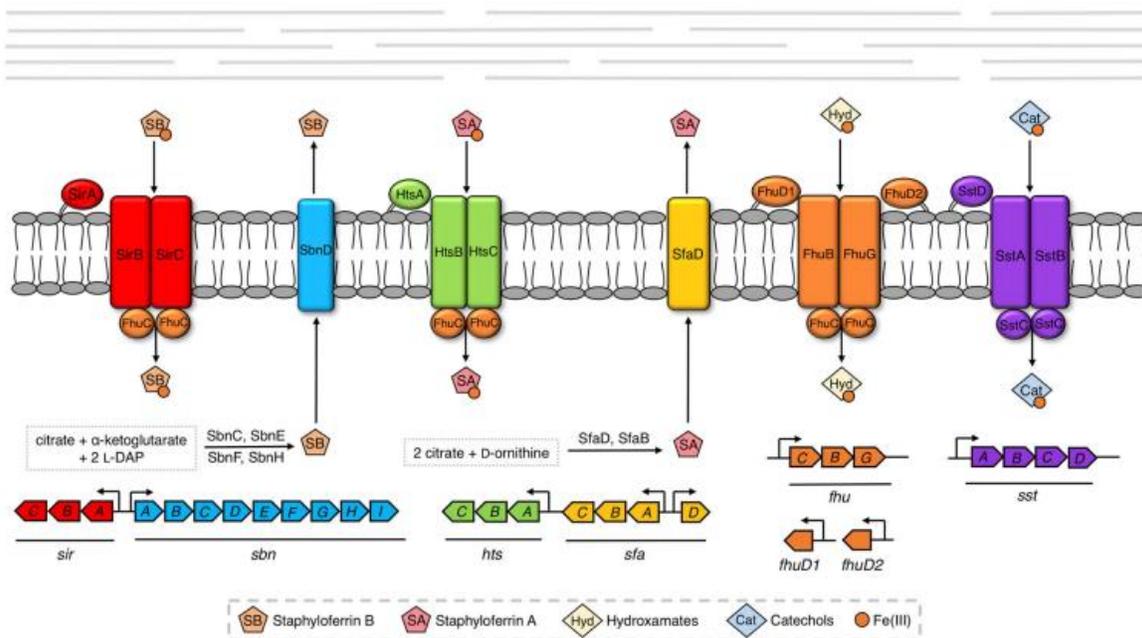
Nota: La invasión a las células epiteliales se produce tras la unión de FnBPA y FnBPB a la fibronectina. Después de esta unión, la integrina $\alpha_5\beta_1$ interactúa con la fibronectina y se produce un puente hacia el interior de la célula (Pietrocola et al., 2017).

Cuando SA infecta al huésped, la bacteria puede enfrentarse a condiciones limitantes de nutrientes. SA usa un amplio abanico de procesos de síntesis y captura de hierro en estas condiciones. Las diferentes rutas encontradas en SA para este fin pueden apreciarse en la Figura 3

Operones de síntesis y captura de hierro en SA. SA utiliza los genes fhuD1 y fhuD2 del operón fhuCBG para captar hidroxamato de hierro. Por otro lado, el ingreso de Catecol-Fe (III) se realiza mediante un transportador tipo ABC codificado por el operón sstABCD, proceso en el que participa el gen sstD.

Figura 3

Operones de síntesis y captura de hierro en SA.

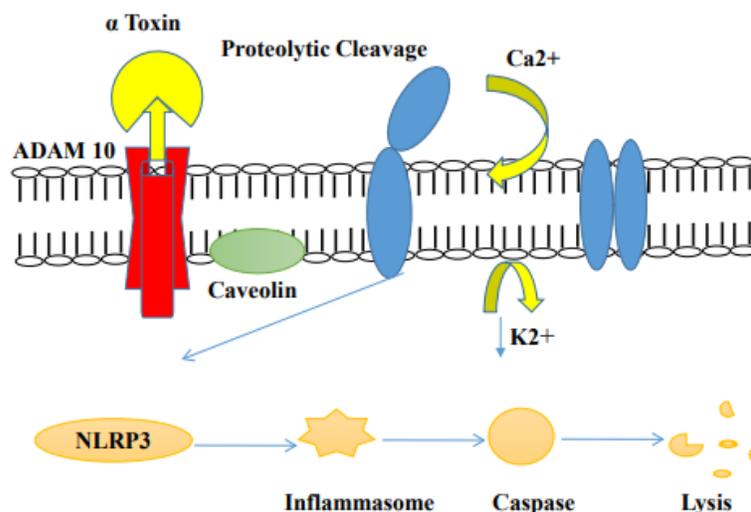


Nota: Los loci *sfa* y *sbn* codifican para las proteínas encargadas de la biosíntesis y transporte de la estafiloferrina A y B, respectivamente. Los sistemas de sideróforos de captura de hierro son HtsABC, SirABC, FhuBGC2-D1/D2 y SstABCD (Conroy et al., 2019).

Dentro del arsenal de toxinas que contiene SA, una de las más importantes es la alfa-toxina. La función de la alfa-toxina es dañar la membrana celular. Lo anterior sucede debido a la formación de poros en la membrana y a través de estos poros pueden escapar metabolitos importantes para el funcionamiento celular (Fang et al., 2020). En la Figura 4, se detalla el mecanismo de esta toxina para causar el rompimiento celular.

Figura 4

Mecanismo de acción de la alfa hemolisina.



Nota. Alfa toxina se une a la membrana celular mediante el receptor ADAM10. Esta unión provoca un poro que resulta en un incremento intracelular del Ca^{2+} y un descenso abrupto de K^{3+} . Luego, la célula reacciona activando la ruta de NLRP3 inflamasoma. Finalmente, el producto de esta ruta es la activación de la caspasa I que propiciará la lisis celular.

Genómica

Los aislados de SA poseen una estructura genómica similar. En general, su genoma mide aproximadamente 2.8 Mbp. Se define como “core” la parte del genoma que comúnmente se encuentra en todos los aislados de SA. Esto contrasta con el genoma accesorio, que es variable entre los aislados (Chua et al., 2013).

El genoma core representa cerca del 75% y está muy conservado entre los aislados. En este conjunto se encuentran los genes housekeeping y también aquellos factores de virulencia necesarios para la colonización y expansión de SA como hla y fnbA. Los eventos de recombinación en los genes del

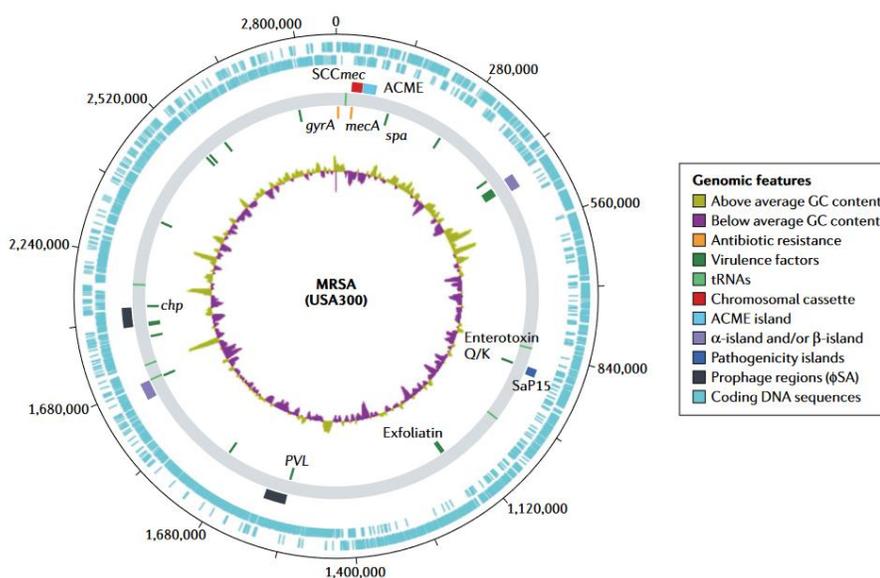
genoma core ocurren 15 veces menos en comparación con los cambios atribuidos a mutaciones puntuales. Esto ha revelado que SA se organiza en una población clonal (Lindsay & Holden, 2004).

El genoma accesorio abarca aproximadamente el 25% y está conformado por islas de patogenicidad, casetes cromosómicos, transposones, plásmidos, bacteriófagos, islas genómicas. Estos elementos comúnmente se asocian a genes de resistencia o virulencia que son transferidos de forma horizontal (Lindsay & Holden, 2004). En la Figura 5

Elementos genéticos típicos encontrados en MRSA puede observar la estructura genética típica de un aislado de MRSA.

Figura 5

Elementos genéticos típicos encontrados en MRSA



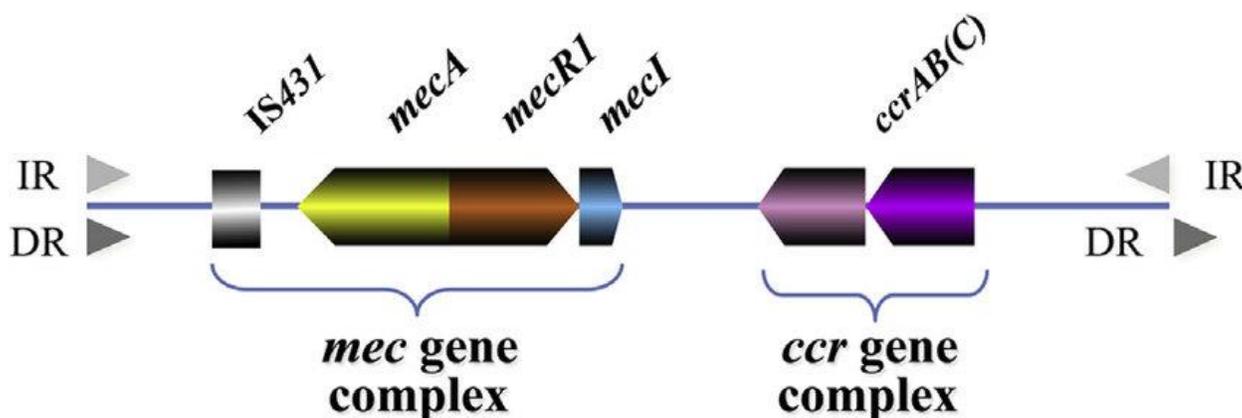
Nota: El círculo interior representa el %GC a lo largo de la secuencia de ADN. De adentro hacia afuera, el círculo 2 muestra genes de resistencia antibiótico y virulencia en naranja y verde, respectivamente. El círculo 3 muestra la ubicación de los tRNAs. El círculo 4 muestra los elementos genéticos móviles, resaltando los cassetes cromosómicos en rojo. El círculo 5 y 6 muestran las secuencias codificantes en azul.

Los Casetes Cromosómicos Estafilocócicos son elementos que se integran en el mismo sitio del cromosoma de SA. El más importante de ellos es el Cassette Cromosómico Estafilocócico mec (SCCmec) que transporta la resistencia a la meticilina. Sin embargo, también puede transportar otros genes (Lindsay & Holden, 2004). En la actualidad, hasta 13 variantes de SCCmec han sido reportadas, pero conforme avanza la secuenciación de aislados de SA, se espera que este número aumente (Giulieri et al., 2020). En la Figura 6

Estructura típica de SCCmec de MRSA.

Figura 6

Estructura típica de SCCmec de MRSA



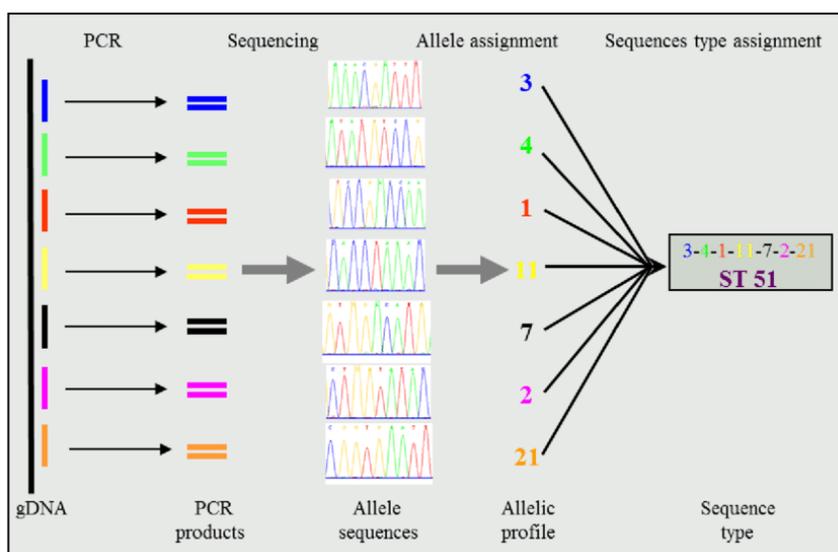
Nota: El complejo mec, está formado por el gen mecA y también sus elementos reguladores mecR1 y mecI. El complejo ccr está constituido por recombinasas que median el movimiento del cassette. Además, se destacan Repeticiones Invertidas (IR) y Repeticiones Directas (DR) (Hiramatsu et al., 2014).

Existen varias técnicas para estudiar la relación entre diferentes aislados de SA. Una de las técnicas más representativas es Tipificación Multilocus de Secuencias (MLST). En la Figura 7 se describe un esquema del proceso para realizarlo. Este método permite estudiar las secuencias de hasta 7 genes housekeeping, de lenta evolución. Los alelos de cada loci definen el perfil alélico del aislado o secuencia tipo. Actualmente, existen bases de datos de MLST curadas que han permitido realizar estudios a gran

escala. Sin embargo, el MLST convencional se ha visto reemplazado durante los últimos años con los métodos basados en la secuenciación de genoma completo, dado la inherente capacidad de resolución aumentada de estos últimos (Humphreys & Coleman, 2019).

Figura 7

Pasos a seguir para realizar el tipaje MLST por el método tradicional.



Nota: El MLST usa la variación genética presente en 7 genes housekeeping. Tras la secuenciación de los fragmentos de PCR correspondientes a dichos genes, se asigna un perfil alélico a cada locus y la combinación de estos perfiles corresponde con una secuencia tipo (Ruppitsch, 2016).

Fundamento Teórico de los Métodos Empleados

Extracción de ADN. La extracción de ADN es un paso fundamental en algunos de los procesos más empleados en la microbiología clínica. Se compone de al menos tres pasos: lisis celular, precipitación, purificación. Existen una variedad de reactivos que permiten rescatar ADN en buenas condiciones. Estas sustancias se describen en detalle en la Tabla 1.

Tabla 1

Reactivos usados comúnmente en la extracción de ADN genómico bacteriano.

Tabla 1

Reactivos usados comúnmente en la extracción de ADN genómico bacteriano.

Reactivo	Función	Referencia
EDTA	Inhibe la degradación del ADN al quelar cationes divalentes importantes para la actividad de nucleasas.	(El-Ashram et al., 2016)
SDS	Es un detergente que permite la remoción de la pared celular y la membrana nuclear que rodea al ADN	(Hasan et al., 2008)
NaCl	Permite remover las proteínas que están adheridas al ADN. Otorga el ambiente iónico necesario para precipitar polisacáridos.	(Heikrujam et al., 2020)
Proteínasa K	Es una proteasa que es responsable de la digestión de las proteínas contenidas en el lisado celular.	(Qamar et al., 2017)
LiCl	Se une selectivamente al ARN y media su precipitación.	(Atashpaz et al., 2010)
Reactivo	Función	Referencia
Cloroformo-Isoamil alcohol	Separa los lípidos y el debris celular hacia la fase orgánica, mientras que posiciona al ADN en la fase acuosa.	(McKiernan & Danielson, 2017)
Isopropanol	Insolubiliza al ADN y lo precipita. Es más efectivo que el etanol.	(Green & Sambrook, 2017)
Etanol	Insolubiliza al ADN y lo precipita.	(Green & Sambrook, 2017)

Buffer TE	Solubiliza al ADN y lo protege de acción de nucleasas.	(Panda et al., 2019)
ARNasa	Degrada el ARN estable presente en la solución.	(Bag et al., 2016)

La extracción de ADN es un proceso cuyo resultado influye en los análisis posteriores. Por ejemplo, cuando se requiere secuenciar, se ha visto que la degradación de DNA arroja parámetros inadecuados para extraer información biológica de la secuencia (McKiernan & Danielson, 2017).

La guía de preparación de Illumina DNA Prep menciona que la presencia de contaminantes puede ocasionar una tagmentación inadecuada. Esto significa fragmentación del ADN poco eficiente. Los contaminantes más comunes y sus posibles orígenes se describen en la Tabla Tabla 2

Lista de contaminantes y sus posibles efectos en la secuenciación Illumina.2. Además, esta guía advierte que se debe monitorear los ratios de calidad 260/280 y 260/230 en un espectrofotómetro. El protocolo de preparación de librerías de Illumina se ha estandarizado para ratios de 260/280 entre: 1,8-2 y ratios de 260/230 entre: 2-2,2. Cualquier valor fuera de este rango indicaría la presencia de contaminantes (Illumina, 2020).

Tabla 2

Lista de contaminantes y sus posibles efectos en la secuenciación Illumina.

Contaminante	Efecto Inhibitorio	Posible Fuente
EDTA	Quelación de iones metálicos	TE buffer
Alcoholes	Desnaturalización y precipitación enzimática	Etanol, isopropanol
Sales	Bloqueo del acceso al ADN	KCl, NaCl, CsCl, NaAc
Fenol: cloroformo	Desnaturalización enzimática	Purificación inapropiada
Polisacáridos	Bloqueo del acceso al ADN	Glicanos, dextranos
Proteína	Bloqueo del acceso al ADN	BSA, PEG
Grasas	Bloqueo del acceso al ADN	Glicerol
Detergentes	Desnaturalización y precipitación enzimática	SDS, Tween 20
Proteasas	Degradación Proteica	Proteasa K, manipulación inadecuada
Nucleasas	Degradación del ADN	Manipulación inadecuada, Enzimas de restricción
DNA/RNA Exógeno	Competición	Manipulación inadecuada
Agarosa	Bloqueo del acceso al ADN	Extracción en gel
Iones metálicos	Reduce especificidad de adición de nucleótidos	Mg ²⁺ de buffer de PCR

Nota: Adaptado de (Illumina, 2020)

Secuenciación de Nueva Generación y Secuenciación Illumina

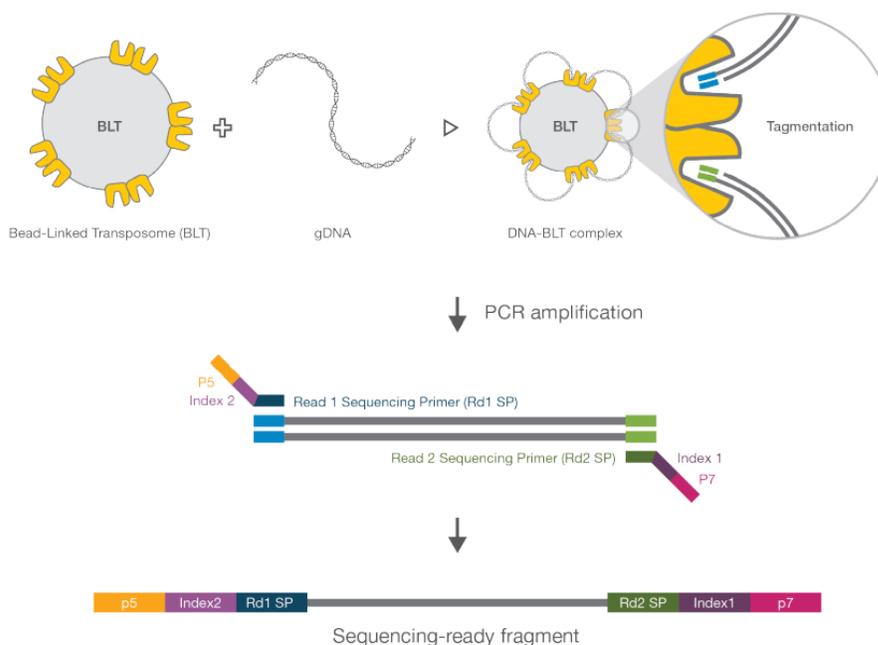
El aspecto fundamental de las tecnologías de secuenciación de nueva generación es la capacidad de realizar secuenciación masiva paralela, esto ha permitido generar datos correspondientes a múltiples secuencias de genomas completos en un mismo instrumento en poco tiempo. La secuenciación de genoma completo de Illumina se compone de tres pasos fundamentales: preparación de librerías, secuenciación, análisis de datos (Clark et al., 2019).

Preparación de Librerías El proceso completo de la preparación de librerías se representa en la Figura 8

Preparación de librerías de ADN según el protocolo Illumina DNA Prep.8. El primer paso es dividir el genoma en fragmentos pequeños. Un método para conseguirlo es usar transposasas Tn5. Estas enzimas se encuentran en *Escherichia Coli* y en esta bacteria se encargan de realizar un corte en la doble cadena de ADN. Luego, transportan el segmento cortado llamado transposón, hacia otro lugar del genoma. En la secuenciación Illumina, los fragmentos que transportan las transposasas Tn5 no son transposones sino adaptadores. Un adaptador es una secuencia de ADN de doble cadena producida mediante síntesis química. Estos adaptadores son agregados a los fragmentos de ADN generados por el corte de la transposasa y sirven como un sitio de unión para el primer de secuenciación. El corte y la adición de adaptadores ocurre en un solo proceso que se conoce como tagmentación (Clark et al., 2019).

Figura 8

Preparación de librerías de ADN según el protocolo Illumina DNA Prep.



Nota: Los adaptadores tienen una secuencia complementaria al oligonucleótido que está fijado en el soporte (p5-p7) el cual es añadido en la tagmentación. También contiene una secuencia índice o barcode para diferenciar entre muestras (multiplexing) y una secuencia que sirve de unión para el primer de secuenciación, que son agregados luego de la Reacción en Cadena de la Polimerasa (PCR) de ciclos limitados (Illumina, 2020).

La tagmentación ocurre a través de transposasas ligados a partículas o cuentas (BLT). Estos complejos contienen secuencias adaptadoras señaladas en la Figura 8 como barras azules y verdes. El ADN genómico es capturado por estas enzimas, fragmentado y en el mismo paso se le añaden secuencias adaptadoras parciales. Este proceso ocurre en el termociclador durante 15 minutos a 55 °C.

Además, en este punto se produce la normalización de las librerías, siempre y cuando se ingrese más de 100 ng de ADN. Esto sucede porque las cuentas solo pueden reaccionar una vez y se saturan con ADN al llegar a un determinado umbral. Los fragmentos de ADN que no se unieron o se unieron de forma errónea son retirados en el lavado. En general, esto contribuye a que el proceso sea más homogéneo en cuanto a la concentración resultante de cada librería y la cobertura de secuenciación. La distancia de un transposoma a otro permite la generación de insertos de longitud uniformes (~ 300 bps). Los cortes son aleatorios y el proceso deriva en fragmentos que poseen secuencias idénticas por cada extremo terminal (Bruinsma et al., 2018).

Luego, se realiza una PCR de ciclos limitados que añade índices específicos cuya secuencia hibridiza en parte con los adaptadores parciales previamente colocados en la tagmentación del ADN. Las nuevas secuencias añadirán ~135 bp de longitud a la librería final. El número de ciclos de PCR es determinado por la cantidad en masa de ADN ingresado en la tagmentación. A mayor input de ADN menos ciclos de PCR son necesarios. El número de ciclos necesarios se resume en la Tabla 3 (Illumina, 2020).

Tabla 3

Número de ciclos recomendados de acuerdo a la masa de entrada de DNA.

ADN total de entrada (ng)	Número de Ciclos de PCR (X)
1-9	12
10-24	8
25-29	6
50-99	5
100-500	5

Nota: Obtenido de: (Illumina, 2020)

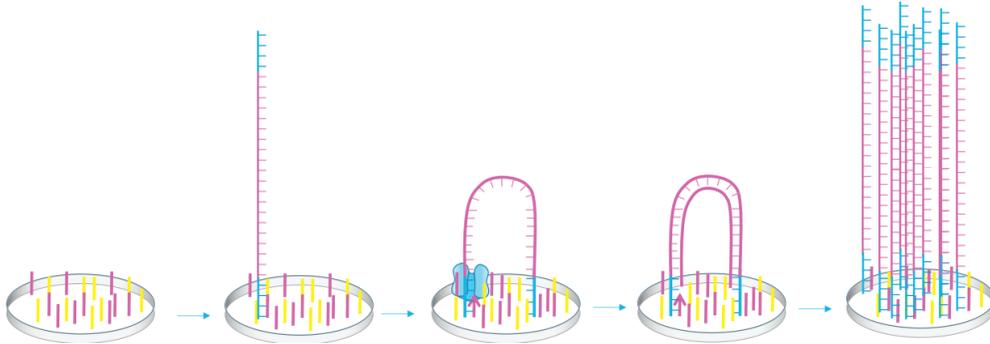
Las librerías amplificadas se “limpian” mediante purificación de doble lado, primero se desechan los fragmentos largos y luego los pequeños. Este paso contribuye a seleccionar al inserto de tamaño más abundante (~ 350 bps), los cuales terminan por unirse a las cuentas. La longitud total de la librería seleccionada incluido inserto y el adaptador completo se extiende en un rango variable, pero tiene como promedio ~600 bps. El ADN es desnaturalizado antes de la secuenciación (Illumina, 2020).

Secuenciación. Los fragmentos individuales de ADN se colocan en puntos discretos de una celda de flujo o flow cell. Esta celda utiliza canales microfluídicos que permiten a los reactivos de secuenciación fluir a través de su superficie. La parte inferior de la celda de flujo posee oligonucleótidos fijos que permiten que los fragmentos individuales de ADN provenientes de la librería se fijen en puntos establecidos (Clark et al., 2019).

El ADN individual fragmentado es convertido en un clúster de secuencias idénticas mediante PCR. Esto se hace para que exista una cantidad suficiente de señal que pueda ser detectada en los pasos siguientes. En la Figura 9 se describe este proceso. Los oligonucleótidos fijados en la flow cell sirven como primers para la amplificación. El primer forward y reverse hacen que el DNA se curve cuando este se hibrida a ellos. Por este motivo, este tipo de PCR se conoce como Amplificación en Puente (Clark et al., 2019).

Figura 9

Amplificación en Puente en Secuenciación Illumina.

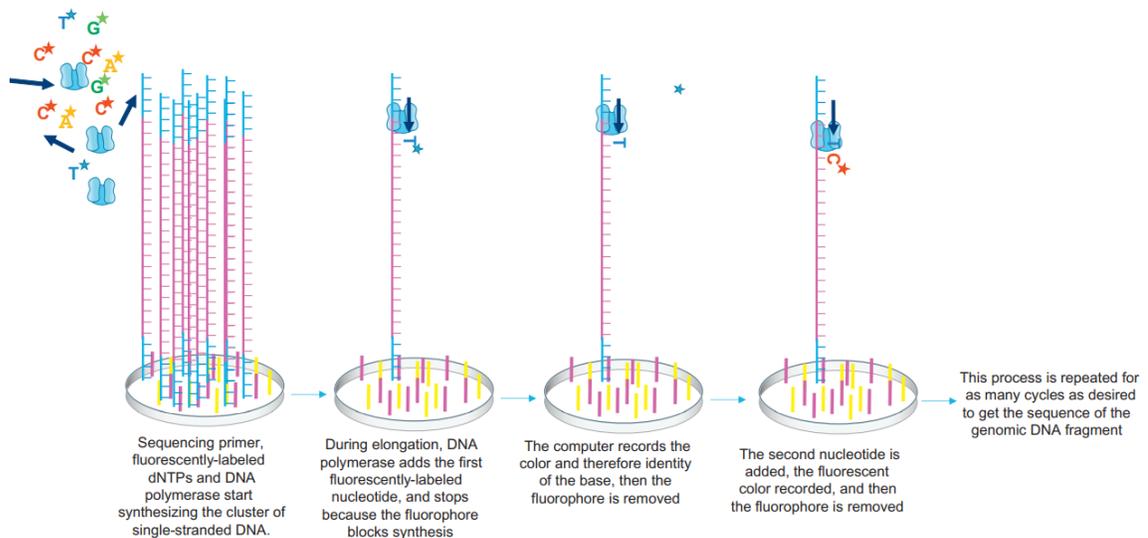


Nota: Tomado de (Clark et al., 2019)

A cada clúster generado se le agrega los primers de secuenciación, dNTPs marcados con fluorescencia y ADN polimerasa. Después de que el primer nucleótido fluorescente se adhiere al primer, el fluoróforo impide que la polimerasa siga añadiendo más nucleótidos. Por ello, se le conoce como grupo de bloqueo. En esta fase se identifica el nucleótido incorporado el cual es lavado permitiendo que la polimerasa agregue otro nucleótido y el ciclo se repita. La secuencia final obtenida mediante este proceso se conoce como lectura (Clark et al., 2019). El proceso se detalla en la Figura 10.

Figura 10

Principio de la Secuenciación por Síntesis en Illumina.



Nota: Se presenta una sola cadena por facilidad. Sin embargo, el mismo proceso ocurre simultáneamente en todas las cadenas idénticas que conforman el clúster. Como existen millones de clústeres se esperan millones de lecturas (Clark et al., 2019).

La secuenciación paired-end (PE) se refiere a la capacidad de secuenciar tanto el fragmento forward como el reverse de la doble cadena de ADN. Una de sus ventajas principales es producir doble cantidad de lecturas como resultado de la preparación de librerías. Además, brindan más información en cuanto a variaciones genéticas e indels. Debido a que la distancia entre los extremos finales de cada fragmento es conocida, se puede alinear estas lecturas a un genoma de referencia con mayor precisión hacia zonas repetitivas o de difícil resolución (Illumina, 2020).

Análisis de Datos derivados de Secuenciación Illumina. La secuenciación PE genera dos archivos con formato FASTQ. Es el formato de intercambio de archivos de secuenciación por excelencia. Este formato se conforma principalmente de la secuencia nucleotídica deducida y un nivel de la calidad asociado a cada base nucleotídica (Cock et al., 2010). La calidad se reporta en términos de PHRED Quality Score, que explica la probabilidad de que una base sea incorrecta. Algunos valores PHRED Quality Score representativos se hallan descritos en la siguiente tabla.

Tabla 4

Valores de Phred Quality Score.

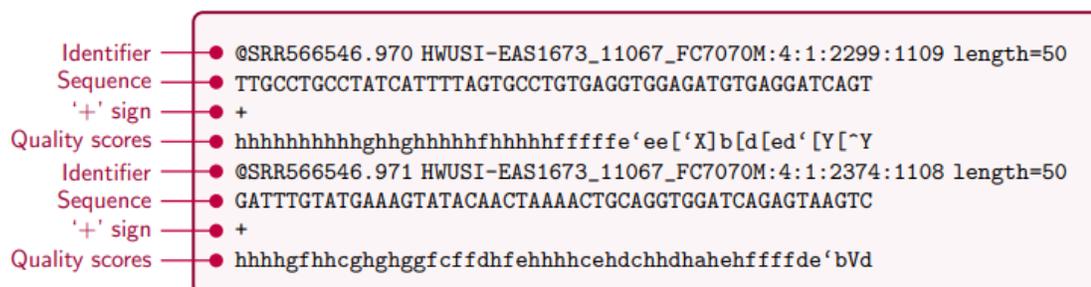
Phred Quality Score	Probabilidad de que la base sea incorrecta
Q40	0.0001
Q30	0.001
Q20	0.01
Q10	0.1

Nota: Obtenido de (Clark et al., 2019)

El formato de este archivo se resume en cuatro líneas diferentes y puede ser apreciado en la Figura 11. La primera línea es del encabezado que usualmente lleva el nombre de la secuencia y otros datos relacionados. La segunda abarca a la secuencia en sí misma. La tercera usualmente sirve como separador, a veces se encuentra únicamente el símbolo + o a su vez, también se puede encontrar repetida la primera línea. La cuarta abarca los valores de Phred Quality Score codificados en código ASCII. (Cock et al., 2010).

Figura 11

Estructura del archivo con formato FASTQ.



Nota: Tomado de: (Hosseini et al., 2016)

Mapear significa alinear las lecturas al genoma de referencia. Las lecturas de Illumina se procesan por defecto usando el software Burrows-Wheeler Alignment tool (BWA) que se basa en la transformación de Burrows–Wheeler (BWT) (H. Li & Durbin, 2009). Este algoritmo utiliza un método computacional que le permite encontrar todos los alineamientos posibles para una lectura con respecto al genoma de referencia en un solo paso. El alineamiento toma en cuenta la calidad de las bases nucleotídicas y la cantidad de mismatches para crear un puntaje de calidad por alineamiento. Toda esta información se almacena en archivos tipo Sequence Alignment Map (SAM) o el tipo Binary Alignment Map (BAM), que es la versión binaria del archivo SAM y puede ser más fácilmente leído por la computadora. El Formato SAM tiene dos partes, un encabezado que empieza con @ y la sección del alineamiento que posee varias columnas. Cada columna brinda información acerca del alineamiento de las lecturas con respecto al genoma de referencia (Hosseini et al., 2016). La estructura del formato puede ser visualizado en la Figura 12.

Figura 12

Estructura del Archivo con Formato Tipo BAM.

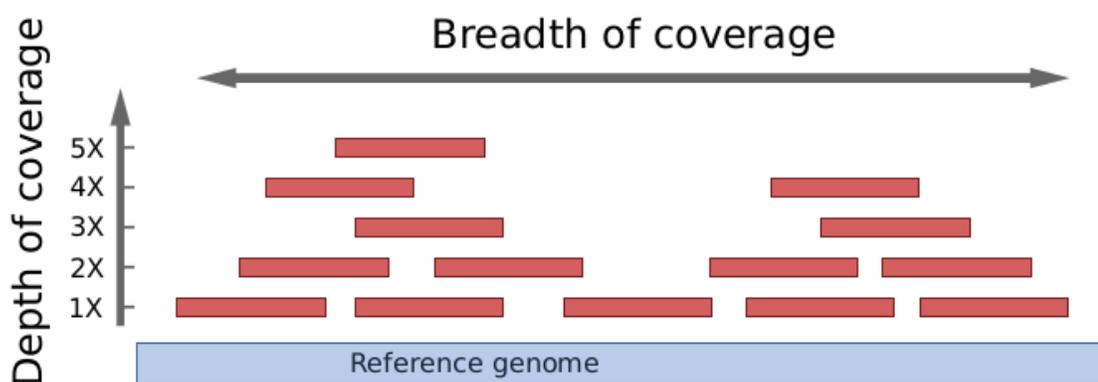
```
@HD VN:1.5 SO:coordinate
@SQ SN:ref LN:45
r001 99 ref 7 30 8M2I4M1D3M = 37 39 TTAGATAAAGGAT *
r002 0 ref 9 30 3S6M1P1I4M * 0 0 AAAAGATAAGG *
r003 0 ref 9 30 5S6M * 0 0 GCCTAAGCT * SA:Z:ref,29,-,6H5M,17,0;
r004 0 ref 16 30 6M14N5M * 0 0 ATAGCTCA *
r003 2064 ref 29 17 6H5M * 0 0 TAGGC * SA:Z:ref,9,+,5S6M,30,1;
r001 147 ref 37 30 9M = 7 -39 CAGCGGC * NM:i:1
```

Nota: Obtenido de (Hosseini et al., 2016)

Del mapeo de lecturas con respecto a un genoma de referencia se puede calcular la cobertura de secuenciación y la profundidad de cobertura. La cobertura se refiere a el porcentaje del genoma representado por el total de lecturas obtenidas de la secuenciación. La profundidad se refiere al número de lecturas que cubre un segmento específico del genoma (Price et al., 2013). Un esquema que resume estos dos conceptos se muestra en la Figura 13.

Figura 13

Diferencia entre grado de cobertura y profundidad de cobertura.

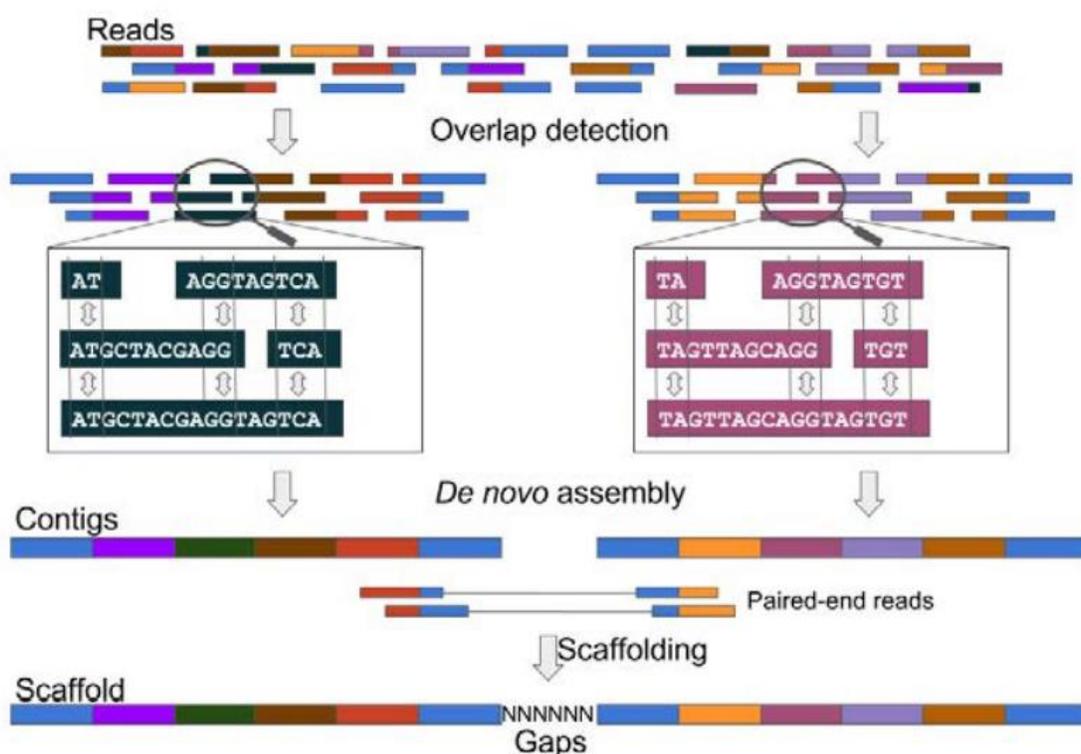


Nota: Tomado de (Clark et al., 2019)

El ensamblaje *de novo* intenta unir las lecturas como si fuese un rompecabezas, la particularidad es que no usa un genoma de referencia. Se basa en la capacidad que tienen las lecturas para solapar. Esto proceso produce uno o más contigs que representan una o múltiples secuencias del genoma y que requieren un post procesamiento para construir el genoma completo. En la Figura 14 se presenta un esquema didáctico de cómo funciona este proceso.

Figura 14

Esquema del ensamblaje de novo.



Nota: Tomado de (Bansal, 2017)

Hipótesis

Las diferencias de expresión de genes relacionados con factores de virulencia en estudio (sdrC, fnbA, fhuD, sstD y hla) en aislados clínicos de SA causantes de bacteriemia se deben a variantes genéticas susceptibles de ser identificadas mediante Next Generation Sequencing.

Capítulo III: Metodología

Zona de estudio

La extracción de ADN de los aislados clínicos se realizó en el Laboratorio de Biotecnología Humana/Nanomedicina y Nanobiología de la Universidad de las Fuerzas Armadas ESPE, sede Sangolquí. La preparación de librerías de ADN genómico se realizó en el Laboratorio de Biosequence S.A.S. El análisis bioinformático se realizó en el Instituto de Microbiología de la Universidad San Francisco de Quito y de forma remota por el estudiante.

Población de estudio

Veinte y seis aislados de *Staphylococcus aureus* fueron obtenidos de una colección de aislados clínicos causantes de infección del torrente sanguíneo. Esta colección se ha construido durante los últimos 3 años en el Laboratorio de Biotecnología Humana/Nanobiología de la Universidad de las Fuerzas Armadas (ESPE) con la cooperación de hospitales y bajo el marco de financiación de SENESCYT y FWO (Flanders Research Foundation). Los aislados de SA previamente a este estudio fueron rescatados, identificados molecularmente, diferenciados de acuerdo a su susceptibilidad a meticilina y el estudio de expresión génica de genes relacionados con factores de virulencia (*fnbA*, *fhuD*, *sstD* y *hla*) fue completado. La presente investigación se realizó específicamente con los aislados que presentaron genes relacionados a factores de virulencia con patrones de expresión significativa (regulación de la expresión génica al alza o a la baja). Las características de estos aislados se encuentran descritas en la Tabla 5.

Tabla 5

Información disponible sobre los aislados de estudio y los factores de virulencia de interés.

	<i>16S</i>	<i>nuc</i>	<i>mecA</i>	<i>hla</i>	<i>fnbA</i>	<i>fhuD</i>	<i>sstD</i>
AA	+	+	-	R	R	R	R
AD	+	+	-	↑	↑	↑	↑
AE	+	+	-	↓	↓	↓	↓
AF	+	+	-	↓	↓	↓	↓
AG	+	+	+	↑	↓	↓	↓
AI	+	+	+	↓	↑	↓	↓
AJ	+	+	-	↓	↓	↓	↓
AL	+	+	+	↓	↑	↓	↓
AK	+	+	+	↓	↓	↓	↓
AO	+	+	+	↑	↓	↑	↓
AP	+	+	+	↓	↑	↑	-
A	+	+	-	↑	↓	↓	↓
B	+	+	-	↑	↑	↑	↑
E	+	+	+	↓	↓	↓	↓
F	+	+	+	↑	↑	↑	↑
G	+	+	-	↑	↑	↑	↑
H	+	+	-	↑	↑	↑	↑
I	+	+	-	↓	↓	↑	↓
J	+	+	-	↓	↓	↓	↓
L	+	+	-	↓	-	↓	-
N	+	+	-	↓	↓	↓	-
T	+	+	-	↓	↓	↓	-
X	+	+	-	↓	↓	↓	↓
Y	+	+	-	↓	↓	↓	↓
Z	+	+	-	↓	↑	↓	↓

Nota: Para los genes expuestos en la tabla + indica presencia, - indica ausencia, ↑ significa expresión significativa, ↓ denota inhibición de la expresión. C significa que el aislado AA fue usado como referencia para comparar la expresión de los factores de virulencia. Tomado de (Noboa, Kirsty, 2020).

Técnicas de Laboratorio

Extracción de ADN

Preparación de la Muestra. Para la preparación de muestras se siguió el protocolo estandarizado en el Laboratorio de Biotecnología Humana/Nanomedicina y Nanobiología. En primer lugar, se sembró los aislados de estudio en caldo nutritivo (Merck) a 36°C durante 18 horas. Luego, se tomó 1,5 mL de dicho cultivo y se lo dispuso en un microtubo de 2 mL. Los tubos con las muestras fueron centrifugados a 14000 rpm durante 10 minutos (Minispin Plus, Eppendorf). El sobrenadante fue desechado y el pellet se utilizó en los siguientes pasos.

Lisis Celular. El pellet se resuspendió en 400µL de Buffer de lisis Tris (50 mM Tris, 50 mM EDTA, 50 mM glucosa, 100 mM NaCl, 1% SDS, 2 mg/mL proteinasa K, LiCl 0.2%). Se mezcló 1 minuto durante vórtex y se llevó al hielo durante 3 minutos. Luego, se realizó un ciclo de vórtex-hielo (1 minuto de cada paso) en cuatro ocasiones. La muestra se incubó en el mismo buffer de lisis durante 20 horas a 55°C, con agitación a 300 rpm.

Precipitación de ADN. El siguiente proceso se trabajó en hielo. En primer lugar, se centrifugó la muestra a 10 000 rpm por 5 min. Luego, se transfirió el sobrenadante a un nuevo tubo con igual volumen de cloroformo- alcohol isoamílico (24:1 vol/vol). Se homogenizó mediante vórtex y se centrifugó a 12000 rpm durante 8 minutos (Minispin Plus, Eppendorf). La fase superior se transfirió a un tubo nuevo.

Purificación de ADN. Para la purificación del ADN extraído, se trabajó en hielo en todo momento. Se añadió 100 µL cloruro de sodio 5M y se mezcló por inversiones delicadas. Luego, se agregó el volumen equivalente de isopropanol frío (-20°C). El isopropanol se agregó gota a gota y se mezcló suavemente para evitar la fragmentación del ADN. Se incubó la mezcla durante 30 min a -20°C. El primer lavado consistió en descartar el sobrenadante y adicionar 500 µL de etanol al 96%

mantenido a 4°C y se centrifugó a 12000 rpm durante 5 minutos. En el lavado 2 se repitió el mismo paso, pero con etanol al 70%. Luego de la centrifugación se descartó el sobrenadante y se dejó secar el pellet a temperatura ambiente. Finalmente, se suspendió el pellet en 50 uL de tampón TE [Tris HCl 10 mM, EDTA 1 mM]. En este punto se cuantificó y evaluó la integridad del ADN (Ver Cuantificación y Conservación del ADN).

Tratamiento con RNAsas. Se adicionó un volumen de RNasa (Invitrogen pH/ con tratamiento de DNAsas a 4mg/ml). Luego se mezcló por inversión y vórtex. Después, se incubó a 37°C por 1 hora con mezclas periódicas. Posteriormente, se adicionó 1/10 (v/v) de cloruro de sodio 5 M (uL, concentración final 0.25 M). Acto seguido, se adicionó 2,5 volúmenes de etanol 100%, se mezcló con inversiones y vórtex. Luego, se incubó el tubo a 80°C durante 30 minutos. Posteriormente, se centrifugó a 12000 rpm por 10 minutos para recolectar el pellet. Se descartó el sobrenadante y se hizo un lavado con 125 uL de etanol al 70% frío. Nuevamente, se centrifugó a 12000 rpm durante 5 minutos. Se descartó el sobrenadante y se dejó secar al aire. Finalmente, se suspendió el pellet en buffer TE [Tris HCl 10 mM, EDTA 1 mM].

Evaluación de la Integridad, Cuantificación y Conservación del ADN. El ADN purificado fue cuantificado mediante el equipo Nanodrop 2000c (Thermo Scientific). Además, se comprobó la integridad del ADN en gel de agarosa (invitrogen) al 2% (v/w). En detalle, la agarosa (invitrogen) se diluyó en buffer TBE 1X. Para cargar las muestras en el gel, se diluyeron 5 ul de producto de extracción con 2 ul de buffer de carga (Blue/Orange Loading Buffer Dye 10X; Promega). Luego, se homogenizó por pipeteo y se cargó 7 ul por pocillo. Se usó 2 ul de marcador molecular de 1 kb DNA Ladder (Invitrogen) con 2 ul de buffer de carga. La corrida de electroforesis se realizó a 80 voltios durante 90 minutos en una cámara de electroforesis horizontal (MGU-402 T, CBS SCIENTIFIC CO) con la fuente de poder (EC-300XL, Thermo Scientific). El ADN se almacenó a -20°C hasta la preparación de librerías, momento en el

que se volvió a cuantificar usando el equipo Quantus™ Fluorometer con el protocolo establecido por el reactivo QuantiFluor® dsDNA System.

Preparación de Librerías para Secuenciación NGS Illumina

Todos los equipos usados en este protocolo fueron de la marca Axygen. Se siguió el protocolo establecido por la guía de referencia de Illumina DNA Prep.

Tagmentación. El primer paso consistió en tagmentar el ADN, lo que significa que el ADN es cortado y modificado con secuencias adaptadoras. En detalle, las muestras fueron diluidas a una concentración inicial de 4 ng/uL. Se agregó 30 ul de muestra de ADN por pocillo de una microplaca de PCR. El reactivo BLT se sometió a vórtex vigorosamente hasta resuspender su contenido. Se preparó una solución master mix de tagmentación multiplicando 11 ul del reactivo BLT por el número de muestras a procesar, a esto se le añadió el resultado de multiplicar 11 ul del reactivo Buffer de Tagmentación 1 (TB1) por el mismo factor. La master mix se sometió a vórtex para lograr una homogenización completa. Luego, usando una micropipeta multicanal se transfirió 20 ul de la master mix a cada pocillo de la microplaca de PCR. A continuación, se pipeteó 10 veces para resuspender cada muestra. Se tapó la placa con un microsello tipo B y se centrifugó a 280 gravedades durante 1 minuto. Se comprobó la ausencia de grumos y burbujas y luego se corrió el programa descrito en la Tabla 6.

Tabla 6

Programa de Tagmentación en el termociclador.

Temperatura	Tiempo (min)
55 °C	15 minutos
10°C	Mantener

Nota: La opción de precalentamiento se fijó en 100°C y el volumen de la reacción ingresado en el programa fue de 50 ul.

Limpieza Post-tagmentación. Para detener la tagmentación, se añadió 10 ul del reactivo buffer de parada de tagmentación (TSB) a cada pocillo. Previamente, el reactivo fue calentado a 37°C durante 10 minutos y mezclado por vórtex. Luego, se pipeteó 10 veces usando la micropipeta multicanal para resuspender las cuentas en la reacción de tagmentación. Después, se selló la placa con un microsello tipo B y se corrió el programa descrito en la Tabla 7

Tabla 7

Programa de Post-Tagmentación en el termociclador

Temperatura	Tiempo (min)
37 °C	15 minutos
10°C	Mantener

A continuación, se colocó la placa de PCR en el soporte magnético y se dejó reposar hasta que el contenido de los pocillos se aclare, aproximadamente 5 minutos. Usando una micropipeta multicanal se retiró el sobrenadante. Luego, se lavó dos veces como sigue a continuación. Se removió la microplaca de PCR del soporte magnético y se pipeteó lentamente 100 ul del reactivo buffer de lavado de la tagmentación (TWB) directamente en las cuentas. Luego, usando una micropipeta microcanal, se homogenizó por pipeteo hasta que las perlas estuvieron resuspendidas. Después, se colocó la microplaca de PCR en el soporte magnético y se esperó hasta que el contenido de los pocillos se aclare, aproximadamente 5 minutos. Finalmente, usando una micropipeta multicanal, se removió el sobrenadante.

Una vez culminados los lavados, se removió la microplaca de PCR del soporte magnético y se pipeteó lentamente 100 ul de reactivo TWB directamente en las cuentas. Luego, se pipeteó lentamente para resuspender las cuentas. Se colocó la microplaca de PCR en el soporte magnético y se dejó reposar.

Amplificación del ADN Tagmentado. En la siguiente etapa se amplificaron las librerías y agregaron adaptadores o índices. En detalle, se preparó una master mix de PCR multiplicando el número de muestras por 22 ul del reactivo master mix de PCR mejorada (EPM) y 22 ul de agua libre de nucleasas respetivamente. Luego, se realizó vórtex y se centrifugó la master mix de PCR a 280 gravedades durante 10 segundos. Se removió y descartó el sobrenadante usando una micropipeta multicanal de 200 ul. A continuación, se removió la placa del soporte magnético e inmediatamente se añadió 40 ul de master mix de PCR directamente en las cuencas de cada pocillo. Inmediatamente, se selló la placa con un microsello tipo B y se sometió a agitación en un shaker a 1600 rpm durante 1 minuto. Luego, se añadió 5 ul de los adaptadores índices i5 e i7 respectivamente a cada muestra. La microplaca de PCR fue sellada con un microfiltro tipo B y se sometió a agitación en un shaker a 1600 rpm durante 1 minuto. Después, se centrifugó a 280 gravedades durante 30 segundos y se corrió el programa de PCR descrito en la Tabla 8.

Tabla 8

Programa de Amplificación de Ciclos Limitados en el Termociclador.

Temperatura	Tiempo (segundos)	Ciclos
68 °C	180	
98°C	180	
98°C	45	5
62°C	30	
68°C	120	
68°C	60	
10 °C	Mantener	

Nota: La opción de precalentamiento se fijó en 100°C y el volumen de la reacción ingresado en el programa fue de 50 ul.

Posteriormente, se comprobó la presencia de librerías mediante un gel de electroforesis (Ver Control de Calidad de Librerías).

Limpieza Post-PCR. El siguiente paso fue la limpieza de librerías. En detalle, se centrifugó la microplaca de PCR a 280 g durante 1 minuto para colectar el contenido en el fondo del pocillo. Luego, se colocó la microplaca en el soporte magnético y se esperó hasta que el contenido esté transparente, aproximadamente 5 minutos. A continuación, se transfirió 45 ul del sobrenadante de cada pocillo hacia un nuevo pocillo correspondiente. Así, se añadió 40 ul de agua libre de nucleasas a cada pocillo que contenía el sobrenadante. Después, se añadió 45 ul de reactivo de purificación de cuentas (SPB) a cada pocillo. Previamente, el reactivo SPB fue dejado a temperatura ambiente durante al menos 30 minutos, luego se sometió a homogenización por vórtex e inversión. La microplaca de PCR fue sellada con un microsello tipo B y sometida a agitación en un shaker a 1600 rpm durante 1 minuto. A continuación, se selló la placa y se incubó a temperatura ambiente durante 5 minutos.

Luego, se colocó la placa en el soporte magnético y se esperó hasta que el contenido se aclare, aproximadamente 5 minutos. El reactivo SPB se sometió a vórtex, luego se agregó 15 ul a los pocillos nuevos, con un número de pocillos igual al de las muestras. Se transfirió 125 ul de sobrenadante de cada pocillo que contenía la muestra a los nuevos pocillos que contenían el reactivo SPB. La microplaca de PCR se selló con un micro sello tipo B y se sometió a agitación en un shaker a 1600 rpm durante 1 minuto. Se incubó, la microplaca a temperatura ambiente durante 5 minutos. Luego, se la colocó en el soporte magnético y se esperó hasta que el contenido esté claro, aproximadamente 5 minutos.

Sin perturbar las cuentas, se removió y descartó sobrenadante. A continuación, se lavó dos veces como se indica. Manteniendo la placa en la cubierta magnética, se agregó 200 ul de etanol al 80% fresco. Luego, se incubó durante 30 segundos. Sin perturbar las cuencas, se removió y descartó todo el sobrenadante. Una vez completados los lavados, se removió todo el etanol residual usando una micropipeta de 20 ul. Se dejó reposar la placa en la cubierta magnética durante 5 minutos para secar el etanol residual. A continuación, se removió la placa de la cubierta magnética y se añadió 32 ul del reactivo buffer de resuspensión (RSB) directamente a las cuencas. El reactivo antes mencionado fue llevado a temperatura ambiente y sometido a vórtex para mezclar. Se resuspendió las cuencas mediante pipeteo. Luego, se incubó a temperatura ambiente durante 2 minutos. Se colocó la microplaca en el soporte magnético y se esperó hasta que el contenido esté claro, aproximadamente 5 minutos. Se transfirió 30 ul de sobrenadante a nuevos pocillos.

Control de Calidad de Librerías. Se definieron dos puntos de control de calidad, el primero justo después de la PCR de ciclos limitados y el segundo, después de la limpieza Post-PCR. En estos puntos se comprobó la presencia de librerías dentro del rango de longitud adecuado mediante un gel de agarosa (Invitrogen) al 2% (v/w). Se diluyó 5 ul de librería con 2 ul de buffer de carga (Blue/Orange Loading Buffer Dye 10X; Promega). Después, se homogenizó por pipeteo y se cargó 7 ul por pocillo. Se usó 2 ul del marcador molecular 100 bp DNA Ladder (Promega). La corrida de electroforesis se realizó a 90 Voltios durante 60 minutos.

Pool de Librerías. El siguiente proceso fue realizar un pool de las librerías preparadas. Se tomó 5 ul de cada pocillo y se mezclaron en un tubo de 1.7 ml. Se realizó un vórtex y luego se centrifugó el contenido del tubo. El pool de librerías fue cuantificado siguiendo el protocolo de QuantiFluor® dsDNA System mencionado anteriormente. Luego, se diluyó a una concentración de 4 nM.

Desnaturalización de las Librerías. Finalmente, siguió el paso de desnaturalización de las librerías. Para esto, el reactivo HT1 se llevó a temperatura ambiente y se almacenó a 4°C hasta su uso. Se calentó la incubadora a 98°C y mientras tanto, se mezcló en un tubo de 1,7 ml, 6ul del pool de librería y 494 ul de HT1. Luego se realizó vórtex a 280 gravedades durante 1 minuto. Se colocó el tubo en el incubador precalentado por dos minutos. Inmediatamente, se enfrió en hielo. Se dejó en hielo durante 5 minutos. Finalmente, se cargó las bibliotecas en cartucho de reactivos MiSeq Reagent Kit v2 y se realizó la secuenciación en el equipo Illumina Miseq.

Análisis Bioinformático

Control de Calidad de las Secuencias Obtenidas. Los archivos FASTQ correspondientes a las lecturas PE se analizaron mediante el software FastQC Versión 0.11.9 (Andrews, 2010). En detalle, se evaluó el Phred Quality Score por base nucleotídica y por lectura. Además, se estudió el Porcentaje guanina-citosina (%GC), la presencia de adaptadores, secuencias duplicadas y/o sobre representadas. Los resultados generados por FASTQC para cada aislado de SA se compilaron y se visualizaron mediante la herramienta MultiQC (Ewels et al., 2016).

Preprocesamiento de las Secuencias Obtenidas. Se utilizó el software Trimmomatic versión 0.39 (Bolger et al., 2014), con un enfoque de “ventana deslizante” de 4 bases, cortando cuando la calidad media de dicho número de bases cae por debajo de un Phred Quality Score = 20. Además, se utilizó el comando “TRAILING” y “LEADING” para remover las bases “N” del inicio y del final de las lecturas. Este proceso cortó las bases nucleotídicas hasta obtener una base cuyo valor de Phred Quality Score =10. Luego, se removieron las lecturas duplicadas con secuencia exacta usando el software Clumpify del paquete de herramientas BBtools Versión 38.36 (Bushnell, s.f).

Identificación de Especie y Evaluación de Presencia de Contaminantes. Se verificó que las lecturas generadas para cada aislado correspondan a *Staphylococcus aureus* mediante la clasificación taxonómica por k-mers del software Kraken 2.0, basado en la librería minikraken (https://ccb.jhu.edu/software/kraken/dl/minikraken_20171019_8GB.tgz) (Davis et al., 2013). En este punto se estableció que las lecturas de cada aislado debían: 1) tener una máxima coincidencia con SA para seguir en el presente estudio y 2) tener menos de 4% de coincidencia con otro taxón en Kraken, tal como lo establece (Raven et al., 2019).

Cobertura y Profundidad de Secuenciación. Se descargó el genoma de referencia de *Staphylococcus aureus* subsp. aureus NCTC 8325 de la base de datos Refseq del NCBI (NC_007795.1). Se alineó las lecturas forward y reverse preprocesadas de cada aislado frente a este genoma de referencia mediante el software BWA Versión 0.7.17 con el algoritmo MEMs (H. Li & Durbin, 2009). Los archivos tipo SAM generados se convirtieron a BAM y se ordenaron de acuerdo a sus coordenadas mediante la herramienta Samtools 1.3.1 (H. Li et al., 2009). El archivo BAM ordenado fue input para el software QualiMap v.2.2.2 (Okonechnikov et al., 2015), el cual determinó la cobertura y profundidad para cada aislado. En este punto se estableció dos condiciones para que un aislado de SA permanezca en el presente estudio: 1) el aislado debió superar un 80% de cobertura con respecto al genoma de referencia y 2) el aislado debió poseer una profundidad de cobertura mediana mínima de 20 X. Estos parámetros fueron establecidos previamente por (Raven et al., 2019).

Ensamblaje *de novo*. El ensamblaje de novo de las lecturas PE para cada aislado se realizó con el software SPAdes Genome Assembler v3.13.0 utilizando el modo “careful” que está diseñado para revisar y corregir errores (Bankevich et al., 2012). Los parámetros de calidad del ensamblaje fueron revisados con el software QAST: Quality Assessment Tool for Genome Assemblies Version: 5.0.2 (Gurevich et al., 2013), de donde se extrajeron parámetros como N50, No. de contigs, longitud del contig más largo, %GC, longitud total. En este punto se estableció que los ensamblajes de cada aislado debieron cumplir: 1) tener un valor de N50 > 40000 kb y un No. de contigs <200 para continuar en este estudio, tal como lo establece (Manara et al., 2018).

Anotación del genoma. Los contigs generados mediante el ensamblaje de novo se anotaron de acuerdo a las secuencias codificantes y sus posibles funciones usando el software Prokka 1.14.6 (Seemann, 2014) con los comandos --usegenus --staphylococcus y --addgenes.

Tipificación de SCCmec. Los contigs ensamblados fueron interrogados por la presencia del elemento genético móvil SCCmec a través de la herramienta SCCmecFinder (Kaya et al., 2018). Esta herramienta también permitió tipificar y en ciertos casos subtificar a SCCmec si estuviese presente en los aislados de SA.

Tipificación MLST. La tipificación MLST se realizó utilizando 7 genes housekeeping (arcC, aroE, glpF, gmk, pta, tpi and yqiL) a través de la herramienta MLST 2.11. Este software comparó los alelos de los genes antes mencionados con respecto a los reportados en la base de datos PubMLST (Jolley & Maiden, 2010) . Luego, se construyó un árbol filogenético para representar gráficamente la variabilidad presente en estos genes. Se incluyeron 6 genomas de referencia de SA con STs conocidos para verificar su distribución dentro del árbol. Las secuencias de *Staphylococcus aureus subsp. aureus* COL ST 250 (GCF_000012045.1), *Staphylococcus aureus subsp. aureus* MRSA252 ST36 (GCF_000011505.1), *Staphylococcus aureus subsp. aureus* MW2 ST1 (GCF_000011265.1), *Staphylococcus aureus subsp.*

aureus N315 ST5 (GCF_000009645.1), *Staphylococcus aureus* Wood 46 ST 97 (GCF_001953735.1), *Staphylococcus aureus* NCTC 8325 (GCF_900475245.1) se descargaron de la base de datos Refseq del NCBI.

En detalle, las secuencias de cada gen se identificaron y extrajeron manualmente a partir de los archivos de anotación generados por Prokka para cada aislado. Después, se realizó un alineamiento múltiple por cada gen mediante el software Mafft V7.490 (Kato & Standley, 2013) con la opción --auto activada para seleccionar automáticamente el mejor algoritmo de alineamiento. A continuación, los alineamientos múltiples de todos los genes se concatenaron mediante el software AliView 1.28 (Larsson, 2014). El modelo de sustitución más apropiado para este alineamiento fue TPM2u+F+I y se determinó usando el software ModelFinder plus (Kalyaanamoorthy et al., 2017) que viene incorporado por defecto en la herramienta IQ-tree 2 (Minh et al., 2020). El software IQ-tree 2 se utilizó para desarrollar el árbol filogenético con un enfoque de máxima verosimilitud. Los nodos tuvieron un valor de soporte resultante de 1000 Ultra Fast Bootstraps (UFBootstraps). Se consideraron significativos los valores de soporte superior a 90. Adicionalmente, se empleó un enfoque bayesiano para corroborar la topología del árbol. Para este caso, se utilizó el software Mr. Bayes 3.2 usando longitud de la cadena MCMC 1000000, frecuencia de muestreo 1000, burn-in 10%. Debido a que Mr. Bayes no dispone del modelo de sustitución TPM2u+F+I, se usó el modelo GTR +F+ I, como se sugiere en un estudio previo (Lecocq et al., 2013). Los árboles resultantes fueron visualizados y anotados manualmente en la herramienta iTOL V5 (Letunic & Bork, 2021).

Core genome SNPs. Para profundizar más en las relaciones genéticas de los aislados se realizó un árbol filogenético basado en cg-SNPs. En síntesis, se utilizó los archivos GFF generados por Prokka como input de la herramienta Roary (Page et al., 2015). Este software permitió determinar el core genome de todos los aislados y como output generó un alineamiento múltiple. Para construir el árbol filogenético se determinó HKY + F como el modelo de sustitución más apropiado y se siguió la misma metodología descrita previamente para la construcción del árbol MLST. Se determinó la cantidad de SNPs de diferencia en base al core-genome para identificar posibles eventos de transmisión asociados a aislados en el presente estudio. Para este fin se utilizó la herramienta SNP-sites 2.5.1-1 (Page et al., 2016). El cut-off utilizado para establecer un potencial evento de transmisión en un período de 6 meses fue de 15 SNPs cómo se estableció previamente para SA (Coll et al., 2020).

Determinación de genes de resistencia y virulencia. Los genes de resistencia y virulencia se rastrearon en los contigs mediante la herramienta Abricate 1.0.1 (Seemann T, Abricate, Github <https://github.com/tseemann/abricate>) la cual compara a nivel de ADN, los contigs con respecto a la base de datos (Comprehensive Antibiotic Resistance Database) CARD (Alcock et al., 2019) y Virulence Factors Database (VFDB) (L. Chen, 2004) para encontrar genes de resistencia y virulencia respectivamente. Se dio como válida la presencia de un determinado gen con el umbral mínimo de 90% de identidad y 85% de cobertura.

Variabilidad genética de los genes (fnbA, fhuD, sstD, hla, sdrC). Se extrajo la secuencia de los factores virulencia de interés (fnbA, fhuD, sstD, hla, sdrC) a partir de los archivos generados por la anotación en Prokka. La identidad y cobertura de la secuencia se corroboró manualmente mediante Basic Local Alignment Search Tool (BLAST) (McGinnis & Madden, 2004). Para determinar si los genotipos identificados de los genes de virulencia de estudio se relacionan con su ST se realizaron árboles filogenéticos a nivel de ADN tal como se describió previamente, con modificaciones relacionadas

al modelo de sustitución para cada gen analizado. Según ModelFinder Plus el mejor modelo de sustitución para el alineamiento múltiple del gen *sdrC* fue GT+F+G4, para *hla* y *fhuD* fue TNF+I, para *fnbA* fue TPM2u+F+I y para *sstD* fue HKY + F. Para estimar el número de variantes genéticas a lo largo del gen, las secuencias fueron comparadas con el genoma de referencia SA NCTC 8235.

Por otro lado, las secuencias genéticas *fnbA*, *sstD* y *hla* fueron indagadas acerca de potenciales variantes genéticas relacionadas a una regulación de la expresión al alza o a la baja. En detalle, se usó el software Mafft como se describió previamente, para construir alineamientos múltiples de las secuencias genéticas que correspondieron a cada patrón de expresión. Los patrones de expresión se pueden visualizar en la Tabla 5

Información disponible sobre los aislados de estudio y los factores de virulencia de interés. En todos los alineamientos se incluyó la secuencia correspondiente al aislado AA, que en un estudio previo sirvió como calibrador para medir dicha expresión (Noboa, Kirsty, 2020). Por ejemplo, para identificar un patrón de variantes genéticas en el gen *hla* relacionadas a regulación de la expresión al alza, se realizó un alineamiento múltiple de las secuencias de *hla* correspondientes a los aislados A, AD, AG, AO, B, F y G; en este alineamiento se incluyó también al aislado AA.

A partir de los alineamientos múltiples se llamó las variantes usando SNP-sites 2.5.1-1 usando el aislado AA como referencia. Para dictaminar la presencia de una variante genética relacionada a un patrón de expresión génica se verificó que exista una coincidencia en todas las secuencias correspondientes al gen analizado para los aislados que presentaron dicho patrón.

Hasta el momento de la realización del presente trabajo, no se dispuso de datos relacionados a patrones de regulación de expresión génica para el gen de virulencia *sdrC* dado que por cuestión de insumos no se ha completado su investigación. Por otro lado, tampoco se dispone de información del patrón de la expresión génica para el gen *fhuD* porque los primers utilizados en el estudio previo son

erróneos, lo que resultó en la amplificación de otro gen que está fuera del interés del presente estudio. Lo anterior fue evidente cuando tras el análisis de la presencia de *fhuD* en los genomas completos de SA no se encontró concordancia con lo reportado previamente (Noboa, Kirsty, 2020). Además, se corroboró lo mencionado mediante la verificación de los primers usados en el estudio previo para todos los genes mediante BLAST. Por estos motivos, el análisis de la variabilidad genética para estos factores de virulencia (*sdrC* y *fhuD*) se limitó únicamente a estudiar los posibles impactos que las mutaciones tengan en el desarrollo de diagnóstico o vacunas.

Capítulo IV Resultados

Extracción de ADN

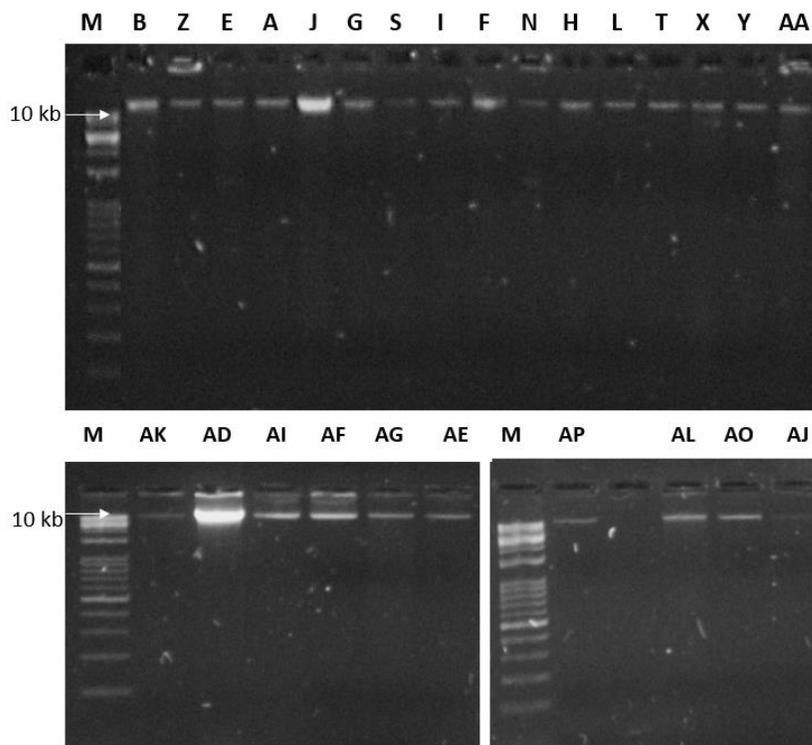
Una adecuada extracción de ADN es fundamental y de suma relevancia para los procesos posteriores. En la Figura 15

Gel de electroforesis que muestra el ADN genómico extraído y posteriormente purificado. se observa una corrida electroforética del ADN extraído correspondiente a cada aislado de estudio. La banda superior a 1 Kb corresponde al ADN genómico obtenido. En términos de integridad, se observó poca o nula degradación de ADN. No se reportó presencia de contaminantes tales como ARN.

La intensidad de cada banda se relaciona con su concentración, por lo que a primera vista se pudo observar una concentración variable de ADN extraído entre los distintos aislados. La concentración se resume en la Tabla 9, calculada por primera vez usando un método espectrofotométrico al culminar la extracción (C1) y luego, justo antes de preparar las librerías mediante un método fluorométrico (C2). Se realizaron dos réplicas por medición y se presentó los valores promedio. Se observó una variación sustancial al comparar los valores de concentración de C1 (promedio [rango], 84,59 [28,1 a 367,1 ng/uL]) y C2 (promedio [rango], 29,4 [6,17 a 168 ng/UL]).

Figura 15

Gel de electroforesis que muestra el ADN genómico extraído y posteriormente purificado.



Nota: M, marcador de peso molecular (1Kb Opti ABM). Las siguientes columnas corresponde al ADN extraído por cada aislado.

Los valores de 260/280 y 260/230 también se exponen en la Tabla 9. En el presente estudio, el valor de 260/280 sobrepasó el rango recomendado como para considerar presencia de ADN puro (promedio [rango], 2,1 [1,97 a 2,28]), mientras que el valor de 260/230 se ajustó a un valor aceptable (promedio [rango], 2,19 [1,92 a 2,46]).

Tabla 9*Concentración de ADN extraído y los valores de calidad 260/280, 260/230*

Aislado	C1	C2	260/280	260/230
AA	111	26	2,09	2,09
AD	367,1	168	2,13	2,32
AE	42,4	21	2,23	1,97
AF	130,7	47	2,14	2,14
AG	71,7	21	2,18	2,05
AI	143,7	40	2,02	2,37
AJ	28,1	6,17	2,15	2,18
AL	90,3	15	2	1,92
AK	57,25	8,1	2,14	2,23
AO	37,4	19	2,03	2,15
AP	78,2	16	2,16	2,06
A	90,3	36	2,28	2,46
B	102,3	56	2,05	2,45
E	39,1	15	2,04	2,23
F	47,6	44	2,02	2,33
G	49,3	25	2,06	2,12
H	64,64	27	2,14	2,31
I	59	15	2,08	2,4
J	225,8	72	2,11	2,29
L	28,8	23	2,03	2,02

S	43,5	9,6	2,16	2,33
N	140,34	12	2,14	2,33
T	35,34	12	2,06	2,17
X	47,75	7,8	2,17	2,01
Y	40,3	12	2,03	2,05
Z	27,3	12	1,97	1,98

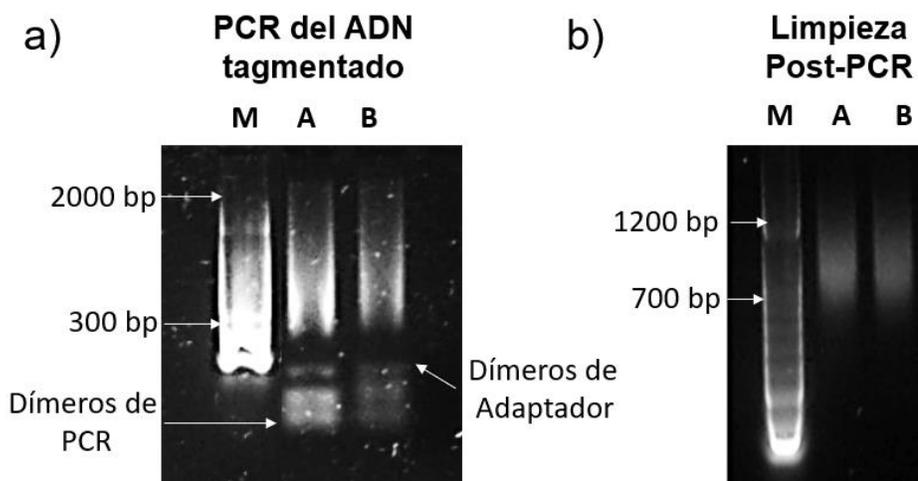
Nota: C1 se refiere a la medición acto seguido de terminar la purificación de ADN, usando un método espectrofotométrico. Por otro lado, C2 se refiere a la medición realizada justo antes de preparar las librerías usando un método fluorométrico.

Preparación de librerías para secuenciación NGS Illumina.

La preparación de librerías se realizó a partir del ADN genómico extraído en el paso previo. En la Figura 16 se muestra las bandas correspondientes a los puntos establecidos como control de calidad de las librerías. Las bandas corresponden a dos muestras representativas durante la etapa de PCR del ADN tagmentado y limpieza Post-PCR. Se encontró diferentes rangos de bandas para cada caso. En detalle, 300-2000 bp para la PCR del ADN tagmentado y 700-1200 bp para la limpieza post-PCR.

Figura 16

Librerías de ADN en gel de electroforesis en distintos puntos del proceso.



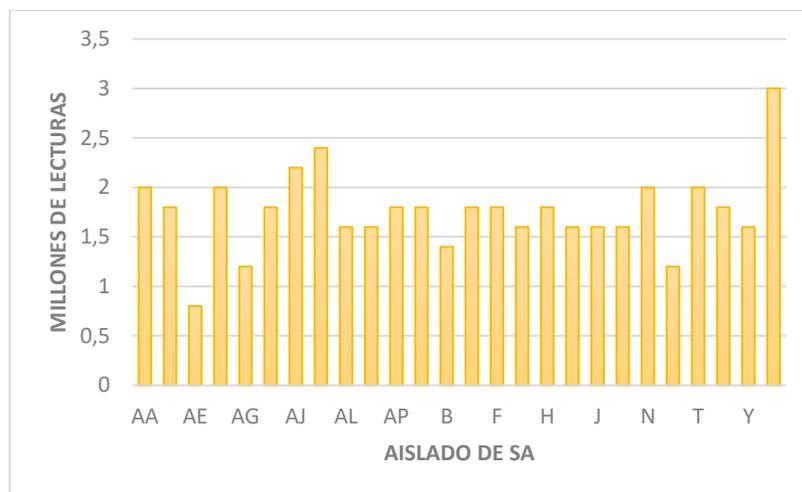
Nota: M, marcador de peso molecular (100bp Promega). Las siguientes columnas corresponden a dos aislados de SA representativos A-B. a) PCR del ADN tagmentado, en este paso se observó una amplificación fragmentos de longitud heterogénea. B) Limpieza Post-PCR, después de esta etapa se encontró un patrón de fragmentos longitud más homogénea.

Análisis Bioinformático

Control de Calidad de las Secuencias Obtenidas. La secuenciación de librerías PE de 26 aislados de *Staphylococcus aureus* generó 45,8 millones de lecturas con un promedio de 150 bp de longitud. El número de lecturas PE para cada aislado fue (promedio [rango], 1,76 millones [0,8 a 3] millones). En la Figura 17 se ilustra la variación de este parámetro por aislado de SA.

Figura 17

Millones lecturas generadas para cada aislado de SA.

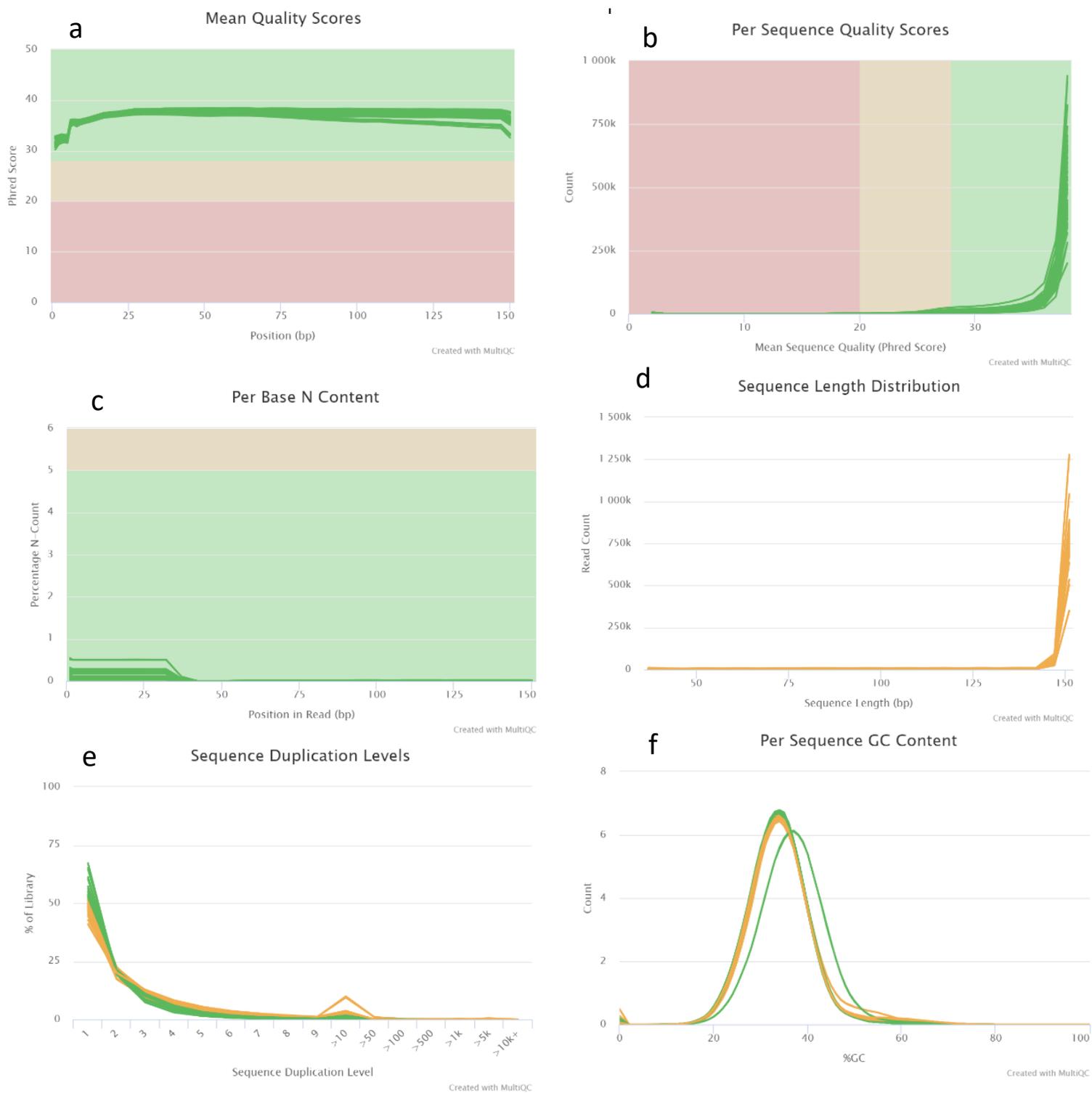


En la Figura 18 se describen los resultados del software FASTQC acerca de los distintos parámetros de calidad de las lecturas. En la Figura 18a se muestra el parámetro “Mean Quality Scores”. FASTQC calculó el valor de calidad promedio en cada posición nucleotídica de todas las lecturas por cada aislado de SA. Por otro lado, MULTIQC utilizó estos resultados para representar gráficamente los valores promedios en todos los aislados de SA. La gráfica resultante es un histograma dividido en secciones de buena calidad (Phred Quality Score > 28), aceptable calidad (20 < Phred Quality Score < 28) y mala calidad (Phred Quality Score < 20). Esta división se corresponde con los colores verde, amarillo y rojo, respectivamente.

En general, los valores promedios de Phred Quality Score superaron el valor de 30, por lo cual las lecturas generadas de la secuenciación Illumina se catalogaron como de buena calidad. Se observó una caída de calidad promedio hacia el final de la longitud de las lecturas. Esto se hizo más evidente inspeccionando el archivo FASTQC individual generado para cada aislado de SA (figura no mostrada). Como resultado de esta inspección se observó que existen bases en esta zona cuya calidad disminuye por debajo de Phred Quality Score de 20.

Figura 18

Gráficos resultantes del análisis de calidad con el programa FASTQC, visualizados a través de MULTIQC.



En la Figura 18 b se muestra el parámetro “Per Sequence Quality Scores”. FASTQC determinó el valor promedio de calidad a lo largo de todas las lecturas por cada aislado de SA. MultiQC representó dichos valores promedios para todos los aislados de SA. Este gráfico permitió identificar si los valores de mala calidad reportados en el apartado previo se asocian a un grupo específico de lecturas. Para el presente caso, se observó que la mayor cantidad de lecturas se encontraron distribuidas en el área de buena calidad. Una buena parte de ellas confluyen en un valor de Phred Score Quality =38. Sin embargo, hubo un pequeño número de lecturas que se clasificaron en la zona amarilla y zona roja, respectivamente.

La Figura 18b describe el parámetro “Per base N Content”. Cuando el secuenciador es incapaz de definir el llamado de una base nucleotídica coloca una letra “N”. Este módulo muestra una advertencia cuando el porcentaje de Ns supera al 5% del contenido de todas las lecturas y registra un error cuando el porcentaje de Ns supera al 20%. En el gráfico se observó que existen Ns en las lecturas generadas para cada aislado. Sin embargo, su cantidad representó menos del 5% del contenido de todas las bases nucleotídicas. Por lo anterior, el programa no emitió alerta alguna para este parámetro.

En la Figura 18c se presenta los resultados del parámetro “Sequence Length Distribution”. FASTQC determinó el tamaño de las lecturas generadas para cada aislado y las clasificó de acuerdo a su abundancia. MULTIQC representó estos valores para todos los aislados de SA. Este módulo reporta una advertencia si es que los tamaños de las lecturas de cada aislado no son iguales. En el presente caso, se observaron lecturas de tamaño variable, esta es la razón por la cual este módulo presentó una alerta en el panel. Sin embargo, claramente se observa que la gran mayoría de lecturas se asociaron con un tamaño de ≈ 150 bp.

En la Figura 18d se muestra el módulo “Sequence Duplications Levels”. FASTQC detectó la cantidad de lecturas con cierto grado de duplicación para cada aislado de SA. MULTIQC representó estos

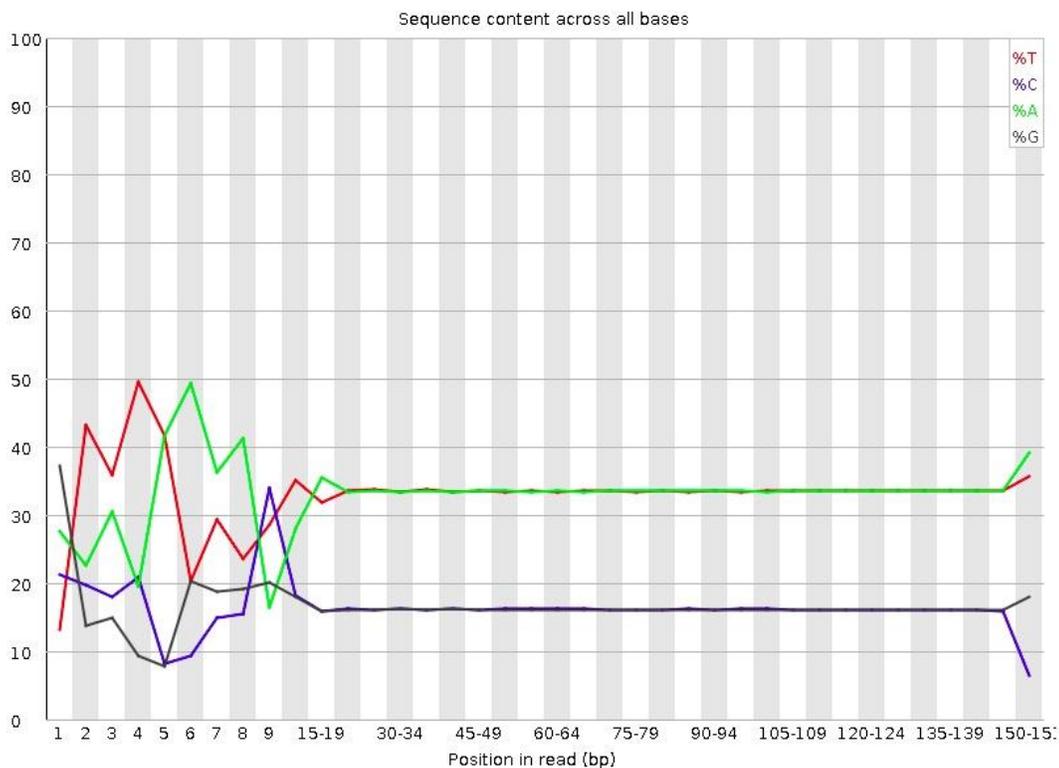
valores para todos los aislados de SA en un solo gráfico. Este módulo presenta una alerta cuando el nivel de lecturas duplicadas supera el 20% del total de lecturas para cada aislado. Así mismo, emite un error cuando se supera el 40%. En el presente estudio, el programa emitió una alerta para 34 archivos FASTQ, los que corresponden a las lecturas forward y reverse de los aislados AA, AD, AF, AI, AJ, AK, AL, AP, A, E, F, H, N, T y Z. Únicamente forward para AO, G, J y L. En promedio, el 31,26% de todas las lecturas generadas para todos los aislados se hallaron duplicadas al menos una vez.

En la Figura 18e se expone el parámetro “Per Sequence GC Content”. FASTQC calculó la distribución del %GC para todas las lecturas de cada aislado. MultiQC representó esta distribución para todos los aislados. Este módulo emite una alerta cuando el %GC no concuerda con el de una distribución normal modelada por el software en al menos un 15%. Este es el caso de 10 archivos FASTQ que corresponden a las lecturas forward y reverse de los aislados AJ, AK, S, X, Z. En general, el promedio de %GC fue 32,68 %. Sin embargo, las lecturas forward y reverse del aislado Y mostraron un pico del 36% de GC.

“Per Sequence base Content”, otro parámetro de relevancia se muestra en la Figura 19. FASTQC calculó el porcentaje de G, C, A, T en cada posición de las lecturas por cada aislado y graficó el porcentaje relativo frente a la posición nucleotídica en la lectura. MULTIQC no arrojó un gráfico de fácil apreciación para este parámetro, por ende, se muestra un gráfico del aislado AJ como ejemplo. Sin embargo, todos los aislados presentaron el mismo comportamiento en este parámetro. Este módulo ofrece una alerta cuando detecta que el porcentaje de cualquier base nucleotídica supera al 10% de diferencia con respecto a las otras en una determinada posición. Además, ofrece un error cuando el porcentaje supera al 20%. Todos los archivos FASTQ mostraron un error en este parámetro, por lo que no aprobaron este módulo de calidad. El motivo es que en las 10-20 primeras y en las últimas 5 bases nucleotídicas de cada lectura existe una adenina o timina 20% de veces más que lo esperado.

Figura 19

Contenido de Base por Secuencia para AJ.



Nota: Se observó un patrón irregular de los nucleótidos A-T en las primeras 10 bases y en las últimas 5 bases, lo que generó un error en este parámetro de calidad.

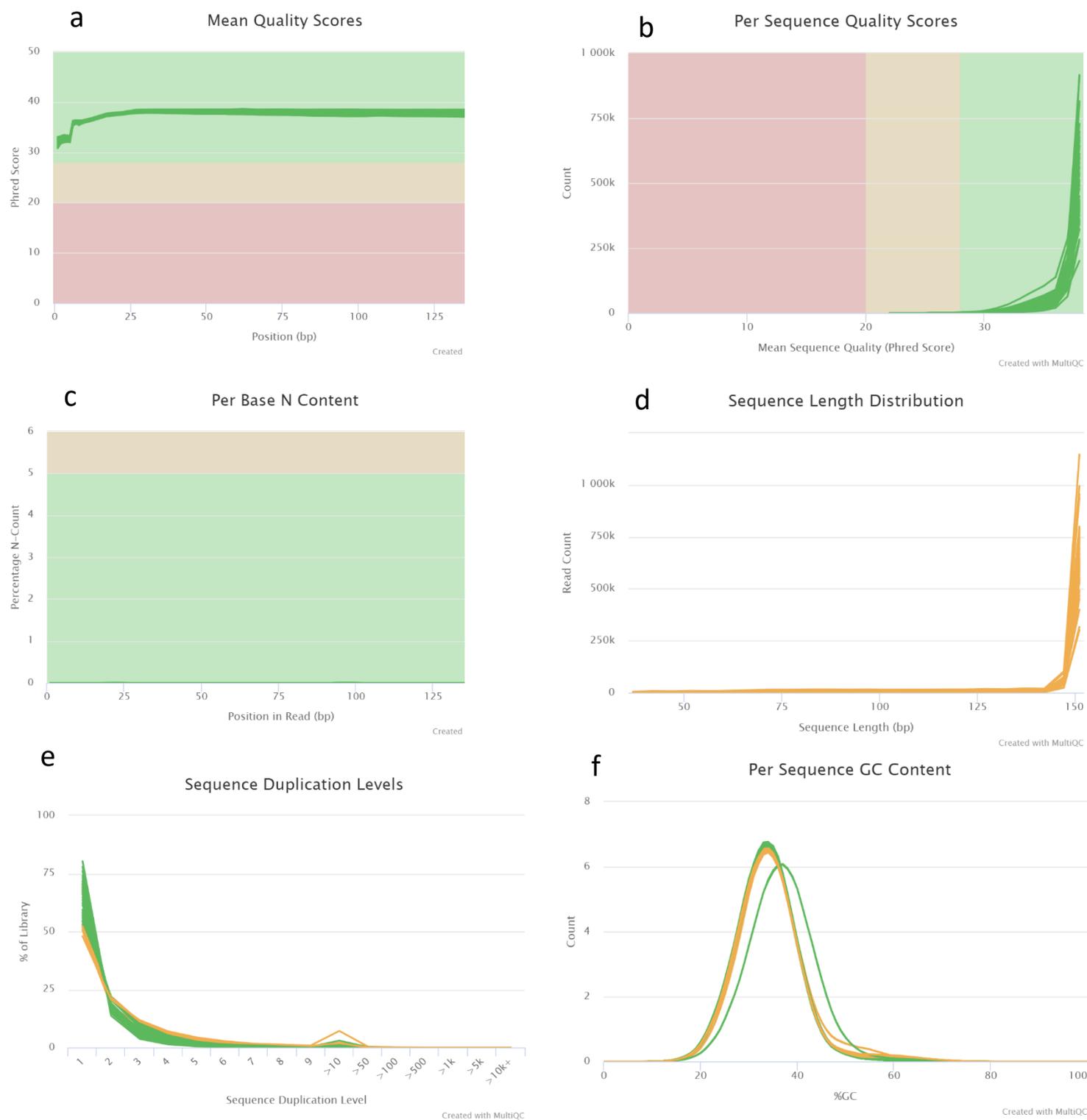
Otros parámetros de interés son “Overrepresented Sequences” que mostró si una lectura es inusualmente abundante en relación al total de lecturas. Este parámetro presenta una alerta si una lectura representa más del 0.1% del total de lecturas. En este caso, 18 archivos FASTQ recibieron esta alerta. Estos archivos corresponden a las lecturas forward y reverse de los aislados AA, AD, AE, AF, AG, AI, AJ, AK, A. Se identificó que la secuencia sobrerrepresentada fue: “NNNNNNNNNNNNNNNNNNNNNNNNNNNN” (figura no mostrada). Finalmente, el último parámetro analizado fue “Adapter Content”. FASTQC buscó la presencia de secuencias correspondientes a los adaptadores usados en la preparación de librerías. Este panel arroja una alerta si la abundancia de adaptadores

supera al 0.1% de las lecturas totales. No se reportó la presencia de adaptadores en este apartado (figura no mostrada).

Preprocesamiento de las Secuencias Obtenidas En base a los resultados previos se realizó un recorte por calidad de secuencia, se eliminó las bases nucleotídicas catalogadas como “N” y se retiró los duplicados. Después del preprocesamiento permanecieron el 93,4% de las lecturas originales. El efecto de estas acciones se analizó nuevamente con el programa FASTQC, lo que se detalla en la Figura 20 y en el texto.

Figura 20

Diagrama de calidad resultante después de realizar el preprocesamiento de las secuencias.



El módulo “Mean Quality Scores” de la Figura 20a mostró que la calidad hacia el final de la lectura se mantiene estable. La inspección de los archivos FASTQC individuales para cada aislado reveló que el Phred Quality Score no disminuyó por un valor debajo de 20. En el parámetro “Per Sequence Quality Scores” de la Figura 20b sugirió que el número de lecturas cuyo Phred Quality Score promedio fue menor a 20 en primera instancia, ahora han sido eliminadas. El número de lecturas empezó a incrementar desde la zona amarilla, concretamente desde un Phred Quality Score =22. En el parámetro “Per Base N Content” de la Figura 20c se apreció que la cantidad de Ns disminuyó incluso hasta el 0% de todas las posiciones de las bases nucleotídicas. En el parámetro “Sequence Duplications Levels” de la Figura 20e se encontró que el porcentaje de lecturas duplicadas disminuyó hasta el 22,62%. De aquí, que 7 archivos FASTQ siguieron emitiendo la alerta. Estos archivos corresponden a las lecturas Reverse de los aislados AA, AF, AJ, AK, N, T y Z. En el módulo “Per Sequence GC Content” de la Figura 20f se observó que el %GC promedio se desplazó desde un 33,62% hasta un 32,38%. A pesar de esto, los aislados AK, S, X, Z continuaron presentando una alerta en este módulo. Además, el aislado Y mantuvo su discrepancia frente al resto de aislados arrojando un valor promedio de 35% de GC. El preprocesamiento de lecturas no afectó el parámetro “Sequence Length Distribution”, Figura 20d y “Per Sequence Base Content” (figura no mostrada).

Identificación de Especie y Evaluación de la Presencia de Contaminantes. En la Tabla 10 se resumen los datos tabulados derivados del análisis por el software Kraken V.2, el cual permitió confirmar que las lecturas generadas para cada aislado coinciden principalmente con *Staphylococcus aureus* (promedio [rango], 93,14% [87,3 a 95,59%]). Como excepción a esta afirmación se encuentra el aislado Y, que dio una coincidencia de 58,92% con *Enterococcus faecalis* y con *Staphylococcus aureus* a penas el 33,07%. Por otro lado, el aislado AK presentó un nivel de coincidencia del 4,51% con Paenibacillaceae. Los

aislados antes mencionados no cumplieron los parámetros establecidos en la sección de metodología y fueron descartados del presente estudio.

Tabla 10

Identificación de especie y presencia de contaminantes mediante el software Kraken V.2

Aislado	% Lecturas asignadas a SA	% Lecturas asignadas al taxón más representativo que no es SA	
AA	95,26	0,41	<i>Proteobacteria</i>
AD	94,07	0,07	<i>Proteobacteria</i>
AE	95,34	0,03	<i>Staphylococcus haemolyticus</i>
AF	94,67	0,23	<i>Proteobacteria</i>
AG	95,59	0,29	<i>Proteobacteria</i>
AI	93,69	0,22	<i>Proteobacteria</i>
AJ	89,6	0,91	<i>Proteobacteria</i>
AK	75,55	4,51	<i>Paenibacillaceae</i>
AL	93	0,13	<i>Staphylococcus warneri</i>
AO	93,03	0,18	<i>Proteobacteria</i>
AP	92,93	0,07	<i>Proteobacteria</i>
A	93,71	0,14	<i>Proteobacteria</i>
B	91,93	0,2	<i>Proteobacteria</i>
E	91,6	0,58	<i>Proteobacteria</i>
F	93,7	0,37	<i>Proteobacteria</i>
G	93,83	0,45	<i>Proteobacteria</i>
H	94,93	0,03	<i>Virus</i>
I	92,7	1,05	<i>Proteobacteria</i>
J	91,66	0,14	<i>Proteobacteria</i>
L	95,55	0,04	<i>Proteobacteria</i>
N	93,81	0,03	<i>Staphylococcus haemolyticus</i>
S	87,3	2,4	<i>Proteobacteria</i>
T	92,72	0,03	<i>Staphylococcus haemolyticus</i>
X	92,77	1,58	<i>Proteobacteria</i>
Y	33,07	58,92	<i>Enterococcus faecalis</i>
Z	92,04	3,03	<i>Proteobacteria</i>

Nota: Se halla resaltado en rojo, los aislados que no cumplieron los parámetros establecidos en la metodología.

Cobertura y Profundidad de Secuenciación. El Software QualiMap generó los datos mostrados en la Tabla 11

Cobertura y Profundidad de Las Lecturas Preprocesadas. Las lecturas mapearon al genoma de referencia en una proporción del (promedio [rango], 92,21% [87,70% a 94,80%]). Se obtuvo una profundidad mediana de cobertura de (promedio [rango], 77,63 [37 AE a 135 Z]). La fracción del genoma de referencia mapeado con al menos una profundidad de cobertura de 50X fue (promedio [rango], 84,28% [11,44 AE a 96,40% F]).

Tabla 11

Cobertura y Profundidad de Las Lecturas Preprocesadas.

Aislado	% Cobertura	Profundidad Mediana de Cobertura	Cobertura \geq 30X	Cobertura \geq 50X
AA	93,60%	91	92,50%	92,04%
AD	93,40%	80	93,60%	92,50%
AE	92,20%	37	75,00%	11,44%
AF	91,30%	93	90,80%	90,27%
AG	92,40%	56	90,00%	66,13%
AI	94,80%	86	93,10%	91,97%
AJ	87,70%	90	90,80%	90,25%
AL	93,80%	71	97,10%	91,88%
AO	93,30%	71	93,50%	90,22%
AP	93,40%	81	93,60%	92,76%
A	91,50%	85	92,30%	91,30%
B	90,80%	62	92,30%	79,34%
E	93,10%	75	93,60%	91,56%
F	94,70%	85	97,10%	96,40%
G	92,70%	70	92,50%	89,51%
H	94,60%	78	92,50%	91,66%
I	90,60%	75	90,50%	88,56%
J	90,50%	69	93,50%	89,87%
L	92,40%	71	90,60%	87,96%
N	93,40%	89	92,00%	91,47%
S	87,70%	47	90,10%	41,38%

T	90,90%	91	90,80%	90,19%
X	90,00%	75	91,80%	88,30%
Z	94,30%	135	95,70%	95,66%

Todos los aislados superaron al menos el 80% de cobertura del genoma de referencia y el 20X de profundidad mediana de cobertura establecido en la metodología. Por tanto, en este punto ningún aislado fue descartado.

Ensamblaje de novo. Se obtuvieron ensamblajes con un número de contigs (promedio [rango], 41,2 [25 a 65]), N50 (promedio [rango], 234,06 [117,2 a 421,7]), contig más largo (Kbp) (promedio [rango], 454,55 [272,6 a 751,9]) y otros parámetros que se resumen en la Tabla 12. En este punto se excluyó del estudio al aislado I (No. contigs 2069 > 200, N50 (kbp) 2,1 < 40), S (No. contigs 206 > 200), X (No. contigs 1293 > 200, N50 (Kbp) 4,8 < 40), Z (No. contigs 2090 > 200). Además, el %GC y longitud total (Mbp) de estos aislados difirió del resto.

Tabla 12

Parámetros de calidad de los ensamblajes obtenidos mediante el software Quast.

Aislado	No. de contigs	Contig más largo (Kbp)	N50 (Kbp)	N75 (Kbp)	L50 (k)	L75 (k)	Longitud total (Mbp)	%GC
AA	32	379,7	243,9	168,2	0	0	2,7	32,71
AD	26	486,2	326,9	156,7	0	0	2,8	32,66
AE	42	403,3	127,2	92,7	0	0	2,8	32,73
AF	44	403,5	127,2	81,4	0	0	2,8	32,67
AG	44	403,5	127,2	92,8	0	0	2,8	32,68
AI	40	563,7	421,7	158	0	0	2,7	32,68
AJ	40	563,7	421,7	158	0	0	2,7	32,78
AL	40	406,3	152,4	107,9	0	0	2,8	32,78
AO	32	488,6	314,6	299,8	0	0	2,8	32,67
AP	25	488,8	326,9	299,8	0	0	2,8	32,73
A	65	421,7	134,3	62,8	0	0	2,8	32,73
B	65	272,6	117,2	60,1	0	0	2,9	32,73
E	54	488,6	300,5	151,3	0	0	2,8	32,78
F	48	751,9	329,9	145,3	0	0	2,8	32,65
G	40	590	243,9	123,4	0	0	2,7	32,72
H	26	501	249,4	129,9	0	0	2,7	32,63
I	2069	38,9	2,1	1,1	0,4	0,9	3,3	32,69
J	58	339,3	239	108	0	0	2,9	32,69
L	32	406,3	152,4	107,9	0	0	2,8	32,67
N	36	406,3	150,9	86,5	0	0	2,8	32,71
S	216	400,6	250,3	108	0	0	3	33,36
T	35	326	174	108	0	0	2,8	32,67
X	1293	43,5	4,8	2,1	0,2	0,4	3,2	33,02
Z	2090	558,9	87,2	1	0	0,4	4,4	40,17

Nota: Se halla resaltado en rojo, los aislados que no cumplieron los parámetros establecidos en la metodología.

Anotación del Genoma La anotación en Prokka permitió identificar los genes, CDSs, rRNAs, tRNAs, tmRNAs presentes en cada aislado. En la Tabla 13 se resume la cantidad promedio de clases anotadas para todos los aislados.

Tabla 13

Número promedio de clases anotadas con Prokka para todos los aislados de SA.

<i>Clases</i>	<i>Cantidad Promedio Anotada</i>
<i>Genes</i>	2581
<i>CDS</i>	2644
<i>rRNA</i>	4
<i>tRNA</i>	57
<i>tmRNA</i>	1

Tipificación de SCCmec. En la Tabla 14 se presentan los tipos de SCCmec encontrados en los aislados de SA analizados. El tipo predominante fue el SCCmec IV, presente en 7 aislados.

Tabla 14

Tipificación SCCmec en los aislados de SA.

<i>Aislado</i>	<i>SCCmec</i>
AA	-
AD	IV*
AE	-
AF	-
AG	-
AI	IVa
AJ	-
AL	IVc
AO	IVa
AP	IVa
A	-
B	-
E	IVa
F	IVc
G	-
H	-
J	-
L	-
N	-
T	-

Nota: *Se consiguió tipificar el cassette, mas no subtificar.

Tipaje MLST El tipaje MLST permitió agrupar a los aislados en secuencias tipo. El más abundante fue ST22 con 7 aislados. Luego, ST2625 con 5 aislados clasificados. Además, se encontró en menor proporción ST5 con 2 aislados, ST30 2 aislados, ST8: 2 aislados, ST25: 1 aislado y ST109:1 aislado. En la Tabla 15 se encuentran los resultados al respecto.

Tabla 15

Tipificación MLST de los aislados de SA en estudio.

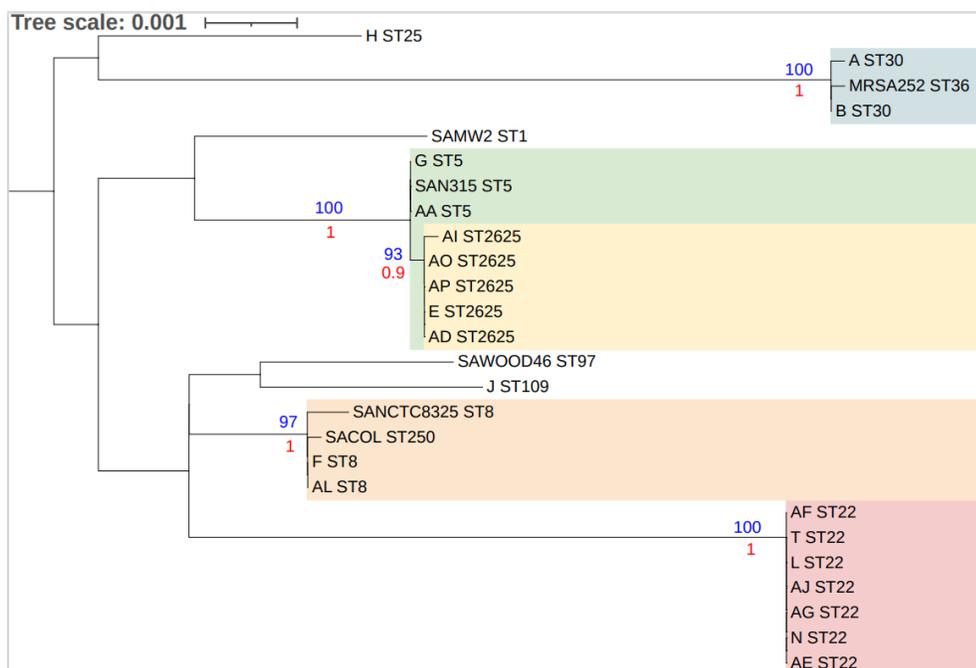
Aislado	ST	arcC	aroE	glpF	Gmk	pta	tpi	yqiL
AA	5	1	4	1	4	12	1	10
A	30	2	2	2	2	6	3	2
AD	2625	1	377	1	4	12	1	10
AE	22	7	6	1	5	8	8	6
AF	22	7	6	1	5	8	8	6
AG	22	7	6	1	5	8	8	6
AI	2625	1	377	1	4	12	1	10
AJ	22	7	6	1	5	8	8	6
AL	8	3	3	1	1	4	4	3
AO	2625	1	377	1	4	12	1	10
AP	2625	1	377	1	4	12	1	10
B	30	2	2	2	2	6	3	2
E	2625	1	377	1	4	12	1	10
F	8	3	3	1	1	4	4	3
G	5	1	4	1	4	12	1	10
H	25	4	1	4	1	5	5	4
J	109	3	27	1	1	1	1	10
L	22	7	6	1	5	8	8	6
N	22	7	6	1	5	8	8	6
T	22	7	6	1	5	8	8	6

En la Figura 21 se muestra el árbol filogenético realizado a partir de las secuencias concatenadas de los 7 genes housekeeping. Como se esperaba, los aislados de SA se agruparon de acuerdo a su ST. Se observó que los genomas de referencia incluidos se posicionan en distintos nodos. Los valores de

UFBoot y posterior probabilities ratifican la topología de este árbol con un alto grado de confianza en los nodos donde se concentra el mismo tipo de ST.

Figura 21

Árbol Filogenético de máxima verosimilitud y bayesiano basado en los 7 genes de MLST.

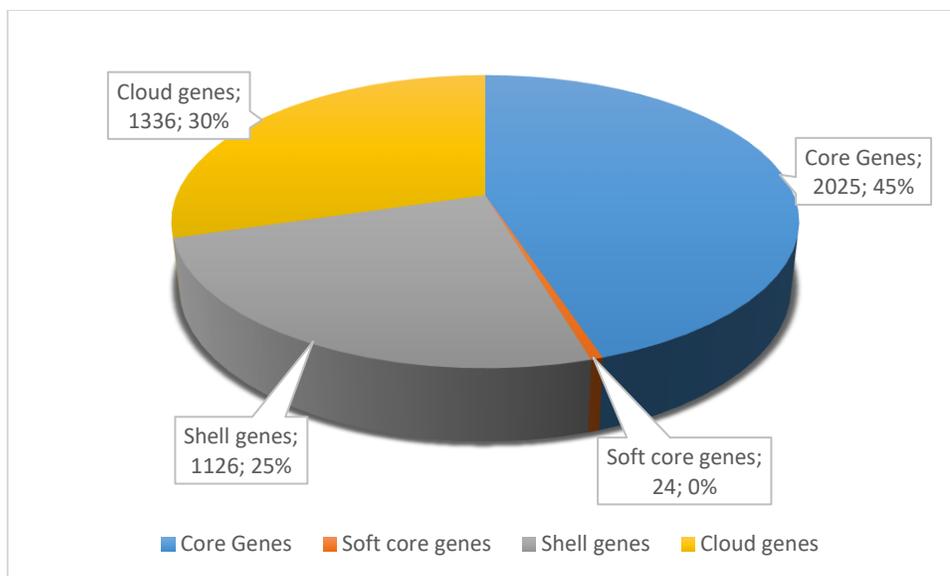


Nota. Valores de soporte UFBoot >90% (ubicados sobre los nodos con color azul) por el software iQ-tree 2: Repeticiones de UFBootstraps=1000, modelo de sustitución TPM2u+F+I. Valores de soporte posterior Probabilities >0.90 (ubicados bajo los nodos con color rojo), calculados con el software Mr. Bayes: modelo de sustitución GTR+F+I, longitud de la cadena MCMC 1000000, frecuencia de muestreo 1000, burn-in 10%. La escala indica la distancia genética calculada. Imagen del árbol filogenético obtenida y editada con iTOL V5.

Core Genome SNPs. La herramienta Roary generó el porcentaje de genes compartidos por los aislados “core genes”, genes que están presentes en dos o más aislados “shell genes”, genes que se encontraron en un único aislado “cloud genes”, genes presentes en el 95% de los aislados, pero no en todos “soft core genes”. Estos datos se resumen en el gráfico tipo pastel de la

Figura 22

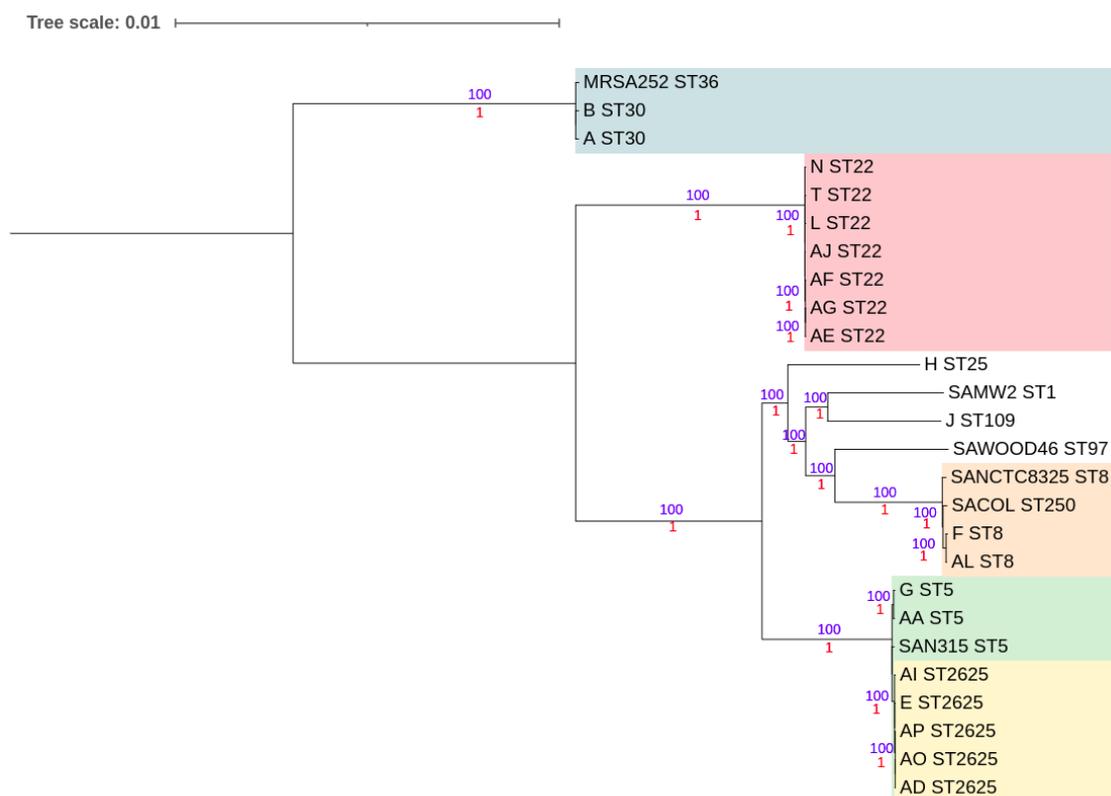
Distribución de los genes según su tipo en el pangenoma.



En la Figura 23 se halla el árbol filogenético que tomó en cuenta 2025 genes, los que conforman el core genome de los aislados del presente estudio. La topología del árbol se asemeja a la obtenida por MLST clasificando a ST30-ST36, ST22, ST8-ST250, ST5, ST2625 respectivamente. El mayor poder de resolución es evidente, dado que existen valores de soporte con el máximo grado de confianza incluso dentro de los grupos de STs antes mencionados.

Figura 23

Árbol Filogenético de máxima verosimilitud y bayesiano basado en el genoma core de los aislados de SA.



Nota. Valores de soporte UFBoot >90% (ubicados sobre los nodos con color azul) por el software iQ-tree

2: Repeticiones de UFBootstraps=1000, modelo de sustitución HKY+F. Valores de soporte posterior

Probabilities >0.90 (ubicados bajo los nodos con color rojo), calculados con el software Mr. Bayes:

modelo de sustitución HKY+F, longitud de la cadena MCMC 1000000, frecuencia de muestreo 1000,

burn-in 10%. La escala indica la distancia genética calculada. Imagen del árbol filogenético obtenida y

editada con iTOL V5.

Se fijó un cut-off de 15 SNPs para comparar el core genoma de los aislados e identificar posibles eventos de transmisión. Tres grupos de aislados se destacaron por cumplir esta condición: ST2625: AO, AP, AD; ST22: N, L, T y ST22: AE, AF, AG. Como ejemplo, en la Tabla 16

Mapa de color basado en el número de SNPs en el genoma core de los aislados pertenecientes a SA ST2625. se muestra el mapa de calor obtenido según las diferencias de SNPs para los aislados ST2625.

Tabla 16

Mapa de color basado en el número de SNPs en el genoma core de los aislados pertenecientes a SA ST2625.

	AI	AO	AP	E	AD
AI	-	77	78	86	77
AO	77	-	1	81	0
AP	78	1	-	82	1
E	86	81	82	-	81
AD	77	0	1	81	-

Búsqueda de genes de resistencia y virulencia La interrogación de los contigs para encontrar genes de resistencia y virulencia reveló que los aislados tuvieron hasta (promedio [rango], 11 [9 a 14] genes de resistencia y (promedio [rango], 55 [50 AJ a 59 AL] genes de virulencia. Los genes de resistencia que compartieron todos los aislados fueron arIS, arIR, tet, mgrA, norA, mepA, mepR. Los aislados de SA compartieron al menos 40 factores de virulencia. El análisis de presencia o ausencia se visualiza en la Tabla 17 y Tabla 18 para los genes de resistencia y virulencia, respectivamente.

Tabla 16

Genes de resistencia encontrados en los aislados clínicos de SA.

Gen	AA	AD	AE	AF	AG	AI	AJ	AL	AO	AP	A	B	E	F	G	H	J	L	N	T	Total
arlS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
arlR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
msrA	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	3
mphC	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
blaZ)	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	18
lnuA	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
tet(38)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
ErmC	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
LmrS	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
FosB	1	1	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	13
mgrA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
norA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
mepA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
mepR	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
tet(L)	0	1	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	5
tetM	0	1	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	5
mecA	0	1	0	0	0	1	0	1	1	1	0	0	1	1	0	0	0	0	0	0	7
AAC(6')- le- APH(2'')- la	0	1	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	5
LmrS	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
tet(K)	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	2
ErmA	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	3
Total	14	14	9	9	9	14	9	12	14	14	11	11	14	11	12	11	11	9	9	9	

Nota: El análisis de presencia-ausencia de genes de resistencia se realizó frente a la base de datos CARD

con un cut-off de 90% de identidad y 85% de cobertura. El "1" representa presencia y el "0" representa ausencia del gen.

sspC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
sspB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
sspA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
isdB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
isdA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
isdC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
isdD	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
isdE	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
isdF	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
srtB	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
isdG	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
clfA	1	0	0	0	0	0	0	1	0	0	1	1	0	1	1	0	0	0	0	6
tsst-1	1	0	1	1	1	0	1	0	0	0	0	1	0	0	1	0	0	1	1	10
Geh	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
Scn	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	19
Chp	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	19
Sak	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	19
Hld	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
sdrE	0	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1	0	1	1	15
sdrD	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	17
Map	0	0	1	1	1	0	1	1	0	0	0	0	0	1	0	0	1	1	1	10
Cna	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	3
Hlb	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
vWbp	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Selk	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Selq	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
lukF-PV	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
lukS-PV	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
clfB	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
cap8H	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2
cap8I	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2
cap8J	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2
cap8K	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2
Sea	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
Eta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Total	57	55	56	55	54	52	50	59	55	55	53	56	56	57	58	56	54	54	55	56

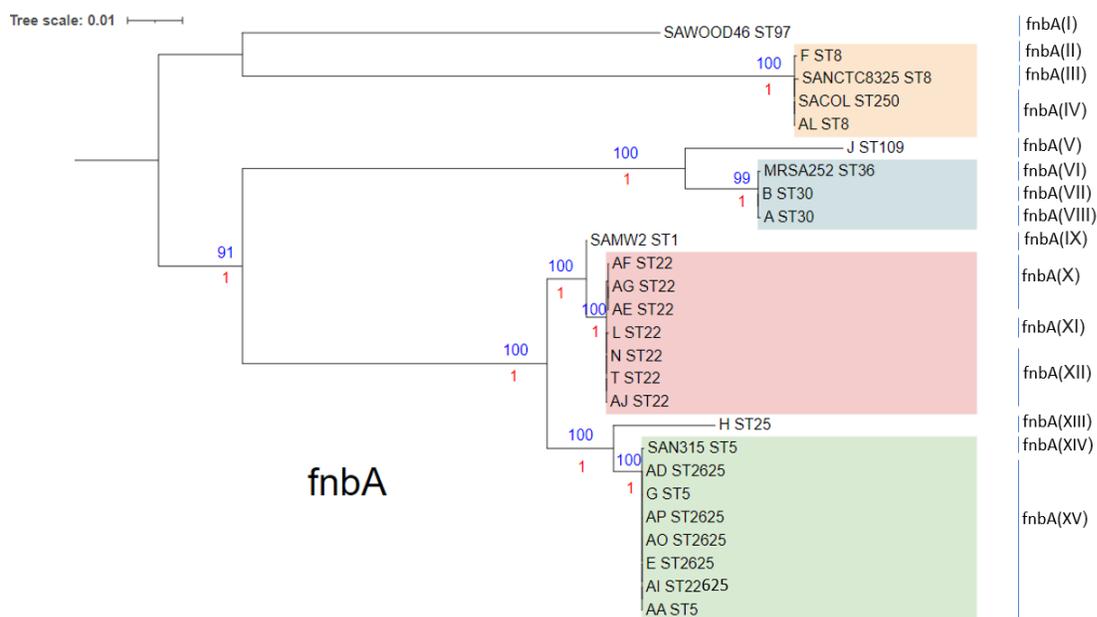
Nota: El análisis de presencia-ausencia de genes de resistencia se realizó frente a la base de datos VFDN

con un cut-off de 90% de identidad y 85% de cobertura. El “1” representa presencia y el “0” representa ausencia del gen.

Variabilidad genética de los genes (*fnbA*, *fhuD*, *sstD*, *hla*, *sdrC*). Se reportó una alta diversidad en el gen *fnbA* encontrándose hasta 687 SNPs de los aislados con respecto al genoma de referencia. En la Figura 24 se muestra el árbol filogenético construido para dicho gen en conjunto con la clasificación por genotipos de *hla* observada. El árbol filogenético reveló una topología similar a la clasificación por MLST y cg-SNPs. Se encontraron varios genotipos de *hla* por ST, por ejemplo, para ST8 se reportó la variante *fnbA* (II) y *fnbA* (IV). Por otro lado, ningún genotipo identificado compartió un ST excepto por *fnbA* (XV) el cual agrupa a ST5 y ST2625 y *fnbA* (IV) que agrupa a SACOL ST250 y AL ST8.

Figura 24

Árbol Filogenético de máxima verosimilitud y bayesiano para fnbA.



Nota. Valores de soporte UFBoot >90% (ubicados sobre los nodos con color azul) por el software iQ-tree 2: Repeticiones de UFBootstraps=1000, modelo de sustitución TPM2u+F+I. Valores de soporte posterior Probabilities >0.90 (ubicados bajo los nodos con color rojo), calculados con el software Mr. Bayes: modelo de sustitución GTR+F+I, longitud de la cadena MCMC 1000000, frecuencia de muestreo

1000, burn-in 10%. La escala indica la distancia genética calculada. Imagen del árbol filogenético obtenida y editada con iTOL V5.

No se encontró una coincidencia de mutaciones en las secuencias de *fnbA* en los aislados para los que previamente se reportó *fnbA* regulado al alza. El cotejamiento para los aislados que presentaron regulación al alza se muestra en la Tabla 19. Tampoco se encontró un patrón al comparar los aislados regulados a la baja (tabla no mostrada).

Tabla 18

Variantes genéticas observadas entre los aislados con regulación al alza de fnbA.

POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia
96	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
166	A	G	0	0	0	0	0	0	1	0	0	1	No coincide
167	C	T	0	0	0	0	0	0	1	0	0	1	No coincide
171	T	C	0	0	0	0	0	0	1	0	0	1	No coincide
177	A	T	0	0	0	0	0	0	0	0	0	1	No coincide
187	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
189	T	A	0	0	0	1	0	0	1	1	0	1	No coincide
202	C	A	0	0	0	0	0	0	1	0	0	0	No coincide
203	A	C	0	0	0	0	0	0	1	0	0	0	No coincide
210	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
216	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
225	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
227	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
236	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
300	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
325	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
347	A	C	0	0	0	1	0	0	1	1	0	0	No coincide
351	A	C	0	0	0	0	0	0	1	0	0	0	No coincide
369	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
376	C	T	0	0	0	1	0	0	1	1	0	0	No coincide
394	A	G	0	0	0	1	0	0	1	1	0	0	No coincide
428	T	C	0	0	0	1	0	0	1	1	0	0	No coincide
443	G	T	0	0	0	1	0	0	1	1	0	0	No coincide

444	G	C,A	0	0	0	1	0	0	2	1	0	0	No coincide
445	A	G	0	0	0	1	0	0	1	1	0	0	No coincide
485	C	T	0	0	0	0	0	0	1	0	0	0	No coincide
514	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
544	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
548	T	C	0	0	0	1	0	0	1	1	0	0	No coincide
549	T	G	0	0	0	1	0	0	1	1	0	0	No coincide
550	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
551	C	A	0	0	0	1	0	0	1	1	0	0	No coincide
554	A	T	0	0	0	1	0	0	1	1	0	0	No coincide
555	A	G	0	0	0	1	0	0	1	1	0	0	No coincide
557	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
558	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
560	A	C	0	0	0	1	0	0	1	1	0	0	No coincide
561	A	G	0	0	0	1	0	0	1	1	0	0	No coincide
564	C	T	0	0	0	1	0	0	1	1	0	0	No coincide
573	T	A,G	0	0	0	1	0	0	2	1	0	0	No coincide
591	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
594	A	G	0	0	0	0	0	0	0	0	0	1	No coincide
595	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
596	G	T,A	0	0	0	1	0	0	2	1	0	0	No coincide
601	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
602	C	A	0	0	0	0	0	0	1	0	0	0	No coincide
603	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
609	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
611	C	G	0	0	0	1	0	0	0	1	0	0	No coincide
612	A	G	0	0	0	1	0	0	1	1	0	0	No coincide
614	C	A	0	0	0	1	0	0	0	1	0	0	No coincide
615	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
616	C	A	0	0	0	1	0	0	1	1	0	0	No coincide
618	A	C	0	0	0	1	0	0	0	1	0	0	No coincide
619	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
620	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
627	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
630	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
631	G	C	0	0	0	0	0	0	1	0	0	0	No coincide

633	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
636	A	T	0	0	0	1	0	0	1	1	0	0	No coincide
641	C	A	0	0	0	0	0	0	1	0	0	0	No coincide
642	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
645	T	A,G	0	0	0	1	0	0	2	1	0	0	No coincide
651	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
653	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
654	C	G,A	0	0	0	1	0	0	2	1	0	0	No coincide
658	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
660	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
663	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
666	C	T	0	0	0	1	0	0	1	1	0	0	No coincide
669	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
670	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
678	C	T	0	0	0	1	0	0	1	1	0	0	No coincide
679	G	C	0	0	0	0	0	0	1	0	0	0	No coincide
680	C	A	0	0	0	1	0	0	1	1	0	0	No coincide
681	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
682	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
687	A	T,G	0	0	0	1	0	0	2	1	0	0	No coincide
690	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
691	A	C	0	0	0	1	0	0	0	1	0	0	No coincide
693	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
694	A	C	0	0	0	1	0	0	0	1	0	0	No coincide
695	G	A,C	0	0	0	1	0	0	2	1	0	0	No coincide
699	A	T,C	0	0	0	1	0	0	2	1	0	0	No coincide
702	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
705	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
711	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
717	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
738	T	G,A	0	0	0	1	0	0	2	1	0	0	No coincide
739	T	C	0	0	0	1	0	0	1	1	0	0	No coincide
744	G	C,T	0	0	0	1	0	0	2	1	0	0	No coincide
747	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
748	T	G	0	0	0	0	0	0	1	0	0	0	No coincide
753	A	T	0	0	0	1	0	0	1	1	0	0	No coincide

754	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
756	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
762	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
765	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
771	G	A,T	0	0	0	1	0	0	2	1	0	0	No coincide
783	C	T,A	0	0	0	1	0	0	2	1	0	0	No coincide
787	G	T	0	0	0	0	0	0	1	0	0	0	No coincide
789	T	C,A	0	0	0	1	0	0	2	1	0	0	No coincide
792	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
798	T	G	0	0	0	1	0	0	0	1	0	0	No coincide
800	C	A	0	0	0	0	0	0	1	0	0	0	No coincide
805	G	C	0	0	0	0	0	0	1	0	0	0	No coincide
807	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
808	A	G	0	0	0	1	0	0	1	1	0	0	No coincide
810	C	T,A	0	0	0	1	0	0	2	1	0	0	No coincide
815	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
816	G	A,T	0	0	0	1	0	0	2	1	0	0	No coincide
817	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
818	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
822	T	A,C	0	0	0	1	0	0	2	1	0	0	No coincide
823	A	C	0	0	0	0	0	0	1	0	0	0	No coincide
824	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
825	C	G,A	0	0	0	1	0	0	2	1	0	0	No coincide
828	A	T	0	0	0	1	0	0	1	1	0	0	No coincide
843	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
844	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
846	C	T	0	0	0	1	0	0	1	1	0	0	No coincide
847	G	T	0	0	0	0	0	0	1	0	0	0	No coincide
849	A	T,C	0	0	0	1	0	0	2	1	0	0	No coincide
852	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
853	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
856	C	G	0	0	0	1	0	0	1	1	0	0	No coincide
858	C	T	0	0	0	1	0	0	1	1	0	0	No coincide
861	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
864	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia

865	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
867	G	T	0	0	0	1	0	0	1	1	0	0	No coincide
870	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
873	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
877	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
879	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
880	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
888	C	T,A	0	0	0	1	0	0	2	1	0	0	No coincide
891	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
897	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
900	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
903	C	T	0	0	0	1	0	0	1	1	0	0	No coincide
920	G	C	0	0	0	1	0	0	0	1	0	0	No coincide
921	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
928	G	A,C	0	0	0	1	0	0	2	1	0	0	No coincide
930	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
935	A	C	0	0	0	1	0	0	1	1	0	0	No coincide
936	G	T,A	0	0	0	1	0	0	2	1	0	0	No coincide
939	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
945	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
947	A	C	0	0	0	1	0	0	0	1	0	0	No coincide
949	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
956	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
957	T	A,C	0	0	0	1	0	0	2	1	0	0	No coincide
958	G	A,C	0	0	0	0	0	0	1	0	0	2	No coincide
961	G	C	0	0	0	1	0	0	0	1	0	0	No coincide
966	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
967	G	T	0	0	0	1	0	0	1	1	0	0	No coincide
968	A	C	0	0	0	1	0	0	1	1	0	0	No coincide
970	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
971	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
972	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
973	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
974	C	A	0	0	0	1	0	0	0	1	0	0	No coincide
975	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia

976	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
978	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
979	C	G	0	0	0	1	0	0	0	1	0	0	No coincide
980	C	A	0	0	0	1	0	0	1	1	0	0	No coincide
981	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
982	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
984	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
985	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
986	T	A,C	0	0	0	1	0	0	2	1	0	0	No coincide
987	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
993	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
994	C	G	0	0	0	1	0	0	1	1	0	0	No coincide
995	C	A	0	0	0	1	0	0	1	1	0	0	No coincide
996	A	T	0	0	0	1	0	0	1	1	0	0	No coincide
1000	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
1003	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1004	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
1005	C	G,A	0	0	0	1	0	0	2	1	0	0	No coincide
1009	A	T,C	0	0	0	1	0	0	2	1	0	0	No coincide
1010	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1011	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
1015	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1017	A	C	0	0	0	1	0	0	0	1	0	0	No coincide
1021	G	T	0	0	0	1	0	0	0	1	0	0	No coincide
1029	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
1032	A	G	0	0	0	1	0	0	1	1	0	0	No coincide
1038	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1041	A	G	0	0	0	0	0	0	1	0	0	1	No coincide
1045	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
1047	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
1052	A	C	0	0	0	1	0	0	0	1	0	0	No coincide
1053	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1054	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
1055	C	A	0	0	0	1	0	0	1	1	0	0	No coincide
1056	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia

1061	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1066	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1068	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1071	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
1072	A	G	0	0	0	1	0	0	1	1	0	0	No coincide
1074	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1077	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
1079	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1086	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
1089	A	T	0	0	0	1	0	0	1	1	0	0	No coincide
1091	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1092	G	T,A	0	0	0	1	0	0	2	1	0	0	No coincide
1098	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1101	C	A	0	0	0	1	0	0	0	1	0	0	No coincide
1102	C	A	0	0	0	1	0	0	1	1	0	0	No coincide
1103	A	C	0	0	0	1	0	0	0	1	0	0	No coincide
1104	G	T	0	0	0	0	0	0	1	0	0	0	No coincide
1105	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1109	C	G	0	0	0	1	0	0	1	1	0	0	No coincide
1113	A	G,T	0	0	0	1	0	0	2	1	0	0	No coincide
1114	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
1116	A	T	0	0	0	1	0	0	1	1	0	0	No coincide
1119	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1122	A	T,G	0	0	0	1	0	0	2	1	0	0	No coincide
1125	G	A,T	0	0	0	1	0	0	2	1	0	0	No coincide
1127	C	T	0	0	0	0	0	0	1	0	0	0	No coincide
1128	G	T	0	0	0	1	0	0	0	1	0	0	No coincide
1131	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
1133	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
1134	T	G,A	0	0	0	1	0	0	2	1	0	0	No coincide
1135	G	A,C	0	0	0	1	0	0	2	1	0	0	No coincide
1140	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
1147	T	C,G	0	0	0	1	0	0	2	1	0	0	No coincide
1148	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
1149	A	G	0	0	0	1	0	0	1	1	0	0	No coincide
POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia

1150	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1151	C	A	0	0	0	1	0	0	1	1	0	0	No coincide
1155	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
1156	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1157	G	A,C	0	0	0	1	0	0	2	1	0	0	No coincide
1158	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1164	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
1166	C	A	0	0	0	1	0	0	1	1	0	0	No coincide
1167	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1172	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1173	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1174	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1178	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1182	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
1185	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
1186	C	T,G	0	0	0	1	0	0	2	1	0	0	No coincide
1188	A	G,T	0	0	0	1	0	0	2	1	0	0	No coincide
1191	G	T,A	0	0	0	1	0	0	2	1	0	0	No coincide
1192	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
1193	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
1194	A	T	0	0	0	1	0	0	1	1	0	0	No coincide
1195	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
1197	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1198	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
1200	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1202	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
1203	A	C,G	0	0	0	1	0	0	2	1	0	0	No coincide
1204	T	A,C	0	0	0	1	0	0	2	1	0	0	No coincide
1206	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1207	C	G	0	0	0	1	0	0	0	1	0	0	No coincide
1209	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1210	C	A	0	0	0	1	0	0	0	1	0	0	No coincide
1212	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1218	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1230	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia

1231	T	A,G	0	0	0	1	0	0	2	1	0	0	No coincide
1232	C	T	0	0	0	0	0	0	1	0	0	0	No coincide
1233	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
1236	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
1238	C	G	0	0	0	0	0	0	1	0	0	0	No coincide
1239	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
1240	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1241	A	C	0	0	0	1	0	0	1	1	0	0	No coincide
1242	C	T,A	0	0	0	1	0	0	2	1	0	0	No coincide
1243	A	C	0	0	0	0	0	0	1	0	0	0	No coincide
1248	C	T,G	0	0	0	1	0	0	2	1	0	0	No coincide
1254	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1257	A	C	0	0	0	1	0	0	0	1	0	0	No coincide
1260	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
1262	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1263	G	T	0	0	0	1	0	0	1	1	0	1	No coincide
1264	G	A,C	0	0	0	1	0	0	2	1	0	0	No coincide
1266	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1270	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
1271	A	G	0	0	0	1	0	0	1	1	0	0	No coincide
1274	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
1275	A	G,T	0	0	0	1	0	0	2	1	0	0	No coincide
1278	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1283	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1284	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
1285	G	T	0	0	0	1	0	0	0	1	0	0	No coincide
1287	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1291	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1293	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
1299	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1302	T	C	0	0	0	1	0	0	1	1	0	0	No coincide
1305	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
1311	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
1315	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1317	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia

1320	T	A,C	0	0	0	1	0	0	2	1	0	0	No coincide
1323	G	T,A	0	0	0	1	0	0	2	1	0	0	No coincide
1324	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
1326	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1335	G	T,A	0	0	0	1	0	0	2	1	0	0	No coincide
1341	C	T,A	0	0	0	1	0	0	2	1	0	0	No coincide
1342	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1347	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
1350	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
1351	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1352	C	A	0	0	0	1	0	0	0	1	0	0	No coincide
1353	A	T,C	0	0	0	1	0	0	2	1	0	0	No coincide
1356	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
1359	A	G,T	0	0	0	1	0	0	2	1	0	0	No coincide
1362	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
1365	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1366	C	A	0	0	0	1	0	0	1	1	0	0	No coincide
1368	A	T	0	0	0	1	0	0	1	1	0	0	No coincide
1371	G	T	0	0	0	1	0	0	1	1	0	0	No coincide
1372	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
1374	A	T,G	0	0	0	1	0	0	2	1	0	0	No coincide
1375	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
1376	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
1378	C	G	0	0	0	1	0	0	1	1	0	0	No coincide
1380	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1383	T	G	0	0	0	0	0	0	1	0	0	0	No coincide
1384	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1386	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
1395	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1396	G	C	0	0	0	1	0	0	1	1	0	0	No coincide
1399	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1401	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1402	T	G,A	0	0	0	1	0	0	2	1	0	0	No coincide
1403	A	T,C	0	0	0	1	0	0	2	1	0	0	No coincide
1404	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia

1407	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1408	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
1410	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
1413	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
1416	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1417	C	A	0	0	0	0	0	0	1	0	0	0	No coincide
1418	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1419	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
1420	G	C	0	0	0	1	0	0	1	1	0	0	No coincide
1421	C	T,G	0	0	0	1	0	0	2	1	0	0	No coincide
1422	A	T,C	0	0	0	1	0	0	2	1	0	0	No coincide
1428	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1431	C	T	0	0	0	1	0	0	1	1	0	0	No coincide
1435	G	T	0	0	0	1	0	0	1	1	0	0	No coincide
1436	T	A	0	0	0	1	0	0	1	1	0	0	No coincide
1438	T	G	0	0	0	1	0	0	0	1	0	0	No coincide
1441	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1442	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
1443	G	A,T	0	0	0	1	0	0	2	1	0	0	No coincide
1446	A	C	0	0	0	0	0	0	1	0	0	0	No coincide
1450	C	A	0	0	0	1	0	0	1	1	0	0	No coincide
1451	G	C	0	0	0	1	0	0	1	1	0	0	No coincide
1452	T	C,A	0	0	0	1	0	0	2	1	0	0	No coincide
1458	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
1470	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
1482	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
1485	C	T	0	0	0	1	0	0	1	1	0	0	No coincide
1494	T	G	0	0	0	1	0	0	0	1	0	0	No coincide
1495	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1497	C	T	0	0	0	0	0	0	1	0	0	1	No coincide
1500	C	A,T	0	0	0	1	0	0	2	1	0	0	No coincide
1501	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
1506	T	G,A	0	0	0	1	0	0	2	1	0	0	No coincide
1510	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
1515	A	T	0	0	0	1	0	0	1	1	0	0	No coincide
POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia

1517	A	C	0	0	0	1	0	0	1	1	0	0	No coincide
1518	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1521	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
1522	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
1524	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
1525	C	G	0	0	0	0	0	0	1	0	0	0	No coincide
1527	A	C	0	0	0	0	0	0	1	0	0	0	No coincide
1528	G	A,T	0	0	0	1	0	0	2	1	0	1	No coincide
1529	A	C	0	0	0	0	0	0	1	0	0	0	No coincide
1530	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
1534	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1536	T	A,C	0	0	0	1	0	0	2	1	0	0	No coincide
1542	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
1544	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
1545	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
1546	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
1547	A	C	0	0	0	0	0	0	1	0	0	0	No coincide
1548	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
1556	C	G	0	0	0	0	0	0	1	0	0	0	No coincide
1557	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1558	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
1559	C	T,G	0	0	0	1	0	0	2	1	0	0	No coincide
1560	A	T	0	0	0	1	0	0	1	1	0	0	No coincide
1563	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
1565	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
1566	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
1567	A	T	0	0	0	0	0	0	1	0	0	0	No coincide
1569	T	A	0	0	0	0	0	0	1	0	0	0	No coincide
1570	A	C	0	0	0	1	0	0	0	1	0	0	No coincide
1572	G	T,C	0	0	0	1	0	0	2	1	0	0	No coincide
1574	G	C	0	0	0	1	0	0	0	1	0	0	No coincide
1575	C	A	0	0	0	1	0	0	0	1	0	0	No coincide
1578	G	T,A	0	0	0	1	0	0	2	1	0	0	No coincide
1581	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1584	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia

1588	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
1589	C	A	0	0	0	1	0	0	0	1	0	0	No coincide
1590	C	G	0	0	0	1	0	0	1	1	0	0	No coincide
1593	G	T	0	0	0	1	0	0	0	1	0	0	No coincide
1594	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
1595	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1597	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1599	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1601	T	C	0	0	0	1	0	0	0	1	0	0	No coincide
1603	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
1604	A	C	0	0	0	1	0	0	0	1	0	0	No coincide
1605	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1606	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1607	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
1608	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1614	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1615	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1617	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1618	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
1620	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1623	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
1624	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1625	A	C	0	0	0	1	0	0	0	1	0	0	No coincide
1626	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1627	A	T	0	0	0	1	0	0	0	1	0	0	No coincide
1630	C	A	0	0	0	1	0	0	0	1	0	0	No coincide
1632	G	T	0	0	0	1	0	0	0	1	0	0	No coincide
1667	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
1668	G	A	0	0	0	1	0	0	1	1	0	1	No coincide
1674	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
1686	G	A	0	0	0	1	0	0	1	1	0	1	No coincide
1689	T	C	0	0	0	1	0	0	1	1	0	1	No coincide
1746	T	A	0	0	0	0	0	0	0	0	0	1	No coincide
1755	T	C	0	0	0	1	0	0	1	1	0	0	No coincide
1761	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia

1767	C	T	0	0	0	0	0	0	0	0	0	1	No coincide
1785	G	A	0	0	0	0	0	0	0	0	0	1	No coincide
1788	C	A	0	0	0	0	0	0	0	0	0	1	No coincide
1794	G	A	0	0	0	0	0	0	0	0	0	1	No coincide
1800	A	C	0	0	0	0	0	0	0	0	0	1	No coincide
1806	A	G	0	0	0	0	0	0	0	0	0	1	No coincide
1812	C	T	0	0	0	0	0	0	0	0	0	1	No coincide
1824	G	T	0	0	0	1	0	0	1	1	0	0	No coincide
1860	T	G	0	0	0	0	0	0	0	0	0	1	No coincide
1872	G	A	0	0	0	1	0	0	1	1	0	0	No coincide
1881	T	C	0	0	0	1	0	0	1	1	0	1	No coincide
1890	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
1893	C	T	0	0	0	1	0	0	1	1	0	0	No coincide
1896	A	G	0	0	0	1	0	0	1	1	0	0	No coincide
1899	A	T	0	0	0	1	0	0	1	1	0	0	No coincide
1902	A	T	0	0	0	1	0	0	1	1	0	0	No coincide
1908	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
1943	A	G	0	0	0	0	0	0	0	0	0	1	No coincide
1971	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
1981	A	G	0	0	0	1	0	0	1	1	0	1	No coincide
1989	T	C	0	0	0	0	0	0	0	0	0	1	No coincide
2004	G	A	0	0	0	1	0	0	1	1	0	1	No coincide
2024	A	G	0	0	0	0	0	0	0	1	0	0	No coincide
2026	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
2031	A	G	0	0	0	1	0	0	0	1	0	1	No coincide
2038	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
2049	A	G	0	0	0	1	0	0	0	1	0	1	No coincide
2067	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
2071	A	G	0	0	0	1	0	0	1	1	0	0	No coincide
2082	T	C	0	0	0	0	0	0	1	0	0	1	No coincide
2139	T	G	0	0	0	1	0	0	0	1	0	0	No coincide
2147	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
2163	C	T	0	0	0	0	0	0	0	0	0	1	No coincide
2175	C	T	0	0	0	0	0	0	0	0	0	1	No coincide
2178	C	T	0	0	0	0	0	0	1	0	0	1	No coincide
POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia

2205	C	T	0	0	0	1	0	0	0	1	0	1	No coincide
2229	C	T	0	0	0	0	0	0	1	0	0	0	No coincide
2259	C	T	0	0	0	0	0	0	0	0	0	1	No coincide
2268	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
2286	T	C	0	0	0	1	0	0	1	1	0	1	No coincide
2310	T	C	0	0	0	0	0	0	0	0	0	1	No coincide
2319	A	T	0	0	0	1	0	0	0	1	0	1	No coincide
2325	T	C	0	0	0	1	0	0	0	1	0	1	No coincide
2340	T	C	0	0	0	1	0	0	0	1	0	1	No coincide
2342	A	C	0	0	0	1	0	0	0	1	0	1	No coincide
2367	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
2373	T	C	0	0	0	0	0	0	0	0	0	1	No coincide
2379	T	A	0	0	0	1	0	0	0	1	0	0	No coincide
2381	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
2403	T	C	0	0	0	0	0	0	1	0	0	1	No coincide
2405	G	T	0	0	0	0	0	0	0	0	0	1	No coincide
2412	C	T	0	0	0	0	0	0	1	0	0	1	No coincide
2445	C	T	0	0	0	0	0	0	1	0	0	0	No coincide
2451	T	C	0	0	0	0	0	0	1	0	0	0	No coincide
2484	G	A	0	0	0	1	0	0	0	1	0	0	No coincide
2487	G	T	0	0	0	1	0	0	0	1	0	0	No coincide
2492	C	T	0	0	0	1	0	0	0	1	0	0	No coincide
2493	G	A	0	0	0	0	0	0	1	0	0	1	No coincide
2499	G	A	0	0	0	1	0	0	0	1	0	1	No coincide
2502	A	G	0	0	0	1	0	0	0	1	0	0	No coincide
2505	G	A	0	0	0	1	0	0	1	1	0	1	No coincide
2508	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
2523	G	A	0	0	0	0	0	0	1	0	0	0	No coincide
2525	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
2526	A	C	0	0	0	0	0	0	1	0	0	0	No coincide
2538	A	G	0	0	0	0	0	0	1	0	0	1	No coincide
2541	G	A	0	0	0	0	0	0	1	0	0	1	No coincide
2544	A	G	0	0	0	0	0	0	1	0	0	0	No coincide
2550	A	G	0	0	0	0	0	0	0	0	0	1	No coincide
2553	G	A	0	0	0	0	0	0	0	0	0	1	No coincide
POS	REF	ALT	AA	AD	AI	AL	AO	AP	B	F	G	H	Coincidencia

2556	G	A	0	0	0	0	0	0	0	1	0	0	1	No coincide
2565	A	G	0	0	0	0	0	0	0	0	0	0	1	No coincide
2567	A	G	0	0	0	0	0	0	0	1	0	0	0	No coincide
2568	A	T	0	0	0	0	0	0	0	1	0	0	0	No coincide
2570	C	A	0	0	0	0	0	0	0	1	0	0	0	No coincide
2571	A	T,G	0	0	0	1	0	0	0	2	1	0	0	No coincide
2580	G	A	0	0	0	0	0	0	0	1	0	0	1	No coincide
2586	A	G	0	0	0	0	0	0	0	0	0	0	1	No coincide
2595	A	G	0	0	0	0	0	0	0	0	0	0	1	No coincide
2625	C	T	0	0	0	1	0	0	0	1	1	0	1	No coincide
2628	A	C,T	0	0	0	1	0	0	0	2	1	0	1	No coincide
2673	A	G	0	0	0	0	0	0	0	0	0	0	1	No coincide
2697	C	T	0	0	0	0	0	0	0	1	0	0	0	No coincide
2731	A	C	0	0	0	0	0	0	0	1	0	0	0	No coincide
2734	G	C,A	0	0	0	1	0	0	0	2	1	0	0	No coincide
2735	C	A	0	0	0	0	0	0	0	1	0	0	0	No coincide
2740	T	G	0	0	0	0	0	0	0	1	0	0	0	No coincide
2832	A	G	0	0	0	0	0	0	0	1	0	0	1	No coincide
2841	C	T	0	0	0	0	0	0	0	0	0	0	1	No coincide
2857	A	C	0	0	0	1	0	0	0	1	1	0	0	No coincide

Nota: En la Tabla se muestra la posición nucleotídica. La columna "REF" indica la base nucleotídica en la secuencia de fnbA del aislado AA, usado como referencia. La columna "ALT" indica 1-3 sustituciones que poseen las secuencias de los aislados en estudio para esa posición. En la Tabla los valores 1-3 se corresponden con dichas sustituciones.

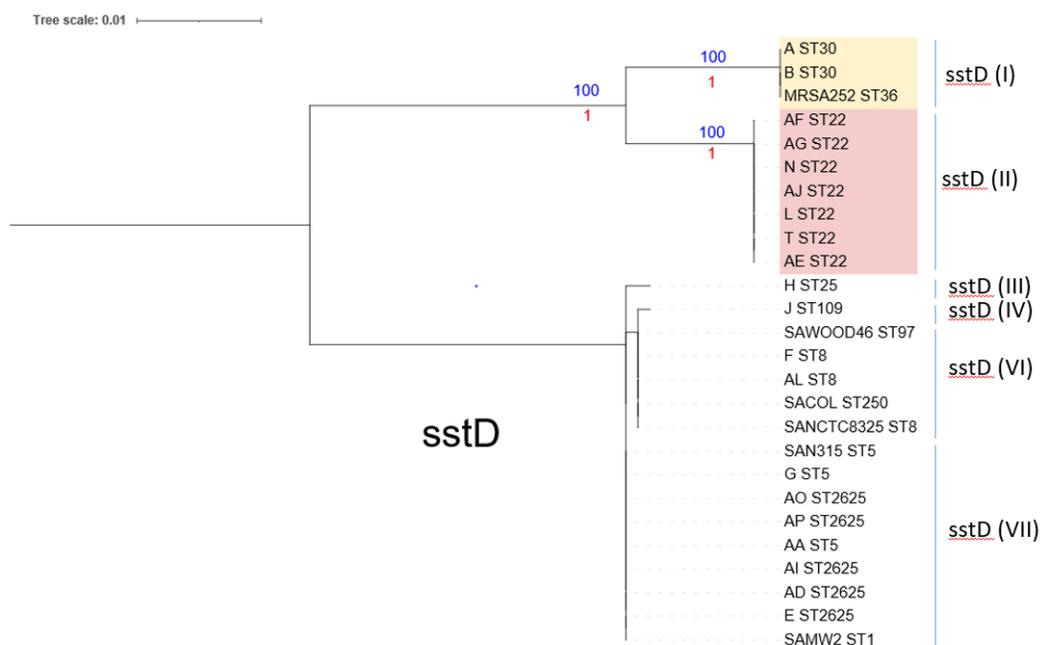
La variabilidad genética reportada para sstD fue menor en relación a fnbA. Se reportaron hasta 74 SNPs en relación a la secuencia de referencia. En la Figura 25

Árbol filogenético de máxima verosimilitud y bayesiano para el gen sstD., se presenta el árbol filogenético obtenido. Además, se presenta la clasificación hasta en 7 genotipos diferentes para sstD. El árbol filogenético únicamente tiene valores de soporte adecuados para dos nodos. El primero conformado por ST30-ST36 y el segundo por ST22. Los otros aislados no pudieron diferenciarse con un grado adecuado de soporte. Varios genotipos compartieron STs como, por ejemplo, ST30-ST36 con

sstD(I); ST97, ST8, ST250 con sstD (VI) y ST5, ST2625, ST1 con sstD (VII). Sin embargo, ningún ST compartió genotipos de sstD.

Figura 25

Árbol filogenético de máxima verosimilitud y bayesiano para el gen sstD.



Nota. Valores de soporte UFBoot >90% (ubicados sobre los nodos con color azul) por el software iQ-tree 2: Repeticiones de UFBootstraps=1000, modelo de sustitución HKY+F. Valores de soporte posterior Probabilities >0.90 (ubicados bajo los nodos con color rojo), calculados con el software Mr. Bayes: modelo de sustitución HKY+F, longitud de la cadena MCMC 1000000, frecuencia de muestreo 1000, burn-in 10%. La escala indica la distancia genética calculada. Imagen del árbol filogenético obtenida y editada con iTOL V5.

No se encontró una coincidencia de mutaciones en las secuencias de sstD en los aislados para los que previamente se reportó sstD regulado al alza. El cotejamiento para los aislados que presentaron regulación al alta se muestra en la Tabla 20Tabla 19

Variantes genéticas observadas entre los aislados con regulación al alza de sstD. Tampoco se encontró un patrón al comparar los aislados regulados a la baja (tabla no mostrada)

Tabla 19

Variantes genéticas observadas entre los aislados con regulación al alza de sstD.

POS	REF	ALT	AA	AD	B	F	G	Coincidencia
28	T	G	0	0	1	0	0	No coinciden
129	T	C	0	0	1	0	0	No coinciden
165	T	C	0	0	1	0	0	No coinciden
168	C	T	0	0	1	0	0	No coinciden
204	A	G	0	0	1	0	0	No coinciden
207	T	C	0	0	1	0	0	No coinciden
225	G	A	0	0	1	0	0	No coinciden
231	C	T	0	0	1	0	0	No coinciden
273	T	C	0	0	1	0	0	No coinciden
297	G	T	0	0	1	0	0	No coinciden
306	T	C	0	0	1	0	0	No coinciden
330	A	G	0	0	1	0	0	No coinciden
345	T	C	0	0	1	0	0	No coinciden
348	A	G	0	0	1	0	0	No coinciden
357	C	T	0	0	1	0	0	No coinciden
366	T	C	0	0	1	0	0	No coinciden
369	A	G	0	0	1	0	0	No coinciden
378	G	T	0	0	1	0	0	No coinciden
384	T	C	0	0	1	0	0	No coinciden
396	T	A	0	0	1	0	0	No coinciden
399	G	A	0	0	1	0	0	No coinciden
402	G	A	0	0	1	0	0	No coinciden
408	C	G	0	0	1	0	0	No coinciden
429	A	C	0	0	1	0	0	No coinciden
438	T	A	0	0	1	0	0	No coinciden
462	C	T	0	0	1	0	0	No coinciden
477	T	A	0	0	1	0	0	No coinciden
483	G	A	0	0	1	0	0	No coinciden
495	T	C	0	0	1	0	0	No coinciden

POS	REF	ALT	AA	AD	B	F	G	Coincidencia
516	C	T	0	0	1	0	0	No coinciden
523	G	A	0	0	1	0	0	No coinciden
552	C	T	0	0	1	0	0	No coinciden
555	T	A	0	0	0	1	0	No coinciden
636	G	T	0	0	1	0	0	No coinciden
639	A	G	0	0	1	0	0	No coinciden
642	C	G	0	0	1	0	0	No coinciden
648	T	C	0	0	1	0	0	No coinciden
652	A	T	0	0	1	0	0	No coinciden
654	T	A	0	0	1	0	0	No coinciden
667	C	T	0	0	1	0	0	No coinciden
675	C	T	0	0	1	0	0	No coinciden
714	T	C	0	0	1	0	0	No coinciden
741	A	T	0	0	1	0	0	No coinciden
747	T	C	0	0	1	0	0	No coinciden
750	A	G	0	0	1	0	0	No coinciden
762	A	C	0	0	1	0	0	No coinciden
768	T	C	0	0	1	0	0	No coinciden
774	C	T	0	0	1	0	0	No coinciden
777	T	C	0	0	1	0	0	No coinciden
780	T	C	0	0	1	0	0	No coinciden
792	T	C	0	0	1	0	0	No coinciden
804	T	C	0	0	1	0	0	No coinciden
805	G	A	0	0	1	0	0	No coinciden
837	G	A	0	0	1	0	0	No coinciden
849	T	A	0	0	1	0	0	No coinciden
852	A	G	0	0	1	0	0	No coinciden
855	G	T	0	0	1	0	0	No coinciden
856	A	G	0	0	1	0	0	No coinciden
936	A	T	0	0	1	0	0	No coinciden
948	T	C	0	0	1	0	0	No coinciden
1005	G	T	0	0	1	0	0	No coinciden
1014	T	A	0	0	1	0	0	No coinciden

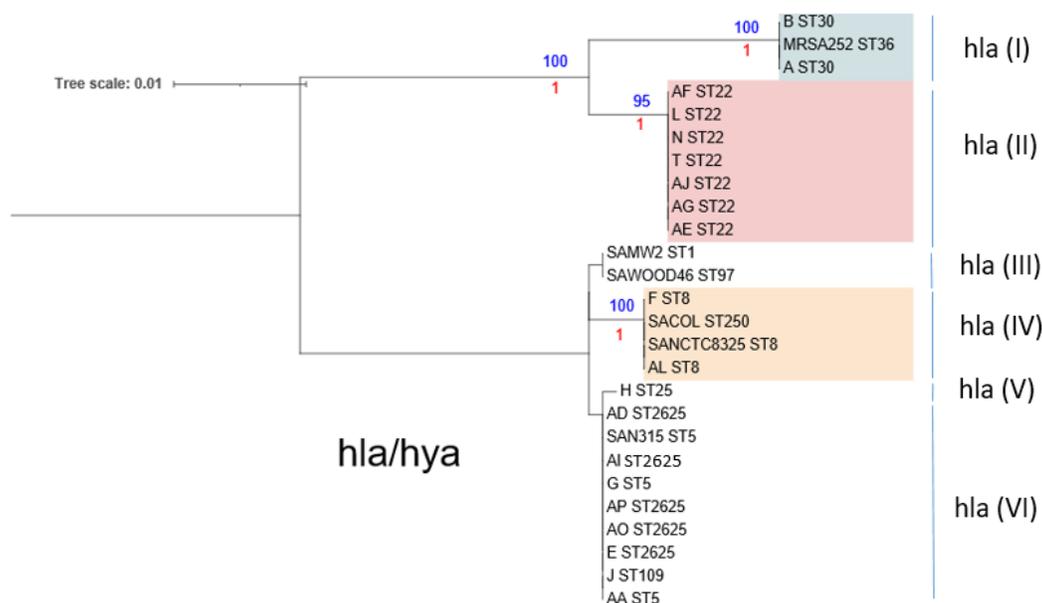
Nota: En la tabla se muestra la posición nucleotídica. La columna “REF” indica la base nucleotídica en la secuencia de sstD del aislado AA, usado como referencia. La columna “ALT” indica 1-3 sustituciones que

poseen las secuencias de los aislados en estudio para esa posición. En la tabla los valores 1-3 se corresponden con dichas sustituciones.

La variabilidad genética de hla aquí reportada es menor en relación con fnbA y sstD. Lo anterior debido a que se detectaron al menos 67 SNPs de los aislados con respecto al genoma de referencia. Esto se traduce en 6 genotipos que pueden ser visualizados en el árbol filogenético presentado en la Figura 26. Se encontró que de los 6 genotipos de hla, un solo genotipo de hla fue encontrado en ST22. Por otro lado, el genotipo más frecuente hla (VI) fue identificado en ST2625, ST5 y ST109.

Figura 26

Árbol filogenético de máxima verosimilitud y bayesiano para el gen hla.



Nota: Valores de soporte UFBoot >90% (ubicados sobre los nodos con color azul) por el software iQ-tree 2: Repeticiones de UFBootstraps=1000, modelo de sustitución TN+F+I. Valores de soporte posterior Probabilities >0.90 (ubicados bajo los nodos con color rojo), calculados con el software Mr. Bayes: modelo de sustitución GTR+F+I, longitud de la cadena MCMC 1000000, frecuencia de

muestreo 1000, burn-in 10%. La escala indica la distancia genética calculada. Imagen del árbol filogenético obtenida y editada con iTOL V5.

No se encontró una coincidencia de mutaciones en las secuencias de hla en los aislados para los que previamente se reportó hla regulado al alza. El cotejamiento para los aislados que presentaron regulación al alta se muestra en la Tabla 21. Tampoco se encontró un patrón al comprar los aislados regulados a la baja (tabla no mostrada)

Tabla 20

Variantes genéticas observadas entre los aislados con regulación al alza de hla.

POS	REF	ALT	AA	G	AD	AO	AG	F	A	B	Coincidencia
51	C	T	0	0	0	0	0	0	1	1	No coincide
120	T	A	0	0	0	0	1	0	1	1	No coincide
134	C	T	0	0	0	0	1	0	0	0	No coincide
222	C	T	0	0	0	0	0	1	0	0	No coincide
232	C	A	0	0	0	0	1	0	1	1	No coincide
234	G	A	0	0	0	0	1	0	1	1	No coincide
235	C	T	0	0	0	0	1	0	1	1	No coincide
240	T	C	0	0	0	0	1	0	1	1	No coincide
243	T	C	0	0	0	0	1	0	1	1	No coincide
270	A	G	0	0	0	0	1	0	1	1	No coincide
294	T	C	0	0	0	0	1	0	1	1	No coincide
300	C	T	0	0	0	0	1	0	1	1	No coincide
306	T	C	0	0	0	0	0	0	1	1	No coincide
327	C	A	0	0	0	0	0	0	1	1	No coincide
337	C	T	0	0	0	0	0	0	1	1	No coincide
366	T	A	0	0	0	0	1	0	1	1	No coincide
372	A	T	0	0	0	0	1	0	1	1	No coincide
384	T	C	0	0	0	0	0	0	1	1	No coincide
387	A	T	0	0	0	0	1	0	1	1	No coincide
396	G	A	0	0	0	0	1	0	1	1	No coincide
405	A	T	0	0	0	0	1	0	1	1	No coincide
411	G	A	0	0	0	0	1	0	1	1	No coincide
423	T	A	0	0	0	0	1	0	1	1	No coincide
447	T	C	0	0	0	0	0	0	1	1	No coincide
450	T	G	0	0	0	0	1	0	1	1	No coincide

463	A	T	0	0	0	0	1	0	1	1	No coincide
465	A	T	0	0	0	0	1	0	1	1	No coincide
480	C	T	0	0	0	0	1	0	1	1	No coincide
481	C	T	0	0	0	0	1	0	1	1	No coincide
483	T	G	0	0	0	0	1	0	1	1	No coincide
501	G	A	0	0	0	0	1	0	1	1	No coincide
516	G	C	0	0	0	0	1	0	1	1	No coincide
531	T	A	0	0	0	0	1	0	1	1	No coincide
552	G	A	0	0	0	0	1	0	1	1	No coincide
555	C	T	0	0	0	0	1	0	1	1	No coincide
576	C	T	0	0	0	0	1	0	1	1	No coincide
585	G	T	0	0	0	0	1	0	1	1	No coincide
624	T	C	0	0	0	0	1	0	0	0	No coincide
627	T	C	0	0	0	0	1	0	1	1	No coincide
645	G	A	0	0	0	0	1	0	1	1	No coincide
651	T	C	0	0	0	0	1	0	0	0	No coincide
675	T	A	0	0	0	0	1	0	1	1	No coincide
681	T	C	0	0	0	0	1	0	1	1	No coincide
684	T	C	0	0	0	0	0	1	0	0	No coincide
687	T	A	0	0	0	0	1	0	1	1	No coincide
702	G	T	0	0	0	0	1	1	0	0	No coincide
714	T	C	0	0	0	0	0	0	1	1	No coincide
733	C	T	0	0	0	0	0	0	1	1	No coincide
747	G	A	0	0	0	0	0	0	1	1	No coincide
750	T	C	0	0	0	0	0	0	1	1	No coincide
762	C	T	0	0	0	0	0	0	1	1	No coincide
786	A	G	0	0	0	0	1	0	1	1	No coincide
793	T	A	0	0	0	0	1	0	1	1	No coincide
805	A	T	0	0	0	0	1	0	0	0	No coincide
846	C	T	0	0	0	0	1	0	1	1	No coincide
850	T	C	0	0	0	0	0	0	1	1	No coincide
852	G	A	0	0	0	0	1	0	1	1	No coincide
855	T	C	0	0	0	0	0	1	0	0	No coincide
882	C	T	0	0	0	0	1	0	0	0	No coincide
902	C	T	0	0	0	0	0	1	0	0	No coincide
903	A	G	0	0	0	0	1	0	0	0	No coincide

Nota: En la tabla se muestra la posición nucleotídica. La columna "REF" indica la base nucleotídica en la secuencia de hla del aislado AA, usado como referencia. La columna "ALT" indica 1-3 sustituciones que

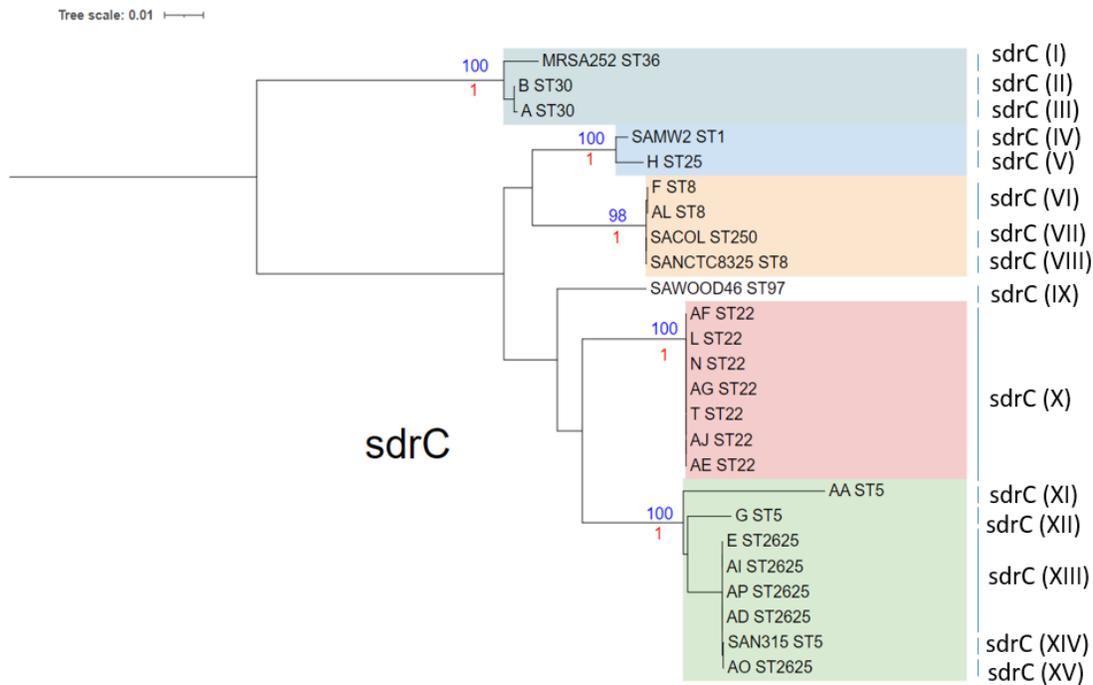
poseen las secuencias de los aislados en estudio para esa posición. En la tabla los valores 1-3 se corresponden con dichas sustituciones.

A diferencia de los demás genes de estudio, *fhuD* fue el único que no se encontró en todos los aislados. La prevalencia de este gen en los aislados de estudio y los genomas de referencia fue del (11/20). Además, *fhuD* fue el gen con menor variación genética reportada. Se observó únicamente dos mutaciones en la secuencia nucleotídica. Las mutaciones fueron C252T para los aislados AE, AF, AG, AJ, L, N y T y T811C para el aislado J. La última mutación señalada es no sinónima y produce la sustitución H271Y. No se generó un árbol filogenético dado la escasa variabilidad genética. No se reportó la búsqueda de un patrón para explicar la expresión génica porque no se dispusieron datos de expresión para dicho gen para los aislados de estudio.

Se reportó una alta diversidad en el gen *sdrC* encontrándose hasta 511 SNPs de los aislados con respecto al genoma de referencia. En la Figura 27 se muestra el árbol filogenético construido para dicho gen en conjunto con la clasificación por los 15 genotipos observados. El árbol filogenético reveló una topología similar a la clasificación por MLST y cg-SNPs. Se encontró un único genotipo por ST para el 60% de STs identificados en el presente estudio (ST36, ST1, ST25, ST250, ST97, ST22). Por otro lado, ningún genotipo identificado compartió STs diferentes. No se analizó la variabilidad genética, dado que no existen datos de expresión para *sdrC*.

Figura 27

Árbol Filogenético de máxima verosimilitud y bayesiano para *sdrC*.



Nota. Valores de soporte UFBoot >90% (ubicados sobre los nodos con color azul) por el software iQ-tree 2: Repeticiones de UFBootstraps=1000, modelo de sustitución GTR+F+G4. Valores de soporte posterior Probabilities >0.90 (ubicados bajo los nodos con color rojo), calculados con el software Mr. Bayes: modelo de sustitución GTR+F+G4, longitud de la cadena MCMC 1000000, frecuencia de muestreo 1000, burn-in 10%. La escala indica la distancia genética calculada. Imagen del árbol filogenético obtenida y editada con iTOL V5.

Capítulo V: Discusión

Durante los últimos años la secuenciación de nueva generación se ha convertido en un análisis estándar en bacteriología. Por ende, diversos estudios han resaltado la necesidad de una extracción de ADN óptima para este proceso (Helmersen & Aamot, 2020). Illumina define como apropiados los ratios de calidad 260/280 entre: 1,8-2 y 260/230 entre: 2-2,2. En el presente estudio, se reportó un valor promedio 260/230 de 2.19, que entra en el rango requerido por Illumina. Por otro lado, el valor promedio 260/280 fue de 2,1 lo que sugiere la presencia de contaminantes. Entre ellos, se ha reportado que el ARN es capaz de aumentar los valores de dicho parámetro (Sanchez et al., 2015).

El protocolo realizado buscó disminuir el impacto del ARN mediante dos enfoques, el primero a través del uso de cloruro de litio, que es conocido por precipitar selectivamente el ARN (Atashpaz et al., 2010). El segundo consistió en emplear RNAsas. La ausencia de ARN detectable en el gel de electroforesis mostrado en la Figura 15 representa la efectividad de estos tratamientos. Sin embargo, no se descarta la presencia de fragmentos de ARN degradado no detectables en el gel, los cuales también son capaces de absorber a 260 nm (Nybo, 2012).

Es bien conocido que extraer ADN de microorganismos Gram positivos presenta un grado de dificultad elevado. Lo anterior debido a que a diferencia de los Gram negativos, las propiedades de la pared celular de los primeros hacen que sea más difícil romperla y exista mayor probabilidad de contaminación (Vingataramin & Frost, 2015). A pesar de esto, el protocolo llevado a cabo presentó mejores ratios de calidad que los obtenidos en un estudio previo que compara diversos métodos de extracción de ADN en SA: Axyprep®: 260/289=1.36, Boiling method; 260/280 = 1.33, columna de sílica: 260/280= 0,18 (Dibbern et al., 2015).

El método de cuantificación de ADN por excelencia recomendado por la guía de Illumina DNA Prep es el fluorométrico. La medición por NanoDrop es una técnica de espectrometría que mide la capacidad de una sustancia química de absorber luz. La cantidad de luz absorbida o transmitida se relaciona con la concentración de una sustancia. Por otro lado, Quantus emplea un método fluorométrico que se basa en el uso de un fluoróforo, el cual interactúa específicamente con una doble cadena de ADN y por ende, la estimación no es afectada por la presencia de otros contaminantes (Khetan et al., 2019). Este hecho explicaría la diferencia de concentración de ADN obtenida en los dos métodos expuesto en la Tabla 9. A esto cabe añadir, la posibilidad de degradación del material genético desde su almacenamiento hasta la preparación de librerías. Teniendo en cuenta estos detalles acerca de la concentración obtenida, se estableció que el método de extracción de ADN aquí desarrollado arrojó un valor promedio suficiente (29,4 ng/ul, basado en el método fluorométrico) para realizar la preparación de librerías de Illumina DNA Prep, la cual requiere inputs tan bajos como 1 ng según las especificaciones del fabricante (Illumina, 2020).

La secuenciación Illumina arrojó un promedio de 1,76 millones de lecturas PE para cada aislado. Lo anterior está acorde a un estudio previo que obtuvo 1,65 millones de lecturas PE siguiendo el protocolo de Illumina Nextera XT DNA library Kit y posterior secuenciación en Illumina Miseq (Mannala et al., 2020). El aislado AE generó 0,8 millones de lecturas, lo que es la mitad del promedio generado para todos los aislados. Hay varias razones por las cuales se puede atribuir este resultado. Una de ellas es la presencia de contaminantes como el EDTA, el cual puede inhibir la actividad de las transposasas. Lo anterior genera una pobre tagmentación que resulta en fragmentos de librería muy largos, los cuales no se adhieren a la flow cell de forma adecuada (Illumina, 2020).

Un estudio que evaluó la calidad de genomas completos de SA secuenciados con tecnología Illumina reveló que el parámetro “Mean per-base Quality” se ha mantenido en un valor de Phred Score

Quality > 35 desde el 2012. El presente estudio concuerda con estos resultados, pues se observa abundantes lecturas que confluyen incluso a un valor de Phred Score Quality = 38 (Petit III & Read, 2018). Por otro lado, es bien conocido que la calidad del llamado de bases tiende a bajar conforme se acerca al final de la longitud de la lectura. Lo anterior se puede apreciar en la Figura 18a y se debe al decaimiento de la señal del fluoróforo (Pfeifer, SP, 2017) .

A pesar de la buena calidad de las secuencias, se decidió realizar un recorte utilizando un enfoque de “ventana deslizante” que detecta si la calidad disminuye por debajo de un umbral cada cierto número de bases. La literatura reporta distintos umbrales para realizar este recorte: Phred Quality Score = 15 (Durand et al., 2018), Phred Quality Score = 20 (Cave et al., 2019), Phred Quality Score = 30 (Capra et al., 2017). Se fijó un Phred Score = 20, dado que se ha reportado que parámetros de recorte muy rigurosos (Phred Quality Score=30) pueden reducir la profundidad de cobertura produciendo lecturas más cortas que serán difíciles de alinear (Bush, 2020).

La Figura 19 mostró que aproximadamente las 10 primeras bases nucleotídicas de cada lectura poseían un contenido de A-T elevados en relación a las otras bases nucleotídicas. Este fenómeno ya ha sido reportado antes y se atribuye a que la transposasa que se utiliza en la tagmentación tiene preferencia para cortar los sitios que poseen A-T (Gunasekera et al., 2021). Particularmente, la transposasa tiene preferencia por un motivo “5'-GTNYWRNAC”. En 2013 una investigación no encontró que esto afecte a los procesos posteriores (Lamble et al., 2013), mientras que otro estudio se contrapuso sugiriendo que los genomas donde exista baja frecuencia del motivo antes mencionado podrían ser sujetos de una tagmentación poco eficiente (Tyler & Christianson, Sara, 2015). En general, si esto afecta a los resultados posteriores aún es tema de debate.

En un experimento de secuenciación las lecturas duplicadas provienen de 3 orígenes principales. Los duplicados naturales son secuencias distintas, pero con gran porcentaje de identidad que pueden

ser percibidas como duplicadas. Los duplicados de PCR se generan debido al paso de la amplificación durante la preparación de la librería. Los duplicados ópticos se producen cuando la señal de un clúster es identificada erróneamente como dos clústeres diferentes (Zhou & Rokas, 2014). Las secuencias duplicadas ópticas y de PCR fueron removidas debido a que su presencia puede afectar los ensamblajes *de novo*. Esto dado que las secuencias duplicadas producen contigs solapados que dan origen a falsos negativos (Gaia et al., 2019). La Figura 20 mostró que en las lecturas preprocesadas no se consiguió eliminar todas las lecturas duplicadas. Esto puede atribuirse a que algunos software pueden subestimar el número de duplicados (Bansal, 2017). De hecho, al software se le dispuso que remueva las lecturas duplicadas que cuenten con 100% de identidad. Por ende, es posible que se hayan escapado lecturas duplicadas que por errores de llamado de bases disminuyan este porcentaje. Por otro lado, el umbral fijado del 100% identidad ayudaría a disminuir la probabilidad de eliminar duplicados naturales.

El porcentaje de GC promedio reportado después del preprocesamiento de las secuencias fue 32,38% que dista en menor medida del porcentaje de GC esperado para SA (~33%). Esta distribución puede ser propiciada por los aislados AK, S, X, Z que emitieron una alerta para este parámetro de calidad. El aislado Y no presentó una alerta en el panel de control, lo que sugirió que el %GC se ajustó a lo esperado según el modelo. Sin embargo, este aislado presentó una divergencia más notoria del resto con un porcentaje GC del 35%. Es bien conocido que las variaciones en este parámetro pueden atribuirse a contaminación (Pfeifer, SP, 2017). Por ende, no es sorpresa verificar que los aislados antes mencionados son los que poseen un mayor grado de contaminación según el software Kraken: Y (58,92%), AK (4,51%), Z(3,03%), S (2,4%) y X (1,58%). La contaminación con lecturas correspondientes a sistemas biológicos ajenos al estudio está bien reportada. De hecho, incluso la base de datos más curada, Refseq posee al menos 1395 genomas con una contaminación significativa (>5%) (Lupo et al.,

2021). La contaminación puede ocurrir en múltiples escenarios. En la mayoría de los casos podrían ser técnicas de cultivo axénico insuficientes (Jeong, Haeyoung & Pan, Jae.Goo, 2016).

En la guía de Illumina para realizar ensamblajes de novo, se establece que una profundidad de cobertura de 50X permite obtener un parámetro de calidad de ensamblaje N50 adecuado y que profundidades más elevadas no contribuyen mayormente a mejorar este parámetro. Entonces, la profundidad de cobertura antes emocionada sería suficiente para realizar la mayoría de tipos de análisis (Illumina, 2010). Por otro lado, diversos estudios han comprobado que para detectar genes de resistencia y virulencia con una confianza superior al 90% se necesita tan solo una profundidad de 15X. Se ha reportado que la cobertura afecta al llamado de variantes (Andreu-Sánchez et al., 2021). Sin embargo, se sabe que a partir de una profundidad de 20X los errores que quizá pudo introducir la polimerasa en el proceso de preparación de librerías se vuelven insignificantes (Bush et al., 2020). En el presente estudio se reportó que en promedio hasta el 84,28% del genoma de referencia fue mapeado con una profundidad de cobertura > 50X. En relación a los valores de profundidad reportados en la literatura, la profundidad encontrada en esta investigación otorga un grado de confianza significativo en la información biológica que derive de estas secuencias.

Los genomas altamente fragmentados representan un problema pues es complicado buscar las secuencias de genes individuales, que pudiesen estar divididos entre distintos contigs (Smits, 2019). Los parámetros que dificultan la construcción de un genoma de novo son las secuencias repetidas, las secuencias cortas, los altos errores de secuenciación (T.-W. Chen et al., 2015). Varios de estos parámetros pudieron contribuir en menor medida a que varios ensamblajes se encontraron altamente fragmentados (#contigs > 200, I: 2069, S: 216, X: 1293 y Z :2090). Curiosamente, los mismos aislados que fueron catalogados como los de mayor presencia de contaminación según Kraken, son los que se corresponden con un ensamblaje altamente fragmentado. Recientemente, un estudio encontró que

todos los ensamblajes que presentaron una fragmentación heterogénea estaban caracterizados por un cierto grado de contaminación. Así, dicha investigación concluyó que la contaminación es uno de las mayores causas de falla en los parámetros de calidad del ensamblaje (Jeong, Haeyoung & Pan, Jae.Goo, 2016). Fuera de este grupo de aislados con esta particularidad, en general se obtuvo genomas de alta calidad con valores N50 que se encuentran en el rango reportado por estudios previos $\approx 122,702$ bp a $806,706$ bp (Bernardy et al., 2019).

MLST es una técnica ampliamente usada para estudiar la diversidad genética de SA. A pesar de que la Electroforesis en gel de campo pulsado otorga una mayor resolución, MLST tiene la ventaja de ser más accesible y reproducible entre laboratorios (Lamers et al., 2011). En Ecuador, una investigación reportó previamente la prevalencia de SA ST22, ST5, ST8, ST30, ST25, ST109 y la relación que estos pudiesen tener con clones epidémicos de relevancia (Zurita et al., 2016). Sin embargo, no se ha reportado hasta la fecha SA ST2625, que en el presente estudio fue el segundo ST más encontrado. En la literatura se encuentra que MRSA ST2625 SCCmec tipo IVa fue responsable de un brote en una Unidad de Cuidado Intensivos de Pediatría en Italia (Ugolotti, E. et al., 2017). La caracterización *in silico* de SCCmec reveló que todos los aislados ST2625 presentaron un SCCmec tipo IVa, por lo que posiblemente están relacionados con los aislados reportados previamente.

Los genomas de referencia se agruparon en distintos nodos del árbol filogenético basado en el MLST. MRSA 252 ST36 se agrupó con los aislados ST30. Lo anterior tiene sentido ya que estudios previos han reportado que MRSA ST36 es un descendiente del clon exitoso MSSA ST30 que ha adquirido el SCCmec tipo IIa (Enright et al., 2002). Así mismo, SA COL ST 250 se agrupó junto con los aislados asignados como ST8. Se ha reportado que ST 250 proviene de ST8 y la diferencia está en una mutación puntual en el alelo *yqiL* (Enright et al., 2002). Finalmente, SAN 315 ST5 y SA NCTC 8235 ST8 se agruparon de manera correcta indicando su similitud con los aislados ST8 identificados en este estudio.

Está bien descrito que MLST es una técnica útil cuando se requiere resolución a gran escala. Sin embargo, si se requiere estudiar la transmisión paciente a paciente MLST es una herramienta que carece de la resolución necesaria (Lamers et al., 2011). Esto fue evidente en la presente investigación, cuando se obtuvo una mayor resolución para clasificar a los aislados utilizando la técnica de core genome-SNPs. A diferencia de MLST que usa solo 7 genes, core-genome-SNPs permite capturar la información contenida en >2000 genes. Debido a esto utilizando la técnica antes mencionada, es posible discriminar entre posibles eventos de transmisión. En la investigación aquí desarrollada se fijó un umbral de 15 SNPs de diferencia en el core genome de distintos aislados para proponer un posible evento de transmisión. Sin embargo, este valor cut-off aún es tema de debate, habiendo diversos estudios que han propuesto distintos valores umbrales SNPs (Ugolotti, E. et al., 2017). También se ha de tener en cuenta las limitaciones de un valor cut-off de SNPs. Por ejemplo, la probabilidad de una transmisión directa no puede ser inferida únicamente conociendo dicho valor. Aún es necesario conocer información epidemiológica clave como visitas al mismo hospital o unidad, personal de salud que estuvo a cargo, etc. Además, el valor cut-off de SNPs no puede ser usado para determinar la direccionalidad de la transmisión (Coll et al., 2020).

Se ha reconocido que *fnbA* es un factor de virulencia de rápida evolución en comparación con los genes housekeeping usados en el MLST (Lamers et al., 2011). Esto se corrobora con el árbol filogenético presentado en la Figura 23. Aquí se observó incluso hasta dos genotipos de *fnbA* dentro de un ST determinado, tal es el caso de ST8, ST30 y ST22. Además, se observó una topología muy similar de árbol a la establecida por MLST, lo que sugiere que la variabilidad de *fnbA* se ajusta a la variabilidad de su contexto genético. Esta afirmación concuerda con un estudio previo en donde se propone que tal como los genes housekeeping de MLST, ciertos factores de virulencia como *fnbA* siguen una evolución clonal, con evidencia mínima de recombinación (Kuhn et al., 2006).

Existen varios métodos de detección rápida de SA basados en evidenciar mediante aglutinación en látex la interacción de la proteína bacteriana fnbA con el fibrinógeno adherido al soporte. Se ha reportado que estas técnicas poseen un porcentaje importante de falsos positivos (Stutz et al., 2011). Lo anterior se explica porque para que esta prueba funcione de forma adecuada, la proteína fnbA de la bacteria debe poseer una afinidad óptima por el fibrinógeno. Sin embargo, esto no se cumple cuando están presentes determinadas variaciones genéticas en la secuencia de ADN de fnbA. Por ejemplo, un estudio previo reportó que las sustituciones GCA-TTA y ACG-GCA en las posiciones 1246-1248 y 1252-1254 pb serían capaces de disminuir considerablemente esta afinidad (Stutz et al., 2011). En el presente estudio, únicamente el genoma de referencia SA WOOD 47 mostró esta sustitución. A pesar de que la frecuencia de esta variante en esta investigación sea baja, es una consideración que debe tomarse en cuenta para la detección de SA de forma apropiada.

En fnbA también se destacaron las sustituciones CAT-CAA (H782Q) y CAT-CAA (K786N), las cuales estuvieron presentes en los aislados AE, AF, AG, AJ, L, N, T y en el genoma de referencia SA MW2. Estas variantes se reportaron previamente en colonizadores nasales y en pacientes que poseían bacteriemia e infección asociada a un dispositivo cardíaco. Sin embargo, la prevalencia fue significativamente mayor en los segundos mencionados. Estas sustituciones provocan una unión por puente de hidrógeno extra de fnbA al fibrinógeno. Al incrementar la afinidad por esta molécula, a SA se le facilita adherirse a las células del huésped. La variedad de ensayos realizados permitió a los investigadores hipotetizar que los colonizadores nasales que poseen SA, cuyas secuencias de fnbA carezcan de estas SNPs probablemente tendrán una mejor esperanza frente a la bacteriemia y la infección asociada a dispositivos cardíacos (Lower et al., 2011). Entonces, los polimorfismos reportados para los aislados antes mencionados podrían ser un blanco útil para predecir el curso de infección de un paciente.

Existe poca información acerca del efecto de las variantes genéticas de sstD en su expresión génica o en su posible función como marcador de diagnóstico y tratamiento. Solo recientemente se reportó una mutación A947G en la secuencia genética de sstD. Se hipotetizó que esta variante genética podría estar relacionada con el nivel de expresión y la disminución de virulencia observada. Sin embargo, no se aportó información concluyente. Cabe recalcar, que ninguno de los aislados en el presente estudio porta dicha variante genética, por lo cual no se puede dar soporte a la hipótesis presentada por el estudio previo (H. Chen et al., 2018).

Previamente, se ha reportado que la inmunización con sstD no limitó la virulencia observada sobre riñones e hígado en un modelo murino. Sin embargo, evitó en gran medida la colonización del corazón. Por este motivo, sstD podría ser considerado para el desarrollo de vacunas multivalentes que pretendan contribuir a disminuir la probabilidad de miocarditis asociada a SA (Beasley et al., 2011). sstD fue el tercer gen con menor diversidad de los analizados en el presente estudio, la cantidad de variaciones fue significativamente diferente a fnbA y sdrC. La variación genética aquí reportada permitirá a los futuros desarrollos de vacunas y tratamientos tener en cuenta dicha variabilidad.

Varias investigaciones han propuesto que hla evoluciona en conjunto con su background genético (Tavares et al., 2014; Xiao et al., 2016). El árbol filogenético para hIA otorgó valores de soporte óptimos para tres grupos ST30-ST36, ST8-ST250 y ST22. Este agrupamiento concuerda con lo registrado en el árbol de MLST y cg-SNPs. Con lo cual, se verifica dicha tendencia y los resultados de este estudio se ajustan con lo antes reportado. Por otro lado, no se consiguió valores de soporte adecuados para ST2625, ST5, ST109, ST1, ST97. Esto podría indicar que hla no evoluciona con su contexto genético para dichos STs. Sin embargo, también puede deberse a un limitado número de muestras correspondientes a cada ST mencionado.

Los aislados A, B y el genoma de referencia MRSA 252 presentaron una sustitución C259T en la secuencia nucleotídica de hla, lo que genera un codón de parada prematuro. Se ha reportado que la presencia de este SNP otorga una reducción significativa de la virulencia y letalidad en un modelo animal(DeLeo et al., 2011). Haciendo un paralelismo, el polipéptido de hla con una sustitución H35L ha sido utilizado como un inmunógeno dado que se ha comprobado la reducción drástica de virulencia para el huésped(Kennedy et al., 2010). Así mismo, el polipéptido cuya secuencia genética tiene una sustitución C259T encontrado en los aislados A, B y MRSA 252 puede ser considerado como un candidato a inmunógeno dada la reducida virulencia en modelos animales.

A diferencia de los demás genes, el presente estudio reportó una baja prevalencia de fhuD1 (11/20). Esto concuerda con varios reportes, por ejemplo, una búsqueda mediante BLASTn de más de 4000 genomas de SA publicados reveló una prevalencia de apenas 35%. Por otro lado, dicha investigación reportó que fhuD2 estuvo presente en todos los genomas analizados (Bacconi et al., 2017). Además, el presente estudio, se contrapone a lo encontrado por (Noboa, Kirsty, 2020) quien reportó que obtuvo amplificación de fhuD1 para todos los aislados aquí investigados. Esta diferencia se debe a que en la investigación antes mencionada se utilizó primers reportados erróneamente por (Jenkins et al., 2015). Mediante la herramienta BLAST se identificó que dichos primers amplifican el gen *FecIII* mas no *fhuD*. Debido a lo anterior, no se poseen datos de expresión génica de *fhuD* para los aislados de interés en esta investigación.

Se sabe que los aminoácidos W197, R199 y W279 son esenciales para la unión del ligando en *fhuD2*. Estos aminoácidos están conservados en *fhuD1* por lo que se espera un rol similar. En el presente estudio, la sustitución aminoacídica H271Y no se corresponde con estos residuos esenciales. Se esperaría entonces, que no exista una alteración significativa de la función de *fhuD1* (Mariotti et al., 2013). La escasa prevalencia de *fhuD1*, su afinidad reducida por el sustrato en comparación con *fhuD2* y

el escaso conocimiento de los sustratos a los cuales es afín fhuD, son razones suficientes para proponer a fhuD2 como un mejor candidato a la fecha para elaborar vacunas. De hecho, fhuD2 ha sido blanco de inmunización en modelos murinos obteniendo resultados prometedores. Entre ellos se destaca una protección inmune en modelos de sepsis y abscesos (Marchetti et al., 2020).

Un estudio previo encontró diferencias en los genotipos de sdrC al comparar distintos linajes (Capra et al., 2017). Aquí se reportó, un genotipo de sdrC para cada ST encontrado, en el 60% de los casos. Se ha propuesto previamente que las adhesinas tipo sdrC sigue una evolución clonal. Sin embargo, otros estudios han encontrado un patrón discordante entre la topología del árbol MLST y la brindada por el gen sdrD. En la presente investigación, la clusterización obtenida a partir del análisis filogenético de sdrC se asemeja con el propiciado por MLST y cg-SNPs. Entonces, aquí se presenta evidencia para apoyar la evolución de sdrC de acuerdo a su contexto genético.

En esta investigación se reportó una alta tasa de variación en las zonas de repeticiones de Serina-Aspartato (SD) de sdrC. Esto tiene implicaciones en la patogenicidad pues, a menos repeticiones de SD tenga el aislado habrá menor cantidad de sitios para glicosilar. Es bien conocido que la glicosilación es reconocida por la respuesta del sistema inmune del huésped. Por tanto, las deleciones de las repeticiones SD podrían ser una estrategia de la evasión de la respuesta inmune por parte de SA. Los aislados que presentaron un menor patrón de SD fueron AO, AP y AD (Ajayi et al., 2018).

No se pudo correlacionar la expresión de los genes de virulencia (hla, fnbA, sstD) en estudio con algún polimorfismo. Estos resultados concuerdan con lo obtenido para hla en un estudio previo. Los de dicha investigación sugirieron que la regulación de hla es multifactorial (Tavares et al., 2014). Los estudios que han determinado variabilidad genética y expresión génica de fnbA tampoco han reportado una correlación aparente (Lower et al., 2011; Stutz et al., 2011) De hecho, se conoce que hla y fnbA son regulados por un complejo mecanismo que surge de la interacción de varios componentes genéticos.

Entre ellos, el gen regulador *Agr*, que a su vez es modificado por rutas de regulación que perciben un estímulo externo y redirigen su respuesta. Los estímulos del medio son importantes reguladores de los genes de virulencia, por ejemplo, en el caso de *sdrC* se ha visto que la temperatura fría aumenta hasta 20 veces la transcripción. En este sentido la temperatura fría ayudaría a la expresión de genes involucrados en la adhesión epitelial y por otro lado, la temperatura caliente regularía al alza los genes involucrados en las infecciones sistémicas como bacteriemia (Bien et al., 2011).

Capítulo VI: Conclusiones

Se extrajo el ADN de 26 aislados clínicos de SA obteniendo valores de concentración y calidad que fueron apropiados para realizar la preparación de librerías y la posterior secuenciación NGS Illumina.

Se prepararon librerías de ADN genómico siguiendo el protocolo de Illumina DNA Prep Kit. Los puntos de control establecidos validaron la presencia de librerías y el tamaño esperado de fragmento esperado para este tipo de aplicación.

Se obtuvo 45,8 millones de lecturas a partir de la secuenciación de genoma completo de 26 aislados clínicos de SA. El control de calidad reveló que más del 90% de lecturas poseían un Phred Quality Score superior a 30. Esto ratifica la credibilidad de la tecnología Illumina en cuanto a llamar una base de forma correcta se refiere.

Se evaluó la presencia de contaminantes mediante el software Kraken y se encontró que al menos 5 aislados contenían lecturas asignadas a un taxón distinto que SA. Estos aislados provocaron parámetros de calidad de ensamblaje deficientes. Por tanto, aquí se propone que evaluar la posible contaminación de las lecturas debe ser considerado una etapa fundamental en el procesamiento de datos de NGS.

Se obtuvo 20 ensamblajes de alta calidad con un valor N50 promedio de 234,06 y un número de contigs promedio de 40. Considerando que SA es una bacteria con presencia importante de regiones repetitivas y que ello puede afectar al ensamblaje *de novo*, los valores antes mencionados estuvieron acorde a lo esperado.

Se reportó una cantidad considerable de genes de resistencia a antibióticos y virulencia en los aislados clínicos. Todos ellos compartieron al menos 11 genes de resistencia y 40 genes de virulencia, lo

cual evidentemente se traducirá en una pérdida de efectividad de tratamiento. Esto indica la relevancia de SA en un entorno clínico.

Se estudió las relaciones genéticas de los aislados de SA usando dos métodos, el primero MLST basado en 7 genes housekeeping. El segundo, cg-SNPs basado en 2043 genes del core genoma. Hubo un alto grado de concordancia entre los dos métodos, sin embargo, cg-SNPs ofreció más resolución.

Se reportó por primera vez el ST 2625 SSmec IVa en Ecuador. Se desconoce la fecha de ingreso de este clon a Ecuador que previamente ha sido vinculado a brotes en Italia. Se llama la atención para que se efectúe una vigilancia epidemiológica oportuna basado en tecnología de vanguardia como NGS.

Se propuso al menos 3 posibles eventos de transmisión de SA porque se encontró aislados que poseyeron diferencias menores a 15 SNPs en su genoma core. Es necesario complementar este hallazgo con información epidemiológica relevante para establecer un grado de probabilidad.

La información descrita en este estudio es relevante de sobre manera para el desarrollo de vacunas, pues algunos de los factores aquí estudiados se encuentran en etapas clínicas de investigación. Además, se propuso que las sustituciones CAT-CAA (H782Q) y CAT-CAA en *fnbA* pueden ser un marcador prognosis del curso de la bacteriemia.

No se encontró una correlación entre el patrón de expresión de genes de virulencia regulado al alza o a la baja en *hla*, *fnbA*, *sstD* y las variantes genéticas encontradas. En concordancia con otros autores se propone que la regulación de la expresión génica de dichos factores es multifactorial y no necesariamente debido a alteraciones en su secuencia.

Capítulo VII: Recomendaciones

Se recomienda disminuir el tiempo entre la extracción de ADN y la preparación de librerías, dado que el transporte de muestras y su almacenamiento en distintas condiciones pueden promover degradación.

Se recomienda secuenciar más genomas completos de SA, específicamente aquellos que se tenga conocimiento *a priori* de su ST para fortalecer el tamaño muestral del presente estudio.

Se recomienda intentar recortar las lecturas contaminadas de los aislados que fueron descartados en este estudio y evaluar la mejora de los parámetros de ensamblaje. Lo anterior se puede realizar mediante técnicas bioinformáticas.

Se recomienda evaluar mediante simulaciones de dinámica molecular el efecto de ciertas variantes encontradas en los factores de virulencia y su afinidad por sus respectivos ligandos.

Se recomienda estudiar los elementos genéticos móviles presentes en los genomas de los aislados del presente estudio y su asociación con determinantes de resistencia y virulencia.

Bibliografía

Abdelbary, M. M. H., Basset, P., Blanc, D. S., & Feil, E. J. (2017). The Evolution and Dynamics of Methicillin-Resistant *Staphylococcus aureus*. En *Genetics and Evolution of Infectious Diseases* (pp. 553–572). Elsevier. <https://doi.org/10.1016/B978-0-12-799942-5.00024-X>

Ajayi, C., Åberg, E., Askarian, F., Sollid, J. U. E., Johannessen, M., & Hanssen, A.-M. (2018). Genetic variability in the *sdrD* gene in *Staphylococcus aureus* from healthy nasal carriers. *BMC Microbiology*, *18*(1), 34. <https://doi.org/10.1186/s12866-018-1179-7>

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H.-K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., ... McArthur, A. G. (2019). CARD 2020: Antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, *gkz935*. <https://doi.org/10.1093/nar/gkz935>

Andreu-Sánchez, S., Chen, L., Wang, D., Augustijn, H. E., Zhernakova, A., & Fu, J. (2021). A Benchmark of Genetic Variant Calling Pipelines Using Metagenomic Short-Read Sequencing. *Frontiers in Genetics*, *12*, 648229. <https://doi.org/10.3389/fgene.2021.648229>

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Atashpaz, S., Khani, S., Barzegari, A., Barar, J., Vahed, S. Z., Azarbajani, R., & Omid, Y. (2010). A robust universal method for extraction of genomic DNA from bacterial species. *Microbiology*, *79*(4), 538–542. <https://doi.org/10.1134/S0026261710040168>

Bacconi, M., Haag, A. F., Chiarot, E., Donato, P., Bagnoli, F., Delany, I., & Bensi, G. (2017). *In Vivo* Analysis of *Staphylococcus aureus*-Infected Mice Reveals Differential Temporal and Spatial Expression Patterns of *fhuD2*. *Infection and Immunity*, *85*(10). <https://doi.org/10.1128/IAI.00270-17>

Bag, S., Saha, B., Mehta, O., Anbumani, D., Kumar, N., Dayal, M., Pant, A., Kumar, P., Saxena, S., Allin, K. H., Hansen, T., Arumugam, M., Vestergaard, H., Pedersen, O., Pereira, V., Abraham, P., Tripathi, R., Wadhwa, N., Bhatnagar, S., ... Das, B. (2016). An Improved Method for High Quality Metagenomics DNA Extraction from Human and Environmental Samples. *Scientific Reports*, *6*(1), 26775. <https://doi.org/10.1038/srep26775>

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. <https://doi.org/10.1089/cmb.2012.0021>

Bansal, V. (2017). A computational method for estimating the PCR duplication rate in DNA and RNA-seq experiments. *BMC Bioinformatics*, *18*(S3), 43. <https://doi.org/10.1186/s12859-017-1471-9>

Beasley, F. C., Marolda, C. L., Cheung, J., Buac, S., & Heinrichs, D. E. (2011). *Staphylococcus aureus* Transporters Hts, Sir, and Sst Capture Iron Liberated from Human Transferrin by Staphyloferrin A, Staphyloferrin B, and Catecholamine Stress Hormones, Respectively, and Contribute to Virulence. *Infection and Immunity*, *79*(6), 2345–2355. <https://doi.org/10.1128/IAI.00117-11>

Bernardy, E. E., Petit, R. A., Moller, A. G., Blumenthal, J. A., McAdam, A. J., Priebe, G. P., Chande, A. T., Rishishwar, L., Jordan, I. K., Read, T. D., & Goldberg, J. B. (2019). Whole-Genome Sequences of *Staphylococcus aureus* Isolates from Cystic Fibrosis Lung Infections. *Microbiology Resource Announcements*, *8*(3). <https://doi.org/10.1128/MRA.01564-18>

Bien, J., Sokolova, O., & Bozko, P. (2011). Characterization of Virulence Factors of *Staphylococcus aureus*: Novel Function of Known Virulence Factors That Are Implicated in Activation of Airway Epithelial Proinflammatory Response. *Journal of Pathogens*, *2011*, 1–13. <https://doi.org/10.4061/2011/601905>

Bitrus, A. A., Peter, O. M., Abbas, M. A., & Goni, M. D. (2018). *Staphylococcus aureus*: A Review of Antimicrobial Resistance Mechanisms. *Veterinary Sciences: Research and Reviews*, 4(2).

<https://doi.org/10.17582/journal.vsrr/2018/4.2.43.54>

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>

Bonar, E. A., Bukowski, M., Hydzik, M., Jankowska, U., Kedracka-Krok, S., Groborz, M., Dubin, G., Akkerboom, V., Miedzobrodzki, J., Sabat, A. J., Friedrich, A. W., & Wladyka, B. (2018). Joint Genomic and Proteomic Analysis Identifies Meta-Trait Characteristics of Virulent and Non-virulent *Staphylococcus aureus* Strains. *Frontiers in Cellular and Infection Microbiology*, 8, 313.

<https://doi.org/10.3389/fcimb.2018.00313>

Bruinsma, S., Burgess, J., Schlingman, D., Czyz, A., Morrell, N., Ballenger, C., Meinholz, H., Brady, L., Khanna, A., Freeberg, L., Jackson, R. G., Mathonet, P., Verity, S. C., Slatter, A. F., Golshani, R., Grunenwald, H., Schroth, G. P., & Gormley, N. A. (2018). Bead-linked transposomes enable a normalization-free workflow for NGS library preparation. *BMC Genomics*, 19(1), 722.

<https://doi.org/10.1186/s12864-018-5096-9>

Bush, S. J. (2020). Read trimming has minimal effect on bacterial SNP-- calling accuracy. *Microbial Genomics*, 13.

Bush, S. J., Foster, D., Eyre, D. W., Clark, E. L., De Maio, N., Shaw, L. P., Stoesser, N., Peto, T. E. A., Crook, D. W., & Walker, A. S. (2020). Genomic diversity affects the accuracy of bacterial single-nucleotide polymorphism-calling pipelines. *GigaScience*, 9(2), giaa007.

<https://doi.org/10.1093/gigascience/giaa007>

Bushnell. (s.f). *BBTools software package*. <https://jgi.doe.gov/data-and-tools/bbtools/>

- Capra, E., Cremonesi, P., Pietrelli, A., Puccio, S., Luini, M., Stella, A., & Castiglioni, B. (2017). Genomic and transcriptomic comparison between *Staphylococcus aureus* strains associated with high and low within herd prevalence of intra-mammary infection. *BMC Microbiology*, *17*(1), 21. <https://doi.org/10.1186/s12866-017-0931-8>
- Cave, R., Misra, R., Chen, J., Wang, S., & Mkrtychyan, H. V. (2019). Whole genome sequencing revealed new molecular characteristics in multidrug resistant staphylococci recovered from high frequency touched surfaces in London. *Scientific Reports*, *9*(1), 9637. <https://doi.org/10.1038/s41598-019-45886-6>
- Chen, H., Wang, Q., Yin, Y., Li, S., Niu, D.-K., & Wang, H. (2018). Genotypic variations between wild-type and small colony variant of *Staphylococcus aureus* in prosthetic valve infectious endocarditis: A comparative genomic and transcriptomic analysis. *International Journal of Antimicrobial Agents*, *51*(4), 655–658. <https://doi.org/10.1016/j.ijantimicag.2017.12.006>
- Chen, L. (2004). VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Research*, *33*(Database issue), D325–D328. <https://doi.org/10.1093/nar/gki008>
- Chen, T.-W., Gan, R.-C., Chang, Y.-F., Liao, W.-C., Wu, T. H., Lee, C.-C., Huang, P.-J., Lee, C.-Y., Chen, Y.-Y. M., Chiu, C.-H., & Tang, P. (2015). Is the whole greater than the sum of its parts? De novo assembly strategies for bacterial genomes based on paired-end sequencing. *BMC Genomics*, *16*(1), 648. <https://doi.org/10.1186/s12864-015-1859-8>
- Chua, K. Y. L., Stinear, T. P., & Howden, B. P. (2013). Functional genomics of *Staphylococcus aureus*. *Briefings in Functional Genomics*, *12*(4), 305–315. <https://doi.org/10.1093/bfgp/elt006>
- Clark, D. P., Pazdernik, N. J., & McGehee, M. R. (2019). DNA Sequencing. En *Molecular Biology* (pp. 240–269). Elsevier. <https://doi.org/10.1016/B978-0-12-813288-3.00008-2>

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, *38*(6), 1767–1771. <https://doi.org/10.1093/nar/gkp1137>

Coll, F., Raven, K. E., Knight, G. M., Blane, B., Harrison, E. M., Leek, D., Enoch, D. A., Brown, N. M., Parkhill, J., & Peacock, S. J. (2020). Definition of a genetic relatedness cutoff to exclude recent transmission of methicillin-resistant *Staphylococcus aureus*: A genomic epidemiology analysis. *The Lancet Microbe*, *1*(8), e328–e335. [https://doi.org/10.1016/S2666-5247\(20\)30149-X](https://doi.org/10.1016/S2666-5247(20)30149-X)

Conroy, B. S., Grigg, J. C., Kolesnikov, M., Morales, L. D., & Murphy, M. E. P. (2019). *Staphylococcus aureus* heme and siderophore-iron acquisition pathways. *BioMetals*, *32*(3), 409–424. <https://doi.org/10.1007/s10534-019-00188-2>

Dadashi, M., Hajikhani, B., Darban-Sarokhalil, D., van Belkum, A., & Goudarzi, M. (2020). Mupirocin resistance in *Staphylococcus aureus*: A systematic review and meta-analysis. *Journal of Global Antimicrobial Resistance*, *20*, 238–247. <https://doi.org/10.1016/j.jgar.2019.07.032>

Davis, M. P. A., van Dongen, S., Abreu-Goodger, C., Bartonicek, N., & Enright, A. J. (2013). Kraken: A set of tools for quality control and analysis of high-throughput sequence data. *Methods*, *63*(1), 41–49. <https://doi.org/10.1016/j.ymeth.2013.06.027>

DeLeo, F. R., Kennedy, A. D., Chen, L., Wardenburg, J. B., Kobayashi, S. D., Mathema, B., Braughton, K. R., Whitney, A. R., Villaruz, A. E., Martens, C. A., Porcella, S. F., McGavin, M. J., Otto, M., Musser, J. M., & Kreiswirth, B. N. (2011). Molecular differentiation of historic phage-type 80/81 and contemporary epidemic *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences*, *108*(44), 18091–18096. <https://doi.org/10.1073/pnas.1111084108>

Dibbern, A. G., Botaro, B. G., Viziack, M. P., Silva, L. F. P., & Santos, M. V. (2015). Evaluation of methods of DNA extraction from *Staphylococcus aureus* in milk for use in real-time PCR. *Genetics and Molecular Research, 14*(1), 227–233. <https://doi.org/10.4238/2015.January.16.6>

Durand, G., Javerliat, F., Bes, M., Veyrieras, J.-B., Guigon, G., Mugnier, N., Schicklin, S., Kaneko, G., Santiago-Allexant, E., Bouchiat, C., Martins-Simões, P., Laurent, F., Van Belkum, A., Vandenesch, F., & Tristan, A. (2018). Routine Whole-Genome Sequencing for Outbreak Investigations of *Staphylococcus aureus* in a National Reference Center. *Frontiers in Microbiology, 9*, 511. <https://doi.org/10.3389/fmicb.2018.00511>

El-Ashram, S., Al Nasr, I., & Suo, X. (2016). Nucleic acid protocols: Extraction and optimization. *Biotechnology Reports, 12*, 33–39. <https://doi.org/10.1016/j.btre.2016.10.001>

Enright, M. C., Robinson, D. A., Randle, G., Feil, E. J., Grundmann, H., & Spratt, B. G. (2002). The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proceedings of the National Academy of Sciences, 99*(11), 7687–7692. <https://doi.org/10.1073/pnas.122108599>

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics, 32*(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>

Fang, B., Liu, B., & Sun, B. (2020). Transcriptional regulation of virulence factors Hla and phenol-soluble modulins α by AraC-type regulator Rbf in *Staphylococcus aureus*. *International Journal of Medical Microbiology, 310*(5), 151436. <https://doi.org/10.1016/j.ijmm.2020.151436>

Gaia, A. S. C., de Sá, P. H. C. G., de Oliveira, M. S., & Veras, A. A. de O. (2019). NGSReadsTreatment – A Cuckoo Filter-based Tool for Removing Duplicate Reads in NGS Data. *Scientific Reports, 9*(1), 11681. <https://doi.org/10.1038/s41598-019-48242-w>

Giulieri, S. G., Tong, S. Y. C., & Williamson, D. A. (2020). Using genomics to understand methicillin- and vancomycin-resistant *Staphylococcus aureus* infections. *Microbial Genomics*, 6(1).

<https://doi.org/10.1099/mgen.0.000324>

Gnanamani, A., Hariharan, P., & Paul-Satyaseela, M. (2017). *Staphylococcus aureus*: Overview of Bacteriology, Clinical Diseases, Epidemiology, Antibiotic Resistance and Therapeutic Approach. En S. Enany & L. E. Crotty Alexander (Eds.), *Frontiers in Staphylococcus aureus*. InTech.

<https://doi.org/10.5772/67338>

Gould, I. M., Reilly, J., Bunyan, D., & Walker, A. (2010). Costs of healthcare-associated methicillin-resistant *Staphylococcus aureus* and its control. *Clinical Microbiology and Infection*, 16(12), 1721–1728. <https://doi.org/10.1111/j.1469-0691.2010.03365.x>

Green, M. R., & Sambrook, J. (2017). Precipitation of DNA with Isopropanol. *Cold Spring Harbor Protocols*, 2017(8), pdb.prot093385. <https://doi.org/10.1101/pdb.prot093385>

Gunasekera, S., Abraham, S., Stegger, M., Pang, S., Wang, P., Sahibzada, S., & O’Dea, M. (2021). Evaluating coverage bias in next-generation sequencing of *Escherichia coli*. *PLOS ONE*, 16(6), e0253440. <https://doi.org/10.1371/journal.pone.0253440>

Guo, Y., Song, G., Sun, M., Wang, J., & Wang, Y. (2020). Prevalence and Therapies of Antibiotic-Resistance in *Staphylococcus aureus*. *Frontiers in Cellular and Infection Microbiology*, 10, 107.

<https://doi.org/10.3389/fcimb.2020.00107>

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUASt: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>

Hasan, S. M. Z., Shafie, M. S. B., & Shah, R. M. (2008). *Efficient method for the extraction of genomic DNA from wormwood (Artemisia capillaris)*. 6.

Heikrujam, J., Kishor, R., & Behari Mazumder, P. (2020). The Chemistry Behind Plant DNA Isolation Protocols. En O.-M. Boldura, C. Baltă, & N. Sayed Awwad (Eds.), *Biochemical Analysis Tools—Methods for Bio-Molecules Studies*. IntechOpen. <https://doi.org/10.5772/intechopen.92206>

Helmersen, K., & Aamot, H. V. (2020). DNA extraction of microbial DNA directly from infected tissue: An optimized protocol for use in nanopore sequencing. *Scientific Reports*, *10*(1), 2985. <https://doi.org/10.1038/s41598-020-59957-6>

Hiramatsu, K., Katayama, Y., Matsuo, M., Sasaki, T., Morimoto, Y., Sekiguchi, A., & Baba, T. (2014). Multi-drug-resistant *Staphylococcus aureus* and future chemotherapy. *Journal of Infection and Chemotherapy*, *20*(10), 593–601. <https://doi.org/10.1016/j.jiac.2014.08.001>

Hosseini, M., Pratas, D., & Pinho, A. (2016). A Survey on Data Compression Methods for Biological Sequences. *Information*, *7*(4), 56. <https://doi.org/10.3390/info7040056>

Humphreys, H., & Coleman, D. C. (2019). Contribution of whole-genome sequencing to understanding of the epidemiology and control of methicillin-resistant *Staphylococcus aureus*. *Journal of Hospital Infection*, *102*(2), 189–199. <https://doi.org/10.1016/j.jhin.2019.01.025>

illumina. (2010). *De Novo Assembly Using Illumina Reads*. https://www.illumina.com/Documents/products/technotes/technote_denovo_assembly_ecoli.pdf

Illumina. (2020). *Illumina DNA Prep Reference Guide (1000000025416)*. https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/illumina_prep/illumina-dna-prep-reference-guide-1000000025416-09.pdf

Jenkins, A., Diep, B. A., Mai, T. T., Vo, N. H., Warrenner, P., Suzich, J., Stover, C. K., & Sellman, B. R. (2015). Differential Expression and Roles of *Staphylococcus aureus* Virulence Determinants during Colonization and Disease. *MBio*, *6*(1). <https://doi.org/10.1128/mBio.02272-14>

- Jeong, Haeyoung & Pan, Jae.Goo. (2016). *Contamination as a major factor in poor Illumina assembly of microbial isolate genomes*. [https://doi.org/doi: https://doi.org/10.1101/081885](https://doi.org/10.1101/081885)
- Jolley, K. A., & Maiden, M. C. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*, *11*(1), 595. <https://doi.org/10.1186/1471-2105-11-595>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Kamel, F. H., & Jarjes, S. F. (2015). *Essentials of Bacteriology and Immunology*. 218.
- Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kaya, H., Hasman, H., Larsen, J., Stegger, M., Johannesen, T. B., Allesøe, R. L., Lemvig, C. K., Aarestrup, F. M., Lund, O., & Larsen, A. R. (2018). SCC *mec* Finder, a Web-Based Tool for Typing of Staphylococcal Cassette Chromosome *mec* in *Staphylococcus aureus* Using Whole-Genome Sequence Data. *MSphere*, *3*(1). <https://doi.org/10.1128/mSphere.00612-17>
- Kennedy, A. D., Wardenburg, J. B., Gardner, D. J., Long, D., Whitney, A. R., Braughton, K. R., Schneewind, O., & DeLeo, F. R. (2010). Targeting of Alpha-Hemolysin by Active or Passive Immunization Decreases Severity of USA300 Skin Infection in a Mouse Model. *The Journal of Infectious Diseases*, *202*(7), 1050–1058. <https://doi.org/10.1086/656043>
- Khetan, D., Gupta, N., Chaudhary, R., & Shukla, J. (2019). Comparison of UV spectrometry and fluorometry-based methods for quantification of cell-free DNA in red cell components. *Asian Journal of Transfusion Science*, *13*(2), 95. https://doi.org/10.4103/ajts.AJTS_90_19

Kim, H. J., Choi, Q., Kwon, G. C., & Koo, S. H. (2021). *Molecular epidemiology and virulence factors of methicillin-resistant Staphylococcus aureus isolated from patients with bacteremia*. 7.

Kuhn, G., Francioli, P., & Blanc, D. S. (2006). Evidence for Clonal Evolution among Highly Polymorphic Genes in Methicillin-Resistant *Staphylococcus aureus*. *Journal of Bacteriology*, 188(1), 169–178. <https://doi.org/10.1128/JB.188.1.169-178.2006>

Lalaouna, D., Desgranges, E., Caldelari, I., & Marzi, S. (2018). MS2-Affinity Purification Coupled With RNA Sequencing Approach in the Human Pathogen *Staphylococcus aureus*. En *Methods in Enzymology* (Vol. 612, pp. 393–411). Elsevier. <https://doi.org/10.1016/bs.mie.2018.08.022>

Lamble, S., Batty, E., Attar, M., Buck, D., Bowden, R., Lunter, G., Crook, D., El-Fahmawi, B., & Piazza, P. (2013). Improved workflows for high throughput library preparation using the transposome-based nextera system. *BMC Biotechnology*, 13(1), 104. <https://doi.org/10.1186/1472-6750-13-104>

Lamers, R. P., Stinnett, J. W., Muthukrishnan, G., Parkinson, C. L., & Cole, A. M. (2011). Evolutionary Analyses of *Staphylococcus aureus* Identify Genetic Relationships between Nasal Carriage and Clinical Isolates. *PLoS ONE*, 6(1), e16426. <https://doi.org/10.1371/journal.pone.0016426>

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22), 3276–3278. <https://doi.org/10.1093/bioinformatics/btu531>

Lecocq, T., Vereecken, N. J., Michez, D., Dellicour, S., Lhomme, P., Valterová, I., Rasplus, J.-Y., & Rasmont, P. (2013). Patterns of Genetic and Reproductive Traits Differentiation in Mainland vs. Corsican Populations of Bumblebees. *PLoS ONE*, 8(6), e65642. <https://doi.org/10.1371/journal.pone.0065642>

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296.

<https://doi.org/10.1093/nar/gkab301>

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

<https://doi.org/10.1093/bioinformatics/btp352>

Li, X., Fang, F., Zhao, J., Lou, N., Li, C., Huang, T., & Li, Y. (2018). Molecular characteristics and virulence gene profiles of *Staphylococcus aureus* causing bloodstream infection. *The Brazilian Journal of Infectious Diseases*, *22*(6), 487–494. <https://doi.org/10.1016/j.bjid.2018.12.001>

Lim, S., Lee, D.-H., Kwak, W., Shin, H., Ku, H.-J., Lee, J., Lee, G. E., Kim, H., Choi, S.-H., Ryu, S., & Lee, J.-H. (2015). Comparative Genomic Analysis of *Staphylococcus aureus* FORC_001 and *S. aureus* MRSA252 Reveals the Characteristics of Antibiotic Resistance and Virulence Factors for Human Infection. *Journal of Microbiology and Biotechnology*, *25*(1), 98–108. <https://doi.org/10.4014/jmb.1410.10005>

Lindsay, J. A., & Holden, M. T. G. (2004). *Staphylococcus aureus*: Superbug, super genome? *Trends in Microbiology*, *12*(8), 378–385. <https://doi.org/10.1016/j.tim.2004.06.004>

Lower, S. K., Lamlerthton, S., Casillas-Ituarte, N. N., Lins, R. D., Yongsunthon, R., Taylor, E. S., DiBartola, A. C., Edmonson, C., McIntyre, L. M., Reller, L. B., Que, Y.-A., Ros, R., Lower, B. H., & Fowler, V. G. (2011). Polymorphisms in fibronectin binding protein A of *Staphylococcus aureus* are associated with infection of cardiovascular devices. *Proceedings of the National Academy of Sciences*, *108*(45), 18372–18377. <https://doi.org/10.1073/pnas.1109071108>

Lupo, V., Van Vlierberghe, M., Vanderschuren, H., Kerff, F., Baurain, D., & Cornet, L. (2021). Contamination in Reference Sequence Databases: Time for Divide-and-Rule Tactics. *Frontiers in Microbiology*, *12*, 755101. <https://doi.org/10.3389/fmicb.2021.755101>

Manara, S., Pasolli, E., Dolce, D., Ravenni, N., Campana, S., Armanini, F., Asnicar, F., Mengoni, A., Galli, L., Montagnani, C., Venturini, E., Rota-Stabelli, O., Grandi, G., Taccetti, G., & Segata, N. (2018). Whole-genome epidemiology, characterisation, and phylogenetic reconstruction of *Staphylococcus aureus* strains in a paediatric hospital. *Genome Medicine*, *10*(1), 82. <https://doi.org/10.1186/s13073-018-0593-7>

Mannala, G. K., Rupp, M., Ghosh, H., Hain, T., Domann, E., & Alt, V. (2020). Complete Genome Sequence of *Staphylococcus aureus* EDCC 5398, a Clinical Isolate from Implant-Associated Bone Infection. *Microbiology Resource Announcements*, *9*(43). <https://doi.org/10.1128/MRA.01136-20>

Marchetti, M., De Bei, O., Bettati, S., Campanini, B., Kovachka, S., Gianquinto, E., Spyrakis, F., & Ronda, L. (2020). Iron Metabolism at the Interface between Host and Pathogen: From Nutritional Immunity to Antibacterial Development. *International Journal of Molecular Sciences*, *21*(6), 2145. <https://doi.org/10.3390/ijms21062145>

Mariotti, P., Malito, E., Biancucci, M., Lo Surdo, P., Mishra, R. P. N., Nardi-Dei, V., Savino, S., Nissim, M., Spraggon, G., Grandi, G., Bagnoli, F., & Bottomley, M. J. (2013). Structural and functional characterization of the *Staphylococcus aureus* virulence factor and vaccine candidate FhuD2. *Biochemical Journal*, *449*(3), 683–693. <https://doi.org/10.1042/BJ20121426>

Mason, F. (2013). Updated Italian checklist of Soldier Flies (Diptera, Stratiomyidae). *ZooKeys*, *336*, 61–78. <https://doi.org/10.3897/zookeys.336.6016>

McGinnis, S., & Madden, T. L. (2004). BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, *32*(Web Server), W20–W25. <https://doi.org/10.1093/nar/gkh435>

McKiernan, H. E., & Danielson, P. B. (2017). Molecular Diagnostic Applications in Forensic Science. En *Molecular Diagnostics* (pp. 371–394). Elsevier. <https://doi.org/10.1016/B978-0-12-802971-8.00021-3>

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>

Murray, C. J., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., ... Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *The Lancet*, S0140673621027240. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)

Noboa, Kirsty. (2020). *Estudio de expresión génica de factores de virulencia de aislados clínicos de Staphylococcus aureus causantes de bacteremia* [Trabajo de Titulación, Universidad de las Fuerzas Armadas ESPE]. <http://repositorio.espe.edu.ec/xmlui/bitstream/handle/21000/22528/T-ESPE-043854.pdf?sequence=1&isAllowed=y>

Nybo, K. (2012). DNase-treated RNA. *BioTechniques*, 52(4), 233–234. <https://doi.org/10.2144/000113836>

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2015). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, btv566. <https://doi.org/10.1093/bioinformatics/btv566>

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, A., & Harris, S. R. (2016). SNP-sites: Rapid efficient extraction of SNPs from multi- FASTA alignments. *Microbial Genomics*, *5*.

Panda, B. B., Meher, A. S., & Hazra, R. K. (2019). Comparison between different methods of DNA isolation from dried blood spots for determination of malaria to determine specificity and cost effectiveness. *Journal of Parasitic Diseases*, *43*(3), 337–342. <https://doi.org/10.1007/s12639-019-01136-0>

Petit III, R. A., & Read, T. D. (2018). *Staphylococcus aureus* viewed from the perspective of 40,000+ genomes [Preprint]. PeerJ Preprints. <https://doi.org/10.7287/peerj.preprints.26890v2>

Pfeifer, SP. (2017). *From next-generation resequencing reads to a high-quality variant data set*. 12. <https://doi.org/doi:10.1038/hdy.2016.102>

Pietrocola, G., Nobile, G., Rindi, S., & Speziale, P. (2017). *Staphylococcus aureus* Manipulates Innate Immunity through Own and Host-Expressed Proteases. *Frontiers in Cellular and Infection Microbiology*, *7*, 166. <https://doi.org/10.3389/fcimb.2017.00166>

Pivard, M., Moreau, K., & Vandenesch, F. (2021). *Staphylococcus aureus* Arsenal To Conquer the Lower Respiratory Tract. *MSphere*, *6*(3). <https://doi.org/10.1128/mSphere.00059-21>

Powers, M. E., & Wardenburg, J. B. (2014). Igniting the Fire: *Staphylococcus aureus* Virulence Factors in the Pathogenesis of Sepsis. *PLoS Pathogens*, *10*(2), e1003871. <https://doi.org/10.1371/journal.ppat.1003871>

Price, J. R., Didelot, X., Crook, D. W., Llewelyn, M. J., & Paul, J. (2013). Whole genome sequencing in the prevention and control of *Staphylococcus aureus* infection. *Journal of Hospital Infection*, 83(1), 14–21. <https://doi.org/10.1016/j.jhin.2012.10.003>

Qamar, W., Khan, M. R., & Arafah, A. (2017). Optimization of conditions to extract high quality DNA for PCR analysis from whole blood using SDS-proteinase K method. *Saudi Journal of Biological Sciences*, 24(7), 1465–1469. <https://doi.org/10.1016/j.sjbs.2016.09.016>

Raven, K. E., Blane, B., Leek, D., Churcher, C., Kokko-Gonzales, P., Pugazhendhi, D., Fraser, L., Betley, J., Parkhill, J., & Peacock, S. J. (2019). Methodology for Whole-Genome Sequencing of Methicillin-Resistant *Staphylococcus aureus* Isolates in a Routine Hospital Microbiology Laboratory. *Journal of Clinical Microbiology*, 57(6). <https://doi.org/10.1128/JCM.00180-19>

Ribeiro, M., Monteiro, F. J., & Ferraz, M. P. (2012). Infection of orthopedic implants with emphasis on bacterial adhesion process and techniques used in studying bacterial-material interactions. *Biomatter*, 2(4), 176–194. <https://doi.org/10.4161/biom.22905>

Ruppitsch, W. (2016). Molecular typing of bacteria for epidemiological surveillance and outbreak investigation / Molekulare Typisierung von Bakterien für die epidemiologische Überwachung und Ausbruchsabklärung. *Die Bodenkultur: Journal of Land Management, Food and Environment*, 67(4), 199–224. <https://doi.org/10.1515/boku-2016-0017>

Sanchez, I., Remm, M., Frascuilho, S., Betsou, F., & Mathieson, W. (2015). How Severely Is DNA Quantification Hampered by RNA Co-extraction? *Biopreservation and Biobanking*, 13(5), 320–324. <https://doi.org/10.1089/bio.2015.0008>

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>

Shettigar, K., & Murali, T. S. (2020). Virulence factors and clonal diversity of *Staphylococcus aureus* in colonization and wound infection with emphasis on diabetic foot infection. *European Journal of Clinical Microbiology & Infectious Diseases*, 39(12), 2235–2246. <https://doi.org/10.1007/s10096-020-03984-8>

Silva, V., Capelo, J. L., Igrejas, G., & Poeta, P. (2020). Molecular Epidemiology of *Staphylococcus aureus* Lineages in Wild Animals in Europe: A Review. *Antibiotics*, 9(3), 122. <https://doi.org/10.3390/antibiotics9030122>

Smits, T. H. M. (2019). The importance of genome sequence quality to microbial comparative genomics. *BMC Genomics*, 20(1), 662, s12864-019-6014–6015. <https://doi.org/10.1186/s12864-019-6014-5>

Stutz, K., Stephan, R., & Tasara, T. (2011). SpA, ClfA, and FnbA Genetic Variations Lead to Staphaurex Test-Negative Phenotypes in Bovine Mastitis *Staphylococcus aureus* Isolates. *Journal of Clinical Microbiology*, 49(2), 638–646. <https://doi.org/10.1128/JCM.01148-10>

Tavares, A., Nielsen, J. B., Boye, K., Rohde, S., Paulo, A. C., Westh, H., Schønning, K., de Lencastre, H., & Miragaia, M. (2014). Insights into Alpha-Hemolysin (Hla) Evolution and Expression among *Staphylococcus aureus* Clones with Hospital and Community Origin. *PLoS ONE*, 9(7), e98634. <https://doi.org/10.1371/journal.pone.0098634>

Thomer, L., Schneewind, O., & Missiakas, D. (2016). Pathogenesis of *Staphylococcus aureus* Bloodstream Infections. *Annual Review of Pathology: Mechanisms of Disease*, 11(1), 343–364. <https://doi.org/10.1146/annurev-pathol-012615-044351>

Tong, S. Y. C., Davis, J. S., Eichenberger, E., Holland, T. L., & Fowler, V. G. (2015). *Staphylococcus aureus* Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management. *Clinical Microbiology Reviews*, 28(3), 603–661. <https://doi.org/10.1128/CMR.00134-14>

Turner, N. A., Sharma-Kuinkel, B. K., Maskarinec, S. A., Eichenberger, E. M., Shah, P. P., Carugati, M., Holland, T. L., & Fowler, V. G. (2019). Methicillin-resistant *Staphylococcus aureus*: An overview of basic and clinical research. *Nature Reviews Microbiology*, *17*(4), 203–218.

<https://doi.org/10.1038/s41579-018-0147-4>

Tyler, A., & Christianson, Sara. (2015). *Comparison of Sample Preparation Methods Used for the Next-Generation Sequencing of Mycobacterium tuberculosis*.

<https://doi.org/doi:10.1371/hournal.phone.0148676>

Ugolotti, E., Di Marco, E., & Bandettini, R. (2017). *Genomic characterization of a paediatric MRSA outbreak by next-generation sequencing*. *98*. <https://doi.org/10.1016/j.jhin.2017.08.009>

Uribe-García, A., Paniagua-Contreras, G. L., Monroy-Pérez, E., Bustos-Martínez, J., Hamdan-Partida, A., Garzón, J., Alanís, J., Quezada, R., Vaca-Paniagua, F., & Vaca, S. (2021). Frequency and expression of genes involved in adhesion and biofilm formation in *Staphylococcus aureus* strains isolated from periodontal lesions. *Journal of Microbiology, Immunology and Infection*, *54*(2), 267–275.

<https://doi.org/10.1016/j.jmii.2019.05.010>

Vingataramin, L., & Frost, E. H. (2015). A single protocol for extraction of gDNA from bacteria and yeast. *BioTechniques*, *58*(3), 120–125. <https://doi.org/10.2144/000114263>

Xiao, M., Zhao, R., Zhang, Q., Fan, X., O'Sullivan, M. V. N., Li, D.-F., Wang, X.-Y., Wu, H.-L., Kong, F., & Xu, Y.-C. (2016). Genotypic Diversity of *Staphylococcus aureus* α -Hemolysin Gene (hla) and Its Association with Clonal Background: Implications for Vaccine Development. *PLOS ONE*, *11*(2), e0149112.

<https://doi.org/10.1371/journal.pone.0149112>

Zhang, L., Gao, J., Barkema, H. W., Ali, T., Liu, G., Deng, Y., Naushad, S., Kastelic, J. P., & Han, B. (2018). Virulence gene profiles: Alpha-hemolysin and clonal diversity in *Staphylococcus aureus* isolates

from bovine clinical mastitis in China. *BMC Veterinary Research*, 14(1), 63.

<https://doi.org/10.1186/s12917-018-1374-7>

Zhou, X., & Rokas, A. (2014). Prevention, diagnosis and treatment of high-throughput sequencing data pathologies. *Molecular Ecology*, 23(7), 1679–1700. <https://doi.org/10.1111/mec.12680>

Zurita, J., Barba, P., Ortega-Paredes, D., Mora, M., & Rivadeneira, S. (2016). Local circulating clones of *Staphylococcus aureus* in Ecuador. *The Brazilian Journal of Infectious Diseases*, 20(6), 525–533. <https://doi.org/10.1016/j.bjid.2016.08.006>