

Análisis del Perfil Molecular de *Staphylococcus aureus* en Estado Plantónico y de *Biofilm*
Empleando Espectrometría de Masas MALDI-TOF y Herramientas de *Machine Learning*

Willy Colán Avendaño

Trabajo de Grado para Optar al Título de Químico

Director

Claudia Cristina Ortiz López

Microbióloga, Doctorado en Biocatálisis.

Codirector

Yuly Andrea Prada Vargas

Química, Doctorado en Química.

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Química

Bucaramanga

2023

Dedicatoria

Este triunfo se lo dedico a mi abuela, Borda María G., tío, Avendaño William, y especialmente a mi querida madre, Avendaño Blanca T., con su frase *“las cosas hay que terminarlas, lléguele el agua a donde le llegue”*. Gracias por nunca haber dudado.

Agradecimientos

Aunque hay triunfos personales, no puedo dejar de lado personas incondicionales que me acompañaron durante este camino Marleny Durán Salazar, Jennifer Andrea Ruiz Durán y Domingo Arturo Ruiz Mejía. Mis compañeros de laboratorio que siempre me hicieron ver el lado positivo de las cosas, mis compañeras de carrera y mis compañeros de *ultimate frisbee*. También agradezco a mi mascota Canela, que me acompañó en las jornadas de estudio y escritura del libro.

Desde luego un agradecimiento a mis evaluadores, directora y codirectora por la confianza, empatía y paciencia, durante el desarrollo de este trabajo de grado. Igualmente, al laboratorio de Espectrometría de masas, al Grupo de Investigación en Bioquímica y microbiología por su apoyo. A la Vicerrectoría de Investigación y extensión, por la financiación del proyecto 3740.

Finalmente, a todos aquellos que hicieron parte de este proceso de formación con una abrazo, una sonrisa o cualquier expresión de apoyo que sin saberlo contribuyó a poder culminar mi carrera profesional.

Tabla de Contenido

		Pág.
Introducción		14
1.	Planteamiento del problema.....	17
2.	Justificación	19
3.	Estado del arte.....	20
4.	Marco Teórico.....	22
4.1.	<i>Staphylococcus aureus</i>	22
4.2.	<i>Biofilm</i>	23
4.3.	Espectrometría de masas MALDI-TOF.....	23
4.4.	Procesamiento de datos de espectrometría de masas	25
4.5.	Quimiometría	25
4.5.1.	Análisis por componentes principales (PCA).....	26
4.5.2.	Métricas de evaluación de modelos predictivos de clasificación	27
4.6.	Aprendizaje automático o Machine Learning.....	28
4.6.1.	Aprendizaje supervisado.....	29
4.6.2.	Aprendizaje no supervisado.....	29
4.6.3.	Algoritmos para Machine Learning	29
4.6.4.	Fases de desarrollo.....	32
5.	Objetivos.....	34
5.1.	Objetivo General.....	34
5.2.	Objetivos Específicos.....	34
6.	Materiales y Métodos.....	35
6.1.	Materiales.....	35

6.2.	Métodos.....	36
6.2.1.	Cultivo de cepas de <i>S. aureus</i> en estado planctónico.....	37
6.2.2.	Formación de biofilm de cepas de <i>S. aureus</i>	38
6.2.3.	Extracción de proteínas para la determinación del perfil molecular.....	39
6.2.4.	Espectrometría de masas MALDI-TOF.....	41
6.2.5.	Pre-procesamiento de datos	41
6.2.6.	Procesamiento de datos.....	43
7.	Resultados y análisis.....	45
7.1.	Obtención de muestras planctónicas y <i>biofilm</i> de <i>S. aureus</i>	45
7.2.	Extracción de proteínas.....	46
7.3.	Análisis por Espectrometría de Masas MALDI-TOF.....	48
7.4.	Análisis por componentes principales	51
7.5.	<i>Machine learning</i>	53
8.	Conclusiones.....	59
	Recomendaciones	60
	Bibliografía.....	61

Lista de Tablas

	Pág.
Tabla 1. <i>Características microbiológicas cepas de estudio.</i>	36
Tabla 2. <i>Número de espectros usados por extracción y estado (biofilm, planctónico).</i>	42
Tabla 3. <i>Número total de espectros de masas usados en la construcción del modelo.</i>	43
Tabla 4. <i>Parámetros de los algoritmos.</i>	44
Tabla 5. <i>Concentración de proteína obtenida para las cepas de estudio en estado planctónico y biofilm según el método de extracción.</i>	47
Tabla 6. <i>Métricas de classification_report de los modelos por protocolo de extracción.</i>	54
Tabla 7 <i>Métricas de cross_val_score de los modelos por protocolo de extracción</i>	55
Tabla 8 <i>Comparación de los valores F1-score generados por las funciones classification_report y cross_val_score.</i>	56

Lista de Figuras

	Pág.
Figura 1. <i>Esquema general de las etapas de la formación de biofilm.</i>	23
Figura 2. <i>Esquema del comportamiento de los datos con dos variables y por tanto dos componentes principales.</i>	27
Figura 3. <i>Matriz de confusión. Vp: verdadero positivo; FP: Falso positivo; FN: falso negativo; VN: verdadero negativo.</i>	28
Figura 4. <i>Esquema de los árboles de decisión en Random forest.</i>	30
Figura 5. <i>Esquema de las capas de una red neuronal.</i>	31
Figura 6. <i>Esquema de validación cruzada k-fold.</i>	32
Figura 7. <i>Esquema de la metodología general del proyecto.</i>	37
Figura 8. <i>Metodología de extracción para la obtención del perfil molecular en espectrometría de masas MALDI-TOF.</i>	39
Figura 9. <i>Esquema de ejecución en Machine Learning.</i>	42
Figura 10. <i>Cinética de crecimiento planctónico de las cepas S. aureus ATCC BAA-44, ATCC 43300 y ATCC 29213 en medio MH a 37°C y agitación a 200 rpm.</i>	45
Figura 11. <i>Cinética de formación del biofilm de las cepas de estudio a las 12, 24, 36, 48, 60 y 72 h de cultivo. A) Cuantificación por CV de la biomasa total; B) Cuantificación por conteo de colonias de las células cultivables.</i>	46
Figura 12. <i>Promedio de las intensidades de las señales del perfil molecular de S. aureus ATCC 43300 en estado planctónico y biofilm. A) extracción HCCl3-MetOH, B) extracción Buffer de lisis, C) extracción AF-ACN.</i>	50

Figura 13. <i>Gráficos de PCA de las extracciones por estados planctónico y biofilm. S. aureus ATCC 29213 (Negro), ATCC 43300 (verde) y ATCC BAA- 44 (Rojo)</i>	52
Figura 14. <i>Componentes principales (azul, biofilm y rojo, planctónico) y varianza explicada para cada extracción.</i>	53
Figura 15. <i>Matriz de confusión de los modelos obtenidos para cada una de las extracciones.</i> ..	57

Lista de Apéndices

	pág.
ANEXO A. Cuantificación de proteínas por el método de Bradford	73
ANEXO B. Rutinas en Python.....	74
ANEXO C. Espectros de masa representativos del perfil molecular de <i>S. aureus</i> en estado planctónico y <i>biofilm</i> empleando diferentes protocolos de extracción de proteínas.	75

Resumen

Título: Análisis del Perfil Molecular de *Staphylococcus aureus* en Estado Plantónico y de *Biofilm* Empleando Espectrometría de Masas MALDI-TOF y Herramientas de *Machine Learning* *

Autor: Willy Colán Avendaño **

Palabras Clave: Espectrometría de Masas MALDI-TOF, perfil molecular, *Staphylococcus aureus*, *Machine learning*, *biofilm*.

Resumen: *Staphylococcus aureus* es un patógeno oportunista de alta incidencia en la salud y la seguridad alimentaria debido a su perfil de resistencia y capacidad de formar *biofilm* en diversas superficies. En este trabajo, se llevó a cabo el análisis del perfil molecular *S. aureus* en estado plantónico y de *biofilm* mediante espectrometría de masas MALDI-TOF. Se emplearon algoritmos de *machine learning* supervisado con el objetivo de encontrar diferencias en los perfiles moleculares obtenidos para estos estados. El perfil molecular hace referencia a las señales representativas del espectro de masas de *S. aureus* en un rango de 2000 *m/z* a 20000 *m/z*. Se utilizaron tres cepas de referencia de *S. aureus* (ATCC 29213, ATCC 43300 y ATCC BAA-44) en estado planctónico (libre flotación) y en estado sésil (formación de *biofilm*). Todas las muestras se trataron con cuatro protocolos de extracción de proteínas: i) Acido fórmico y acetonitrilo, ii) BugBuster®, iii) Cloroformo - Metanol y iv) buffer de lisis. Los espectros de masas adquiridos se ajustaron y normalizaron para realizar un análisis por componentes principales. Este análisis evidenció que la extracción ácido fórmico y acetonitrilo agrupó mejor los datos según su estado (plantónico o *biofilm*) en comparación con las demás extracciones. La evaluación de los modelos

generados por los algoritmos Máquina de Soporte Vectorial, *XGboost*, *Random Forest* y Redes Neuronales Artificiales, de cada extracción mostró que los modelos de los algoritmos Máquina de Soporte Vectorial y Redes Neuronales Artificiales alimentados con datos de la extracción ácido fórmico y acetonitrilo, presentaron valores de *F1-score* de 0.94 y 0.92, respectivamente. Los modelos predictivos generados con los datos provenientes de la extracción AF-ACN y los algoritmos SVC y ANN son los más adecuados para continuar los estudios realizados en este trabajo, debido a que presentaron valores superiores a 0.92 en la métrica *F1-score*. En la matriz de confusión los dos algoritmos presentaron dos falsas identificaciones del estado planctónico y ninguna falsa identificación del estado de *biofilm*. en general, ambas asociadas al estado planctónico. Los resultados de este trabajo evidenciaron la viabilidad de implementar modelos predictivos para la identificación del estado de *biofilm*.

* Trabajo de Grado

** Facultad de Ciencias. Escuela de química. Pregrado en Química. Director: Claudia Cristina Ortiz López. Microbióloga, Doctora en biocatálisis. Codirector: Yuly Andrea Prada Vargas. Química, Doctora en Química

Abstract

Title: Molecular Profile Analysis of *Staphylococcus aureus* in Planktonic and Biofilm States Using MALDI-TOF Mass Spectrometry and Machine Learning Tools *.

Author: Willy Colán Avendaño **

Keywords: MALDI-TOF mass spectrometry, molecular profiling, *Staphylococcus aureus*, Machine learning, biofilm.

Abstract: *Staphylococcus aureus* is an opportunistic pathogen of high incidence in health and food safety due to its resistance profile and ability to form biofilm on various surfaces. In this work, the analysis of the molecular profile of *S. aureus* in planktonic and biofilm states was carried out by MALDI-TOF mass spectrometry. Supervised machine learning algorithms were used to find differences in the molecular profiles obtained for these states. The molecular profile refers to signals representative of the mass spectrum of *S. aureus* in a range from 2000 m/z to 20000 m/z. Three *S. aureus* reference strains (ATCC 29213, ATCC 43300, and ATCC BAA-44) in planktonic (free-floating) and sessile (biofilm-forming) states were used. All samples were treated with four protein extraction protocols: i) Formic acid and acetonitrile, ii) BugBuster®, and iii) Chloroform - Methanol. and iv) lysis buffer. The acquired mass spectra were adjusted and normalized to perform a principal component analysis. This analysis showed that the formic acid and acetonitrile extraction grouped the data better according to their state (planktonic or biofilm) compared to the other extractions. The evaluation of the models generated by the algorithms Support Vector Machine, XGboost, Random Forest, and Artificial Neural Networks of each extraction showed

that the models of the algorithms Support Vector Machine and Artificial Neural Networks fed with data from the formic acid and acetonitrile extraction presented F1-score values of 0.94 and 0.92, respectively. The predictive models generated with the data from the AF-ACN extraction and the SVC and ANN algorithms are the most suitable to continue the studies carried out in this work because they presented values higher than 0.92 in the F1-score metric. In the confusion matrix, the two algorithms presented two false identifications of the planktonic state and no false identification of the biofilm state, both associated with the planktonic state. The results of this work showed the feasibility of implementing predictive models for biofilm state identification.

* Undergraduate Work

** Faculty of Sciences. School of Chemistry. Undergraduate in Chemistry. Director: Claudia Cristina Ortiz López. Microbiologist, PhD in biocatalysis. Co-director: Yuly Andrea Prada Vargas. Chemistry, PhD in Chemistry

Introducción

La espectrometría de masas con ionización/desorción láser asistida por una matriz y un analizador de tiempo de vuelo (MALDI-TOF MS). Consiste en una técnica analítica de ionización suave ampliamente utilizada en la caracterización de macromoléculas como ácidos nucleicos, péptidos y proteínas (Alizadeh et al., 2021; Hajduk et al., 2016; Karas & Hillenkamp, 1988). En esta técnica el analito se co-cristaliza con una matriz que tiene la capacidad de absorber radiación pulsada por un láser, provocando la desorción y ionización por transferencia de iones o cargas. Las moléculas cargadas del analito son separadas en un analizador de tiempo de vuelo y detectadas en función de su relación masa-carga (m/z) (Karas et al., 1985).

La aplicación de la espectrometría de masas MALDI-TOF acompañada del uso de herramientas de análisis computacional y bases de datos han permitido la identificación de microorganismos a nivel de género y especie por medio del perfil molecular característico de cada microorganismo (Schulthess *et al.*, 2014). Estos perfiles son determinados a partir de un conjunto de señales diferenciables, presentes en un rango de masas del espectro del microorganismo de interés (Złoch *et al.*, 2020a). Una vez recopilados los espectros de masas, estos son analizados en *softwares* de tratamiento de datos para ordenar, clasificar, y discriminar la información. Una de las herramientas computacionales utilizadas en este tipo de análisis es el aprendizaje automático o “*Machine Learning*”; esta rama de la inteligencia artificial permite el análisis de datos mediante el uso de algoritmos multivariados que pueden generar modelos predictivos capaces de aprender, discernir y mejorar a medida que procesa dichos datos (Dematheis et al., 2022; H. Y. Wang et al., 2018; Wei et al., 2021).

En los últimos años, varios investigadores han explorado la capacidad de la espectrometría de masas MALDI-TOF en la identificación de microorganismos a partir de perfiles moleculares. A pesar de ser una técnica altamente sensible y resolutive, la preparación de la muestra es un punto

crítico debido a que puede dar lugar a perfiles moleculares diferentes de una misma muestra (Monedeiro et al., 2021; Złoch et al., 2020a). En este sentido, se han estudiado el efecto de las condiciones de crecimiento y los métodos de extracción de macromoléculas, principalmente de proteínas, para la determinación del perfil molecular de un microorganismo (Monedeiro *et al.*, 2021). Otros estudios se han enfocado en la predicción de la resistencia antimicrobiana a partir del perfil molecular mediante herramientas de *Machine learning* (Weis *et al.*, 2022), la diferenciación entre perfiles moleculares de cepas formadoras y no formadoras de *biofilm*, entre otros (Asghari et al., 2021; Caputo et al., 2018a; Rodríguez-Temporal et al., 2023a). Estas investigaciones han demostrado que existe la necesidad de recopilar información sobre microorganismos patógenos altamente resistentes como *Staphylococcus aureus*, el cual está catalogado por la Organización Mundial de la Salud (OMS) como un microorganismo de prioridad elevada para la investigación y desarrollo de nuevos antibióticos (World Health Organization, 2021).

Al igual que otras bacterias, *S. aureus* se caracteriza por formar una macroestructura llamada *biofilm*, en la que los microorganismos se encuentran agrupados en estado sésil y recubiertos por una matriz de exopolisacáridos (S. J. Kim et al., 2018). El *biofilm*, es una estrategia de supervivencia que permite la adaptación del microorganismo a las condiciones desfavorables del medio, aumentar su virulencia y su resistencia frente a antimicrobianos (Grace & Fetsch, 2018; Vani et al., 2023). Por lo anterior, este trabajo se enfocó en el análisis de perfiles moleculares de diferentes cepas de colección de *S. aureus* con la finalidad de diferenciar los estados planctónico y *biofilm*, en el entorno *JupyterLab* con lenguaje de programación *Python*. El Análisis por Componentes Principales (PCA) sobre las metodologías de extracción, determinó que la extracción ácido fórmico y acetonitrilo agrupó mejor los datos según la etiqueta de estado (planctónico y *biofilm*) en comparación con las demás extracciones. Los datos de cada extracción se usaron para la generación de cuatro modelos predictivos usando los algoritmos, Máquina de

soporte Vectorial, *XGboost*, *Random Forest* y Redes Neuronales Artificiales. El puntaje de *F1-score* del modelo predictivo, para la extracción ácido fórmico y acetonitrilo, de cada algoritmo fue de 0.94 (Máquina de soporte Vectorial), 0.60 (*XGboost*), 0.78 (*Random Forest*) y 0.92 (Redes Neuronales Artificiales). En el mismo orden, el puntaje *F1-score* para la extracción cloroformo metanol fue de 0.89, 0.66, 0.85 y 0.85. Todos los modelos de la extracción buffer de lisis tuvieron valores de *F1-score* inferiores a 0.78. Los modelos para extracción ácido fórmico y acetonitrilo fueron de 0.95 ± 0.09 y 0.95 ± 0.05 . En conclusión, los modelos de los algoritmos Máquina de soporte Vectorial y Redes Neuronales Artificiales, construidos con datos de la extracción ácido fórmico y acetonitrilo realizaron las mejores predicciones. Estos resultados contribuye al desarrollo de futuras estrategias para el uso de MALDI-TOF-MS y de *Machine Learning*, como una herramienta potencial y versátil en el diagnóstico de infecciones asociadas al *biofilm* y la investigación de posibles blancos moleculares para el desarrollo de estrategias terapéuticas.

1. Planteamiento del problema

S. aureus es un microorganismo patógeno relacionado con infecciones intrahospitalarias, que puede presentar resistencia a diferentes antibióticos y tiene una alta capacidad de formación de *biofilm*. En el 2017, la Organización Mundial de la Salud (OMS) clasificó a *S. aureus* como un microorganismo prioritario para la investigación y desarrollo de nuevos antimicrobianos, metodologías diagnósticas y vacunas (World Health Organization, 2017). Para el año 2021, la OMS lo catalogó como un patógeno multirresistente de prioridad elevada porque se ha demostrado su resistente frente a diversas generaciones de antibióticos, representando un alto riesgo a la salud pública (World Health Organization, 2021).

Por su parte, el *biofilm* es una macroestructura formada por uno o varios tipos de microorganismos en estado sésil adheridos a una superficie (biótica o abiótica), recubierto(s) por una matriz de exopolisacáridos (Kranjec *et al.*, 2021). El *biofilm* causa infecciones como la fibrosis quística, endocarditis y otras afecciones asociadas al uso de dispositivos médicos; las cuales se caracterizan por ser persistentes, progresivas y por la baja respuesta a tratamientos clínicos convencionales (Hall-Stoodley *et al.*, 2004). La dificultad para realizar un diagnóstico oportuno de los microorganismos en estado *biofilm* con lleva a la ineficacia de los antibióticos administrados, sumado a la baja disposición de una muestra representativa y los extensos periodos de tiempo requeridos en el análisis por cultivo celular (Høiby *et al.*, 2015; Silva *et al.*, 2021). El Instituto Nacional de Salud de Estados Unidos demostró que cerca del 80% de las infecciones microbianas en el cuerpo humano se asocian a la formación de *biofilm* (Cámara *et al.*, 2022). Adicionalmente, la mayoría de los tratamientos antibióticos de uso clínico han sido diseñados y evaluados sobre matrices bacterianas en estado plantónico (Iskandar *et al.*, 2022).

Dada la alta prevalencia de *S. aureus* en estado de *biofilm* y la baja efectividad de los antibióticos convencionales, es necesaria la implementación de métodos eficientes que permitan

la tipificación del microorganismo, y la identificación de factores de virulencia tales como la resistencia a antimicrobianos y la formación de *biofilm*. Estudios basados en espectrometría de masas MALDI-TOF y el uso de herramientas computacionales constituyen una alternativa confiable para este tipo análisis (Silva et al., 2021; Yu et al., 2022a). Un claro ejemplo es el *software Biotyper*, el cual fue desarrollado por la empresa Bruker, para la identificación de microorganismos mediante los perfiles moleculares obtenidos por MS-MALDI-TOF (Lovison et al., 2023; Schulthess et al., 2014). El *software Biotyper* realiza una comparación del espectro de masas de una muestra con la base de datos de patrones únicos de microorganismos de referencia, permitiendo la identificación a nivel de género o especie (Ashfaq et al., 2022; Krásný et al., 2013).

Teniendo en cuenta la necesidad de desarrollar nuevos modelos de análisis de datos, que faciliten el diagnóstico o la identificación de posibles blancos terapéuticos relacionados al *biofilm* de cepas patógenas, se planteó la siguiente pregunta de investigación, ¿Es posible identificar señales diferenciales entre los perfiles moleculares de cepas de *Staphylococcus aureus* en estado platónico y de *biofilm* mediante espectrometría de masas MALDI-TOF y *Machine Learning*?

2. Justificación

En la naturaleza, *S. aureus* se encuentra en estado planctónico y de *biofilm*. El *biofilm* constituye un tipo de macroestructura que le confiere a la bacteria protección frente a condiciones de estrés externo, persistencia y proliferación en medios bióticos y abióticos (Vázquez-Sánchez and Rrodríguez-López, 2018). El *biofilm* de *S. aureus* se ha detectado en dispositivos médicos como: catéteres implantados, válvulas cardíacas, huesos y prótesis articulares, también en productos alimenticios e instalaciones de procesamiento de alimentos (Pietrocola *et al.*, 2022). Sin embargo, no existe un diagnóstico específico y de rutina para determinar la presencia de *biofilm* (Høiby *et al.*, 2015; Silva, Marques and Röder, 2021). Teniendo en cuenta el riesgo que representa *S. aureus* y las propiedades que le confiere la formación de *biofilm*, se planteó el uso de MS-MALDI-TOF y *Machine Learning* para explorar la posibilidad de generar un modelo predictivo que permita diferenciar el perfil molecular de cepas de *S. aureus* en estado planctónico y la formación de *biofilm*.

3. Estado del arte

El análisis de los espectros de masas adquiridos por MALDI-TOF empleando herramientas de *Machine Learning*, se ha aplicado a diferentes áreas de la investigación de patógenos microbianos. Debido al gran impacto que ha tenido el desarrollo de la plataforma BIOTYPER y la venta de equipos de MALDI enfocadas a laboratorios clínicos diagnósticos (como el VITEK), muchas investigaciones de la última década se han enfocado en la construcción de bases de datos y algoritmos que fortalezcan para la identificación de microorganismos a nivel de especie y cepa (Tsuchida, 2018). Estudios como el de Rodríguez-Temporal y colaboradores, publicado recientemente, permiten la identificación rápida de subespecies de *Mycobacterium abscessus* con una exactitud del 96% a partir del espectro proteico obtenido por MALDI-TOF (Rodríguez-Temporal, Herrera, *et al.*, 2023). Este enfoque no solo se limita a las áreas clínicas, otros autores han enfocado sus esfuerzos en la identificación de patógenos de aves o de plantas (Wigmann *et al.*, 2019; Wang *et al.*, 2023).

De igual manera, otra área de aplicación importante es la detección de marcadores moleculares asociados a la resistencia antibiótica (Florio *et al.*, 2020; Weis *et al.*, 2022). Weis y colaboradores desarrollaron una aproximación en *machine learning* para predecir la resistencia a antimicrobianos directamente de datos derivados de espectros de masas de MALDI-TOF de aislados clínicos de *S. aureus*, *K. pneumoniae* y *E. coli* con diferentes perfiles de resistencia (Weis *et al.*, 2022). Este estudio también realizó un análisis retrospectivo, demostrando que en 51 de 63 casos clínicos la aplicación del algoritmo hubiera mejorado la rapidez del diagnóstico y la eficacia de los tratamientos. Estudios similares realizados por diversos autores entre los que se destacan (Wu *et al.*, 2022; Yu *et al.*, 2022a; Zvezdanova *et al.*, 2022; Iskender *et al.*, 2023), han aportado valiosas metodologías para la detección de resistencia en *M. tuberculosis*, *Kebsiella pneumoniae*, *Aspergillus fumigatus* y *S. aureus*, respectivamente.

Por otro lado, se han publicado trabajos relacionados con la caracterización fenotípica de los microorganismos, ya sea con fines de desarrollo biotecnológico o para aplicaciones en salud (Van Oosten and Klein, 2020; Weis *et al.*, 2020; Rodríguez-Temporal, Díez, *et al.*, 2023b). En un estudio desarrollado por Caputo y colaboradores en el año 2018, los autores realizaron una aproximación preliminar utilizando Biotyper para la clasificación de 18 aislados clínicos de *S. epidermidis* en productores o no productores de *biofilm* (Caputo *et al.*, 2018^a). Sin embargo, los autores no presentaron datos del modelo y análisis estadísticos utilizados, y solo realizaron una breve asociación a la expresión de proteínas de adhesión identificables dentro del espectro.

Recientemente, en un estudio publicado por Rodríguez-Temporal *et al.* En el 2023, se demostró la capacidad del uso de MALDI-TOF para diferenciar entre cepas capaces de producir una abundante cantidad de *biofilm* de aquellas que presentaban poca formación, clasificando correctamente el 81% de las cepas de *S. aureus* y el 74% de las cepas de *C. albicans* (Rodríguez-Temporal, Díez, *et al.*, 2023b).

Otros investigadores han optado por caracterizar la formación de *biofilm* dependiendo de la superficie de adhesión. Por un lado, asghari y colaboradores, se enfocaron en la identificación y caracterización de cepas de *P. aeruginosa* a partir de *biofilm* adheridos a diferentes superficies; mientras que Pereira en el 2015, caracterizó la formación *biofilm* de *P. aeruginosa* en vidrio y plástico en función del tiempo (Pereira *et al.*, 2015; Asghari *et al.*, 2021b).

4. Marco Teórico

4.1. *Staphylococcus aureus*

S. aureus es una bacteria Gram positiva ubicua de tejidos mucosos de diferentes especies animales, de la microbiota de la piel humana y el medio ambiente en general (Grace and Fetsch, 2018). Por otra parte, *S. aureus* se constituye como uno de los patógenos humanos más importantes, debido a su capacidad de colonizar tejidos, causar infección y evitar eficientemente el sistema inmune del hospedero (Leme et al., 2021). Para la organización mundial de la salud (OMS) este microorganismo tiene un elevado grado de letalidad debido al aumento de su farmacoresistencia representando un riesgo a la salud pública (World Health Organization, 2021). Actualmente se realiza la identificación basada en el cultivo microbiano o técnicas moleculares como qPCR. Sin embargo, estas técnicas no permiten conocer el estado metabólico del microorganismo ni los mecanismos de virulencia y resistencia que se encuentra expresando, además de ser técnicas que requieren de tiempos largo de procesamiento, profesional altamente capacitado y altos costos (Sanchini, 2022).

Uno de los factores de virulencia más importante, pero poco identificado en la práctica clínica es la capacidad del microorganismo de adherirse a superficies y formar *biofilm* (Kranjec et al., 2021). La formación de *biofilm* confiere propiedades biológicas y químicas que dificultan la efectividad de los tratamientos convencionales, ya que estos suelen estar diseñados para cultivos planctónicos o de libre flotación (Iskandar et al., 2022). Estudios epidemiológicos recientes, han determinado que el *biofilm* es la causa más común de infecciones asociadas a mucosas y dispositivos médicos (Pietrocola et al., 2022).

4.2. Biofilm

El *biofilm* es una comunidad microbiana en estado sésil embebidas en una matriz de exopolisacáridos que permite el intercambio de material genético, nutrientes y moléculas de comunicación celular (Donlan, 2002). Suelen encontrarse naturalmente adheridos a superficies bióticas y abióticas mediante interacciones hidrofóbicas, electrostáticas o la participación de algunas moléculas de superficie como la autolisina, ácidos teicoicos o proteínas de adhesión (Cincarova et al., 2016; Vázquez-Sánchez & Rodríguez-López, 2018). La formación del *biofilm* se realiza en cinco etapas (Figura 1): i) adhesión reversible, ii) adhesión irreversible, iii) agregación de células en microcolonias, iv) maduración del *biofilm*, finalmente v) dispersión (Hall-Stoodley et al., 2004).

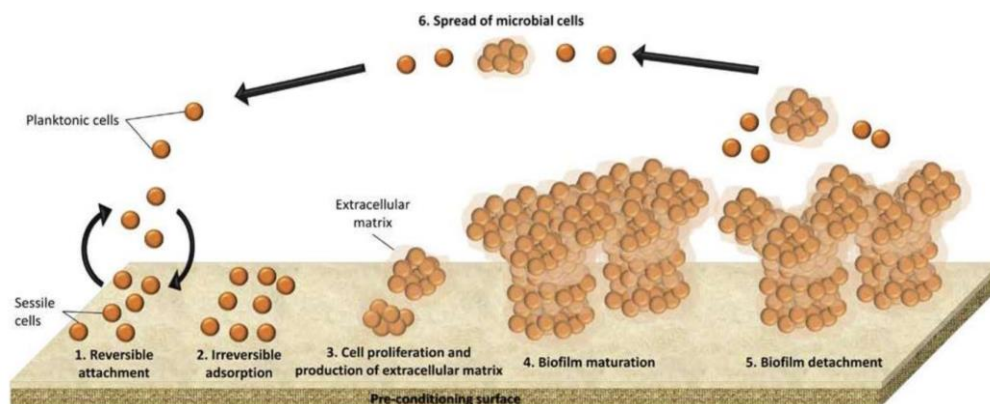


Figura 1. Esquema general de las etapas de la formación de biofilm.

Nota. Adaptado de *Staphylococcus aureus*, Chapter 5 (p. 89), Vázquez-Sánchez D., Rodríguez-López P., 2018, Elsevier.

4.3. Espectrometría de masas MALDI-TOF

La ionización/desorción láser asistida por matriz (MALDI), es una técnica de ionización suave o de baja energía que permite el estudio de macromoléculas de altos pesos moleculares o termolábiles (Karas and Hillenkamp, 1988). En esta ionización, el analito se co-cristaliza con una

matriz que tiene la capacidad de absorber radiación ultravioleta generada por un láser pulsante (comúnmente de Nd:YAG, $\lambda = 355$ nm) (Karas et al., 1985). La absorción de radiación se debe a la presencia de cromóforos y grupos funcionales que extiende el sistema conjugado de electrones Pi (π) en la estructura molecular de la matriz ionizándola. Una vez ionizada la matriz, se da ionización secundaria en la que se transfieren electrones, protones y cationes al analito, el cual es dirigido al analizador de tiempo de vuelo mediante un campo electrostático (Jurgen H. Gross, 2004).

En el analizador de tiempo de vuelo TOF se separan los iones impulsados por la aplicación de un diferencial de potencial definido, dentro de un tubo de vuelo de longitud conocida (Grix *et al.*, 1988). Debido a que el potencial acelera los iones, esta energía eléctrica se transforma en energía cinética que se puede expresar en función de la velocidad (ecuación 1) y a su vez en función del tiempo (ecuación 2-3) (Jürgen H. Gross, 2004).

Energía cinética en función de la velocidad (e , representa la carga del electrón; z , un número entero de cargas; U , voltaje; m_i , masa):

$$v = \sqrt{\frac{2ezU}{m_i}} \quad (1)$$

Tiempo en función de la energía cinética (en el numerador d representa la distancia y en el denominador la energía cinética en función de la velocidad):

$$t = \frac{d}{\sqrt{\frac{2ezU}{m_i}}} \quad (2)$$

$$t = \frac{d}{\sqrt{2eU}} \sqrt{\frac{m_i}{z}} \quad (3)$$

Dada la correspondencia entre el tiempo de vuelo y la relación masa-carga (m/z), el detector mide los intervalos de tiempo de llegada de los analitos y así establece la relación m/z de cada uno de los iones.

4.4. Procesamiento de datos de espectrometría de masas

Un set de datos obtenidos por espectrometría de masas es el conglomerado de espectros adquiridos para realizar un estudio, estos a su vez contienen un conjunto de señales que representan la información característica del analito. Para el análisis de muestras microbiológicas estas señales representan información biológica potencialmente relevante, usualmente asociada a proteínas o metabolitos (Coombes et al., 2007) de un microorganismo. Antes de realizar el análisis del set de espectros de masas, se debe llevar a cabo un preprocesamiento que usualmente involucra: i) corrección de la línea base, ii) suavizado (*smoothing*), iii) normalización, iv) alineamiento de picos, y v) detección de picos (Albrethsen, 2007; Coombes et al., 2007; Peris-Díaz & Krężel, 2021). Existen diferentes aproximaciones matemáticas para realizar los pasos mencionados anteriormente; muchas de las cuales ya se encuentran incluidas dentro de los softwares propios de los espectrómetros de masas (*i.e.* FlexAnalysis Bruker Daltonics), o se pueden realizar a través de paquetes estadísticos en entornos como Python, R, o MatLab.

4.5. Quimiometría

La quimiometría suele definirse como la disciplina que utiliza métodos matemáticos y estadísticos para seleccionar, diseñar procedimientos y experimentos que permitan obtener información relevante a partir de datos químicos (Alciaturi et al., 2010; Peris-Díaz & Krężel, 2021). Actualmente, se puede incluir en esta definición que la quimiometría también proporciona la máxima información química mediante el análisis de datos químicos. Debido a las características multidimensionales de los diferentes datos analíticos, se suelen organizar en matrices de n objetos (filas) y p características (columnas). Los objetos pueden ser muestras,

moléculas, materiales, entre otros; mientras que las características son patrones, espectros, características estructurales o propiedades físicas (Alciaturi *et al.*, 2010).

4.5.1. Análisis por componentes principales (PCA)

Los análisis de grandes volúmenes de datos multivariados generan matrices de correlación de cientos de valores, lo cual dificulta el reconocimiento de patrones. El PCA es uno de los métodos de reducción de variables más utilizados, con el que se busca reducir la cantidad de datos basándose en la correlación entre ellos (Miller and Miller, 2010). Usualmente se utiliza como herramienta exploratoria en el análisis de las observaciones de un conjunto de ‘ p ’ variables a un número ‘ n ’ muestras, construyendo una matriz ‘ X ’ de dimensiones ‘ $n \times p$ ’ con el fin de encontrar una combinación lineal (X_a) de las columnas que contenga la máxima varianza posible (Jolliffe and Cadima, no date; Miller and Miller, 2010). Las combinaciones lineales en los modelos de PCA vienen dadas por la ecuación 1, donde ‘ a ’ es un vector de constantes ‘ a_1, a_2, \dots, a_p ’ y la varianza de cualquier combinación lineal se expresa como la ecuación 2, donde ‘ a^T ’ es la transposición del vector ‘ a ’ y ‘ S ’ es la matriz de covarianza del conjunto muestral (Jolliffe and Cadima, 2016).

$$\sum_{j=1}^p a_j x_j = X_a \quad (1)$$

$$\text{var}(X_a) = a^T S a \quad (2)$$

La Figura 2a ilustra el comportamiento de los datos cuando solo hay dos variables y por ende solo dos componentes principales, las cuales son ortogonales entre si; mientras que la Figura 2b muestra los puntos asociados a los nuevos ejes (CP1, CP2) y sus respectivas proyecciones,

mostrando que la mayoría de los puntos quedan sobre PC1. Debido a esta nueva agrupación, esta componente contiene la mayor variación permitiendo reducir a una dimensión en lugar de trabajar en dos dimensiones.

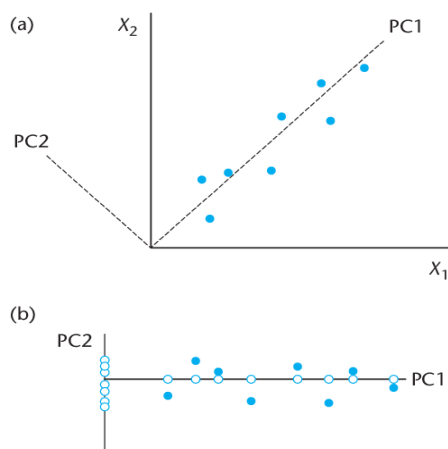


Figura 2. Esquema del comportamiento de los datos con dos variables y por tanto dos componentes principales.

Nita. A) Diagrama ilustrativo de los componentes principales, CP1 y CP2, para las dos variables X_1 y X_2 . B) puntos referidos a los ejes de las componentes principales.

4.5.2. Métricas de evaluación de modelos predictivos de clasificación

La matriz de confusión es una herramienta que permite evaluar un modelo de clasificación, de manera intuitiva, porque solo requiere de una inspección visual de las predicciones del modelo que se está evaluando. La matriz de confusión consta de cuatro elementos. Verdaderos positivos, son los elementos positivos que el modelo predice como positivos. Falso positivo, son los elementos negativos que el modelo predice como positivos. Falso negativo, son los elementos positivos que el modelo predice como negativos. Verdadero negativo, son los elementos negativos que el modelo predice como negativos y se organizan como se muestra en la Figura 3.

		Clase Real	
		Positivo	Negativo
Clase Predicha	Positivo	VP	FP
	Negativo	FN	VN

Figura 3. Matriz de confusión. *Vp*: verdadero positivo; *FP*: Falso positivo; *FN*: falso negativo; *VN*: verdadero negativo.

Otra herramienta para evaluar un modelo predictivo es mediante los parámetros de, precisión, *Recall* y *F1-score* que la librería *scikit-learn* puede calcular mediante el comando *classification_report*. La precisión representa la relación entre la cantidad de predicciones positivas y cuáles de esas son realmente positivas. El *Recall*, es la relación entre las predicciones positivas y la totalidad de verdaderos positivos. *F1-score*, es la media armónica de la precisión y el *Recall*, es muy útil cuando las clases no poseen la misma cantidad de datos.

4.6. Aprendizaje automático o Machine Learning

Machine Learning es una rama de la inteligencia artificial enfocada en estudiar, diseñar y mejorar modelos matemáticos que puedan inferir y tomar decisiones mediante el uso de datos acordes al contexto requerido. Los algoritmos usados para mejorar dichos modelos combinan conceptos de informática con probabilidad, estadística y optimización, de modo que la cantidad y calidad de los datos de referencia son un punto crítico en el desempeño del aprendizaje (Subasi, 2020). El uso de *Machine learning* en el análisis de las señales de los espectros masas, surge por la necesidad de procesar grandes cantidades de datos con fines diagnósticos, identificación y predicción (Bonaccorso, 2017). Dado que la adquisición de espectros de masas mediante MALDI-

TOF es rápida y muestra las señales de iones cargados con una alta resolución, los datos que se pueden extraer de los espectros pueden alimentar un algoritmo de aprendizaje (Wei *et al.*, 2021). Dependiendo del tamaño de los datos, sus características y el tipo de estudio que se desea realizar, se pueden plantear diferentes escenarios de aprendizaje.

4.6.1. Aprendizaje supervisado

Es un escenario donde el algoritmo se alimenta con un conjunto de datos de entrenamiento. Estos datos están compuestos por un valor de entrada y un valor de salida esperado, a partir de esta información el algoritmo empieza a realizar correcciones hasta que la diferencia entre el valor predicho y el esperado sea cercano a cero (Desaire et al., 2022). Al finalizar, el algoritmo se alimenta con un set de datos de prueba, los cuales son datos conocidos pero que no se usaron en fase de entrenamiento, de esta manera se evalúa el rendimiento del aprendizaje y su capacidad para realizar predicciones (Desaire et al., 2022).

4.6.2. Aprendizaje no supervisado

Aquí solo le damos las características al algoritmo, nunca las etiquetas. Queremos que nos agrupe los datos que le dimos según sus características. El algoritmo solo sabe que como los datos comparten ciertas características, de esa forma asume que pueda que pertenezcan al mismo grupo.

4.6.3. Algoritmos para Machine Learning

4.6.3.1 Máquinas de vector soporte. Es un algoritmo de aprendizaje supervisado usado para clasificar y realizar regresiones. Este algoritmo se centra en encontrar un hiperplano que separe de la mejor manera dos clases dentro de un conjunto de datos. Dado que no todos los datos se pueden separar linealmente, este algoritmo puede utilizar funciones para transformar las características del hiperplano y asignar los datos a espacios dimensionales diferentes en los cuales resulte más fácil separar las clases. Estas funciones se designan en el parámetro kernel y puede

ser: función de base radial o gaussiana, lineal, polinómica y sigmoide (Nalepa and Kawulok, 2019).

4.6.3.2 XGBoost. El algoritmo refuerzo de gradiente extremo (XGBoost) se basa en árboles de decisión que se van ajustando para corregir los errores del modelo existente. Inicia con un árbol de decisión basado en reglas condicionales, el cual por sí solo es muy débil, por lo tanto, el algoritmo crea secuencialmente arboles de decisión donde cada uno corrige los errores anteriores y utiliza un gradiente descendiente con el fin minimizar la función de pérdida que mide la diferencia entre las predicciones y los valores reales. Para evitar el sobre entrenamiento del modelo se aplican técnicas de penalización y poda de árboles (Li *et al.*, 2023).

4.6.3.3 Random Forest. Es un algoritmo de aprendizaje en conjunto, que se basa en la creación de árboles de decisión en subconjuntos aleatorios del conjunto total de los datos como se observa en la Figura 4.

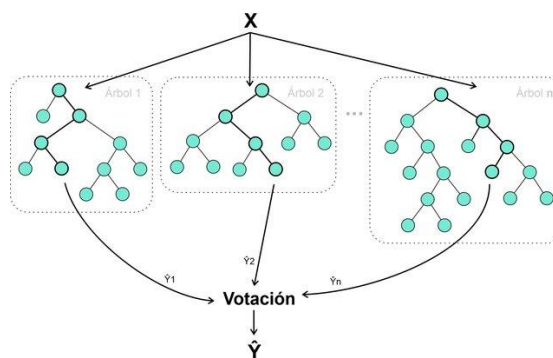


Figura 4. Esquema de los árboles de decisión en Random forest.

Nota. Tomado de (Gonzales, 2021)

Cada uno de los árboles en su respectivo subconjunto realiza una votación o una predicción y el resultado se obtiene por votación o promedio respectivamente. La aleatoriedad de los subconjuntos evita el sobre entrenamiento, reduce la varianza del modelo y permite conocer las características más importantes para clasificar o predecir (Parmar, Katariya and Patel, 2019).

4.6.3.4 Redes Neuronales artificiales. Se basa en la estructura y funcionamiento del cerebro humano, la unidad básica de las redes neuronales artificiales es la neurona artificial y es capaz de tomar múltiples entradas, ponderarlas por pesos, las suma y emplea una función para producir una salida. Como se muestra en la Figura 5, estas neuronas se organizan por capas dentro de la red neuronal en: capa de entrada, la cual se reciben las características del conjunto de datos; capas ocultas, realizan transformaciones no lineales de las entradas y capas de salida, encargadas de producir la respuesta. La asociación entre neuronas tiene un peso específico el cual determina la importancia de una entrada en particular para la salida de la neurona, la suma de las entradas ponderadas (con mayor peso) activa una función de salida. Este algoritmo aprende automáticamente mediante datos y el ajuste interactivo de los pesos según las conexiones entre neuronas (Wu and Feng, 2018).

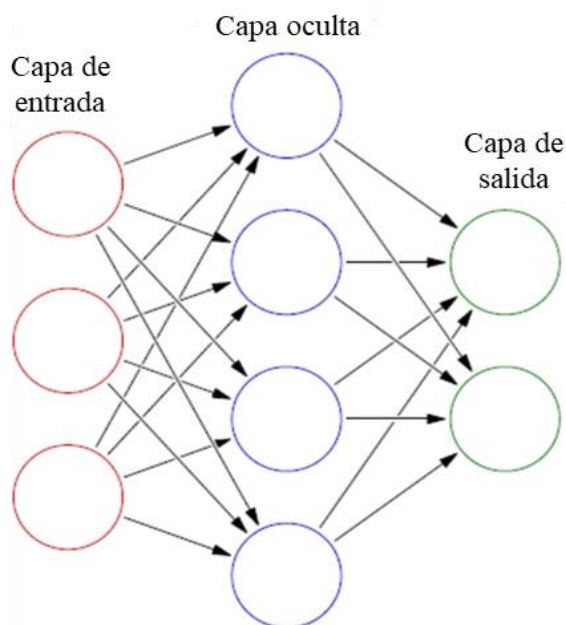


Figura 5. Esquema de las capas de una red neuronal.

Nota. Tomado y adaptado de (Wu and Feng, 2018).

4.6.4. Fases de desarrollo

4.6.4.1 Fase de entrenamiento. En esta fase se tiene una cantidad enorme de datos, de la cual se separa una parte para entrenar al algoritmo y darle toda esta información para que encuentre los patrones necesarios y después pueda hacer predicciones.

4.6.4.2 Fase de prueba. El resto de los datos que quedan, se van a usar para hacer las pruebas. Así le podemos hacer preguntas al algoritmo y evaluar si las respuestas están bien o mal, y saber si está aprendiendo o no. Si vemos que no coinciden los datos, tendremos que agregar más datos o cambiar el método que estamos utilizando. Pero si se observa que hay entre un 80% a 90% de respuestas correctas, podemos decir que hay un buen grado de aprendizaje y poder utilizar ese algoritmo.

4.6.4.3 Validación cruzada. La validación cruzada es un proceso mediante el cual se puede evaluar el rendimiento de un modelo predictivo y se busca disminuir la probabilidad de tener un sobre entrenamiento del modelo. En la Figura 6 se muestra el esquema de la técnica *k-fold*.

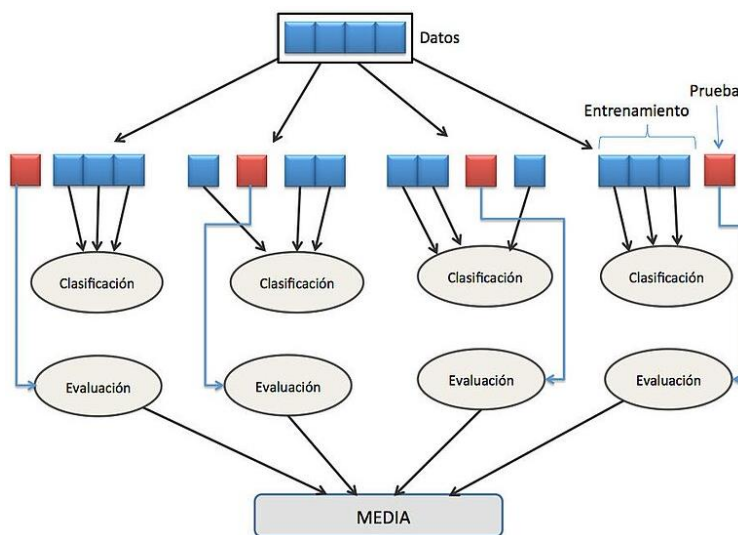


Figura 6. Esquema de validación cruzada *k-fold*.
Nota. Tomado y modificado de (Pal and Patel, 2020).

La validación cruzada *k-fold* divide los de datos en un grupo de entrenamiento con el que se construye el modelo y un grupo de datos de prueba, con los que se prueba el modelo creado con los datos de entrenamiento. Sin embargo, *k-fold* realiza otro tipo de separación de los datos, agrupa aleatoriamente todos los datos en k subconjuntos con tamaños similares que se denominan “fold”, el modelo se entrena y valida en cada iteración de los fold como se muestra en la Figura 6, al final se promedian las métricas de evaluación de cada fold proporcionando una estimación más robusta y confiable (Pal and Patel, 2020).

5. Objetivos

5.1. Objetivo General

Implementar una metodología de análisis mediante espectrometría de masas MALDI-TOF y *Machine Learning* que detecte señales diferenciales del perfil molecular relacionadas con la formación de *biofilm* de tres cepas de *Staphylococcus aureus* ATCC 29213, ATCC BAA 44 y ATCC 43300.

5.2. Objetivos Específicos

Establecer un método de preparación de muestras de *Staphylococcus aureus* en estado planctónico y de *biofilm* que permita adquirir espectros de masas de para la identificación de perfiles moleculares mediante Espectrometría de Masas MALDI-TOF.

Implementar parámetros instrumentales del espectrómetro de masas MALDI-TOF en la adquisición de espectros de masas de muestras en estado planctónico y sésil de cepas de *S. aureus*.

Proponer un modelo predictivo basado en los perfiles moleculares de cepas de *S. aureus* que permita la diferenciación entre el estado planctónico y sésil usando la herramienta de aprendizaje automatizado *Machine learning*.

6. Materiales y Métodos

6.1. Materiales

Todos los solventes utilizados en este trabajo fueron grado HPLC, a menos que se indique lo contrario. El agua utilizada para todos los procedimientos fue tipo I ($\text{pH} \approx 7$, conductividad $0.09 \mu\text{S}/\text{cm}$) obtenida de un equipo Smart2pure (ThermoScientific, MA, USA) del laboratorio GIBIM. Se utilizó material de plástico libre de RNAsa, DNAsa, proteasas, el cual fue esterilizado adecuadamente por calor húmedo en un autoclave 50X-120V (All-American, VA, USA). El material de vidrio se usó exclusivamente en la experimentación proteómica, se mantuvo libre de detergentes y esterilizado en el autoclave. Los reactivos acetonitrilo (ACN), cloroformo (HCCL_3), ácido fórmico (AF), metanol, isopropanol, ácido trifluoroacético (TFA), ácido tricloroacético (TCA), etanol, acetona, cloruro de sodio, glucosa, cloruro de potasio, fosfato de potasio monobásico, fosfato de sodio dibásico se adquirieron a MERCK (Darmstadt, Germany). Los medios de cultivo Müller Hinton (MH), infusión cerebro corazón (BHI) y tripticasa soya (TSB) se adquirieron a MERCK (Darmstadt, Germany). La matriz de ácido alfa-ciano-4-hidroxicinámico (HCCA) y el hidrocloreuro de guanidina (HCG) se adquirieron de calidad biología molecular a SigmaAldrich (San Luis, MI, Estados Unidos).

Las cepas de *S. aureus* utilizadas son de referencia de ATCC (American Type Culture Collection) y fueron adquiridas comercialmente a LabCare (ANT, Colombia). Estas cepas presentan diferentes características fisiológicas, de patogenicidad y de resistencia (Tabla 1). Todas las cepas del estudio se encuentran conservadas en el laboratorio de Bioquímica y Microbiología a $-20 \text{ }^\circ\text{C}$. Previo a la realización de experimentos, las bacterias se inocularon en caldo de cultivo tripticasa soya (TSB) e incubados por 16 -24 h a $37 \text{ }^\circ\text{C}$, seguidos de un repique en agar Infusión cerebro corazón (BHI) para su reactivación y aislamiento.

Tabla 1.*Características microbiológicas cepas de estudio.*

Cepa de <i>Staphylococcus aureus</i>	Características
ATCC 29213	Cepa sensible a antibióticos Productor leve de β -lactamasas
ATCC 43300	Cepa resistente a la meticilina y oxacilina (SARM); Genotipo: pvl negativo; MRSA – SCCmec tipo II Clon ibérico multirresistente.
ATCC BAA-44	Cepa resistente a: Azitromicina, cefoxitina, ciprofloxacina, clindamicina, doxiciclina, eritromicina, gentamicina, penicilina, rifampicina, tetraciclina; Genotipo: pvl negativo; MRSA – SCCmec tipo I

6.2. Métodos

Este trabajo se llevó a cabo utilizando el esquema metodológico presentado en la Figura 7. Se inició el cultivo de las cepas de *S. aureus* (ATCC 29213, ATCC 43300, ATCC BAA 44) tanto para el estado planctónico, como para la cinética de formación de formación de *biofilm* en el tiempo, seguido de la extracción de proteínas mediante cuatro metodologías diferentes para las muestras en estado planctónico y de *biofilm*. Posteriormente se prepararon las muestras para su análisis mediante espectrometría de masas MALDI-TOF y la adquisición del set de espectros de masas. Mediante el software *FlexAnalysis* se ajustó la línea base y se exportaron los datos en formato de texto (.txt), finalmente se cargaron los datos en el ambiente de *Python* con el fin de evaluar el desempeño de cuatro algoritmos de *machine learning* en la generación de un modelo predictivo que permita diferenciar entre el estado planctónico y de *Biofilm*.

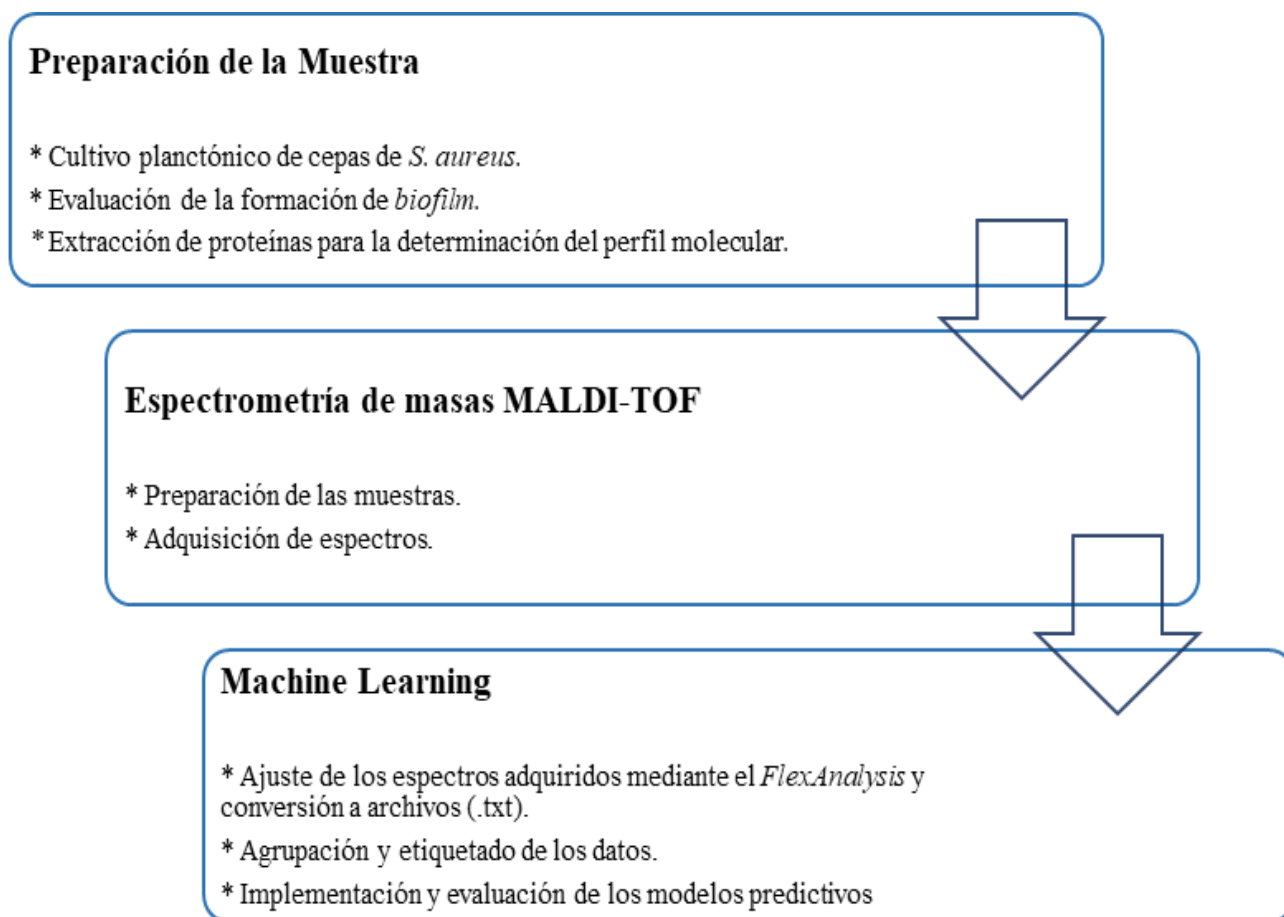


Figura 7. Esquema de la metodología general del proyecto.

6.2.1. Cultivo de cepas de *S. aureus* en estado planctónico

Las cepas se repicaron en medio sólido Agar BHI y se incubaron por 24 h a 37 °C. A partir de este cultivo fresco se preparó un inóculo en 30 mL de medio líquido MH ajustando a 4 en la escala de McFarland (Densidad óptica de 0.123-0.14 a 595 nm; $\sim 12 \times 10^8$ UFC/mL) el cual se incubo por dos horas a 37°C con agitación orbital a 200 rpm en un equipo MaxQ 4450 (ThermoScientific, MA, USA) para garantizar que todas las células se encuentran es estado planctónico. Una vez finalizado el tiempo de cultivo, se recuperaron las células planctónicas por centrifugación a 10000 g por 15 minutos a 4 °C en una centrifuga refrigerada IEC CL31R

multispeed (ThermoScientific, MA, USA). Una vez finalizada la centrifugación se descartó el sobrenadante y el pellet se lava tres veces utilizando Buffer Fosfato Salino (PBS, 100 mM, pH 7.2).

6.2.2. Formación de biofilm de cepas de *S. aureus*

Para este proyecto se evaluó inicialmente la cinética de formación de *biofilm* de las cepas de *S. aureus* en microplacas fondo en U de 96 pozos utilizando el protocolo publicado por Martínez *et al.* con algunas modificaciones (Martínez *et al.*, 2023). Brevemente, las cepas se repicaron en medio BHI agar y se incuban por 24 h a 37 °C. A partir de este cultivo se tomaron tres colonias para preparar un preinóculo en 5 mL de medio líquido TSB, el cual se incubo por 12 h a 37°C en agitación orbital a 200 rpm en un equipo MaxQ 4450 (ThermoScientific, MA, USA). Una vez finalizado el tiempo incubación se ajustó el preinóculo a 0.5 de Macfarland (Densidad óptica de 0.08-0.100 a 595 nm; $\sim 1.5 \times 10^8$ UFC/mL) en medio TSB 2X suplementado con glucosa al 4% (TSB-G). A partir del inóculo ajustado se realizó una dilución 1/50 en medio TSB-G, la cual se utilizará como inóculo. A cada pozo se añadieron 100 μ L de PBS y 100 μ L del inóculo de la cepa correspondiente, obteniendo una concentración inicial de células en cada pozo de $1-5 \times 10^5$ UFC/mL. Como control de esterilidad se dispondrá de pozos con PBS y el medio de cultivo sin inóculo. Todas las placas se incubaron a 37°C, sin agitación y cada 12 h se realizó la estimación de la formación de *biofilm* a través de la metodología de Cristal violeta (CV) y por método “*Drop plate counting*” o conteo por goteo en placa (Herigstad *et al.*, 2001; O’Toole, 2011).

Una vez determinado el punto máximo de formación de *biofilm* para cada cepa, se repitió la metodología anterior incubando por 24 y 36 h para la extracción de proteínas. Brevemente, una vez completadas las 24 o 36 h, según corresponda, se vaciaron los pozos y se lavaron 3 veces con buffer PBS para retirar el medio de cultivo y las células planctónicas. Para recuperar las células del *biofilm*, se agregaron 50 μ L de PBS a cada pozo y se sónico la placa en agua estéril por 1 minuto en un baño ultrasónico E30H (ELMA, Freiburg, Alemania). Posteriormente, se adicionaron

150 μ L adicionales de PBS y se realizó un raspado a cada pozo con una punta de micropipeta estéril. Con una micropipeta se tomaron los 200 μ L del PBS con las células del *biofilm* y se adicionaron a un eppendorf de 2.0 mL, el cual se centrifugo a 10000 g por 15 minutos a 4 °C en una centrifuga refrigerada IEC CL31R multispeed (ThermoScientific, MA, USA) para eliminar el sobrenadante. El pellet bacteriano obtenido se utilizó para la extracción de proteínas.

6.2.3. Extracción de proteínas para la determinación del perfil molecular

Se evaluaron cuatro protocolos de extracción de proteínas para determinar su efecto sobre el perfil molecular de las cepas de *S. aureus* (Figura 8). Todas las extracciones fueron secadas en un equipo SpeedVac (ThermoScientific, MA, USA) a 35°C por 2 h y resuspendido en HCG para cuantificar y dejar en conservación a -20°C hasta su uso.

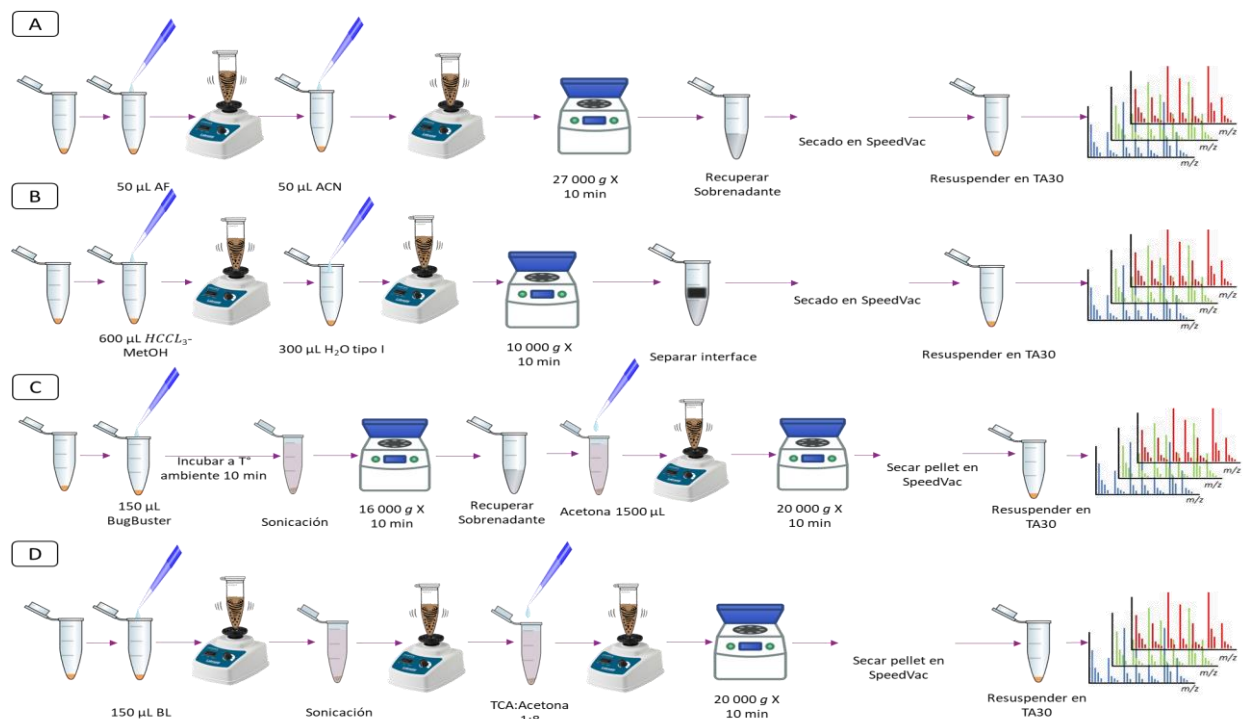


Figura 8. Metodología de extracción para la obtención del perfil molecular en espectrometría de masas MALDI-TOF.

Nota: A) ACN-AF; B) Cloroformo - metanol ($CHCl_3$ -MetOH) ; C) BugBuster® ;D) Búfer de lisis.

Adicionalmente, a todas las extracciones se les cuantificó la concentración de proteína utilizando el método de Bradford (Bradford, 1976). La curva de cuantificación se realizó con un patrón de albúmina de suero bovino (BSA) y concentraciones desde 0.1 mg/mL hasta 2 mg/mL (anexo A). Adicionalmente se calculó el límite de cuantificación (0.17 mg/mL) y el límite de detección (0.02 mg/mL).

6.2.3.1 Extracción Ácido Fórmico Acetonitrilo (AF-ACN). El esquema de la metodología de AF-ACN se presenta en la Figura 8A. Brevemente, se adicionaron 100 μ L de AF sobre cada uno los pellets celulares obtenidos para el cultivo plantónico (sección 6.2.1) y para el *biofilm* (sección 6.2.2.), seguido de agitación fuerte a 3200 rpm por 5 min en un vortex (VWR); una vez homogenizado el pellet en el AF, se adicionaron 100 μ L de ACN, seguido de agitación fuerte con vortex. A continuación, la muestra se centrifugo a 27000 g por 5 min a 4°C, y se recuperó el sobrenadante.

6.2.3.2 Extracción Cloroformo metanol (HCCl₃-MetOH). Como se muestra en la Figura 8B, primero se adiciona 600 μ L de una solución de Cloroformo y metanol en una proporción 1:4 v/v sobre el pellet celular obtenido para el cultivo plantónico y el *biofilm*. A continuación, se agitó en un vortex, se agregó 300 μ L de agua tipo I y se centrifugo a 10000 g por 10 min a 4°C, para separar la interfase de la fase acuosa y orgánica. Después de centrifugar se retira la fase orgánica y acuosa, se realizan dos lavados con metanol frío con el fin de recuperar el pellet de proteínas.

6.2.3.3 Extracción BugBuster® (BB). Para esta extracción se adicionaron 150 μ L de BugBuster® sobre cada uno los pellets celulares obtenidos para el cultivo plantónico (sección 6.2.1) y para el *biofilm* (sección 6.2.2), se agitó en un vortex y se dejó incubar a temperatura ambiente por 10 min; una vez finalizado el tiempo de incubación se realizó una ruptura mecánica de las células mediante un sonicador de punto (Cole-Parmer, USA), para precipitar las proteínas

se adicionó 1500 μL de acetona fría y se centrifugó a 20000 g por 10 min a 4°C en una centrifuga SL 8R.

6.2.3.4 Extracción con búfer de lisis (BL). Como se observa en la Figura 8D, se adicionaron 150 μL de Buffer de lisis sobre el pellet celular obtenido para el cultivo platónico (sección 6.2.1) y para el *biofilm* (sección 6.2.2.), se agitó en un vortex y se realizó una ruptura mecánica de las células con un sonicador de punto. A continuación, se precipitaron las proteínas con 1500 μL de acetona fría y se centrifugó a 20000 g por 10 min a 4°C en una centrifuga SL 8R.

6.2.4. Espectrometría de masas MALDI-TOF

Las muestras extraídas fueron diluidas en TA30 (30:70 [v/v], ACN-TFA 0.1%) en una relación 1:100, y se sembraron en un porta muestras para MALDI (Plate MTP 384 marca Bruker) por el método de doble capa (matriz/muestra/matriz). Brevemente, se agregó un μL de matriz HCCA a una concentración de 20 mg/mL la cual se secó a temperatura ambiente. Una vez seca la primera capa de matriz se adicionó un μL de la muestra respectiva, se dejó secar. Finalmente se agregó la última capa correspondiente a un μL de matriz. Los perfiles moleculares se adquirieron usando un espectrómetro de masa Ultraflex II MALDI-TOF-TOF MS/MS marca Bruker Daltonics en modo lineal de iones positivos en un rango de masas de 2000 m/z a 20000 m/z , bajo los siguientes parámetros: 3.0 kV de potencial de aceleración, intensidad del láser del 90%, y aproximadamente 10 disparos de laser por *spot*.

6.2.5. Pre-procesamiento de datos

Los espectros de masas adquiridos mediante MALDI-TOF-MS fueron suavizados, corregidos en su línea base y exportados en formato texto (.txt) con el programa Bruker Daltonics FlexAnalysis V3.3.80. Una vez exportados se renombraron y organizaron por grupos de análisis.

En la Figura 9 se presenta el esquema general utilizado para el procesamiento de los datos y la construcción del modelo predictivo. Script utilizado para el análisis Anexo B.

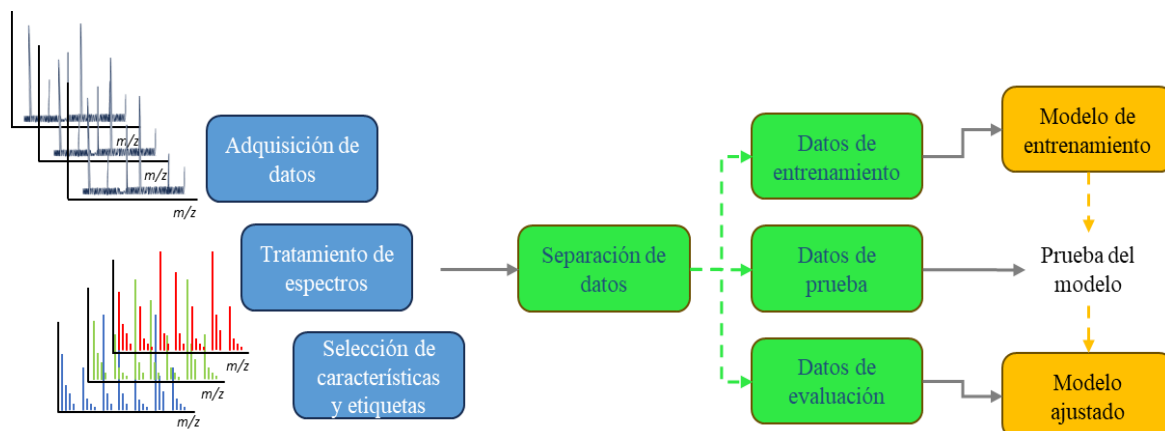


Figura 9. Esquema de ejecución en Machine Learning.

Se realizó una inspección inicial de los datos mediante un PCA, con el fin de observar el comportamiento de los datos de cada extracción, cuando se encuentran etiquetados por cepa. Es decir, todas las cepas de una misma extracción y estado hacían parte de un set de análisis. Esta organización se presenta en la Tabla 2.

Tabla 2.

Número de espectros usados por extracción y estado (biofilm, planctónico).

AF-ACN		
Cepa de <i>S. aureus</i>	Biofilm (n)	Planctónico (n)
ATCC BAA-44	7	3
ATCC 43300	9	6
ATCC 29213	9	6
HCC13-MetOH		
Cepa de <i>S. aureus</i>	Biofilm (n)	Planctónico (n)
ATCC BAA-44	7	3
ATCC 43300	9	6
ATCC 29213	9	6
BL		
Cepa de <i>S. aureus</i>	Biofilm (n)	Planctónico (n)
ATCC BAA-44	7	3
ATCC 43300	7	6
ATCC 29213	7	6

A partir de los resultados obtenidos del PCA anterior, todos los espectros de masas de las tres cepas se agruparon de acuerdo con el protocolo de extracción, y se etiquetaron por estado (planctónico y *biofilm*) como se muestra en la Tabla 3. Esta organización de los datos fue la utilizada para los análisis posteriores.

Tabla 3.

Número total de espectros de masas usados en la construcción del modelo.

Extracción	<i>S. aureus</i>		Espectros totales para el modelo (<i>n</i>)
	<i>Biofilm</i> (<i>n</i>)	Planctónico (<i>n</i>)	
AF-ACN	25	15	40
HCCl ₃ - MetOH	25	15	40
BL	21	15	36

6.2.6. *Procesamiento de datos*

Los archivos .txt se cargaron al entorno *JupyterLab*, dónde se agruparon en dos variables: una contiene las relaciones m/z , y otra la intensidad de las señales para cada relación m/z . La variable de la intensidad de las señales se normalizaron mediante la herramienta *StandardScaler* de la librería *scikit-learn*, con el fin que los datos presenten una distribución normal. Script utilizado para el análisis Anexo B.

Debido a que la cantidad de datos usados en la generación de modelos predictivos es inferior a 40 espectros de masas por extracción, se realizó una validación cruzada mediante *k-fold*. Esta técnica se aplicó a cada uno de los algoritmos. Los parámetros usados se muestran en la Tabla 4.

Tabla 4.
Parámetros de los algoritmos.

Algoritmo	Parámetros
<i>k-fold</i>	<i>n_splits=4, shuffle=True y random_state=123</i>
<i>SVC</i>	<i>(C=60, kernel='rbf', random_state=123)</i>
<i>XGBoost</i>	<i>(random_state=123)</i>
<i>RF</i>	<i>(n_estimators=100, random_state=113)</i>
<i>ANN</i>	<i>(hidden_layer_sizes=(20,20), learning_rate_init=0.01, solver='lbfgs', max_iter=5000, random_state=4)</i>

Esta técnica se usa comúnmente cuando la cantidad de muestras son iguales o menores 100 datos. La evaluación y comparación de los modelos se realizó mediante las métricas obtenidas por *classification_report*, *cross_val_score* y matrices de confusión.

7. Resultados y análisis

7.1. Obtención de muestras planctónicas y *biofilm* de *S. aureus*

Con el fin de obtener las muestras planctónicas, fue necesario garantizar que los microorganismos estudiados (*S. aureus* ATCC BAA-44, ATCC 43300 y ATCC 29213) estuvieran en la fase exponencial de crecimiento y de libre flotación. Por lo tanto, se utilizó un cultivo fresco en un medio no suplementado como el MH, con una agitación constante, para disminuir la probabilidad de adhesión a la superficie y un periodo de incubación alrededor de la mitad de la fase exponencial. En la Figura 10 se observa que el crecimiento de las tres cepas es similar, iniciando la fase de crecimiento alrededor de dos horas y llegando a un estado estacionario o latencia en la hora 8. Teniendo en cuenta las observaciones mencionadas anteriormente, se escogió como tiempo adecuado para la obtención de las muestras en estado planctónico, 2 horas de cultivo en un medio MH.

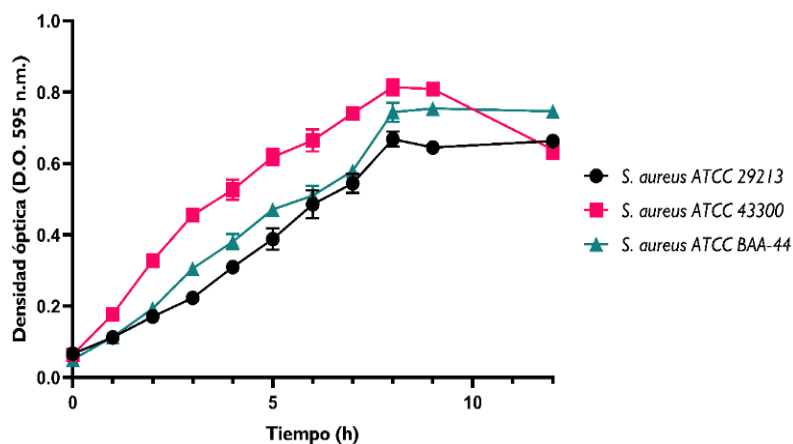


Figura 10. Cinética de crecimiento planctónico de las cepas *S. aureus* ATCC BAA-44, ATCC 43300 y ATCC 29213 en medio MH a 37°C y agitación a 200 rpm.

En cuanto a la formación de *biofilm* en el laboratorio, a diferencia de lo que podemos encontrar en la naturaleza, ésta se ve limitada por los nutrientes y la deshidratación del medio de cultivo utilizado (Moormeier and Bayles, 2017). Por lo tanto, fue necesario realizar una

caracterización preliminar de la cinética de formación de *biofilm* para cada una de las cepas de estudio. En la Figura 11, se observa que es estable en términos de biomasa y células viables entre las 24 y 36 horas de cultivo para todas las cepas, por lo consiguiente se seleccionó la hora 36 como punto para realizar la extracción de las biomoléculas. La Figura 11^a, evidencia que la cepa de *S. aureus* ATCC 43300 es un alto productor de *biofilm* con absorbancias superiores a 4; Sin embargo, la concentración de células viables se mantuvo entre 8 y 20 x10⁷ UFC/cm². Por otro lado, *S. aureus* ATCC 29213 se caracterizó como la cepa con menor producción de *biofilm*, seguido por *S. aureus* ATCC BAA-44, como productor de *biofilm* moderado en comparación con los resultados ya descritos para las demás cepas. Esta clasificación se realizó con los resultados de la semi-cuantificación del *biofilm* por el método de cristal violeta, como otros autores también lo reportan (Caputo *et al.*, 2018a; Latorre *et al.*, 2022).

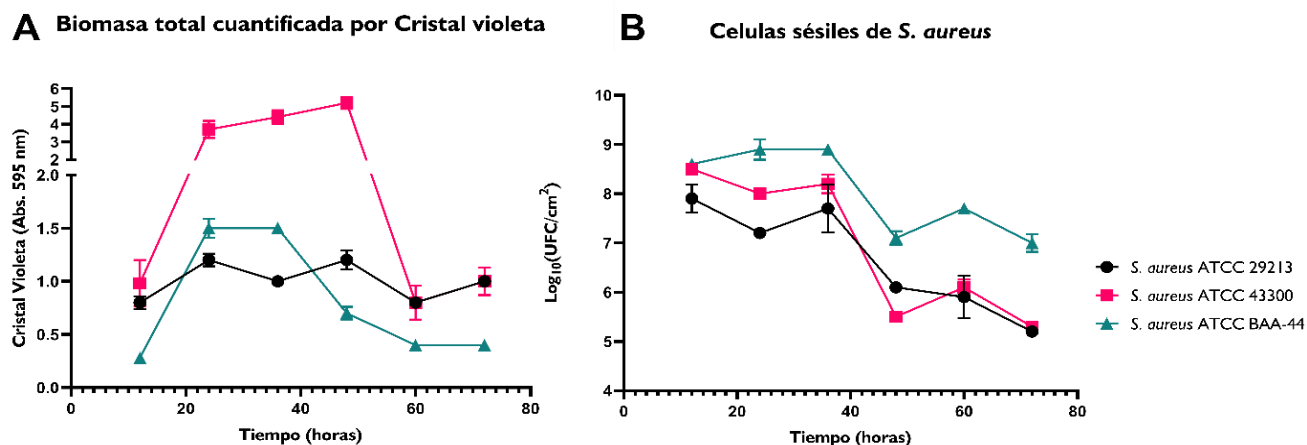


Figura 11. Cinética de formación del *biofilm* de las cepas de estudio a las 12, 24, 36, 48, 60 y 72 h de cultivo. A) Cuantificación por CV de la biomasa total; B) Cuantificación por conteo de colonias de las células cultivables.

7.2. Extracción de proteínas

Se aplicaron cuatro protocolos de extracción de proteínas (AF-ACN, HCCl₃-MetOH y BL) a las muestras obtenida en la fase de exponencial del estado planctónico y del *biofilm* de las cepas de *S. aureus* ATCC BAA-44, ATCC 43300 y ATCC 29213. La cuantificación se realizó mediante

el método de Bradford y una curva de cuantificación realizada con un patrón de Albúmina de Suero Bovino (BSA) presentada en el Anexo A. En la Tabla 5, se muestran los resultados de la determinación de la concentración de proteína, según cada uno de los protocolos de extracción (AF-ACN, HCCl₃-MetOH), el estado de cultivo y la cepa.

Tabla 5.

Concentración de proteína obtenida para las cepas de estudio en estado planctónico y biofilm según el método de extracción.

	<i>S. aureus</i> ATCC 29213		<i>S. aureus</i> ATCC 43000		<i>S. aureus</i> ATCC BAA-44	
	Planctónico mg/mL	Biofilm mg/mL	Planctónico mg/mL	Biofilm mg/mL	Planctónico mg/mL	Biofilm mg/mL
AF-ACN.	0.63 ± 0.01	< LC	0.82 ± 0.03	0.34 ± 0.08	0.87 ± 0.09	0.17 ± 0.1
BugBuster	0.47 ± 0.06	< LC	0.50 ± 0.09	0.19 ± 0.07	0.38 ± 0.06	< LC
HCCl₃-MetOH	1.1 ± 0.1	0.19 ± 0.07	1.10 ± 0.04	0.43 ± 0.04	1.0 ± 0.3	0.21 ± 0.08
BL	2.3 ± 0.7	0.18 ± 0.09	2.5 ± 0.9	0.43 ± 0.07	1.9 ± 0.2	0.3 ± 0.1

Nota. LC, hace referencia al límite cuantificación de la curva (0.17 mg/mL).

Se observa que las concentraciones más altas de proteína se presentan en el estado planctónico y el protocolo de extracción BL, seguido de HCCl₃-MetOH, con concentraciones de proteína superiores 1 mg/mL. Por el contrario, el protocolo de extracción BugBuster[®] solo presentó una concentración máxima de proteína 0.5 mg/mL y el protocolo de extracción AF-ACN permitió obtener concentraciones entre 0.63 mg/mL y 0.87 mg/mL. Para el estado de *biofilm* las concentraciones de proteína en general disminuyeron en comparación con los resultados del estado planctónico, incluso las extracciones AF-ACN y Bugbuster[®] fueron tan bajas que no fue posible cuantificar la proteína de la cepa ATCC 29213.

Las diferencias en las concentraciones de proteína obtenida utilizando cada protocolo de extracción se debe probablemente a su capacidad para la disrupción de las células bacterianas, tanto sésiles como planctónicas. Se conoce que la metodología de AF-ACN es una lisis suave en la cual se obtienen proteínas principalmente de membrana y ribosomales (Zhu *et al.*, 2015). El protocolo de BL presenta una alta concentración de proteínas debido a que realiza una lisis química

y mecánica de las células bacterianas, acompañada por una solubilización principalmente de las proteínas debido a presencia de agentes caotrópicos, detergentes y anfólitos (Rogers and Bomgarden, 2016). Adicionalmente, este protocolo posee una etapa de precipitación que permite la concentración y purificación de proteínas (Peach, Marsh and MacPhee, 2012). Por otro lado, la extracción de proteínas en el protocolo de HCCl_3 -MetOH se basa en una lisis química, seguida de una concentración de estas en la interfaz entre la fase acuosa y la orgánica (Burnum-Johnson *et al.*, 2017).

Podemos observar que las muestras planctónicas obtuvieron una mayor concentración de proteína en todas las extracciones evaluadas, probablemente relacionado a la alta células (2 a 6×10^9 UFC/mL) (Tabla 5). Por otro lado, las muestras en estado de *biofilm* presentan una baja concentración de células viables, lo cual explica la disminución de la proteína extraída con respecto al estado planctónico (Hall-Stoodley, Costerton and Stoodley, 2004). En el caso de la extracción correspondiente a BugBuster®, se presentaron concentraciones por debajo del límite de detección (0.17 mg/mL), por lo que se demostró que esta metodología no es adecuada para la extracción de proteínas de muestras con baja biomasa. Utilizando la metodología de AF-ACN se obtuvieron concentraciones cercanas al límite de detección, con un máximo de 0.4 mg/mL; mientras que BL y HCCl_3 -MetOH concentraciones de $0.18 - 0.43$ mg/mL y $0.19 - 0.43$ mg/mL.

7.3. Análisis por Espectrometría de Masas MALDI-TOF

Se obtuvieron 71 y 45 espectros de masas de los estados *biofilm* y planctónico, respectivamente. En la Figura 12 se presentan la comparación de los espectros promedios del estado planctónico y *biofilm* para la cepa de *S. aureus* ATCC 43300, con los diferentes protocolos de extracción: HCCl_3 -MetOH (Figura 12A), BL (Figura 12B) y AF-ACN (Figura 12C). De los espectros adquiridos es importante resaltar que la región donde se evidenció la presencia de señales fue de 2000 a 8000 m/z . Estos resultados son consistentes con lo presentado por otros

investigadores para perfiles moleculares de *S. aureus*, en los cuales se observan un comportamiento similar de las señales (Yu et al., 2022b; Złoch et al., 2020b). La comparación de los espectros promedios del estado planctónico y *biofilm* para las cepa de *S. aureus* ATCC 29213 y BAA-44, con los diferentes protocolos de extracción: HCCl_3 -MetOH, BL y AF-ACN, presentaron resultados similares (ANEXO C).

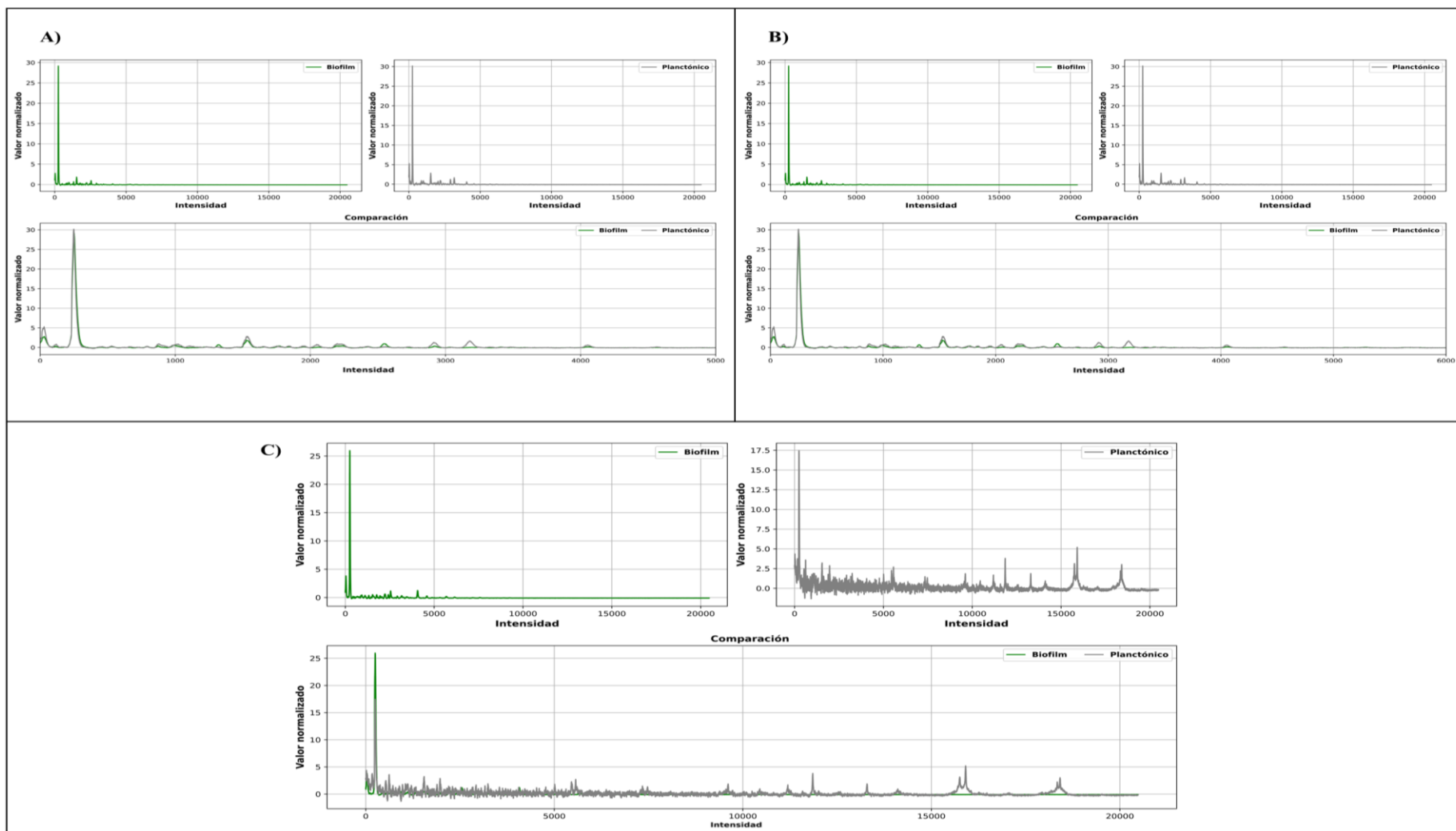


Figura 12. Promedio de las intensidades de las señales del perfil molecular de *S. aureus* ATCC 43300 en estado planctónico y biofilm. A) extracción $HCCl_3$ -MetOH, B) extracción Buffer de lisis, C) extracción AF-ACN.

Los espectros asociados a la extracción BugBuster[®] del estado planctónico presentaron señales con intensidades inferiores a 1×10^1 de abundancia, la cuales pueden ser debidas a la presencia de interferencias o a la pérdida de la muestra. La baja intensidad de señales o ausencia de estas no se puede atribuir a un exceso de proteína que afectara el proceso de ionización, ya que esta extracción fue la que presentó una menor concentración de proteína en todas las cepas. Debido a la sensibilidad del equipo, es poco probable que estos resultados se relacionen solamente con una baja cantidad de proteína. Por lo tanto, se atribuye este fenómeno a una pérdida o degradación de la proteínas durante la preparación, o a la presencia de contaminantes no proteicos que afecten la ionización.

7.4. Análisis por componentes principales

Para el análisis por componentes principales se usaron 40 espectros por cada uno de los protocolos de extracción. En la inspección inicial mediante un PCA, se observó que el comportamiento de los datos de cada extracción cuando se encuentran etiquetados por cepa es disperso. Solamente los datos pertenecientes a la cepa ATCC BAA-44 obtenidos con BL, se agruparon cuando se encontraban en estado de *biofilm* separándose de las demás cepas. sin embargo, esto no se observó en el estado planctónico. Debido a que la mayoría de los datos no presentó una agrupación por cepa, es posible utilizarlos en conjunto para ensayos posteriores. Esto es congruente con lo encontrado por otros autores, los cuales evidencian la dificultar de diferenciar entre diferentes cepas de microorganismos, ya que la expresión proteica tiende a ser muy similar entre ellos (Caputo et al., 2018b; Rodríguez-Temporal et al., 2023b; J. Wang et al., 2021; Złoch et al., 2020b).

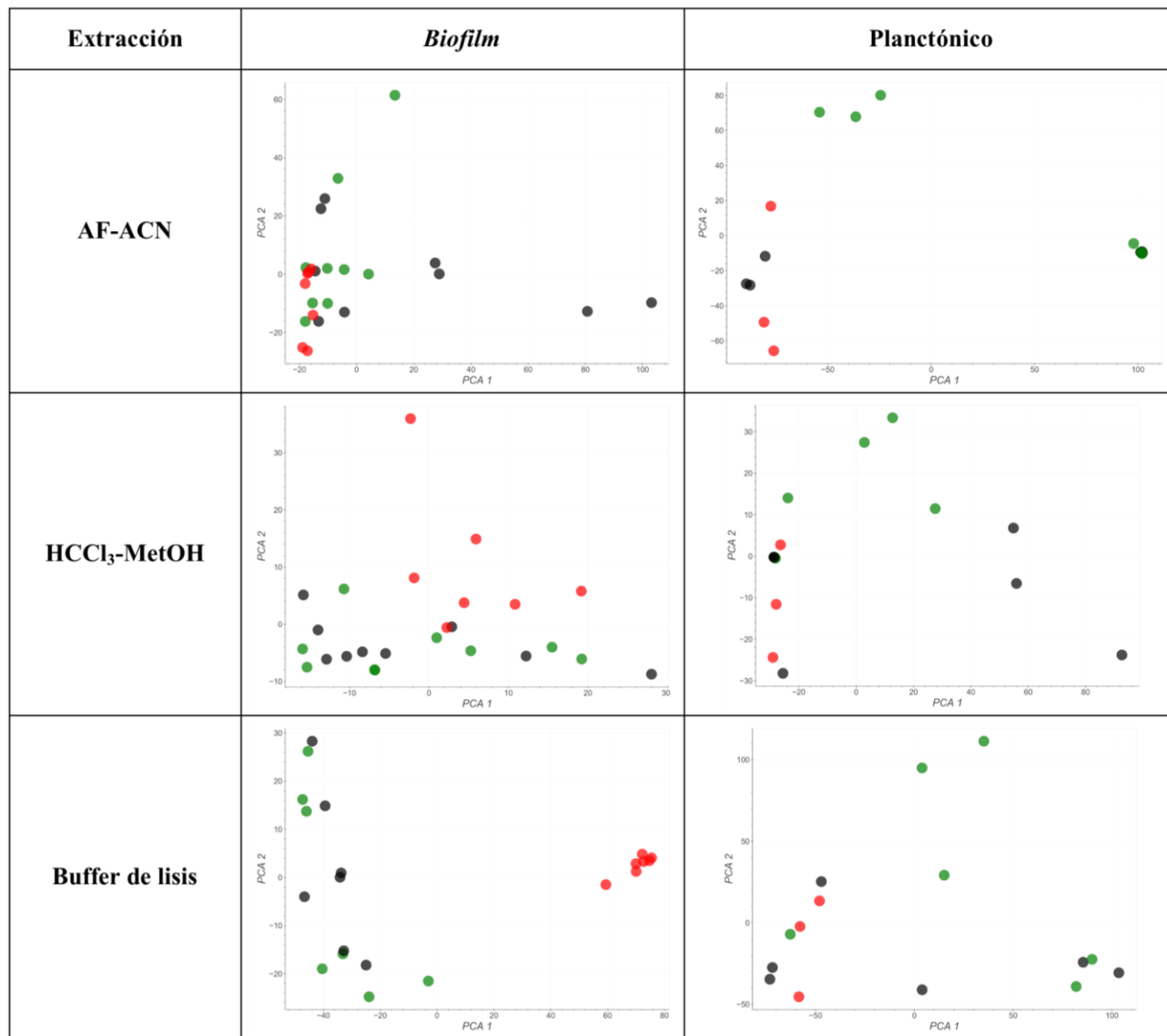


Figura 13. Gráficos de PCA de las extracciones por estados planctónico y biofilm. *S. aureus* ATCC 29213 (Negro), ATCC 43300 (verde) y ATCC BAA- 44 (Rojo).

En la Figura 14 se observan los resultados obtenidos para los espectros etiquetados según su estado (*azul, biofilm* y *rojo, planctónico*). Esta información se complementa con las gráficas de PCA de la Figura 14, en la cual la extracción AF-ACN separa mejor los datos de *biofilm* en color azul de los datos planctónico de color rojo. La varianza explicada para AF-ACN, HCCl₃-MetOH y BL fue del 95.29%, 95.81% y 98.3%, respectivamente.

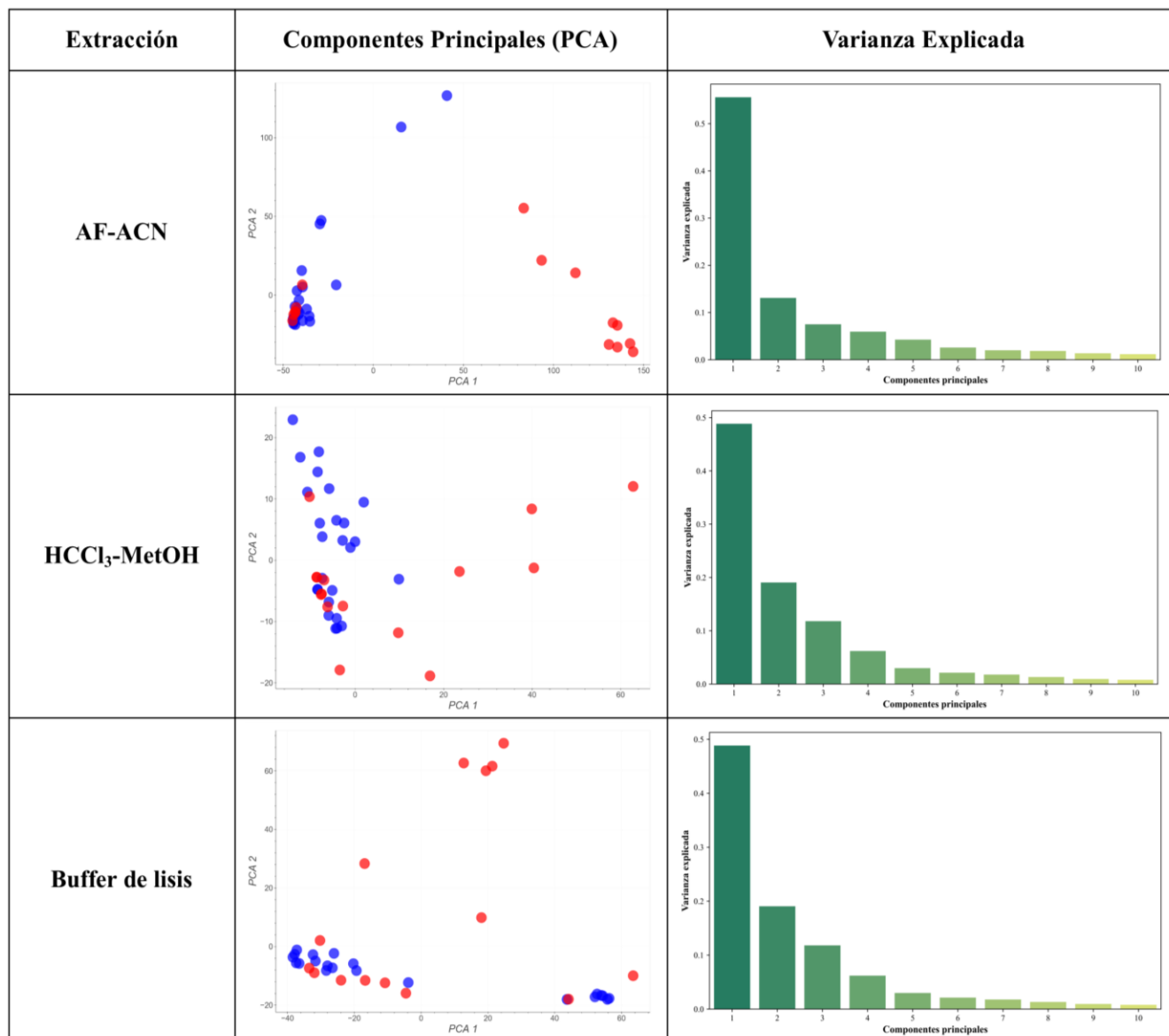


Figura 14. Componentes principales (azul, biofilm y rojo, planctónico) y varianza explicada para cada extracción.

7.5. Machine learning

En la Tabla 6 se observa que todos los modelos construidos con datos de la extracción HCCl₃-MetOH presentan los valores de precisión, *Recall* y *F1-score* más altos, con puntuaciones de *F1-score* en predicciones de *biofilm* y planctónico superiores al 0.92 y 0.90, respectivamente, con excepción del modelo XGB.

Tabla 6.

Métricas de classification_report de los modelos por protocolo de extracción.

Extracción AF-ACN					
Modelo	Clase	Precisión	Recall	F1-score	F1-score ponderado
SVC	<i>Biofilm</i>	0.93	1.00	0.96	0.95
	Planctónico	1.00	0.87	0.93	
XGB	<i>Biofilm</i>	0.76	0.88	0.81	0.74
	Planctónico	0.73	0.53	0.62	
RF	<i>Biofilm</i>	0.81	1.00	0.89	0.84
	Planctónico	1.00	0.60	0.75	
ANN	<i>Biofilm</i>	0.93	1.00	0.96	0.95
	Planctónico	1.00	0.87	0.93	
Extracción CHCl ₃ -MetOH					
Modelo	Clase	Precisión	Recall	F1-score	F1-score ponderado
SVC	<i>Biofilm</i>	0.96	0.96	0.96	0.95
	Planctónico	0.93	0.93	0.93	
XGB	<i>Biofilm</i>	0.86	1.00	0.93	0.9
	Planctónico	1.00	0.73	0.85	
RF	<i>Biofilm</i>	0.96	0.92	0.94	0.93
	Planctónico	0.88	0.93	0.9	
ANN	<i>Biofilm</i>	1.00	0.88	0.94	0.93
	Planctónico	0.83	1.00	0.91	
Extracción BL					
Modelo	Clase	Precisión	Recall	F1-score	F1-score ponderado
SVC	<i>Biofilm</i>	0.82	0.86	0.84	0.8
	Planctónico	0.79	0.73	0.76	
XGB	<i>Biofilm</i>	0.83	0.90	0.86	0.83
	Planctónico	0.85	0.73	0.79	
RF	<i>Biofilm</i>	0.80	0.95	0.87	0.83
	Planctónico	0.91	0.67	0.77	
ANN	<i>Biofilm</i>	0.76	0.90	0.83	0.77
	Planctónico	0.82	0.60	0.69	

Se podría inferir que la extracción HCCL₃-MetOH es la más apropiada para alimentar un modelo predictivo debido a que obtuvo un mejor desempeño general entre los modelos. Para los datos que provienen de una extracción de AF-ACN, los algoritmos que probablemente realicen las mejores predicciones serian SVC y ANN. Los datos provenientes de la extracción BL no son los más adecuados para la construcción de un modelo predictivo ya que cuenta con valores menores de *F1-score* ponderado. Sin embargo, los resultados observados en la Tabla 6 no son totalmente confiables. Esto se debe a que las dos clases o etiquetas (*biofilm* - planctónico) no tienen el mismo

número de datos, por lo tanto, hay un desbalance entre la cantidad de datos afectando especialmente la precisión reportada por la herramienta *classification_report*. El *F1-score* busca dar una mejor perspectiva del rendimiento de un modelo combinando y equilibrando, al relacionar los valores de precisión y *Recall*. Sin embargo, no se pueden usar en este caso porque provienen de la misma herramienta (*classification_report*). Para poder analizar el valor de *F1-score* teniendo en cuenta el desbalance de las clases se usó la función *cross_val_score* con el parámetro *scoring='f1'*; de esta manera se obtiene una métrica para cada *fold* de la validación cruzada *k-fold* con su respectivo promedio y desviación como se muestra en la Tabla 7.

Tabla 7

Métricas de cross_val_score de los modelos por protocolo de extracción

AF-ACN						
Modelo	fold-1	fold-2	fold-3	fold-4	F1-score	Desviación
SVC	1.00	1.00	0.75	1.00	0.94	0.11
XGB	0.66	0.50	0.50	0.73	0.60	0.10
RF	1.00	1.00	0.33	0.80	0.78	0.27
ANN	1.00	0.80	0.88	1.00	0.92	0.08
HCCl₃-MetOH						
Modelo	fold-1	fold-2	fold-3	fold-4	F1-score	Desviación
SVC	0.66	1.00	1.00	0.91	0.89	0.14
XGB	0.00	1.00	0.75	0.91	0.66	0.40
RF	0.50	1.00	1.00	0.91	0.85	0.21
ANN	0.50	1.00	1.00	0.92	0.85	0.21
BL						
Modelo	fold-1	fold-2	fold-3	fold-4	F1-score	Desviación
SVC	0.66	0.66	1.00	0.57	0.73	0.16
XGB	0.80	0.57	0.88	0.86	0.78	0.12
RF	0.80	0.00	1.00	0.86	0.66	0.40
ANN	0.50	0.40	1.00	0.57	0.61	0.23

Al comparar los valores de *F1-score* del *classification_report* con los obtenidos para el *cross_val_score* (Tabla 8), podemos observar una mayor diferencia relativa en los datos relacionados con las extracciones HCCl₃-MetOH y BL. Esto implica que los *F1-score* del

classification_report no son confiables para estas extracciones. Por el contrario, los modelos de SVC y ANN para datos que provienen de una extracción AF-ACN, mantienen sus valores de *F1-score* de 0.94 y 0.92, respectivamente.

Tabla 8

Comparación de los valores *F1-score* generados por las funciones *classification_report* y *cross_val_score*.

Extracción AF-ACN			
Modelo	F1-score ponderado	F1-score	Diferencia Relativa [%]
SVC	0.95	0.94	1.1
XGB	0.740	0.6	18.9
RF	0.840	0.78	7.1
ANN	0.950	0.92	3.2
Extracción HCl ₃ -MetOH			
Modelo	F1-score ponderado	F1-score	Diferencia Relativa [%]
SVC	0.95	0.89	6.3
XGB	0.9	0.66	26.7
RF	0.93	0.85	8.6
ANN	0.93	0.85	8.6
Extracción BL			
Modelo	F1-score ponderado	F1-score	Diferencia Relativa [%]
SVC	0.8	0.73	8.8
XGB	0.83	0.78	6.0
RF	0.83	0.66	20.5
ANN	0.77	0.61	20.8

La Figura 15 presenta las matrices de confusión de cada modelo según el protocolo de extracción. Los modelos generados a partir de los datos de la extracción BL presentan en su diagonal secundaria (falsos positivos y falsos negativos) los valores más altos en comparación con las demás extracciones, indicando un bajo rendimiento en la predicción de los estados planctónico y *biofilm*. Los modelos de SVC y ANN de la extracción AF-ACN solo presentan dos falsos positivos en el estado planctónico e identifica la totalidad de las etiquetas del *biofilm*. Por el contrario, los modelos RF y XGB presentaron 6 y 10 falsas identificaciones, respectivamente. Los resultados obtenidos para el modelo de SVC en la extracción HCl₃-MetOH son más equilibrados,

presentando únicamente dos falsas identificaciones, una de *biofilm* y otra del estado planctónico. Los modelos de XGB y RF presentaron más de 3 falsas identificaciones y el modelo ANN solo presenta 3 falsas identificaciones de *biofilm*.

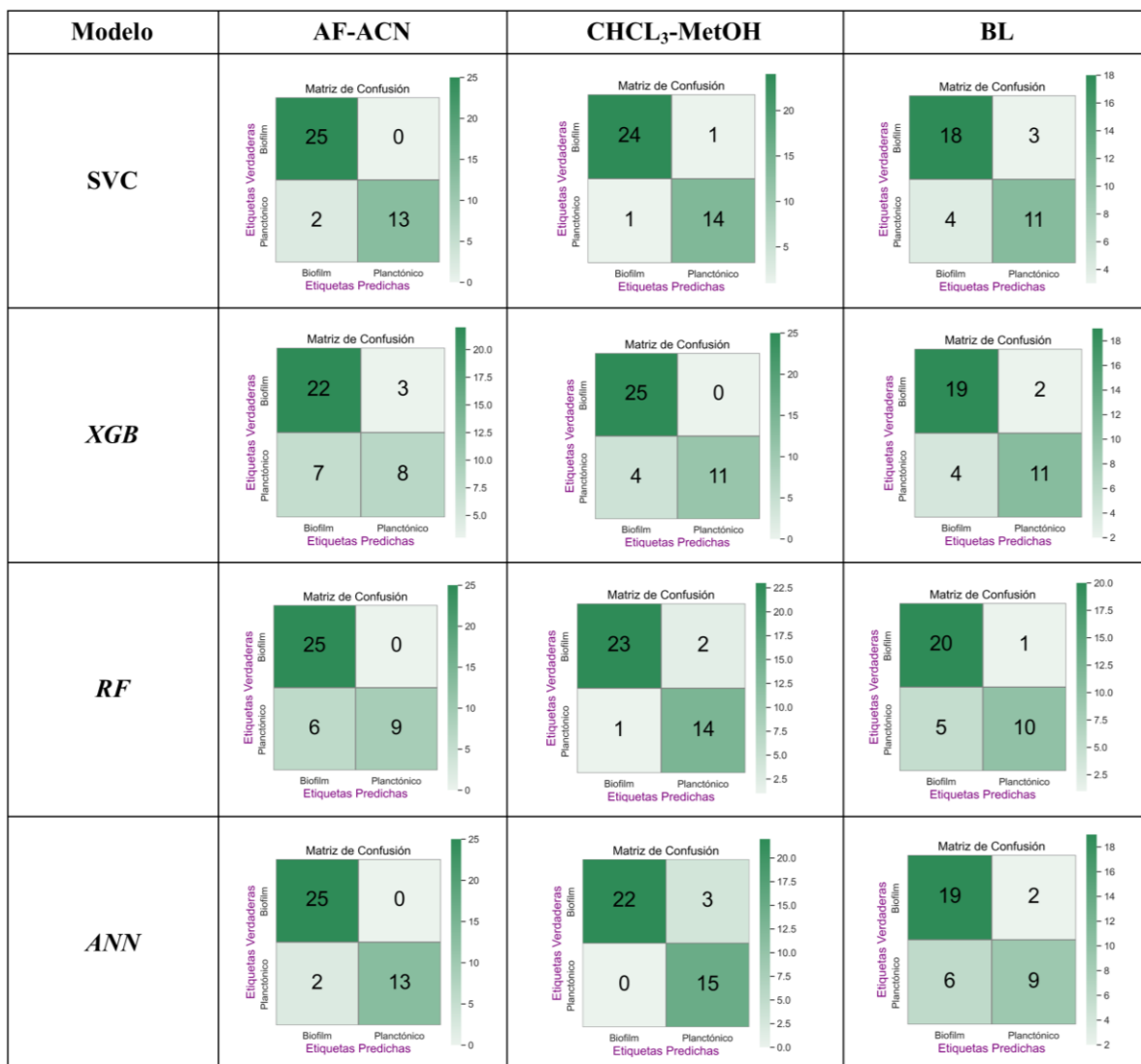


Figura 15. Matriz de confusión de los modelos obtenidos para cada una de las extracciones.

Los resultados presentados anteriormente demuestran que los modelos de SVC y ANN alimentados con los datos del protocolo de extracción de proteínas AF-ACN permitieron una mejor predicción del estado de *biofilm* y planctónico. Sin embargo, la extracción HCL₃-MetOH genera datos prometedores que se podrían llegar optimizar para obtener mejores modelos.

Finalmente se observó que cada uno de los algoritmos usados en la generación de los modelos predictivos, puede mejorarse modificando los parámetros asociados a cada uno de ellos y de esta manera ajustar el modelo con el fin de obtener mejores resultados. Sin embargo, esta no era la finalidad de este trabajo de investigación y los parámetros usados se seleccionaron exclusivamente en función de la cantidad de datos disponibles.

8. Conclusiones

Los protocolos HCCl₃-MetOH, AF-ACN permitieron la extracción de proteínas de los estados planctónico y de *biofilm* de las cepas de *S. aureus* ATCC BAA-44, ATCC 43300 y ATCC 292131, de manera que se pudo realizar la adquisición de espectros de masas relacionados con el perfil molecular de cada cepa.

El protocolo BugBuster® presentó concentraciones de proteína inferiores a 0.50 mg/mL en estado planctónico y 0.19 mg/mL en el estado de *biofilm*. La extracción HCCl₃-MetOH (*Biofilm* \geq 0.19 mg/mL y Planctónico \geq 1.0 mg/mL) y BL (*Biofilm* \geq 0.18 mg/mL y Planctónico \geq 1.9 mg/mL).

Los modelos predictivos se SVC y ANN alimentados con los espectros obtenidos a partir de la extracción AF-ACN permitieron la predicción de los estados de planctónico y *biofilm* valores superiores a 0.92 en la métrica *FI-score*, presentando solo dos falsas identificaciones en general, ambas asociadas al estado planctónico.

Recomendaciones

Como el presente trabajo de investigación fue una aproximación preliminar, se recomienda aumentar la cantidad de muestras por estado (planctónico y *biofilm*) para los protocolos de extracción de proteínas AF-ACN y HCCl_3 -MetOH. De esta manera seleccionar solo un protocolo y un solo algoritmo con el fin de optimizar el modelo predictivo. También es importante contemplar la inclusión de aislados clínicos, las cuales permitirán una mayor comprensión de las señales asociadas a la formación de *biofilm*, y mejorarán la robustez del modelo predictivo.

Las Extracciones BL y el kit BugBuster[®] probablemente requieran de un diseño experimental que permita mejorar la reproducibilidad de los espectros de masas adquiridos. De esta manera evaluar diferentes algoritmos y determinar si los datos provenientes de estas extracciones no permiten construir un modelo que diferencie el estado planctónico del de *Biofilm*. También se recomienda aumentar la cantidad de datos, así como el uso de otros algoritmos y modelos de clasificación con el fin de evaluar su potencial en la obtención de señales proteicas únicas asociadas al *biofilm*.

Iniciar una fase de calibración de los modelos obtenidos en este trabajo para evaluar el desempeño y de esta manera continuar el estudio de señales con miras a la identificación de posibles blancos terapéuticos o marcadores moleculares.

Bibliografía

Albrethsen, J. (2007) 'Reproducibility in Protein Profiling by MALDI-TOF Mass Spectrometry'. Available at: <https://doi.org/10.1373/clinchem.2006.082644>.

Alciaturi, C. *et al.* (2010) 'Aplicación de métodos matemáticos y estadísticos en la industria', *Revista Técnica de la Facultad de Ingeniería Universidad del Zulia*, 33(2), pp. 164–168. Available at: http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0254-07702010000200008&lng=es&nrm=iso&tlng=es (Accessed: 25 November 2023).

Alizadeh, M. *et al.* (2021) 'MALDI-TOF Mass Spectroscopy Applications in Clinical Microbiology', *Advances in Pharmacological and Pharmaceutical Sciences*. Hindawi Limited. Available at: <https://doi.org/10.1155/2021/9928238>.

Asghari, E. *et al.* (2021a) 'Identification of microorganisms from several surfaces by maldi-tof ms: P. aeruginosa is leading in biofilm formation', *Microorganisms*, 9(5). Available at: <https://doi.org/10.3390/microorganisms9050992>.

Asghari, E. *et al.* (2021b) 'Identification of microorganisms from several surfaces by maldi-tof ms: P. aeruginosa is leading in biofilm formation', *Microorganisms*, 9(5). Available at: <https://doi.org/10.3390/microorganisms9050992>.

Ashfaq, M.Y., Da'na, D.A. and Al-Ghouti, M.A. (2022) 'Application of MALDI-TOF MS for identification of environmental bacteria: A review', *Journal of Environmental Management*, 305, p. 114359. Available at: <https://doi.org/10.1016/J.JENVMAN.2021.114359>.

Bonaccorso, G. (2017) *Machine learning algorithms: reference guide for popular algorithms for data science and machine learning*.

Bradford, M.M. (1976) 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.', *Analytical biochemistry*, 72, pp. 248–54. Available at: <https://doi.org/10.1006/abio.1976.9999>.

Burnum-Johnson, K.E. *et al.* (2017) 'MPLEx: a method for simultaneous pathogen inactivation and extraction of samples for multi-omics profiling', *The Analyst*, 142(3), pp. 442–448. Available at: <https://doi.org/10.1039/C6AN02486F>.

Cámara, M. *et al.* (2022) 'Economic significance of biofilms: a multidisciplinary and cross-sectoral challenge', *npj Biofilms and Microbiomes*, 8(1). Available at: <https://doi.org/10.1038/s41522-022-00306-y>.

Caputo, P. *et al.* (2018a) 'Use of MALDI-TOF MS to discriminate between biofilm-producer and non-producer strains of staphylococcus epidermidis', *International Journal of Environmental Research and Public Health*, 15(8). Available at: <https://doi.org/10.3390/ijerph15081695>.

Caputo, P. *et al.* (2018b) 'Use of MALDI-TOF MS to discriminate between biofilm-producer and non-producer strains of staphylococcus epidermidis', *International Journal of Environmental Research and Public Health*, 15(8). Available at: <https://doi.org/10.3390/ijerph15081695>.

Cincarova, L. *et al.* (2016) 'Changes in the Expression of Biofilm-Associated Surface Proteins in Staphylococcus aureus Food-Environmental Isolates Subjected to Sublethal Concentrations of Disinfectants', *BioMed Research International* [Preprint]. Available at: <https://doi.org/10.1155/2016/4034517>.

Coombes, K.R., Baggerly, K.A. and Morris, J.S. (2007) 'Pre-processing mass spectrometry data', *Fundamentals of Data Mining in Genomics and Proteomics*, pp. 79–102. Available at: https://doi.org/10.1007/978-0-387-47509-7_4/COVER.

Dematheis, F. *et al.* (2022) 'Machine Learning Algorithms for Classification of MALDI-TOF MS Spectra from Phylogenetically Closely Related Species *Brucella melitensis*, *Brucella abortus* and *Brucella suis*', *Microorganisms*, 10(8), p. 1658. Available at: <https://doi.org/10.3390/MICROORGANISMS10081658/S1>.

Desaire, H., Go, E.P. and Hua, D. (2022) 'Advances, obstacles, and opportunities for machine learning in proteomics', *Cell Reports Physical Science*, 3(10), p. 101069. Available at: <https://doi.org/10.1016/J.XCRP.2022.101069>.

Donlan, R.M. (2002) 'Biofilms: Microbial life on surfaces', *Emerging Infectious Diseases*, pp. 881–890. Available at: <https://doi.org/10.3201/eid0809.020063>.

Florio, W. *et al.* (2020) 'Detection of Antibiotic-Resistance by MALDI-TOF Mass Spectrometry: An Expanding Area', *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A. Available at: <https://doi.org/10.3389/fcimb.2020.572909>.

Gonzales, J.I. (2021) *Detección de fallas en vehículos aéreos multirrotor mediante aprendizaje estadístico*. Universidad de Buenos Aires.

Grace, D. and Fetsch, A. (2018) 'Staphylococcus aureus a foodborne pathogen: epidemiology, detection, characterization, prevention, and control: an overview', in A. Fetsch (ed.). San Diego, CA: Andre G. Wolff.

Grix, R. *et al.* (1988) 'A time-of-flight mass analyzer with high resolving power', *Rapid Communications in Mass Spectrometry*, 2(5), pp. 83–85. Available at: <https://doi.org/10.1002/RCM.1290020503>.

Gross, Jurgen H. (2004) 'Matrix-Assisted Laser Desorption/Ionization', in *Mass Spectrometry*, pp. 411–440.

Gross, Jürgen H. (2004) 'Time-of-Flight Instruments', in *Mass Spectrometry*. Springer B. Germany, pp. 113–130.

Hajduk, J., Matysiak, J. and Kokot, Z.J. (2016) 'Challenges in biomarker discovery with MALDI-TOF MS', *Clinica Chimica Acta*. Elsevier B.V., pp. 84–98. Available at: <https://doi.org/10.1016/j.cca.2016.04.033>.

Hall-Stoodley, L., Costerton, J.W. and Stoodley, P. (2004) 'Bacterial biofilms: From the natural environment to infectious diseases', *Nature Reviews Microbiology*, pp. 95–108. Available at: <https://doi.org/10.1038/nrmicro821>.

Herigstad, B., Hamilton, M. and Heersink, J. (2001) 'How to optimize the drop plate method for enumerating bacteria', *Journal of Microbiological Methods*, 44(2), pp. 121–129. Available at: [https://doi.org/10.1016/S0167-7012\(00\)00241-4](https://doi.org/10.1016/S0167-7012(00)00241-4).

Høiby, N. *et al.* (2015) 'ESCMID guideline for the diagnosis and treatment of biofilm infections 2014', *Clinical Microbiology and Infection*, 21(S1), pp. S1–S25. Available at: <https://doi.org/10.1016/J.CMI.2014.10.024>.

Iskandar, K. *et al.* (2022) 'Antibiotic Discovery and Resistance: The Chase and the Race', *Antibiotics*. MDPI. Available at: <https://doi.org/10.3390/antibiotics11020182>.

Iskender, S. *et al.* (2023) 'Rapid determination of colistin resistance in *Klebsiella pneumoniae* by MALDI-TOF peak based machine learning algorithm with MATLAB', *Diagnostic Microbiology and Infectious Disease*, 107(4), p. 116052. Available at: <https://doi.org/10.1016/J.DIAGMICROBIO.2023.116052>.

Jolliffe, I.T. and Cadima, J. (no date) 'Principal component analysis: a review and recent developments'. Available at: <https://doi.org/10.1098/rsta.2015.0202>.

Karas, M., Bachmann, D. and Hillenkamp, F. (1985) 'Influence of the Wavelength in High-Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules', *Analytical Chemistry*, 57(14), pp. 2935–2939. Available at: https://doi.org/10.1021/AC00291A042/ASSET/AC00291A042.FP.PNG_V03.

Karas, M. and Hillenkamp, F. (1988) 'Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons', *Analytical Chemistry*, 60(20), pp. 2299–2301. Available at: https://doi.org/10.1021/AC00171A028/ASSET/AC00171A028.FP.PNG_V03.

Kim, S.J. *et al.* (2018) 'Surface proteins and the formation of biofilms by *Staphylococcus aureus*', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1860(3), pp. 749–756. Available at: <https://doi.org/10.1016/j.bbamem.2017.12.003>.

Kranjec, C. *et al.* (2021) 'Staphylococcal Biofilms: Challenges and Novel Therapeutic Perspectives', *Antibiotics 2021, Vol. 10, Page 131*, 10(2), p. 131. Available at: <https://doi.org/10.3390/ANTIBIOTICS10020131>.

Krásný, L., Hynek, R. and Hochel, I. (2013) 'Identification of bacteria using mass spectrometry techniques', *International Journal of Mass Spectrometry*, 353, pp. 67–79. Available at: <https://doi.org/10.1016/j.ijms.2013.04.016>.

Latorre, M.C. *et al.* (2022) 'The classification of *Staphylococcus aureus* strains by biofilm production differs depending on the method used', *Enfermedades Infecciosas y Microbiología Clínica*, 40(3), pp. 134–137. Available at: <https://doi.org/10.1016/j.eimc.2020.11.002>.

Leme, R.C.P., Bispo, P.J.M. and Salles, M.J. (2021) 'Community-genotype methicillin-resistant *Staphylococcus aureus* skin and soft tissue infections in Latin America: a systematic

review’, *Brazilian Journal of Infectious Diseases*. Elsevier Editora Ltda. Available at: <https://doi.org/10.1016/j.bjid.2021.101539>.

Li, L. *et al.* (2023) ‘Dynamic response prediction of underwater explosive vessel based on LOO-XGBoost model’, *Neural Computing and Applications*, 35(36), pp. 25057–25067. Available at: <https://doi.org/10.1007/s00521-023-08613-x>.

Lovison, O.A. *et al.* (2023) ‘Validation of a MALDI-TOF MS Method for SARS-CoV-2 Detection on the Bruker Biotyper and Nasopharyngeal Swabs: A Brazil—UK Collaborative Study’, *Diagnostics* 2023, Vol. 13, Page 1470, 13(8), p. 1470. Available at: <https://doi.org/10.3390/DIAGNOSTICS13081470>.

Martínez, A. *et al.* (2023) ‘Effect of Essential Oil from *Lippia origanoides* on the Transcriptional Expression of Genes Related to Quorum Sensing, Biofilm Formation, and Virulence of *Escherichia coli* and *Staphylococcus aureus*’, *Antibiotics*, 12(5). Available at: <https://doi.org/10.3390/antibiotics12050845>.

Miller, J.N. (James N.) and Miller, J.C. (Jane C. (2010) ‘Statistics and Chemometrics for Analytical Chemistry, 6th Edition’, p. 268.

Monedeiro, F. *et al.* (2021) ‘Influence of culture medium on bacterial molecular profiles in different ionization modes with the use of computational methods’, *International Journal of Mass Spectrometry*, 466. Available at: <https://doi.org/10.1016/j.ijms.2021.116614>.

Moormeier, D.E. and Bayles, K.W. (2017) ‘*Staphylococcus aureus* biofilm: a complex developmental organism’, *Molecular Microbiology*, 104(3), pp. 365–376. Available at: <https://doi.org/10.1111/mmi.13634>.

Nalepa, J. and Kawulok, M. (2019) ‘Selecting training sets for support vector machines: a review’, *Artificial Intelligence Review*, 52(2), pp. 857–900. Available at: <https://doi.org/10.1007/s10462-017-9611-1>.

Van Oosten, L.N. and Klein, C.D. (2020) ‘Machine Learning in Mass Spectrometry: A MALDI-TOF MS Approach to Phenotypic Antibacterial Screening’, *Journal of Medicinal Chemistry*, 63(16), pp. 8849–8856. Available at: https://doi.org/10.1021/ACS.JMEDCHEM.0C00040/ASSET/IMAGES/LARGE/JM0C00040_0004.JPEG.

O’Toole, G.A. (2011) ‘Microtiter Dish Biofilm Formation Assay’, *Journal of Visualized Experiments*, (47), pp. 10–11. Available at: <https://doi.org/10.3791/2437>.

Pal, K. and Patel, Biraj. V. (2020) ‘Data Classification with k-fold Cross Validation and Holdout Accuracy Estimation Methods with 5 Different Machine Learning Techniques’, in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, pp. 83–87. Available at: <https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00016>.

Parmar, A., Katariya, R. and Patel, V. (2019) ‘A Review on Random Forest: An Ensemble Classifier’, in Springer, Cham, pp. 758–763. Available at: https://doi.org/10.1007/978-3-030-03146-6_86.

Peach, M., Marsh, N. and MacPhee, D.J. (2012) ‘Protein Solubilization: Attend to the Choice of Lysis Buffer’, in Humana Press, Totowa, NJ, pp. 37–47. Available at: https://doi.org/10.1007/978-1-61779-821-4_4.

Pereira, F.D.E.S. *et al.* (2015) ‘Use of MALDI-TOF mass spectrometry to analyze the molecular profile of *Pseudomonas aeruginosa* biofilms grown on glass and plastic surfaces’,

Microbial Pathogenesis, 86, pp. 32–37. Available at: <https://doi.org/10.1016/j.micpath.2015.07.005>.

Peris-Díaz, M.D. and Krężel, A. (2021) ‘A guide to good practice in chemometric methods for vibrational spectroscopy, electrochemistry, and hyphenated mass spectrometry’, *TrAC Trends in Analytical Chemistry*, 135, p. 116157. Available at: <https://doi.org/10.1016/J.TRAC.2020.116157>.

Pietrocola, G. *et al.* (2022) ‘Colonization and Infection of Indwelling Medical Devices by *Staphylococcus aureus* with an Emphasis on Orthopedic Implants’, *International Journal of Molecular Sciences*. MDPI. Available at: <https://doi.org/10.3390/ijms23115958>.

Rodríguez-Temporal, D., Díez, R., *et al.* (2023a) ‘Determination of the ability of matrix-assisted laser desorption ionization time-of-flight mass spectrometry to identify high-biofilm-producing strains’, *Frontiers in Microbiology*, 13. Available at: <https://doi.org/10.3389/fmicb.2022.1104405>.

Rodríguez-Temporal, D., Díez, R., *et al.* (2023b) ‘Determination of the ability of matrix-assisted laser desorption ionization time-of-flight mass spectrometry to identify high-biofilm-producing strains’, *Frontiers in Microbiology*, 13. Available at: <https://doi.org/10.3389/fmicb.2022.1104405>.

Rodríguez-Temporal, D., Herrera, L., *et al.* (2023) ‘Identification of *Mycobacterium abscessus* Subspecies by MALDI-TOF Mass Spectrometry and Machine Learning’, *Journal of Clinical Microbiology*. Edited by M.B. Miller, 61(1), pp. e0111022–e0111022. Available at: <https://doi.org/10.1128/jcm.01110-22>.

Rogers, J.C. and Bomgardner, R.D. (2016) 'Sample Preparation for Mass Spectrometry-Based Proteomics; from Proteomes to Peptides', in: Springer, Cham, pp. 43–62. Available at: https://doi.org/10.1007/978-3-319-41448-5_3.

Sanchini, A. (2022) 'Recent Developments in Phenotypic and Molecular Diagnostic Methods for Antimicrobial Resistance Detection in *Staphylococcus aureus*: A Narrative Review', *Diagnostics*, 12(1). Available at: <https://doi.org/10.3390/DIAGNOSTICS12010208/S1>.

Schulthess, B. *et al.* (2014) 'Evaluation of the Bruker MALDI Biotyper for Identification of Gram-Positive Rods: Development of a Diagnostic Algorithm for the Clinical Laboratory', *Journal of Clinical Microbiology*. Edited by R. Patel, 52(4), pp. 1089–1097. Available at: <https://doi.org/10.1128/JCM.02399-13>.

Silva, N.B.S., Marques, L.A. and Röder, D.D.B. (2021) 'Diagnosis of biofilm infections: current methods used, challenges and perspectives for the future', *Journal of Applied Microbiology*, 131(5), pp. 2148–2160. Available at: <https://doi.org/10.1111/JAM.15049>.

Subasi, A. (2020) *Practical Machine Learning for Data Analysis Using Python, Practical Machine Learning for Data Analysis Using Python*. Elsevier. Available at: <https://doi.org/10.1016/B978-0-12-821379-7.00008-4>.

Tsuchida, S. (2018) 'Application of MALDI-TOF for Bacterial Identification', in *The Use of Mass Spectrometry Technology (MALDI-TOF) in Clinical Microbiology*. Elsevier, pp. 101–112. Available at: <https://doi.org/10.1016/B978-0-12-814451-0.00007-1>.

Vani, S., Vadakkan, K. and Mani, B. (2023) 'A narrative review on bacterial biofilm: its formation, clinical aspects and inhibition strategies', *Future Journal of Pharmaceutical Sciences*, 9(1). Available at: <https://doi.org/10.1186/s43094-023-00499-9>.

Vázquez-Sánchez, D. and Rrodríguez-López, P. (2018) 'Chapter 5 - Biofilm formation of *S. aureus*', in. Elsevier Inc., pp. 87–103.

Wang, H.Y. *et al.* (2018) 'A new scheme for strain typing of methicillin-resistant *Staphylococcus aureus* on the basis of matrix-assisted laser desorption ionization time-of-flight mass spectrometry by using machine learning approach', *PLoS ONE*, 13(3). Available at: <https://doi.org/10.1371/journal.pone.0194289>.

Wang, J. *et al.* (2021) 'Evaluation of three sample preparation methods for the identification of clinical strains by using two MALDI-TOF MS systems', *Journal of Mass Spectrometry*, 56(2). Available at: <https://doi.org/10.1002/jms.4696>.

Wang, Z. *et al.* (2023) 'Machine learning combined with MALDI-TOF MS has the potential ability to identify serotypes of the avian pathogen *Riemerella anatipestifer*', *Journal of Applied Microbiology*, 134(2). Available at: <https://doi.org/10.1093/JAMBIO/LXAC075>.

Wei, Y. *et al.* (2021) 'Machine-learning-enhanced time-of-flight mass spectrometry analysis', *Patterns*, 2(2), p. 100192. Available at: <https://doi.org/10.1016/J.PATTER.2020.100192>.

Weis, C. *et al.* (2020) 'Topological and kernel-based microbial phenotype prediction from MALDI-TOF mass spectra', *Bioinformatics*, 36(Supplement_1), pp. i30–i38. Available at: <https://doi.org/10.1093/BIOINFORMATICS/BTAA429>.

Weis, C. *et al.* (2022) 'Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning', *Nature Medicine*, 28(1), pp. 164–174. Available at: <https://doi.org/10.1038/s41591-021-01619-9>.

Wigmann, É.F. *et al.* (2019) 'MALDI-TOF MS fingerprinting for identification and differentiation of species within the *Fusarium fujikuroi* species complex', *Applied Microbiology*

and Biotechnology, 103(13), pp. 5323–5337. Available at: <https://doi.org/10.1007/S00253-019-09794-Z/FIGURES/5>.

World Health Organization (2017) *La OMS publica la lista de las bacterias para las que se necesitan urgentemente nuevos antibióticos*. Available at: <https://www.who.int/es/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed> (Accessed: 4 May 2023).

World Health Organization (2021) *Patógenos multirresistentes que son prioritarios para la OMS*. Available at: <https://www.paho.org/es/noticias/4-3-2021-patogenos-multirresistentes-que-son-prioritarios-para-oms> (Accessed: 5 May 2023).

Wu, X. *et al.* (2022) ‘Prediction of Mycobacterium tuberculosis drug resistance by nucleotide MALDI-TOF-MS.’, *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*, 121, pp. 47–54. Available at: <https://doi.org/10.1016/j.ijid.2022.04.061>.

Wu, Y. chen and Feng, J. wen (2018) ‘Development and Application of Artificial Neural Network’, *Wireless Personal Communications*, 102(2), pp. 1645–1656. Available at: <https://doi.org/10.1007/s11277-017-5224-x>.

Yu, J. *et al.* (2022a) ‘Rapid Identification of Methicillin-Resistant Staphylococcus aureus Using MALDI-TOF MS and Machine Learning from over 20,000 Clinical Isolates’, *Microbiology Spectrum*, 10(2). Available at: https://doi.org/10.1128/SPECTRUM.00483-22/SUPPL_FILE/SPECTRUM00483-22_SUPP_1_SEQ4.PDF.

Yu, J. *et al.* (2022b) ‘Rapid Identification of Methicillin-Resistant Staphylococcus aureus Using MALDI-TOF MS and Machine Learning from over 20,000 Clinical Isolates’, *Microbiology Spectrum*, 10(2). Available at: <https://doi.org/10.1128/spectrum.00483-22>.

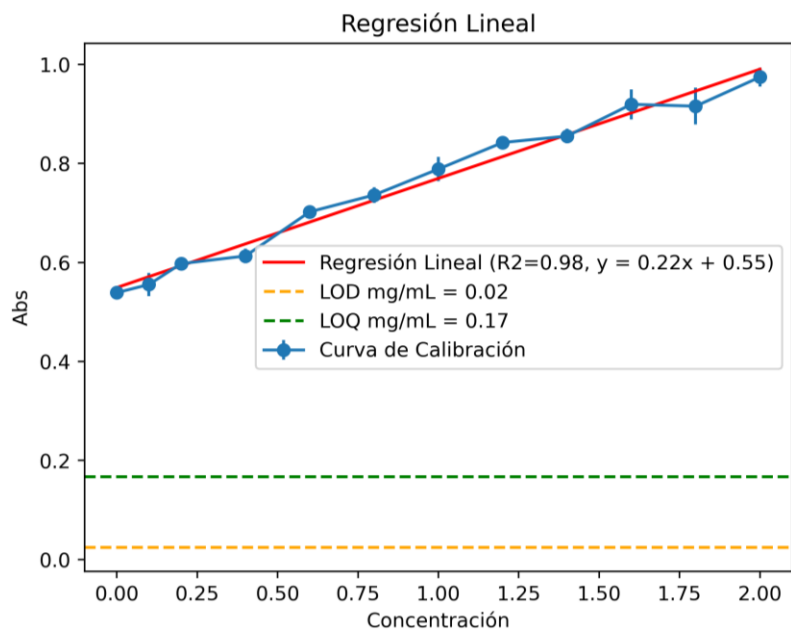
Zhu, W. *et al.* (2015) 'Evaluation of the Biotyper MALDI-TOF MS system for identification of Staphylococcus species', *Journal of Microbiological Methods*, 117, pp. 14–17. Available at: <https://doi.org/10.1016/j.mimet.2015.07.014>.

Złoch, M. *et al.* (2020a) 'Study on Molecular Profiles of Staphylococcus aureus Strains: Spectrometric Approach', *Molecules*, 25(21). Available at: <https://doi.org/10.3390/molecules25214894>.

Złoch, M. *et al.* (2020b) 'Study on Molecular Profiles of Staphylococcus aureus Strains: Spectrometric Approach', *Molecules*, 25(21). Available at: <https://doi.org/10.3390/molecules25214894>.

Złoch, M. *et al.* (2020c) 'Study on Molecular Profiles of Staphylococcus aureus Strains: Spectrometric Approach', *Molecules*, 25(21). Available at: <https://doi.org/10.3390/molecules25214894>.

Zvezdanova, M.E. *et al.* (2022) 'Detection of azole resistance in *Aspergillus fumigatus* complex isolates using MALDI-TOF mass spectrometry', *Clinical Microbiology and Infection*, 28(2), pp. 260–266. Available at: <https://doi.org/10.1016/J.CMI.2021.06.005>.

ANEXO A. Cuantificación de proteínas por el método de Bradford

Límite de detección: 0.02 mg/mL

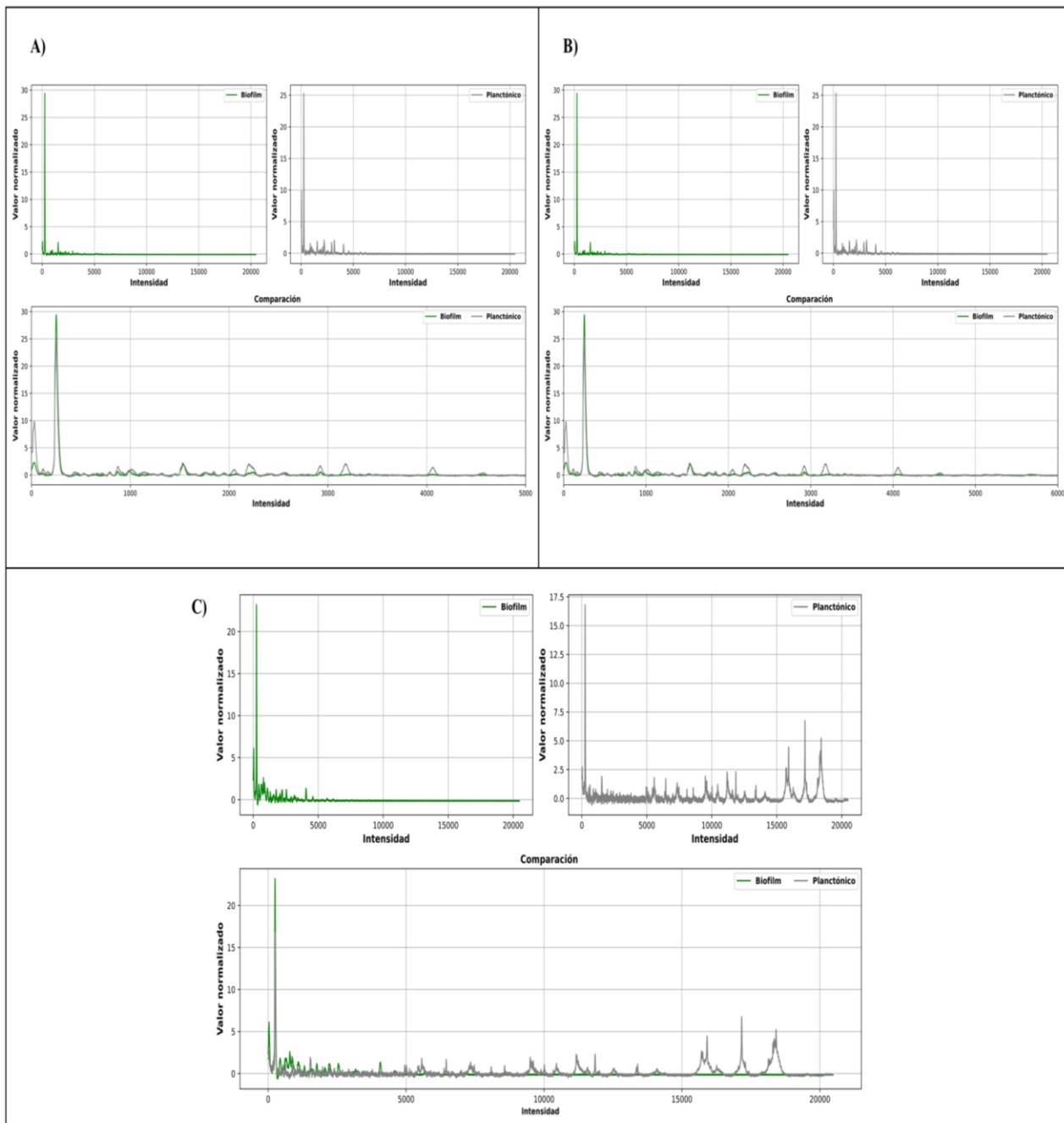
Límite de cuantificación: 0.17 mg/mL

ANEXO B. Rutinas en Python

Archivo en línea: <https://github.com/Darkilljack>

ANEXO C. Espectros de masa representativos del perfil molecular de *S. aureus* en estado planctónico y *biofilm* empleando diferentes protocolos de extracción de proteínas.

Espectros de masa representativos del perfil molecular de *S. aureus* ATCC 29213 en estado planctónico y *biofilm*. A) extracción HCC13-MetOH, B) extracción Buffer de lisis, C) extracción AF-ACN.



Espectros de masa representativos del perfil molecular de *S. aureus* ATCC BAA-44 en estado planctónico y biofilm. A) extracción HCC13-MetOH, B) extracción Buffer de lisis, C) extracción AF-ACN.

