

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE FARMACIA



TESIS DOCTORAL

**Aplicación de técnicas de diagnóstico emergentes
(espectrometría de masas MALDI-TOF y espectroscopía de
infrarrojos FT-IR) a la caracterización de microorganismos
de origen clínico**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Ana Candela González

Directoras

Belén Rodríguez Sánchez
Patricia Muñoz García

Madrid

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE FARMACIA



TESIS DOCTORAL

**APLICACIÓN DE TÉCNICAS DE DIAGNÓSTICO
EMERGENTES (ESPECTROMETRÍA DE MASAS
MALDI-TOF Y ESPECTROSCOPIA DE
INFRARROJOS FT-IR) A LA CARACTERIZACIÓN
DE MICROORGANISMOS DE ORIGEN CLÍNICO**

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

Ana Candela González

DIRECTORAS

Belén Rodríguez Sánchez
Patricia Muñoz García

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

PROGRAMA DE DOCTORADO EN MICROBIOLOGÍA Y PARASITOLOGÍA



TESIS DOCTORAL

**APLICACIÓN DE TÉCNICAS DE DIAGNÓSTICO EMERGENTES
(ESPECTROMETRÍA DE MASAS MALDI-TOF Y ESPECTROSCOPIA
DE INFRARROJOS FT-IR) A LA CARACTERIZACIÓN DE
MICROORGANISMOS DE ORIGEN CLÍNICO**

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR:

ANA CANDELA GONZÁLEZ

DIRECTORAS:

BELÉN RODRÍGUEZ SÁNCHEZ

PATRICIA MUÑOZ GARCÍA

AGRADECIMIENTOS

Esta tesis, no es una tesis al uso. Es una tesis entre dos tierras. El doble de difícil de llevar a cabo, pero el doble de gratificante. Varias personas me habían dicho que estas serían las páginas más difíciles de escribir, pero no creo que sea verdad, porque tengo el doble de gente a la que agradecer, y eso lo hace más fácil.

Primero, a los de allí.

Mi familia adoptiva durante 5 años, tan lejos de casa, pero que muchas veces fuisteis hogar.

Gracias primeramente a mis directoras, por haberme ayudado a crecer, tanto científica como personalmente.

A Belén, por ser directora, amiga y confidente. Repito las palabras que te dediqué el día de mi despedida de la residencia: Me llevo a una amiga. Gracias por haberme acompañado y guiado en cada paso de esta tesis, aprendiendo y descubriendo a mi lado las maravillas del mundo MALDI-TOF y más recientemente del FTIR y *Machine Learning*. Por haber tenido siempre disponible un oído para escucharme y un hombro sobre el que llorar. Espero que podamos seguir compartiendo ciencia interhospitalaria y congresos muchos años más. Como tú dices, este año es “año par” y vamos a ir a por todas. Empecemos por esta tesis.

Y a mí otra directora, la Dra. Patricia Muñoz, gracias por brindarme la oportunidad de desarrollar esta tesis en el excelente servicio de Microbiología del Hospital Gregorio Marañón, que me vio crecer como microbióloga y científica novel, y al que siempre llevaré en la cabeza y en el corazón, allá donde vaya.

A David, que aunque no seas director sobre el papel, si lo has sido en la vida real. Gracias por todas tus correcciones y paciencia. Trabajar contigo al lado lo ha hecho todo mucho más fácil.

Gracias también a todo el equipo de Clover; realmente habéis sido un trébol de cuatro hojas en el desarrollo de esta tesis, que sin vosotros no hubiera tenido lugar.

Y a todos mis compañeros del servicio de microbiología del HGM, técnicos, residentes y adjuntos, por acompañarme en las largas tardes de lectura y análisis de MALDI y tomaros unos cuantos cafés conmigo. Mer y Emilia, gracias por vuestra paciencia infinita, vuestros consejos y por tener siempre la puerta del despacho abierta para mí. Carmen, gracias por todos los tirones de orejas para que me pusiera a escribir.

Después, a los de aquí.

A mis nuevos compañeros del CHUAC. Gracias por acogerme y empujarme a terminar esta tesis. En esta carrera de fondo compartida con muchos de vosotros, parece que ya se ve la luz al final del túnel. Cris y Paula, gracias por los múltiples cafés terapéuticos en los que siempre arreglábamos el mundo. Marina, gracias por tu confianza y apoyo desde el minuto uno. Espero que nos queden muchos años más de ciencia juntas.

Finalmente, a los de siempre.

A los que siempre están. A todos aquellos que se preocuparon por mí, me dieron “tuppers” de mantenimiento y me preguntaron por cómo iba mi tesis, aun sin entender muy bien de qué iba ese rayo láser que analizaba bacterias a modo de Star Wars.

A mis padres, por haberme inculcado desde pequeña la autoexigencia y capacidad de superación. No sería quien soy sin vosotros a mi lado, gracias por acompañarme en cada paso del camino, dejándome tomar mis propias decisiones y apoyándolas aunque no siempre estuvierais de acuerdo. A mi madre, que me lo ha pedido así: “Por aguantarme”. No negaré que tiene su mérito. Y a mi padre, gracias por obligarme a terminar esta tesis y ser tan pesado preguntándome cómo iba la cosa. Os debo todo lo que tengo.

Y por último, a ti, Gonzalo, mi compañero en este viaje. Gracias por creer en mí cuando ni yo lo hacía. Esta tesis es tanto tuya como mía.

Abuelos, espero que desde arriba estéis muy orgullosos de mí.

Ya casi soy doctora...

*And last but not least, I want to thank me. I want to thank me for believing in me;
I want to thank me for doing all this hard work;
I want to thank me for having no days off;
I want to thank me for never quitting.*

PUBLICACIONES CIENTÍFICAS ORIGINALES Y COMUNICACIONES A CONGRESOS RELACIONADAS CON ESTA TESIS

PUBLICACIONES CIENTÍFICAS ORIGINALES

1. **Candela A**, Arroyo MJ, Sánchez-Molleda Á, Méndez G, Quiroga L, Ruiz A, Cercenado E, Marín M, Muñoz P, Mancera L, Rodríguez-Temporal D, Rodríguez-Sánchez B. Rapid and Reproducible MALDI-TOF-Based Method for the Detection of Vancomycin-Resistant *Enterococcus faecium* Using Classifying Algorithms. *Diagnostics (Basel)*. 2022 Jan 27; 12(2):328. doi: 10.3390/diagnostics12020328. PMID: 35204419; PMCID: PMC8871047.
2. **Candela A**, Guerrero-López A, Mateos M, Gómez-Asenjo A, Arroyo MJ, Hernández-García M, Del Campo R, Cercenado E, Cuénod A, Méndez G, Mancera L, Caballero JD, Martínez-García L, Gijón D, Morosini MI, Ruiz-Garbajosa P, Egli A, Cantón R, Muñoz P, Rodríguez-Temporal D, Rodríguez-Sánchez B. Automatic Discrimination of Species within the *Enterobacter cloacae* Complex Using Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry and Supervised Algorithms. *J Clin Microbiol*. 2023 Apr 20; 61(4):e0104922. doi: 10.1128/jcm.01049-22. Epub 2023 Apr 4. PMID: 37014210; PMCID: PMC10117122.
3. **Candela A**, Arroyo MJ, Sánchez-Cueto M, Marín M, Cercenado E, Méndez G, Muñoz P, Mancera L, Rodríguez-Temporal D, Rodríguez-Sánchez B. Rapid discrimination of *Pseudomonas aeruginosa* ST175 isolates involved in a nosocomial outbreak using MALDI-TOF Mass Spectrometry and FTIR Spectroscopy coupled with Machine Learning. *Transbound Emerg Dis*, vol. 2023, Article ID 8649429, 11 pages, doi: 10.1155/2023/8649429 Epub 2023 Sept 7.

COMUNICACIONES A CONGRESOS

1. **A Candela**, L Quiroga, MJ Arroyo, A Ruiz, E Cercenado, G Méndez, M Marín, P Muñoz, L Mancera, B Rodríguez-Sánchez. Rapid MALDI-TOF MS-based method for vancomycin-resistant *Enterococcus faecium* detection. **ECCMID, Paris (2020)**. Abstract aceptado (Congreso cancelado por la pandemia de COVID-19).
2. **A Candela**, L Quiroga-Maroto, MJ Arroyo, M Marín-Arriaza, E Cercenado, S Poza, P Muñoz, L Mancera, B Rodríguez-Sánchez. Utilidad de MALDI-TOF para la discriminación de aislados del clon de alto riesgo ST175 de *Pseudomonas aeruginosa*. **SEIMC, Zaragoza (2020)**. Abstract aceptado (Congreso cancelado por la pandemia de COVID-19).
3. **A Candela**, L Quiroga, MJ Arroyo, M Marín, E Cercenado, G Méndez, P Muñoz, L Mancera, L Pérez-Lago, B Rodríguez-Sánchez. MALDI-TOF MS utility for the discrimination of the ST175 *Pseudomonas aeruginosa* high risk clone in a hospital outbreak. **ECCMID, Viena (2021)**. Póster (Presentado online por la pandemia de COVID-19).
4. **A Candela**, MJ Arroyo, L Quiroga, A Ruiz, E Cercenado, G Méndez, M Marín, P Muñoz, L Mancera, B Rodríguez-Sánchez. A rapid and reproducible MALDI-TOF-based method for vancomycin resistance detection in *Enterococcus faecium* isolates. **ECCMID, Viena (2021)**. Póster (Presentado online por la pandemia de COVID-19).
5. **A Candela**, MJ Arroyo, A Martín, G Méndez, P Muñoz, L Mancera, L Alcalá, M Marín, B Rodríguez-Sánchez. Rapid *Clostridioides difficile* ribotyping based on MALDI-TOF MS and peak analysis. **MSACL EU Virtual (2021)**. Comunicación oral.
6. **A Candela**, M Mateos, A Gómez-Asenjo, M Blázquez-Sánchez, MJ Arroyo, M Hernández-García, R del Campo, E Cercenado, G Méndez, L Mancera, J de Dios

- Caballero, L Martínez-García, D Gijón, MI Morosini, P Ruiz-Garbajosa, R Cantón, P Muñoz, D Rodríguez-Temporal, B Rodríguez-Sánchez. Discrimination of species within the *Enterobacter cloacae* complex using MALDI-TOF Mass Spectrometry and Machine Learning. **MSACL EU Virtual (2021)**. Comunicación oral.
7. M Esteban, **A Candela**, M Sánchez-Cueto, E Cercenado, M Marín, P Muñoz, B Rodríguez-Sánchez, D Rodríguez-Temporal. Analysis of an outbreak caused by extensively drug-resistant *Pseudomonas aeruginosa* using Fourier Transform InfraRed Spectroscopy (FTIR). **ECCMID, Lisboa (2022)**. Póster.
 8. **A Candela**, D Rodríguez-Temporal, MJ Arroyo, M Marín, A Martin, L Jiménez-Navarro, L Alcalá, G Méndez, P Muñoz, L Mancera, B Rodríguez-Sánchez. Rapid *Clostridioides difficile* ribotyping based on MALDI-TOF MS and machine learning. **ECCMID, Lisboa (2022)**. Comunicación oral.
 9. **A Candela**, D Rodríguez-Temporal, MJ Arroyo, M Marín, A Martin, L Jiménez-Navarro, L Alcalá, G Méndez, P Muñoz, L Mancera, B Rodríguez-Sánchez. Método Rápido de Tipado de *Clostridioides difficile* Basado en Espectrometría de Masas MALDI-TOF y Machine Learning. **SEIMC, Granada (2022)**. Póster.
 10. **A Candela**, M Esteban, M Sánchez-Cueto, E Cercenado, P Muñoz, B Rodríguez-Sánchez, D Rodríguez-Temporal. Detección de Aislados de *Enterococcus faecium* Resistentes a Vancomicina Mediante Espectroscopía Infrarroja por Transformada de Fourier (FTIR) y Machine Learning. **SEIMC, Granada (2022)**. Póster.
 11. **A Candela**, M Esteban, M Sánchez-Cueto, E Cercenado, M Marín, P Muñoz, B Rodríguez-Sánchez, D Rodríguez-Temporal. Análisis de un Brote Causado por *Pseudomonas aeruginosa* XDR Mediante Espectroscopía Infrarroja de Transformada de Fourier (FTIR). **SEIMC, Granada (2022)**. Póster.

- 12. A Candela**, D Rodríguez-Temporal, MJ Arroyo, M Marín, L Jiménez-Navarro, A Martín, L Alcalá, P Muñoz, G Bou, L Mancera, M Oviaño, B Rodríguez-Sánchez. High molecular weight differentiation of clinically relevant *Clostridioides difficile* ribotypes using MALDI-TOF and Machine Learning. **ECCMID (2023)**. Póster.
- 13. M Blázquez-Sánchez**, D Rodríguez-Temporal, **A Candela**, M Sánchez-Cueto, MJ Arroyo, M Marín, A Martín, L Jiménez-Navarro, L Alcalá, G Méndez, P Muñoz, L Mancera, B Rodríguez-Sánchez. *Clostridioides difficile* ribotyping by Fourier-Transform Infrared Spectroscopy and Machine Learning. **ECCMID (2023)**. Póster.
- 14. D Rodríguez-Temporal**, M Blázquez-Sánchez, A Guerrero-López, **A Candela**, L Jiménez-Rosillo, M Sánchez-Cueto, MJ Arroyo, A Martín, L Jiménez-Navarro, V Gómez-Verdejo, PM Olmos, G Méndez, L Mancera, P Muñoz, L Alcalá, M Marín, B Rodríguez-Sánchez. Desarrollo de modelos de clasificación automática basados en espectrometría de masas MALDI-TOF y Machine Learning para el tipado rápido de *Clostridioides difficile*. **SEIMC (2023)**. Comunicación oral.

ÍNDICE

LISTADO DE ACRÓNIMOS	1
RESUMEN	7
ABSTRACT	9
INTRODUCCIÓN	11
1 IMPORTANCIA DE LA CORRECTA CARACTERIZACIÓN DE MICROORGANISMOS DE INTERÉS CLÍNICO	13
2 ESPECTROMETRÍA DE MASAS: HISTORIA Y APLICACIONES	18
3 TECNOLOGÍA MALDI-TOF Y PROTEÓMICA: FUNDAMENTOS TEÓRICOS	21
3.1 Aplicación de MALDI-TOF para la identificación de microorganismos	24
A) A partir de cultivo	25
B) A partir de muestra directa.....	26
C) Retos en la identificación de microorganismos con MALDI-TOF.....	28
3.2 Aplicación de la tecnología MALDI-TOF para la detección de resistencias antimicrobianas	32
A) Análisis directo del perfil proteico bacteriano.....	33
B) Estudio de cambios de masa en el antibiótico mediante la detección de la actividad enzimática bacteriana	34
C) Estudio del perfil proteico bacteriano tras la incubación con el antibiótico...	36
3.3 Aplicación de la tecnología MALDI-TOF para el tipado de cepas clínicas	38
4 ESPECTROSCOPIA INFRARROJA POR TRANSFORMADA DE FOURIER (FTIR)	42
5 BREVE INTRODUCCIÓN A LOS MODELOS AUTOMATIZADOS O MACHINE LEARNING	46
5.1 Modelos supervisados	47
5.2 Modelos no supervisados	48
5.3 Aplicaciones de Machine Learning en microbiología y espectrometría de masas	51
JUSTIFICACIÓN Y OBJETIVOS	55
TRABAJOS CIENTÍFICOS	63
Capítulo 1. Método rápido y reproducible basado en MALDI-TOF para la detección de <i>Enterococcus faecium</i> resistentes a vancomicina empleando algoritmos de clasificación	65

MATERIAL SUPLEMENTARIO CAPÍTULO 1	85
Capítulo 2. Discriminación automática de las especies dentro del complejo <i>Enterobacter cloacae</i> empleando espectrometría de masas MALDI-TOF y algoritmos supervisados	95
MATERIAL SUPLEMENTARIO CAPÍTULO 2	109
Capítulo 3. Discriminación rápida de aislados de <i>Pseudomonas aeruginosa</i> ST175 involucrados en un brote nosocomial empleando espectrometría de masas MALDI-TOF y espectroscopía FTIR combinados con <i>Machine Learning</i>	113
MATERIAL SUPLEMENTARIO CAPÍTULO 3	129
Capítulo 4. Ribotipado de <i>Clostridioides difficile</i> empleando MALDI-TOF y algoritmos de <i>Machine Learning</i>.....	141
RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN.....	149
CONCLUSIONES	185
BIBLIOGRAFÍA.....	189

LISTADO DE ACRÓNIMOS

ANN: Artificial Neural Network (Redes neuronales artificiales)

ADN: Ácido Desoxirribonucleico

ARNr: Ácido Ribonucleico Ribosomal

ASO-PCR: Allele-Specific Oligonucleotide Polymerase Chain Reaction (Reacción en cadena de la polimerasa con oligonucleótidos específicos de alelo)

AST: Antimicrobial Susceptibility Testing (Estudios de sensibilidad antimicrobiana)

AUC: Area Under the Curve (Área bajo la curva)

BLEE: β -Lactamasa de Expectro Extendido

BHI: Medio de cultivo Brain Heart Infusion (Infusion cerebro-corazón)

C3G: Cefalosporinas de 3^a Generación

CDC: Centers for Disease Control and Prevention (Centros para el control y prevención de enfermedades)

cg-MLST: Core Genome Multi Locus Sequencing Typing (Tipado por secuenciación de regiones multilocus del genoma central)

Clover MSDAS: Plataforma informática de análisis de espectros Clover Mass Spectrometry Data Analysis Software.

CRAB: Carbapenem-Resistant Acinetobacter baumannii (Acinetobacter baumannii resistente a carbapenémicos)

CV: Coeficiente de Variación

Curva ROC: Receiver Operating Characteristic Curve (Curva característica operativa del receptor)

Da: Daltons (unidad de masa atómica)

LISTADO DE ACRÓNIMOS

DBLR-FS: Algoritmo de Machine Learning “Dual Bayesian Logistic Regression with Feature Selection” (Regresión Logística bayesiana dual con selección de características)

DOT-MGA: Direct-On-Target Microdroplet Growth Assay (Estudio de crecimiento directo de microgotas sobre placa)

DTR: Difficult to treat resistance (Bacterias resistentes de difícil tratamiento)

ECC: Enterobacter cloacae Complex (Complejo *Enterobacter cloacae*)

EM: Espectrometría de Masas

ESKAPE: Acrónimo para las bacterias *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* y el Complejo *Enterobacter cloacae*

ESCAPE: Nuevo acrónimo propuesto para las bacterias *Enterococcus faecium*, *Staphylococcus aureus*, *Clostridioides difficile*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* y Enterobacterias

ESI: Electrospray Ionization (Ionización por Electrospray)

FA-VAE: Algoritmo de Machine Learning “Factor Analysis-Variational AutoEncoder” (Auto-Codificador Variacional de Análisis de Factores)

FTIR: Fourier Transformed Infrared Spectroscopy (Espectroscopía infrarroja por transformada de Fourier)

HCCA: Ácido α -ciano-4-hidroxicnámico

HCA: Hierarchical Clustering Analysis (Análisis de agrupamiento jerárquico)

HCDGU: Hospital Central de la Defensa “Gómez Ulla”

HGM: Hospital Gregorio Marañón

IA: Inteligencia Artificial

ID: Identificación

KDa: KiloDalton (10^3 Da)

KPC: *Klebsiella pneumoniae* Carbapenemasa

KNN: Algoritmo de Machine Learning K-Nearst Nighbor (k-vecino más proximo)

LDA: Algoritmo de Machine Learning Linear Discriminant Aalysis (Análisis de discriminación lineal)

LDI: Laser Desorption/Ionization (Desorción/Ionización por láser)

Light-GBM: Algoritmo de Machine Learning Light Gradient Boosting Machine (Máquina de potenciación de gradiente ligero)

LCR: Líquido Cefalorraquídeo

LOO: Método de validación cruzada “LeaOe Out” o “Dejar a uno fuera”

MALDI-TOF MS: Matrix-Assisted Laser Desorption/Ionization-Time-Of-Flight Mass Spectrometry (Espectrometría de masas por ionización/desorción asistida por matriz con tiempo de vuelo)

MDR: Multi Drug Resistant (Resistencia a múltiples antibióticos)

MGIT: Medio de cultivo Mycobacterial Growth Indicator Tube (Indicador de crecimiento de micobacterias)

ML: Machine Learning (Aprendizaje automatico)

MLST: Multi Locus SequenIce Typing (Tipado por secuenciación de regiones de múltiples locus)

MLSA: Multi Locus SequenAlysis (Análisis de secuenciación de múltiples locus)

MLVA: Multi Locus Variant Alysis (Análisis de variantes de múltiples locus)

MS: Mass Spectrometry (Espectrometría de masas)

MSI database: Mass Spectrometric Identification database (Base de datos de identificación de espectrometría de masas)

m/z: Relación masa/carga

NGS: Next Generation Sequencing (Secuenciación de nueva generación)

LISTADO DE ACRÓNIMOS

OMS: Organización Mundial de la Salud

PBP: Penicillin Binding Proteins (Proteínas fijadoras de penicilina)

PCA: Algoritmo de Machine Learning Pincipal Component Analysis (Análisis de componentes principales)

PCR: Polymerase Chain Reaction (Reacción en cadena de la polimerasa)

PDR: Pan-Drug Resistant (Resistente a todos los antibióticos)

PFGE: Pulsed Field Gel Electrophoresis (Electroforesis de campo pulsado)

PLS-DA: Algoritmo de Machine Learning Partial Least Squares Discriminant Analysis (Análisis discriminante de mínimos cuadrados parciales)

qPCR: PCR cuantitativa (quantitative-PCR) o PCR a tiempo real

RAST: Rapid Antimicrobial Susceptibility Testing (Estudios rápidos de sensibilidad antimicrobiana)

RF: Algoritmo de Machine Learning Random Forest (Bosque aleatorio)

RH: Ratio de Hidrólisis

SARM: Staphylococcus aureus Resistente a Meticilina

SASM: Staphylococcus aureus Sensible a Meticilina

SDS: Detergente Sodium Dodecyl Sulfate (Dodecilsulfato sódico)

SLD: Soft Laser Desorption (Desorción láser suave)

SNPs: Single Nucleotide Polymorphisms (polimorfismos de un solo nucleótido)

SVM: Algoritmo de Machine Learning Support Vector Machine (Máquina de vectores de soporte)

TIC: Normalización Total Ion Current (Corriente iónica total)

UHRC: Hospital Universitario Ramón y Cajal

UHB: Hospital Universitario Basel

VSE: Vancomycin Susceptible Enterococcus (*Enterococcus* sensible a vancomicina)

VRE: Vancomycin Resistant Enterococcus (*Enterococcus* resistente a vancomicina)

UFC: Unidades Formadoras de Colonias

VPN: Valor Predictivo Negativo

VPP: Valor Predictivo Positivo

WGS: Whole Genome Sequencing (Secuenciación de genoma completo)

wg-MLST: Whole Genome MLST (tipado por secuenciación de regiones de múltiples locus del genoma completo)

XDR: Extrremely Drug Resistant (Extremadamente resistente a antibióticos)

RESUMEN

La espectrometría de masas MALDI-TOF es una tecnología emergente que ha irrumpido de lleno en el laboratorio de microbiología clínica, suponiendo un cambio en el paradigma de la identificación de microorganismos y permitiendo una identificación rápida en apenas minutos en comparación con las técnicas bioquímicas y moleculares disponibles previamente. A su vez, la tecnología MALDI-TOF comienza a aplicarse también para usos alternativos a la identificación microbiana, como la detección de mecanismos de resistencia o el tipado de aislados clínicos. Por otro lado, la inteligencia artificial también se encuentra en auge dentro del campo sanitario, sirviendo de ayuda en el procesamiento y análisis de datos.

En esta tesis se presentan cuatro trabajos que emplean la tecnología MALDI-TOF junto con algoritmos de *Machine Learning* para el análisis de espectros de masas, para usos alternativos a la identificación rutinaria de microorganismos, a saber: la diferenciación entre *Enterococcus faecium* sensible (VSE) o resistente a vancomicina (VRE), la identificación de especies dentro del complejo *Enterobacter cloacae* (ECC), el tipado de un brote de *Pseudomonas aeruginosa* ST175 (junto con un análisis mediante espectroscopía FTIR), y el ribotipado de *Clostridioides difficile*.

En el estudio de *E. faecium*, se realizó un estudio inicial de reproducibilidad, y se observó que los espectros analizados mostraron menor variabilidad con los datos normalizados y para los picos que se encontraban dentro del rango de 3000-9000 *m/z*. Se consiguió la separación entre aislados VSE y VRE aplicando diferentes algoritmos de *Machine Learning* con la plataforma Clover MSDAS, con un acierto del 80,9% con SVM, 79,2% con RF y 77,5% con PLS-DA. La mejor separación entre los aislados VanA y VanB se obtuvo mediante PLS-DA, con una correcta clasificación en el 86,6% de los casos.

En cuanto al estudio de ECC, inicialmente, mediante la aplicación de algoritmos no supervisados se observó una clara separación entre *Enterobacter hormaechei* y *Klebsiella aerogenes* y una tendencia a la diferenciación entre el resto de las especies de ECC. Posteriormente se emplearon modelos supervisados en Clover MSDAS que se

RESUMEN

validaron externamente con un porcentaje de acierto a nivel de especie 100% correcta para *E. asburiae* y *E. kobei*, y entre 91,2% y 98,0% para las especies ECC restantes.

El estudio de *P. aeruginosa* permitió la diferenciación de aislados pertenecientes a un brote con la misma precisión que la obtenida mediante PFGE, suponiendo un ahorro en tiempo y recursos. Se encontraron tres picos biomarcadores específicos de brote (5169 *m/z*, 6915 *m/z* y 7236 *m/z*) en los espectros analizados con los que se realizó un estudio con algoritmos no supervisados (PLS-DA, RF, SVM y KNN), que permitieron la correcta clasificación del 100% de los aislados en comparación con los resultados obtenidos por PFGE. La tecnología FTIR mostró un poder de discriminación similar al PFGE y también alcanzó una correcta discriminación de los diferentes ST analizados.

Y por último, en el estudio de ribotipado de *C. difficile*, se confirmó la presencia de 7 picos biomarcadores que permitieron la diferenciación de los ribotipos RT027 y RT181 del resto de ribotipos (2463 *m/z*, 4933 *m/z*, 4993 *m/z*, 3353 *m/z*, 6187 *m/z*, 6651 *m/z* y 6710 *m/z*). Estos resultados se validaron a tiempo real con aislados de dos brotes independientes procedentes de hospitales de Madrid. Además de la aplicación de algoritmos de Machine Learning de Clover MSDAS, se desarrolló un software libre (AutoCdiff) para el análisis de espectros, que permite la clasificación automática de los aislados en “RT027”, “RT181” y “Otros Ribotipos”.

MALDI-TOF ha demostrado ser una valiosa herramienta que, en combinación con algoritmos de *Machine Learning* puede ayudar a la rápida diferenciación entre microorganismos sensibles y resistentes, la identificación de especies problemáticas y el tipado bacteriano. Esta tecnología se encuentra en constante desarrollo y presenta un futuro prometedor, en el que, con la ayuda de la inteligencia artificial, permitirá obtener resultados rápidos y fiables que permitan optimizar el flujo de trabajo en el laboratorio de microbiología clínica, suponiendo un ahorro en tiempo y recursos.

ABSTRACT

MALDI-TOF mass spectrometry is an emerging technology that has fully erupted in the clinical microbiology laboratory, representing a change in the paradigm of microorganism identification and allowing rapid identification in just minutes compared to biochemical and molecular techniques previously available. At the same time, MALDI-TOF technology is also being studied for alternative applications such as detection of resistance mechanisms or typing of clinical isolates. On the other hand, artificial intelligence is also blooming in the healthcare field, helping in data processing and analysis.

This thesis presents four studies that apply MALDI-TOF mass spectrometry technology coupled with Machine Learning algorithms for the analysis of mass spectra, for four alternative uses other than routine identification of microorganisms: the differentiation between vancomycin-susceptible *Enterococcus faecium* (VSE) from vancomycin-resistant *E. faecium* (VRE), species identification within the *Enterobacter cloacae* complex (ECC), typing of a *Pseudomonas aeruginosa* ST175 outbreak (combined with analysis by FTIR spectroscopy), and ribotyping of *Clostridioides difficile*.

In the *E. faecium* study, a reproducibility study was carried out and it was observed that the analyzed spectra presented less variability with normalized data and for the peaks within the range of 3000-9000 *m/z*. Separation between VSE and VRE isolates was achieved by applying different Machine Learning algorithms with the Clover MSDAS platform, with a success rate of 80.9% with SVM, 79.2% with RF and 77.5% with PLS-DA. The best differentiation between VanA and VanB isolates was obtained using PLS-DA, with a correct classification in 86.6% of the cases.

Regarding the ECC study, initially, through the application of unsupervised algorithms a clear separation between *Enterobacter hormaechei* and *Klebsiella aerogenes* and a tendency towards differentiation between the rest of the ECC species was observed. Subsequently, supervised models were applied in Clover MSDAS which were externally validated with a success rate at the species level of 100% for *E. asburiae* and *E. kobei*, and between 91,2% and 98,0% for the rest of species studied.

ABSTRACT

The *P. aeruginosa* study allowed the differentiation of outbreak isolates with the same precision as PFGE, saving time and resources. Three outbreak-specific biomarker peaks (5169 *m/z*, 6915 *m/z* and 7236 *m/z*) were found in the analyzed spectra which were applied for the classification of isolates using unsupervised algorithms (PLS-DA, RF, SVM and KNN). These peaks allowed the correct classification of 100% of the isolates compared to the results obtained by PFGE. FTIR technology showed a discriminatory power similar to PFGE and also achieved correct discrimination of the different STs analyzed.

Finally, in the *C. difficile* ribotyping study, the presence of 7 biomarker peaks was confirmed which allowed the differentiation of ribotypes RT027 and RT181 from the rest of the ribotypes (2463 *m/z*, 4933 *m/z*, 4993 *m/z*, 3353 *m/z*, 6187 *m/z*, 6651 *m/z* and 6710 *m/z*). These results were validated in real time with isolates from two independent outbreaks from hospitals in Madrid. In addition to the application of Clover MSDAS Machine Learning algorithms, a free software (AutoCdiff) was developed for spectra analysis, which allows the automatic classification of isolates into “RT027”, “RT181” and “Other Ribotypes” categories.

MALDI-TOF has proven to be a valuable tool which, in combination with Machine Learning algorithms, can help in the rapid differentiation between susceptible and resistant microorganisms, the correct identification of problematic species and bacterial typing. This technology is in constant development and presents a promising future, in which, with the aid of artificial intelligence, will provide quick and reliable results which will help in the optimization of the workflow in the clinical microbiology laboratory, saving time and resources.

INTRODUCCIÓN

1 IMPORTANCIA DE LA CORRECTA CARACTERIZACIÓN DE MICROORGANISMOS DE INTERÉS CLÍNICO

En el laboratorio de microbiología clínica, la identificación y caracterización rápida y precisa de los microorganismos causantes de infección en el ser humano es una prioridad, y una variable en constante evolución. En las últimas décadas se ha vivido una revolución diagnóstica, en la que cada vez más técnicas rápidas se encuentran disponibles y cuyo objetivo principal es el de acortar al máximo el tiempo que transcurre entre la toma de la muestra clínica y la identificación y caracterización de los microorganismos [1].

El cultivo tradicional de muestras infecciosas ha sido durante mucho tiempo el método estándar en el laboratorio clínico para el aislamiento, la identificación, y el estudio de la sensibilidad de los microorganismos. Este abordaje sigue siendo uno de los pilares fundamentales en los laboratorios de microbiología clínica, ya que permite, además de la identificación microbiana, el estudio de la sensibilidad antibiótica (AST por sus siglas en inglés, Antimicrobial Susceptibility Testing) [2].

Sin embargo, el cultivo microbiológico se basa en la capacidad de los microorganismos para crecer y multiplicarse en los medios de cultivo adecuados. Esto está limitado por el tiempo de crecimiento inherente a cada microorganismo -y en ocasiones, como sucede con los virus o algunas bacterias como las espiroquetas, por la ausencia del mismo-. Otros factores pueden afectar también al crecimiento de los microorganismos, como por ejemplo, que se trate de una muestra proveniente de un paciente tratado con antimicrobianos en el momento de la toma de muestra, que la muestra no se tome de la manera adecuada, o que se demore demasiado su transporte y recepción en el laboratorio [3]. Además, se debe orientar el crecimiento según el tipo de infección y bacteria que se sospeche, empleando medios de cultivo selectivos -con componentes que inhiben el crecimiento de determinados microorganismos-; diferenciales -con componentes que permiten distinguir microorganismos según sus características de crecimiento-; y en ocasiones, medios enriquecidos -con componentes para facilitar el crecimiento de microorganismos

INTRODUCCIÓN

fastidiosos-. También se deben emplear diferentes condiciones y temperaturas de incubación según el microorganismo que se pretende aislar. Por lo que, si todo esto no confluye correctamente, se pueden obtener cultivos negativos en casos de infección [4, 5]. Además, La mayoría de patógenos clínicos causantes de infección en humanos tardan en crecer un mínimo de 12-24 horas hasta ser identificables mediante técnicas fenotípicas convencionales. En ocasiones, con determinados microorganismos como las bacterias anaerobias, micobacterias, u hongos, este proceso se demora y tardan días e incluso meses en crecer.

La identificación mediante técnicas bioquímicas se basa en el estudio del metabolismo de la bacteria, con una batería de pruebas que tardan al menos otras 24 horas en aportar resultados. Estas técnicas bioquímicas son por lo general lentas y laboriosas, y en ocasiones, no arrojan resultados concluyentes [5-7]. La microscopía óptica combinada con diversas tinciones según la sospecha clínica puede aportar resultados preliminares que orienten al tratamiento, pero no permite una identificación concreta de género y especie necesaria para la instauración de un correcto tratamiento antibiótico dirigido.

En el flujo de trabajo convencional del laboratorio, una vez identificada la bacteria, se procede a realizar un estudio de su sensibilidad mediante técnicas de difusión en agar con discos o tiras de gradiente, o microdilución en caldo. Estas técnicas tardan 24 horas más en ser interpretables. Por tanto, el proceso completo de identificación y estudio de sensibilidad se demoraría al menos 72 horas hasta la emisión de un informe definitivo [6-8].

En la era de la emergencia de resistencias antimicrobianas, la identificación rápida de un microorganismo, así como la determinación de su sensibilidad a los antibióticos, son aspectos críticos para la instauración de un tratamiento antimicrobiano guiado, evitando así la emergencia de nuevas resistencias. Además, se ha demostrado que un tratamiento antibiótico erróneo, o retrasado en el tiempo, aumenta la morbi-mortalidad en los pacientes, así como los costes relacionados con la asistencia sanitaria [9-11]. La tendencia actual de los laboratorios de microbiología es a la automatización para la obtención de resultados fiables, en ausencia de errores

humanos, y en el menor tiempo posible, permitiendo también el ahorro en recursos de personal [12].

En la segunda mitad del S. XX, con el descubrimiento de la estructura del ADN, el desarrollo de la técnica de reacción en cadena de la polimerasa (PCR) en los años 80, y posterior PCR a tiempo real (qPCR), se produjo la primera revolución en el diagnóstico microbiológico. La introducción de técnicas moleculares de detección y amplificación de ácidos nucleicos supuso una mejora en cuanto a la rapidez de diagnóstico [13]. La PCR permite, además de identificar microorganismos a partir de secuencias conservadas de su ADN, identificar también posibles genes relacionados con la resistencia a determinados antibióticos, como la detección del gen *mecA* que implica resistencia a meticilina en *Staphylococcus*, los genes *vanA/B* de resistencia a la vancomicina en *Enterococcus*, o la detección de diversas enzimas β -lactamasas y carbapenemasas que hidrolizan los antibióticos β -lactámicos, haciendo que las bacterias se vuelvan resistentes a ellos.

Sin embargo, la detección de ADN bacteriano no siempre se correlaciona con viabilidad del microorganismo, ni con expresión enzimática. Se puede dar el caso de detectar ADN de una bacteria que no se estuviera replicando y no fuera la causante de la infección, así como detectar mecanismos de resistencia que no se estuvieran expresando, o, al tratarse de una técnica dirigida a una secuencia genética concreta, no detectar otras bacterias o mecanismos de resistencia que sí estuvieran presentes. Por tanto, siempre es recomendable correlacionar los resultados de PCR con pruebas fenotípicas y bioquímicas [14].

Otra desventaja de las técnicas moleculares es que precisan de un laboratorio adaptado para su realización con diferentes salas para evitar contaminaciones cruzadas, instrumentos y material especializado de alto coste y, por lo general, personal entrenado para su realización [11, 15]. Actualmente existen plataformas automatizadas multiplex de detección de ácidos nucleicos a tiempo real que han conseguido eliminar la necesidad de personal entrenado en biología molecular y reducir el tiempo hasta la obtención de resultados. Sin embargo, su coste económico es elevado, mayor que el de la PCR convencional [12, 15].

INTRODUCCIÓN

La aparición de la secuenciación masiva, secuenciación de genoma completo (WGS en inglés, Whole Genome Sequencing) o secuenciación de nueva generación (NGS en inglés, Next Generation Sequencing) ha supuesto una nueva revolución en el campo de la medicina en general y de la microbiología en particular. Esta novedosa técnica permite conocer todo el genoma de una bacteria, analizando tanto ADN cromosómico como accesorio, en un único experimento en el que en paralelo se llevan a cabo millones de reacciones de secuenciación [16].

La secuenciación masiva aporta una elevada cantidad de información, la cual en ocasiones es difícilmente interpretable por personal no especializado en la materia. El análisis de toda la información obtenida mediante este tipo de secuenciación es un proceso complejo, que requiere de personal entrenado en bioinformática así como de equipos informáticos de gran potencia y capacidad de almacenamiento. Además, precisa de secuenciadores y reactivos no disponibles en todos los laboratorios por su elevado coste, y de personal altamente cualificado en técnicas de biología molecular para la extracción de ácidos nucleicos y la preparación de librerías previas al proceso de secuenciación [16-18].

En la actualidad, la secuenciación masiva es la técnica más completa para la caracterización de microorganismos, ya que permite además de identificarlos, conocer de manera relativamente rápida sus mecanismos de resistencia, sus genes de virulencia y patogenicidad, así como monitorizar su evolución a lo largo del tiempo, y comparar sus similitudes o discrepancias con otras bacterias para el control de la diseminación de brotes. Sin embargo, por todos sus requerimientos técnicos, económicos y de personal, no es una técnica accesible a todos los laboratorios de microbiología clínica y actualmente está restringida a aquellos más punteros y con mayores recursos económicos.

Es por todo esto, que los laboratorios de microbiología clínica necesitaban del desarrollo de una alternativa rápida, fiable y coste-eficaz para la caracterización de microorganismos de interés clínico. Es ahí donde la tecnología de espectrometría de masas MALDI-TOF MS (Matrix-assisted Laser Desorption/Ionization-Time-of-Flight Mass Spectrometry) ha encontrado su nicho. La identificación rápida -en minutos- de

los microorganismos causantes de infección ha conllevado una importante mejora en la correcta instauración del tratamiento antimicrobiano, mejorando a su vez la asistencia sanitaria a los pacientes, y disminuyendo la morbi-mortalidad [15, 19, 20].

Además del impacto clínico, la instauración de MALDI-TOF en los laboratorios supone una mejora en rentabilidad en los centros hospitalarios, reduciendo significativamente el gasto de los laboratorios tras la inversión inicial del equipo [21]. El empleo de MALDI-TOF para una rápida identificación de microorganismos y el conocimiento de sensibilidad antibiótica rápida cuando sea posible, supone un ahorro en la duración de la estancia hospitalaria así como en consumo antibiótico y los costes que esto conlleva [22, 23]. En la Figura 1 se representa la mejora en ahorro de tiempos que ha supuesto la instauración de la tecnología MALDI-TOF para el diagnóstico de las infecciones en el laboratorio de microbiología clínica.

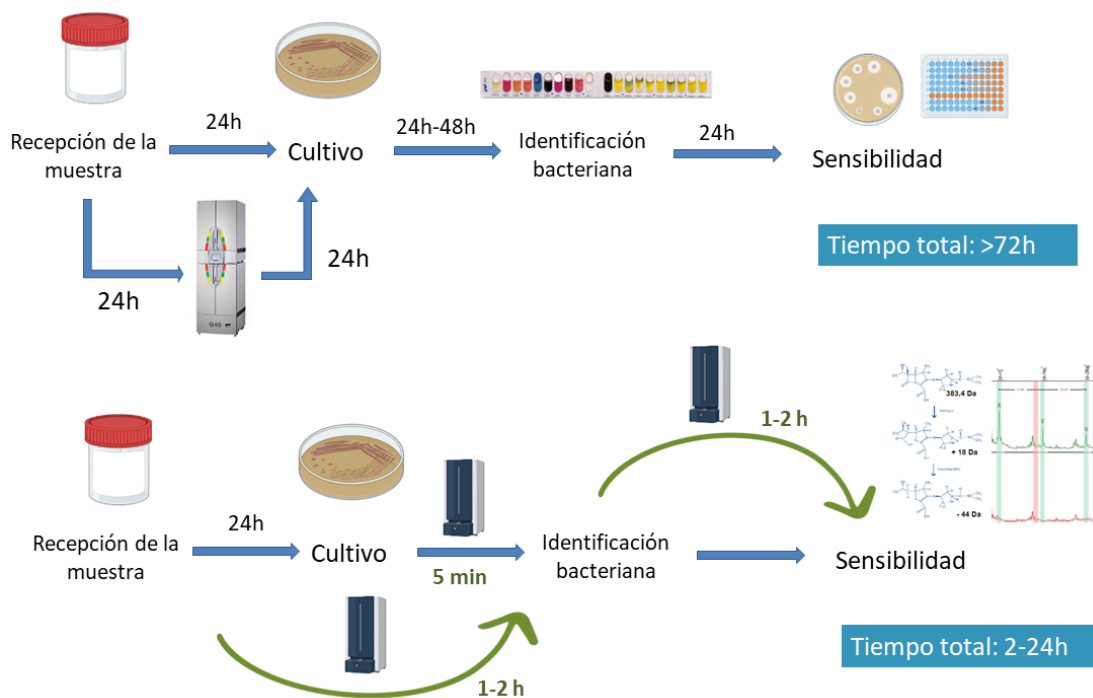


Figura 1. Reducción de tiempos en el procesamiento de laboratorio tras la instauración de MALDI-TOF

2 ESPECTROMETRÍA DE MASAS: HISTORIA Y APLICACIONES

La Espectrometría de Masas (EM) o Mass Spectrometry (MS) es una técnica analítica de determinación estructural, que posibilita la clasificación e identificación de compuestos químicos en función de su masa (peso molecular) y su carga eléctrica. Permite analizar la composición de diferentes elementos químicos al medir los iones que los conforman y separarlos en función de su relación masa/carga (m/z).

La base de la espectrometría de masas se fundamenta en la ionización de los compuestos de interés, seguida de su separación y clasificación en función de su relación m/z obteniéndose lo que se conoce como “huella química” del compuesto, que es la representación de estos iones por orden creciente de masa frente a su abundancia relativa (intensidad). La espectrometría de masas se puede emplear para conocer la composición química de una sustancia, determinar su estructura química, conocer su peso molecular exacto o cuantificar compuestos químicos ya conocidos [24].

El funcionamiento de un espectrómetro de masas se divide en tres procesos fundamentales. Primeramente, en el proceso de ionización, una fuente de ionización transforma los analitos de la muestra en iones gaseosos. Esto puede realizarse con diferentes tipos de ionizadores, como la ionización por electrospray (ESI) o la desorción/ionización por láser asistida por matriz (MALDI), que son las dos técnicas de ionización suave que permiten el análisis de proteínas. Una vez ionizados, estos iones se separan según su relación m/z mediante un analizador de masas, que en el análisis proteómico puede ser de varios tipos, como analizadores de tiempo de vuelo (TOF), analizadores de cuadrúpolo (Q), analizadores de trampa iónica (QIT, LIT) o analizador de masas de resonancia de ciclotrón de iones por transformada de Fourier (FTICR). Por último, tras esta separación, los iones viajan a distintas velocidades según su masa a través del tubo de vacío (los de menor tamaño viajan a mayor velocidad que los más grandes). Llegan a un detector que los registra en función de la relación m/z , de su abundancia relativa y del tiempo que han tardado en impactar contra él [25]. El resultado es un espectro en el que se representan todos los iones registrados, donde el

eje de abscisas es la relación m/z , y el eje de ordenadas la intensidad o abundancia relativa de cada ion (reflejado en unidades arbitrarias, u. a.).

La espectrometría de masas fue descubierta de manera casual por el físico británico J.J. Thomson -descubridor también del electrón-, a comienzos del siglo XX, cuando, con la ayuda de E. Everett, trabajó en la creación de un instrumento para la medición de la masa de los rayos catódicos mientras estudiaban la transmisión de electricidad a través de los gases [26]. Thomson y Everett desarrollaron un instrumento que permitía la detección de iones con trayectoria parabólica tras atravesar un campo eléctrico paralelo a otro magnético, y su posterior detección en una placa fotográfica. Previo a este descubrimiento, a finales del S. XIX, el físico alemán Wilhelm Wien, había descubierto que las partículas cargadas eléctricamente podían ser desviadas cuando se les aplicaba un campo magnético, sentando las bases de la espectrometría de masas [24].

Durante la primera mitad de siglo XX, el instrumento desarrollado por J.J. Thomson fue perfeccionándose y aumentando su poder de resolución, lo que permitió su uso para la detección de isótopos, siendo el descubrimiento más notable la detección de un isótopo más estable del Neón, con masa 22 y no 20 [24]. Hacia mediados de siglo, se empezó a utilizar la espectrometría de masas para la detección de hidrocarburos en empresas petrolíferas, pero existía el problema de no poder detectar compuestos de alto peso molecular por sus fragmentaciones y descomposiciones al depender la ionización de una colisión de los electrolitos con partículas cargadas.

Catherine Fenselau (Figura 2) fue la primera mujer dedicada a la espectrometría de masas y la primera persona en dirigir su aplicación al campo de la medicina. Fue la encargada de desarrollar el área de espectrometría de masas orientada a la biomedicina en la universidad Johns Hopkins de Maryland. Sus estudios se centraron en el área del metabolismo humano y la investigación de compuestos alquilantes para el tratamiento del cáncer [27]. Fue pionera en emplear la espectrometría de masas para la identificación de microorganismos en 1975 [28], técnica que posteriormente sería perfeccionada mediante el empleo de la tecnología SLD (Soft Laser Desorption).

INTRODUCCIÓN

El descubrimiento de una energía laser suave que permitiera que compuestos previamente considerados termolábiles se descompusieran en iones sin llegar a desintegrarse totalmente les valió el premio Nobel de química a Koichi Tanaka y John Fenn en 2002 por el descubrimiento de la tecnología de desorción-ionización por láser (LDI) y ionización por electrospray (ESI), respectivamente (Figura 2) [29]. Al contrario de lo que se cree, Tanaka no fue el desarrollador de la tecnología MALDI-TOF. Aunque ambos métodos emplean un láser para la ionización, el MALDI utiliza una matriz orgánica mientras que la tecnología desarrollada por Tanaka empleaba una suspensión de nanopartículas metálicas en glicerol, que absorbían la energía del láser [30].

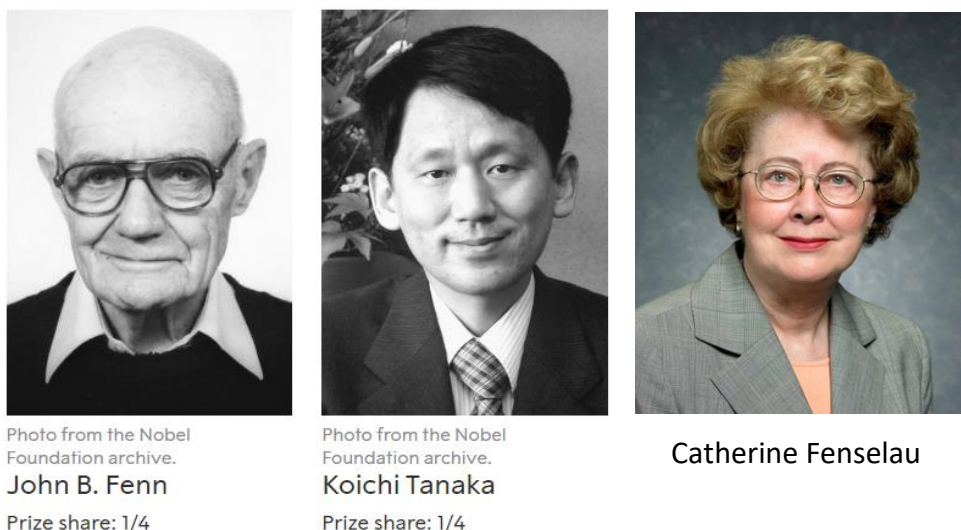


Figura 2. John Fenn y Koichi Tanaka, ganadores del premio Nobel en química en 2002, y Catherine Fenselau

La adición de la matriz orgánica fue desarrollada posteriormente por Karas y Hillenkamp, que añadieron una matriz orgánica, inicialmente de ácido nicotínico, y acuñaron el término de 'MALDI' (Matrix-Assisted Laser Desorption/Ionization). Esta nueva técnica, aumentó la sensibilidad del método desarrollado por Tanaka y permitió que las muestras cristalizaran de manera uniforme con la matriz, evitando así su degradación cuando son excitadas con el láser [31, 32].

Este nuevo descubrimiento permitió la aplicación de MALDI-TOF en el campo de la microbiología clínica para la identificación de microorganismos. Gracias a la energía suave del láser y la adición de matriz orgánica, con la tecnología MALDI-TOF las proteínas ribosómicas bacterianas, de mayor peso molecular que lo analizado hasta la fecha, se podían ionizar sin desintegrarse, emitiendo un espectro de masas característico para cada microorganismo analizado, que a su vez, podía ser contrastado con bases de datos previamente creadas con perfiles proteicos representativos de cada microorganismo, permitiendo así su identificación por comparación.

3 TECNOLOGÍA MALDI-TOF Y PROTEÓMICA: FUNDAMENTOS TEÓRICOS

MALDI-TOF MS es el acrónimo para las palabras en inglés “Matrix-assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry”, es decir, espectrometría de masas basada en la tecnología de ionización por desorción láser asistida por matriz, con detección mediante tiempo de vuelo. Aunque lleva años empleándose en centros de investigación, este tipo de espectrometría ha incrementado su popularidad en entornos hospitalarios en los últimos años, haciéndose a día de hoy prácticamente imprescindible para la identificación de microorganismos causantes de infecciones, entre otras muchas aplicaciones.

Los equipos de MALDI-TOF empleados en el ámbito hospitalario constan de tres componentes principales (Figura 3) [33]:

- 1) Un láser de nitrógeno (N_2) como fuente de ionización, que produce luz ultravioleta que irradia la muestra a una longitud de onda de 337nm y es capaz de sublimarla e ionizarla sin desintegrarla.
- 2) Un analizador con vacío que aplica un campo electromagnético a los iones y permite que vuelen a través de él a distintas velocidades en función de su masa.
- 3) Un detector de tiempo de vuelo en la parte superior, que recibe los iones y los transforma en una señal eléctrica que es procesada y enviada a un ordenador.

INTRODUCCIÓN

Esto da lugar a un espectro de masas donde los iones se clasifican en función de su ratio masa-carga (m/z) y su abundancia relativa (intensidad). Dado que la mayoría de los iones formados poseen una sola carga ($z=1$), la relación m/z es equivalente a la masa de cada ion.

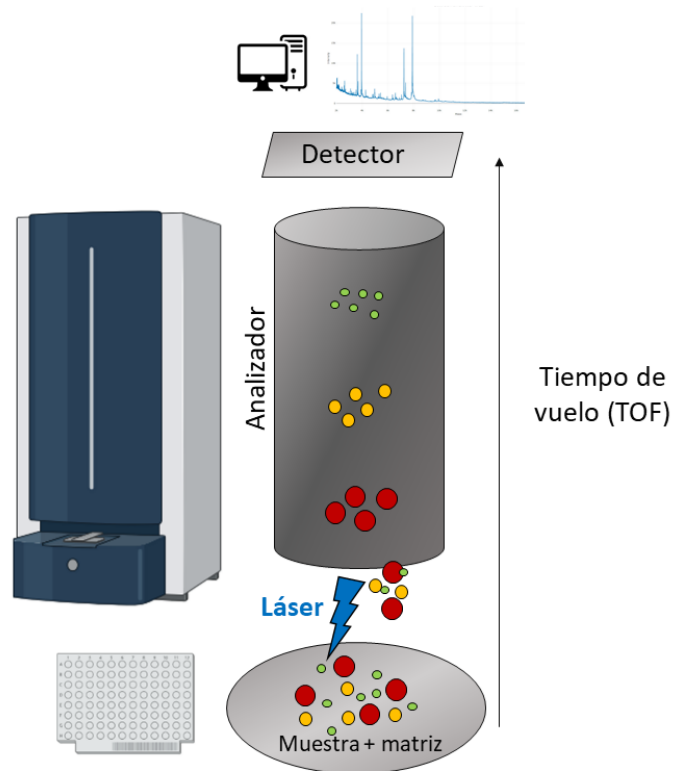


Figura 3. Espectrómetro de masas MALDI-TOF

El microorganismo a analizar debe aplicarse de manera homogénea en una posición de la placa metálica de MALDI con la ayuda de un palillo de madera o una punta de pipeta, y posteriormente mezclarse con $1\mu\text{l}$ de matriz orgánica que se ionice con facilidad.

Hay distintos tipos de matrices orgánicas según las moléculas que se quieran analizar, pero se ha observado que en la identificación de microorganismos intactos y proteínas, las más adecuadas son las matrices derivadas del ácido cinámico [34, 35]. En concreto en los laboratorios de microbiología clínica se emplea la matriz de HCCA (Ácido α -ciano-4-hidroxicinámico). Esta matriz cristaliza de manera homogénea con la

muestra y al ser irradiada por el láser absorbe la radiación UV, transmitiéndole protones a la muestra e impidiendo la fragmentación de los iones.

Esto permite la detección de iones de gran tamaño en fase gaseosa, por lo general con relaciones m/z entre 2000 y 20000 Daltons (Da), región en la cual se encuentran la mayoría de proteínas ribosómicas de los microorganismos, y que permiten su diferenciación.

Cuando el pocillo es irradiado con el láser, la matriz se sublima arrastrando consigo la muestra a estudio, y pasa a estado gaseoso directamente sin pasar por estado líquido, ionizándose inmediatamente. La mayor parte de la energía es absorbida por la matriz, lo que permite que la muestra se ionice sin fragmentarse en exceso.

Los distintos iones de la muestra vuelan a través de la columna del analizador, aceleradas por un campo magnético hasta impactar con el detector, que registra el tiempo de vuelo en función de la masa de cada ion. Esto da lugar a un espectro de masas en el que cada pico es un ion, y que representa, en el eje de ordenadas la intensidad, que simboliza la abundancia relativa de cada ion que ha impactado contra el detector, y en el eje de abscisas la relación m/z de cada ion. Para la identificación microbiana, este espectro de masas es después analizado y comparado con una base de datos, que tiene en cuenta diferentes picos biomarcadores de cada especie y permite la identificación del microorganismo, si este se encuentra presente y representado en la base de datos. Si un microorganismo no se encuentra en la base de datos, por muy buena que sea la calidad de espectro obtenido, este no se identificará correctamente [33].

En los últimos años han salido al mercado numerosos equipos que emplean la tecnología MALDI-TOF para la identificación de microorganismos en los laboratorios de microbiología clínica, siendo los más distribuidos los equipos MALDI-TOF de Bruker Daltonics (MALDI Biotyper Smart, MALDI Biotyper Sirius, MALDI Microflex, MALDI Ultraflex...) y de Biomerieux (VITEK MS y VITEK MS Prime). Otras compañías como ASCEND, Shimadzu, AuTOF-MS1000 y Zybio se van abriendo camino actualmente en el mercado.

3.1 Aplicación de MALDI-TOF para la identificación de microorganismos

La principal aplicación actual de la tecnología MALDI-TOF en los laboratorios de microbiología clínica es la identificación de microorganismos aislados de muestras clínicas que se reciben en el laboratorio. Una identificación rápida es fundamental para el manejo adecuado del paciente, ya que permite una instauración correcta del tratamiento antimicrobiano [33, 36].

El MALDI-TOF empezó a emplearse en los laboratorios de microbiología clínica durante los años 2009-2010. Su instauración supuso una mejora muy significativa en dichos laboratorios en cuanto al tiempo para la obtención de resultados de identificación bacteriana, reduciéndolo de 24-72 horas a minutos. A día de hoy, la identificación de microorganismos mediante MALDI-TOF se encuentra perfectamente estandarizada en los laboratorios clínicos, y se emplea de manera rutinaria [19, 36].

Como ejemplo real, en el laboratorio de microbiología clínica del Hospital General Universitario Gregorio Marañón (HGM), hospital donde se desarrolló esta tesis, el MALDI-TOF se instaló y se comenzó a implementar para identificación bacteriana en el año 2011. Actualmente se identifican todo tipo de microorganismos con esta tecnología de manera rutinaria. Principalmente se identifican bacterias crecidas en medio sólido, pero también hongos y levaduras, micobacterias, microorganismos a partir de hemocultivo directo y en ocasiones directamente a partir de BHI, orina y líquido cefalorraquídeo cuando la muestra lo permite y en función de las necesidades del paciente. En el año 2023, en este laboratorio se analizaron un total de 66175 muestras mediante MALDI-TOF, con una media de 5578 muestras por mes, y ya se trata de un equipo completamente integrado en la rutina asistencial del laboratorio de microbiología.

La identificación de microorganismos de relevancia clínica puede realizarse de dos maneras, a partir de un cultivo crecido en medio sólido o líquido, o directamente a partir de muestra clínica en determinadas ocasiones que la muestra lo permita.

A) A partir de cultivo

A partir de una colonia crecida en medio de agar sólido, es posible hacer una identificación empleando MALDI-TOF en apenas minutos. La colonia debe colocarse en una posición de la placa metálica para su análisis y cubrirse con 1µl de matriz orgánica de HCCA. Este proceso puede mejorarse con una extracción de proteínas “on-target” directamente en la placa añadiendo 1µl de ácido fórmico puro tras la aplicación de la colonia y previo a la adición de la matriz. En condiciones en las que los microorganismos sean de difícil identificación debido al grosor de su pared, como es el caso de los hongos, micobacterias o algunas bacterias Gram-positivas, es posible hacer una extracción proteica previa con etanol, ácido fórmico y acetonitrilo, que facilite la disrupción de la pared celular y permita obtener espectros de calidad óptima [37].

También es posible realizar la identificación bacteriana a partir de cultivos líquidos en medio tioglicolato o BHI, dos medios de cultivo enriquecidos que mejoran el crecimiento de las bacterias. La manera rutinaria de proceder con estos cultivos es incubándolos previamente durante al menos 16-24 horas tras la inoculación de la muestra clínica para permitir el crecimiento, y posteriormente cuando se observa turbidez, dar un pase a medio de cultivo sólido a partir del cual se identifica tras otra incubación (lo que añade otras 16-24 horas más al tiempo de identificación). La ventaja de identificar directamente a partir de cultivo líquido con MALDI-TOF es que permite ahorrar el tiempo de la segunda incubación, pudiendo identificar directamente cuando se observe crecimiento suficiente (turbidez) en el medio líquido.

En este proceso de identificación se incluyen pasos de centrifugaciones previas para la limpieza de la muestra y la eliminación de residuos que pudieran interferir. Se puede añadir un paso intermedio de sonicación para mejorar la extracción de la muestra. Como resultado, se obtiene un pellet que permite la identificación directamente en la placa de MALDI-TOF [38].

Esta identificación puede ser útil en el caso de microorganismos de difícil crecimiento, que no sean capaces de crecer en placas de agar convencionales y con las condiciones que se emplean rutinariamente en los laboratorios de microbiología, como en el caso de muestras procedentes de pacientes con meningitis, infección

INTRODUCCIÓN

protésica o endocarditis, y también con bacterias anaerobias. Tiene la desventaja de no funcionar correctamente en el caso de muestras polimicrobianas, pudiendo identificar con éxito hasta un máximo de dos de los microorganismos presentes en la muestra.

El proceso de identificación directa a partir de medio líquido también se ha empleado con éxito en la identificación de micobacterias a partir de medio MGIT con un paso previo de inactivación de las micobacterias una vez crecidas. Esto ha supuesto una mejora en los tiempos de identificación con respecto al cultivo en medio sólido, además de un ahorro en costes en relación a la identificación convencional empleada hasta el momento por métodos moleculares, convirtiéndolo en el método de elección para la identificación de micobacterias en muchos laboratorios clínicos [39, 40].

B) A partir de muestra directa

Determinadas muestras biológicas permiten una identificación directa a partir de la muestra clínica sin tener que esperar al crecimiento en medios de cultivo. Esto es posible solo en casos concretos en los que se cumplan una serie de requisitos. MALDI-TOF identifica correctamente el patógeno presente en hemocultivos positivos de manera inequívoca en prácticamente todos los casos de infección monomicrobiana, especialmente cuando se trata de una bacteria Gram-negativa. La interpretación de los resultados en caso de infecciones polimicrobianas no es tan sencilla, ni siquiera disponiendo de un módulo específico para ellos, como en el caso de utilizar el kit Sepsityper [41, 42]. Esta característica puede conocerse de antemano con la realización de una tinción de Gram directa de la muestra. Por otra parte, la carga bacteriana de la muestra clínica debe ser alta, de al menos 10^5 UFC/ml [43], cosa que no se cumple en la mayoría de las infecciones, siendo necesaria una incubación previa para aumentar el número de bacterias. Además, es preciso disponer de un volumen de muestra suficiente para poder concentrar las bacterias por centrifugación y obtener un pellet suficiente para la identificación. En todo caso, debe tratarse de muestras biológicas que en condiciones normales sean estériles, para evitar interferencias en la identificación del microorganismo causante de infección [33, 44].

Se han realizado estudios de identificación de bacterias en muestra directa de sangre, orina, líquido pleural, líquido articular y líquido cefalorraquídeo con resultados prometedores [45-48]. Quizás, el menos aconsejado de ellos sea el análisis directo de LCR, por las pequeñas cantidades de muestra que se obtienen, siendo preferible destinar la muestra a un panel molecular de PCR multiplex si estuviera disponible, ya que estos tienen límites de detección inferiores y son más sensibles en el caso de una baja carga bacteriana.

La identificación de microorganismos a partir de muestra directa sí se encuentra implantada en el caso de hemocultivos positivos para el diagnóstico de bacteriemia y sepsis. La presencia de bacterias en sangre se trata de un proceso con una elevada tasa de mortalidad en el que se ha demostrado que el tiempo de actuación para la instauración de un correcto tratamiento antibiótico es decisivo para la supervivencia del paciente, por lo que es crucial tener la identificación del microorganismo lo antes posible [49, 50]. Actualmente se han diseñado y se encuentran a la venta kits comerciales para la identificación de microorganismos a partir de hemocultivo directo como Sepsityper de Bruker Daltonics, RapidBacPro de Nittobo o Vitek MS Blood Culture de Biomerieux [51, 52].

El método “casero” o “*in-house*” de identificación a partir de hemocultivo directo consiste en un proceso de lisis y centrifugaciones seriadas y lavados para concentrar la muestra y obtener un sedimento o pellet a partir del cual realizar la identificación como si se tratara de una colonia crecida en medio sólido (la primera centrifugación se centra en la eliminación de las células eucariotas presentes en la muestra, y las siguientes, en la concentración del microorganismo y la eliminación de otras sustancias que puedan interferir) [53]. En el caso de no obtener resultados de esta manera, se puede intentar mejorar el procedimiento con una extracción con etanol y ácido fórmico o realizar un subcultivo corto (3-5 horas) de una gota del hemocultivo positivo para que haya crecimiento suficiente que permita la identificación [54, 55]. También se puede realizar una identificación empleando tubos de extracción de suero con gel separador, en el que, tras centrifugar, las bacterias se quedarán pegadas a este gel y se podrán identificar de su superficie [56, 57].

El método de centrifugaciones seriadas presenta una ventaja adicional, ya que a partir de este pellet también se podrán realizar estudios de sensibilidad convencionales de microdilución sin tener que esperar al crecimiento en cultivo [58].

C) Retos en la identificación de microorganismos con MALDI-TOF

Tras la adquisición con el MALDI-TOF de Bruker Daltonics (el sistema MALDI-TOF empleado en el desarrollo de esta tesis doctoral) se obtiene un espectro media que es el sumatorio de los espectros obtenidos en las distintas mediciones realizadas para una única posición. Por lo general, para identificación se realizan disparos de láser de 40 en 40 hasta un sumatorio de 240 disparos. Este espectro media, que se realiza en la región de 2 a 20 kDa, se enfrenta a una base de datos y los resultados de esta comparación se miden con una escala de puntuación logarítmica que va de 0 a 3 según la similitud del espectro problema con espectros de referencia presentes en dicha base de datos. Con este tipo de MALDI-TOF, una identificación con una puntuación $\geq 2,0$, es fiable a nivel de género y especie; un score entre 1,7 y 2,0 da un resultado fiable a nivel de género; y un resultado $< 1,7$ daría como resultado una identificación no fiable o ausencia de identificación por comparación con la base de datos.

La ventaja que presenta este sistema es que la base de datos no es cerrada y se puede ampliar con microorganismos correctamente caracterizados mediante métodos moleculares para mejorar su funcionamiento. Esto se ha llevado a cabo sobre todo con especies de *Nocardia*, micobacterias, hongos y bacterias anaerobias, que estaban poco representados en las bases de datos iniciales de Bruker [59-62].

Hay determinados microorganismos como las micobacterias que presentan dificultad en su identificación con MALDI-TOF, por lo que se ha llegado a un consenso en la comunidad científica de ampliar la puntuación considerada como válida al nivel especie a $\geq 1,6$ si esta identificación se repite en todas las identificaciones que aporta el MALDI-TOF [63, 64]. En la Figura 4 se resume el flujo de trabajo para la identificación de microorganismos empleando MALDI-TOF.



Figura 4. Proceso de identificación de microorganismos mediante MALDI-TOF

Pese a que la identificación mediante MALDI-TOF se encuentra estandarizada en la mayoría de laboratorios clínicos, hay determinadas especies o complejos que todavía suponen un reto para su correcta identificación con MALDI-TOF, y que hay que tener en cuenta a la hora de informarlos en el laboratorio. Este es el caso de la diferenciación entre *Escherichia coli* y *Shigella*, las especies incluidas dentro de los complejos de *Enterobacter cloacae* o *Citrobacter freundii*, o la distinción de *Streptococcus pneumoniae* frente a *Streptococcus* del grupo *mitis* o *bovis*. Las principales dificultades en la identificación de especies bacterianas con el MALDI se resumen en la Tabla 1.

Muchos de estos problemas de identificación se han ido solucionando con las actualizaciones de las librerías comerciales, y frente a los que no, se han publicado varios artículos científicos aportando soluciones para la diferenciación entre estas bacterias problemáticas [65-70].

Hay diversos factores que pueden afectar a la correcta identificación, entre ellos, la cantidad de microorganismo disponible para analizar, la antigüedad o edad del propio cultivo, el medio de cultivo y el protocolo de extracción de proteínas empleado [71-73]. Es preferible emplear cultivos que se encuentren en fase logarítmica de crecimiento para evitar esporas, productos metabólicos o autólisis y degradación de las proteínas que se pretenden analizar. Otros factores como la presencia de cápsula o

INTRODUCCIÓN

pared celular también pueden entorpecer la correcta identificación, siendo necesario un paso de extracción previa con ácido fórmico y acetonitrilo [37, 74].

Tabla 1. Principales especies bacterianas que suponen un problema en la identificación con MALDI-TOF de Bruker

MICROORGANISMOS	PROBLEMA
<i>Escherichia coli</i> y <i>Shigella</i> spp.	No diferencia entre ambos microorganismos
<i>Enterobacter cloacae</i> complex	No diferencia entre las especies pertenecientes al complejo
<i>Citrobacter freundii</i> complex	No diferencia entre las especies pertenecientes al complejo
<i>Achromobacter</i> spp.	No diferencia entre las especies <i>A. ruhlandi</i> y <i>A. xylosoxidans</i>
<i>Brucella</i> spp.	Confusión con el género <i>Octrobactrum</i>
<i>Burkholderia cepacia</i> complex	No diferencia entre las especies pertenecientes al complejo
<i>Burkholderia mallei</i> y <i>B. pseudomallei</i>	Solo identificación a nivel de género
<i>Klebsiella pneumoniae</i> complex	No diferencia correctamente entre las especies <i>K. pneumoniae</i> , <i>K. quasipneumoniae</i> , <i>K. variicola</i> y <i>K. quasivariicola</i>
<i>Neisseria meningitidis</i>	Puede confundirla con otras especies de <i>Neisseria</i> comensales de la nasofaringe
<i>Salmonella</i> spp.	No permite identificar <i>S. typhi</i>
Grupo <i>Bacillus cereus</i>	No diferencia entre las especies pertenecientes al grupo. Confusión con <i>B. anthracis</i>
<i>Streptococcus pneumoniae</i> y <i>Streptococcus grupo mitis</i>	No diferencia entre ambos microorganismos

Independientemente de todo esto, por muy buena que sea la calidad del espectro obtenido, si este microorganismo no se encuentra presente en la base de datos, su identificación no será posible [75]. Por último, para garantizar una correcta identificación de los picos de cada espectro, es preciso calibrar el MALDI con frecuencia empleando un estándar interno [76]. En el caso de Bruker se emplea un estándar comercial denominado BTS, por sus siglas Bacterial Test Standard, que

contiene una suspensión de una cepa de *E. coli* liofilizada, comercializada por el distribuidor con picos de masa claramente definidos para permitir la calibración.

En resumen, un espectro de masas de buena calidad es aquel que presenta un elevado número de picos biomarcadores en la región estudiada que permitan su identificación, con una buena resolución (intensidad elevada y bajo ruido), y una reproducibilidad técnica y biológica precisa. Estos aspectos se resumen en la Figura 5, extraída del artículo de Cuènod A. et al, 2021 [71].

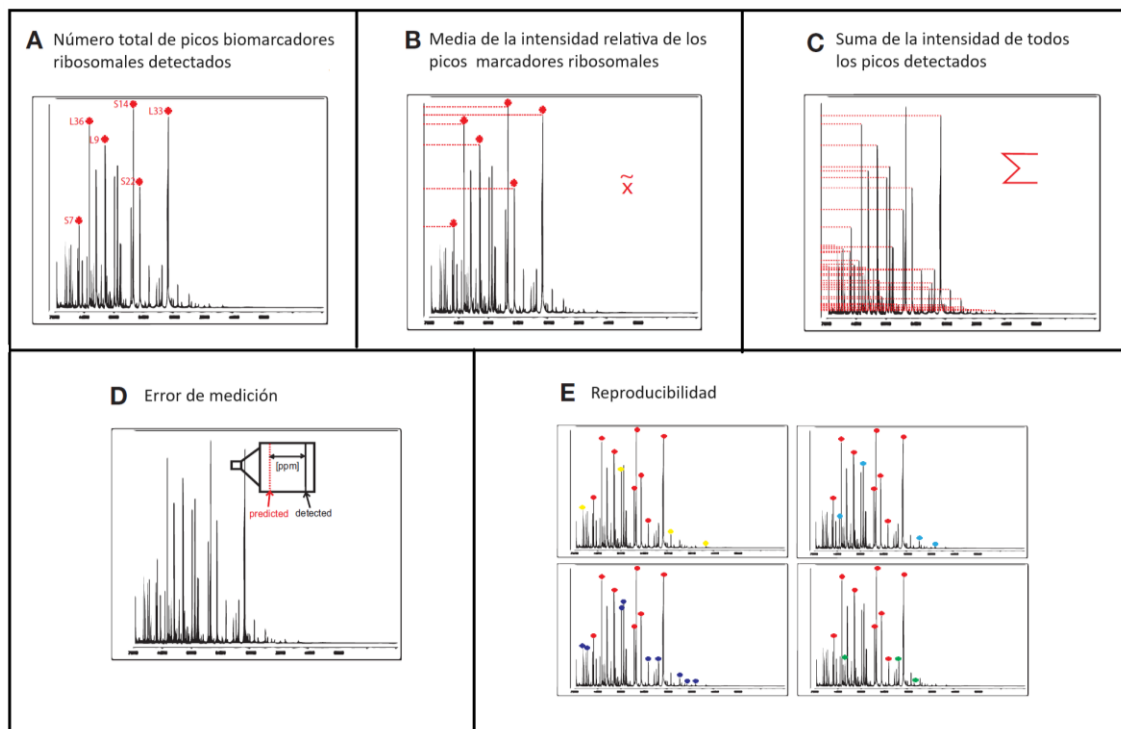


Figura 5. Parámetros que influyen en la calidad de un espectro de masas

Más allá de la identificación bacteriana, desde la implantación de MALDI-TOF en los laboratorios de microbiología clínica, se ha intentado emplear esta tecnología para otros fines diferentes a la identificación, que permitan una caracterización más rápida de los microorganismos, de su epidemiología y de su perfil de sensibilidad. Se ha empleado el MALDI-TOF en estudios de sensibilidad y resistencia a diversos antimicrobianos, así como en estudios de tipado bacteriano para control epidemiológico.

3.2 Aplicación de la tecnología MALDI-TOF para la detección de resistencias antimicrobianas

La determinación convencional de sensibilidad por microdilución en caldo o mediante tiras de gradiente o discos de antibiótico es un proceso que conlleva entre 18 y 24 horas para la obtención de resultados concluyentes. Actualmente se encuentra bajo estudio el método RAST (Rapid Antimicrobial Susceptibility Testing) de detección rápida de sensibilidad, que permite intuir los resultados de sensibilidad mediante difusión con discos en agar al empezar a observarse visualmente crecimiento del microorganismo, entre las 4 y las 8 horas [77, 78].

La ventaja de MALDI-TOF frente a esta técnica es que permite obtener resultados todavía más rápidos para ciertos antibióticos, lo que aporta un valor añadido al clínico junto con la identificación del microorganismo para instaurar un correcto tratamiento antibiótico. Además, tras la inversión inicial de adquisición del equipo, la determinación de sensibilidad mediante espectrometría de masas MALDI-TOF no supone un sobrecoste añadido más que el de los reactivos de matriz, ácido fórmico, y en algunos casos, el coste del antibiótico, lo que lo hace más accesible que determinados métodos de estudio rápido de sensibilidad como son los paneles moleculares multiplexados o incluso los estudios fenotípicos.

El análisis de sensibilidad y resistencia a un antimicrobiano se puede estudiar mediante MALDI-TOF de tres maneras distintas: 1) Analizando directamente el espectro de cada bacteria en busca de picos biomarcadores de resistencia o comparando el espectro completo de bacterias resistentes y sensibles, 2) Analizando los cambios que se producen en el antibiótico tras una corta exposición a bacterias resistentes, y 3) Analizando los cambios en las bacterias comparando espectros previos y posteriores a la exposición al antibiótico [79].

A) Análisis directo del perfil proteico bacteriano

El abordaje más sencillo para el análisis de resistencias antibióticas mediante MALDI-TOF consiste en la búsqueda de picos biomarcadores que puedan estar asociados a la expresión de un determinado gen de resistencia y permitan diferenciar entre bacterias sensibles y resistentes. Esto puede hacerse mediante la comparación de espectros entre sí, o con espectros de referencia de los cuales su sensibilidad ya es conocida y está comprobada mediante métodos moleculares. Se pueden analizar posibles picos biomarcadores asociados a la expresión de alguna proteína que marque un determinado fenotipo de resistencia o analizar el espectro completo en búsqueda de diferencias que permitan la clasificación de las cepas [80].

La desventaja de este método es que puede llegar a ser poco reproducible, y es necesario validarlo internamente con los aislados de cada institución o zona geográfica, ya que los picos que diferencian las cepas no necesariamente pueden deberse a mecanismos de resistencia, sino a clones concretos que lleven esa resistencia asociada. Se han publicado estudios con este abordaje para la diferenciación de SARM y SASM [81, 82], la detección de *Bacteroides fragilis* portador de *cfiA* [83, 84], diferenciación entre *Enterococcus faecium* resistentes y sensibles a vancomicina [85], detección de *Klebsiella pneumoniae* portadora de carbapenemasa tipo KPC [86, 87], etc. Al igual que se han publicado otros rebatiendo la falta de estandarización del método y su variabilidad según la zona geográfica [88-90].

Este enfoque para la detección de resistencias, además de validarse en cada lugar geográfico, debe estandarizarse y llevarse a cabo siguiendo siempre la misma metodología, ya que se ha demostrado que hay factores externos y ajenos a la bacteria que afectan a la presencia o ausencia de picos, como puede ser el medio de cultivo empleado o la edad del cultivo bacteriano [80, 91].

Basándose en muchos de estos estudios, Bruker ha desarrollado el módulo “MALDI Biotyper subtyping software”, que permite la detección de determinados patrones de resistencia tras la identificación de la bacteria si ésta se identifica con un score $\geq 2,0$. Este es el caso de SARM que presentan un pico en 2415 Da, determinadas

INTRODUCCIÓN

Enterobacterias productoras de KPC que presentan el pico 11109 Da o cepas de *B. fragilis* productoras de *cfiA* (división II) que se distinguen de las no productoras (división I) según una serie de picos que se encuentran presentes o ausentes entre 4000 y 5000 Da [33, 92, 93].

En los últimos años se han aplicado tecnologías de *Machine Learning* para el análisis automatizado de picos que permiten un estudio más minucioso y estandarizado que la observación a simple vista de presencia o ausencia de picos. Mediante algoritmos matemáticos, esta metodología permite clasificar las cepas según un determinado patrón de picos específicos del grupo de microorganismos que se pretenda estudiar [94-96].

B) Estudio de cambios de masa en el antibiótico mediante la detección de la actividad enzimática bacteriana

Una de las técnicas más fiables para detectar resistencia mediante MALDI-TOF, es el estudio de la inactivación del antibiótico mediante hidrólisis por parte de enzimas bacterianos, cuando ambos se enfrentan en disolución. Es una técnica muy precisa ya que se estudian antibióticos con masas conocidas y fijas, cuyos picos varían al exponerse a la bacteria si esta es resistente y los hidroliza. Además, permite observar la actividad enzimática de la bacteria en tiempo real.

Este procedimiento se desarrolló inicialmente para analizar la resistencia a antibióticos β -lactámicos, ya que lo que se estudia es la ruptura del enlace amida del anillo β -lactámico tras la adición de una molécula de agua [97]. La adición de esta molécula de agua hace que varíe la masa del antibiótico, lo que permite distinguir entre aislados sensibles y resistentes. Para este tipo de ensayos se han empleado penicilinas (ampicilina, piperacilina), cefalosporinas de tercera generación (cefotaxima, ceftazidima y ceftriaxona) para detectar actividad β -lactamasa, y antibióticos carbapenémicos (ertapenem, meropenem e imipenem) para detectar actividad carbapenemasa. En la Tabla 2 se resumen los picos de masa específicos de estos antibióticos, así como los cambios de masa tras su hidrólisis [97, 98].

INTRODUCCIÓN

De igual manera que los ensayos de hidrólisis de β -lactámicos, se han intentado llevar a cabo estudios de resistencia con especies bacterianas productoras de enzimas modificadoras de aminoglucósidos de tipo AAC(6')-Ib-cr [99, 100]. Esta enzima es una acetiltransferasa que añade un grupo acetilo al antibiótico volviéndolo inactivo. El MALDI-TOF en este caso, también detecta un cambio de masas en el antibiótico al ser acetilado, ya que esta modificación supone un aumento en la masa de 43 Da.

Es importante resaltar que mediante los estudios de hidrólisis solo se detecta resistencia antibiótica mediada por enzimas de resistencia, y que las bacterias pueden ser resistentes mediante otros mecanismos que no se estén detectando, como la presencia de bombas de expulsión de antibióticos, la pérdida de porinas que impide que el antibiótico acceda a la bacteria para ser efectivo, o mutaciones en genes de la propia bacteria que impidan la acción del antibiótico.

Esta técnica de determinación de resistencia, además de llevarse a cabo a partir de un cultivo bacteriano puro, también se puede estudiar a partir de muestras directas, como las orinas o los hemocultivos, añadiendo un paso previo de lisis para formar un pellet bacteriano sobre el que realizar el estudio [101, 102]. Bruker ya dispone de un kit automatizado, el MBL STAR Carba (para detección de carbapenemasas) y MBL STAR BL (para detección de β -lactamasas), que analiza de manera automatizada los espectros, arrojando resultados al momento de la identificación sin la necesidad de análisis posteriores [103, 104].

C) Estudio del perfil proteico bacteriano tras la incubación con el antibiótico

Otra aproximación para la detección de resistencia a antimicrobianos consiste en la incubación de la propia bacteria con el antibiótico durante un tiempo suficiente que permita observar su crecimiento. Posteriormente se analizan los cambios en el perfil proteico, midiendo la ratio del área bajo la curva de los picos de la bacteria en presencia de antibiótico entre los picos de la bacteria en ausencia de antibiótico [105]. Bruker ha desarrollado también un software comercial, el MBT-ASTRA, que indica sensibilidad o resistencia directamente tras la medición del espectro cuando se realiza

este ensayo. También se han publicado estudios de marcaje con isótopos no radioactivos para la cuantificación del crecimiento, pero dada su complejidad estos no se emplean en la actualidad [106-108].

En el estudio publicado por Idelevich et al [109] se desarrolla una metodología simplificada para la detección de aislados resistentes de *K. pneumoniae* y *Pseudomonas aeruginosa*, el DOT-MGA, en la que se estudia la acción del antibiótico sobre el crecimiento de la bacteria en la propia placa metálica de MALDI-TOF. Tras la incubación de la bacteria con el antimicrobiano, se retira el medio de cultivo con un papel absorbente y se añade matriz orgánica (HCCA) para analizar cada posición de la placa mediante MALDI-TOF. Si se trata de un microorganismo sensible se observa un espectro no identificable por MALDI-TOF, ya que el antibiótico ha actuado de manera efectiva contra la bacteria. Por el contrario, en el caso de que se trate de una bacteria resistente, se observará un espectro con picos que sí que podrá identificarse, ya que el antibiótico no es activo frente a esta bacteria. En la Figura 6 se observa un ejemplo extraído del trabajo de Idelevich [109] en el que se representa el espectro obtenido de un bacteria resistente (A) y de una sensible (B).

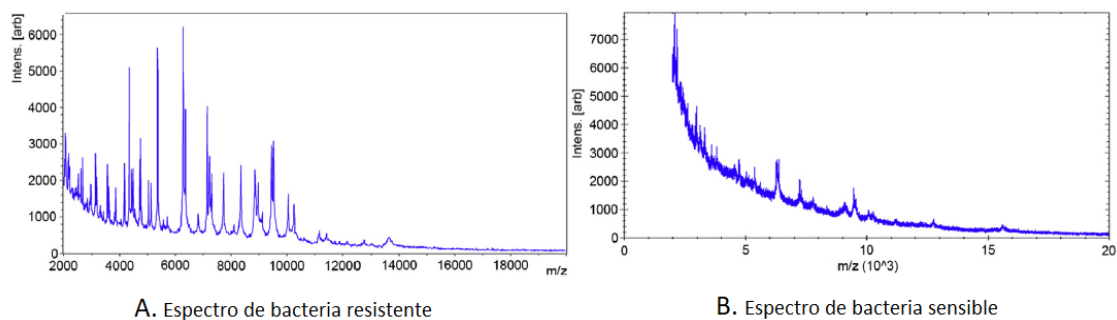


Figura 6. Ejemplo de espectros obtenidos en el ensayo DOT-MGA [109]

Inicialmente el modelo se desarrolló con meropenem, pero en estudios posteriores se probaron diferentes antibióticos, con resultados de concordancia entre el 70 y 100% con los ensayos convencionales de sensibilidad [110]. También se ha evaluado este ensayo con bacterias Gram-positivas para diferenciar SARM de SASM y *S. pneumoniae* de *S. mitis* [111, 112].

INTRODUCCIÓN

La principal desventaja de este método es que no puede considerarse un método rápido per se, ya que es necesario que el período de incubación sea el suficiente para permitir el crecimiento o la inhibición bacteriana. La ventaja es que permite la detección de mecanismos de resistencia independientes de la actividad enzimática bacteriana, como la presencia de bombas de eflujo o la pérdida de porinas, que evitan que el antibiótico actúe y permiten el crecimiento bacteriano, lo que se puede medir mediante MALDI-TOF.

3.3 Aplicación de la tecnología MALDI-TOF para el tipado de cepas clínicas

La tecnología MALDI-TOF puede emplearse también como técnica rápida y de primera línea para el tipado de bacterias y la detección de brotes, por su rapidez en el tiempo de obtención de resultados y su bajo coste.

El tipado bacteriano permite diferenciar subtipos dentro de una misma especie y estudiar la clonalidad entre aislados. Se puede realizar en dos momentos diferenciados: retrospectivamente para entender la diseminación de clones, y prospectivamente para la detección de brotes y el diagnóstico de pacientes colonizados, para poder implementar medidas frente a la diseminación y control de la infección para la prevención de brotes [113].

En la actualidad se emplean diferentes métodos para el tipado bacteriano en el laboratorio de microbiología. La electroforesis en gel por campo pulsado (PFGE) y el tipado por secuenciación de múltiples locus (MLST) se han considerado las técnicas *Gold Estándar* hasta hace poco tiempo que se ha empezado a utilizar la secuenciación masiva y se ha empezado a cuestionar. En la técnica de PFGE, el ADN de las cepas a estudio se fragmenta con enzimas de restricción y posteriormente estos fragmentos se revelan en un gel de agarosa con campos eléctricos desde diferentes ángulos, lo que permite observar distintos *pulsotipos* para la clasificación de los aislados. En el MLST se realiza una PCR de 7-8 genes *housekeeping* o conservados que permite identificar polimorfismos en las bacterias clasificándolas según su ST (*Sequence type*). Ambas técnicas son largas y costosas, y en algunos casos no presentan suficiente poder discriminatorio para distinguir entre cepas [114].

También se pueden realizar estudios de tipado mediante WGS con las técnicas de core genome MLST (cg-MLST) y whole genome MLST (wg-MLST), estudios de secuenciación masiva que mejoran el MLST convencional y amplían el número de genes estudiados para el tipado buscando alelos en multitud de genes conservados del genoma core (cg-MLST, wg-MLST) y accesorio (wg-MLST) de cada especie. Estas dos técnicas además permiten calcular las distancias entre cepas según sus similitudes y diferencias en estos genes, algo que con el PFGE y el MLST convencional no es posible. Son técnicas robustas y con alto poder de resolución, sin embargo suponen también un elevado coste a nivel instrumental y de reactivos, y una necesidad de personal formado en bioinformática para un análisis especializado y complejo de las secuencias [114].

Es por todo esto que se ha planteado el uso del MALDI-TOF para el tipado bacteriano. Principalmente porque el instrumento ya se encuentra instalado en la mayoría de laboratorios, y además porque tras la inversión inicial en el equipo no supone un gasto añadido elevado, ya que los reactivos que se emplean son los mismos que se utilizan en la identificación bacteriana. Es una tecnología más rápida que las mencionadas anteriormente, lo que permite obtener resultados preliminares para orientar a los clínicos en el control de la infección. Sin embargo, el tipado con MALDI-TOF es un método menos robusto que la identificación, por lo que hay una serie de requisitos que es necesario tener en cuenta y cumplir a la hora de analizar aislados para tipar.

La metodología a seguir en el tipado con MALDI-TOF se basa en la creación de un modelo inicial (*training set*) con una colección de aislados bien caracterizada molecularmente (mediante PFGE, MLST, cg-MLST...) en la que se identifiquen picos biomarcadores que permitan diferenciar las bacterias mediante herramientas informáticas (FlexAnalysis, ClinProTools, BioNumerics, MassUp, CloverBiosoft...). Posteriormente este modelo ha de validarse con una colección independiente de cepas también caracterizadas, de 5 a 10 veces mayor en número. Se deben realizar cálculos de sensibilidad, especificidad, VPP y VPN y una vez se obtengan los valores adecuados, se puede empezar a aplicar el modelo con cepas clínicas. El procedimiento a seguir se resume en la Figura 7 [113].

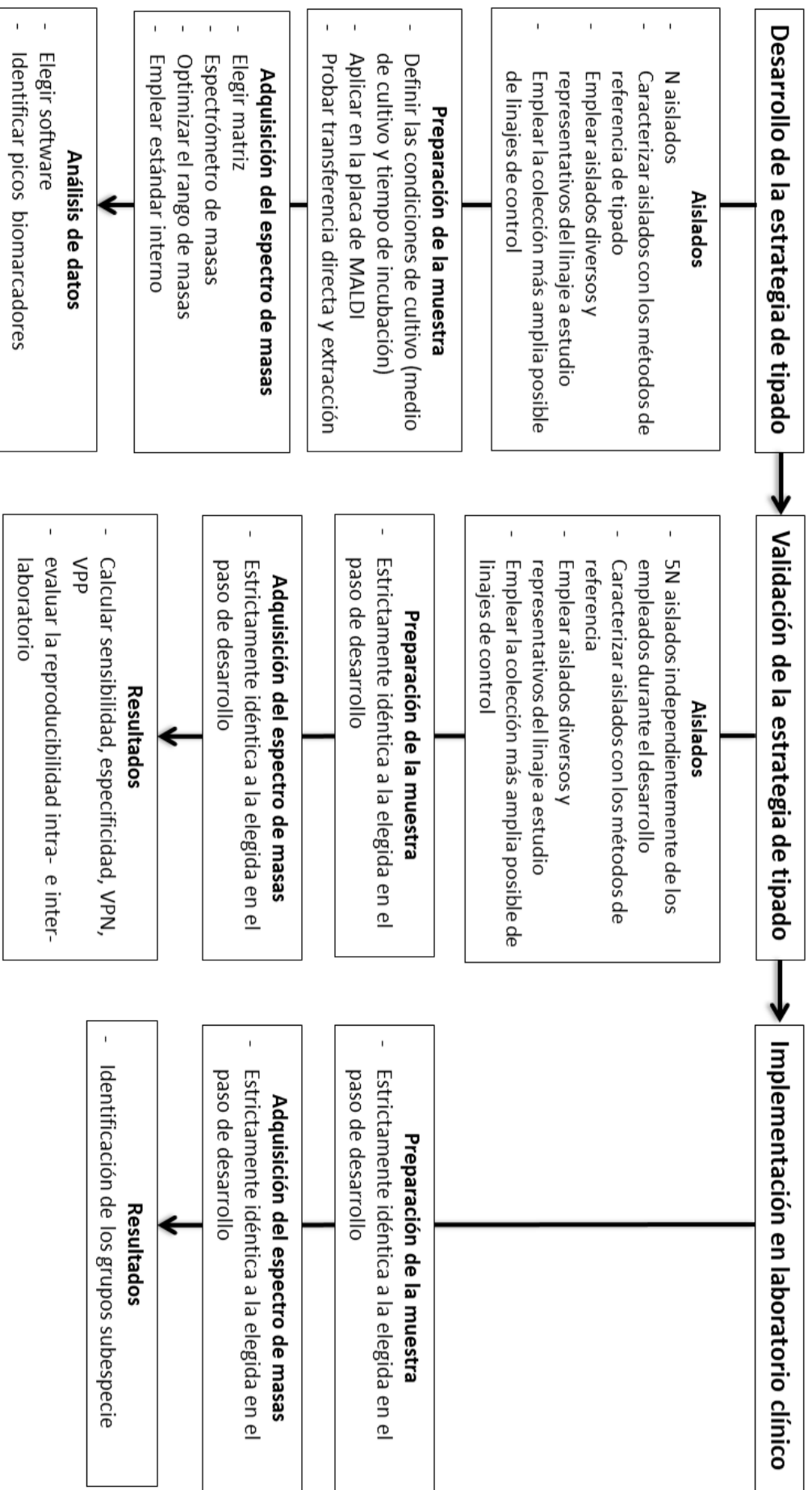


Figura 7. Metodología recomendada en el tipado de cepas con MALDI-TOF. Adaptado de Sauguet et al, 2017

Para obtener reproducibilidad a la hora de tipar con MALDI-TOF es imprescindible que las condiciones de cultivo bacteriano sean las mismas para todas las bacterias analizadas, ya que variaciones en el medio de cultivo o en el tiempo de incubación pueden afectar a la expresión proteica y por consiguiente al tipado [80, 115, 116]. También el protocolo de preparación de la muestra puede alterar el espectro de masas según el tipo de extracción empleado, los solventes utilizados, y la concentración y composición de la matriz. La potencia del láser, el nivel de desgaste y la calidad de la muestra en el pocillo también pueden alterar los valores m/z . Se deben optimizar las condiciones de cultivo, extracción proteica y adquisición de espectros para cada método y mantenerlos cada vez que se analicen nuevas cepas [113, 117].

Igualmente, se recomienda hacer análisis de reproducibilidad técnica (analizar varios pocillos de la placa) y biológica (analizar la misma cepa días consecutivos), obteniendo varios espectros para una misma cepa, para así asegurar que se obtienen resultados correctos y reproducibles [118].

En definitiva, el tipado con MALDI-TOF es una técnica prometedora pero que necesita del desarrollo de guías internacionales por ahora no disponibles, con la descripción de protocolos sencillos y reproducibles por todos los laboratorios, para una correcta aplicación clínica e interpretación de resultados permitiendo evitar al máximo la variabilidad intercentro.

Al no haber guías disponibles, es preciso poner la técnica a punto antes de su aplicación clínica probando varios medios de cultivo, matrices y métodos de extracción, así como la utilización de estándares internos y un correcto post-tratamiento de los espectros. En ocasiones puede ser necesario un cambio en el rango de masas a estudiar, aunque por lo general, los picos más discriminatorios se encuentran en la región entre 4000 y 10000 Da, y casi nunca por encima de 15000 [119]. Antes del procesamiento informático es necesario alinear los espectros con picos conservados ya descritos de la propia especie para evitar la variabilidad intrínseca de los espectros de masas. Idealmente, se deben validar los estudios de tipado con MALDI-TOF de manera multicéntrica, con centros de localizaciones

INTRODUCCIÓN

geográficas diversas, para poder contrastar si los resultados que se observan son reproducibles a gran escala.

Hasta el momento, se han realizado estudios de tipado con MALDI-TOF para diferenciar subespecies de *Mycobacterium abscessus* [120, 121], clones de alto riesgo de *P. aeruginosa* [122-124], ribotipos de *Clostridioides difficile* [125, 126], complejos clonales y STs asociados con resistencia a meticilina de *Staphylococcus aureus* [127-131], clones de *K. pneumoniae* [132, 133], tipos capsulares de *Haemophilus influenzae* [134], cepas de *Achromobacter xylosoxidans* aisladas de pacientes con fibrosis quística [65], serovares de *Salmonella enterica* [135], serotipos de *S. pneumoniae* [136], subtipos de *Streptococcus agalactiae* [137, 138], filogrupos de *E. coli* [139-141]... con diversidad en los resultados debido a la falta de estandarización entre laboratorios [142].

Es presumible que el tipado con MALDI-TOF no sustituirá a la secuenciación masiva o a las técnicas genotípicas convencionales, pero es una herramienta rápida y coste-efectiva, que puede ser usada de manera complementaria como primera línea de tipado para la implementación de medidas de control de la infección y para prevenir la diseminación de patógenos en el ámbito hospitalario.

4 ESPECTROSCOPIA INFRARROJA POR TRANSFORMADA DE FOURIER (FTIR)

Como técnica complementaria a la tecnología MALDI-TOF para la detección y tipado de brotes hospitalarios, se está comenzando a utilizar en el laboratorio clínico la tecnología de espectroscopía infrarroja por transformada de Fourier (FTIR). Esta tecnología presenta utilidad en el tipado molecular y en la detección y seguimiento de brotes a tiempo real, aportando resultados en menos de 3 horas.

La tecnología FTIR se desarrolló mayoritariamente en la década de 1990, donde se observó que cada bacteria al ser irradiada emitía un espectro infrarrojo característico que permitía identificarla y diferenciarla de las demás [143-145]. Sin

embargo, esta tecnología no se comenzó a aplicar hasta hace unos años, ya que la inexactitud de los sistemas de clasificación bacteriana existentes en aquel momento, bien por las dudas en el posicionamiento taxonómico de los microorganismos o por la metodología poco desarrollada para el tipado, junto con la ausencia de bases de datos y la falta de estandarización metodológica, no hacían posible saber si los resultados obtenidos con el FTIR eran concordantes y correctos [146].

Con la aparición y mejora de las técnicas moleculares y la secuenciación masiva en concreto, se ha comprobado que las caracterizaciones moleculares obtenidas por FTIR se correlacionan con los resultados de secuenciación, y se ha comenzado a utilizar esta herramienta en el laboratorio clínico. A pesar de que la espectroscopía FTIR podría emplearse en microbiología para la identificación de microorganismos, al estar ya instaurada la tecnología MALDI-TOF para este fin, el FTIR ha encontrado su nicho en el tipado bacteriano, ya que permite diferenciar entre cepas muy relacionadas entre sí.

Los espectros obtenidos con FTIR se comparan entre sí para observar sus diferencias o similitudes y poder clasificarlos como próximos o distantes, infiriendo así si se trata o no un brote o un conjunto de cepas clonales. Es una tecnología muy precisa y selectiva, que permite diferenciar entre microorganismos estrechamente relacionados, por lo que es el complemento perfecto al MALDI-TOF para el tipado bacteriano.

La espectroscopía FTIR se basa en la absorción y emisión de luz infrarroja por enlaces moleculares específicos presentes en los distintos grupos de moléculas (lípidos, proteínas, polisacáridos y ácidos nucleicos) de la sustancia que se va a medir. En la espectroscopía FTIR la muestra es irradiada con un haz de luz infrarroja y los diferentes compuestos químicos de la célula bacteriana se excitan y vibran emitiendo radiación a diferentes longitudes de onda. Estas ondas vibracionales son sometidas a un proceso matemático conocido como transformada de Fourier, que descompone las señales complejas de los distintos grupos moleculares en todas sus frecuencias y origina un espectro media, en el que se representa cada grupo molecular y por tanto la composición del microorganismo o su “huella dactilar” [147].

INTRODUCCIÓN

En la espectroscopía FTIR se mide la absorbancia y emisión de vibración de las muestras al ser incididas por el haz de luz y se registra un espectro cuyo rango completo va de 4000 cm^{-1} a 400 cm^{-1} [146, 148]. En él se distinguen distintas regiones de bandas vibracionales asociadas a los diferentes componentes de las bacterias (las diferentes regiones se representan en la Figura 8):

- Región 1 (de 3000 a 2800 cm^{-1}): Absorción de lípidos, principalmente ácidos grasos.
- Región 2 (de 1800 a 1500 cm^{-1}): Absorción de péptidos y proteínas, compuesta de dos bandas, la banda de Amida I y la banda de Amida II.
- Región 3 (de 1500 a 1200 cm^{-1}): Región mixta de absorción de proteínas, ácidos grasos y grupos fosfato de los ácidos nucleicos.
- Región 4 (de 1200 a 900 cm^{-1}): Absorción de polisacáridos o hidratos de carbono que componen la pared celular.
- Región 5 o región *fingerprint* (huella dactilar, de 900 a 700 cm^{-1}): Región que presenta patrones específicos de bandas, pero a la que todavía no se le ha asociado un componente celular concreto.

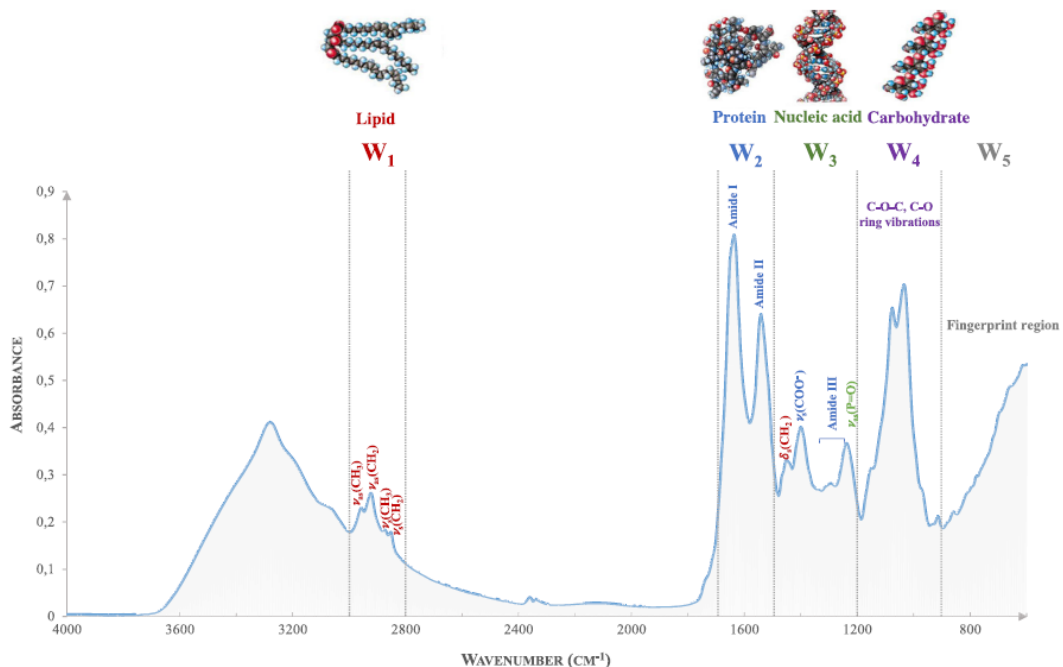


Figura 8. Espectro de infrarrojo donde se observan las distintas regiones vibracionales empleadas para el análisis [146, 148]

Tradicionalmente se asocia un mayor poder discriminatorio a las regiones 3 y 4, por corresponderse con el tipado tradicional de los antígenos polisacáridos O (somático), H (flagelar) y K (capsular) presentes en la pared celular, aunque esto no se cumple en todos los casos y en ocasiones son otras regiones las que presentan mejor poder discriminatorio [148]. En cualquier caso, con el espectroscopio FTIR se puede elegir en qué región de estas bandas se trabaja según los intereses del estudio.

El procedimiento de análisis de muestras con FTIR es sencillo y rápido en comparación con otras técnicas de subtipado. Inicialmente, de un cultivo puro se obtiene una cantidad pequeña de colonias que se resuspende de manera homogénea en un vial conteniendo etanol y agua. Posteriormente esta suspensión se trasfiere a una placa de sílice y se deja secar. Una vez seca, la placa es introducida para su lectura en el instrumento, obteniéndose un espectro.

Los espectros en crudo de cepas relacionadas entre sí son muy similares y no permiten la diferenciación a simple vista, por lo que, para su análisis es posible aplicar la primera o segunda derivada de los espectros, para que las diferencias se hagan más notables. Cada espectro obtenido se compara con el resto, y el resultado de estas comparaciones da un valor de distancia espectral. Cuanto mayor sea la similitud entre los espectros, menor será la distancia espectral (dos espectros idénticos tendrán una distancia espectral de valor 0) y cuanto más diferentes sean los espectros, mayor será su distancia espectral. La visualización de estas diferencias se puede observar mediante dendrogramas y/o matrices de distancia y/o algoritmos matemáticos en 2D o 3D supervisados (LDA) y no supervisados (PCA y HCA).

Al igual que con el tipado con MALDI-TOF, el tipado con FTIR debe realizarse primero generando un modelo o *training set* (conjunto de entrenamiento) con aislados conocidos bien caracterizados molecularmente para luego validarlo, primero con aislados conocidos y después desconocidos. Es preciso que todas las muestras de un mismo estudio se analicen en las mismas condiciones, ya que cualquier variación en el medio de cultivo o tiempo de incubación puede variar los resultados e invalidar el estudio. Esta tecnología, además, ya de manera inicial precisa de la realización de 3 réplicas por cada aislado analizado, para así poder obtener reproducibilidad.

INTRODUCCIÓN

En los últimos años se han llevado a cabo numerosos estudios de tipado con FTIR, así como de detección de brotes por bacterias multirresistentes. La tecnología FTIR se ha aplicado sobre todo en microorganismos relevantes en el campo de la industria alimentaria, para la detección de brotes de especies como *S. enterica* [149, 150], *E. coli* [151, 152], *Bacillus* spp. [153], *Campylobacter* spp. [154], *Yersinia* spp. [155, 156] y *Listeria monocytogenes* [157, 158]. También se han realizado estudios en el ámbito hospitalario [159] para el tipado de *Legionella pneumophila* [160, 161], *K. pneumoniae* [162-164] y *Klebsiella oxytoca* [165], *Enterobacter cloacae* complex [166], *Acinetobacter baumannii* [167, 168], *P. aeruginosa* [169], *S. pneumoniae* [170], *S. aureus* [171-173] y *M. abscessus* [174].

5 BREVE INTRODUCCIÓN A LOS MODELOS AUTOMATIZADOS O MACHINE LEARNING

El *Machine Learning* (ML) o aprendizaje de modelos automatizados es una rama de la inteligencia artificial (IA) centrada en el uso de datos y algoritmos para imitar la forma en que los humanos aprenden, mejorando gradualmente su precisión [175].

Consiste en el aprendizaje mediante el uso de métodos estadísticos por parte de sistemas a los que nosotros previamente administramos información (en forma de datos) para que, posteriormente, tomen decisiones independientes y sin intervención humana. Este aprendizaje se realiza de manera continua y, a medida que se proporciona mayor información, las predicciones automatizadas serán cada vez mejores, es decir, la máquina “aprende” a través de la experiencia y el ensayo-error. Es una herramienta innovadora que se está comenzando a utilizar en todos los campos de la ciencia, con especial interés en el campo de la medicina. Todavía se encuentra en fase experimental y de desarrollo pero plantea un futuro muy prometedor [176].

El *Machine Learning* puede ser “clásico” o “profundo” (lo que se conoce como *Deep Learning*) y ambos se diferencian en la manera de aprender de cada algoritmo. En el primero, hay una necesidad de intervención humana para establecer las

características que diferencien los datos a analizar, es decir, se precisa de información previa al análisis. El *Deep Learning*, sin embargo, basa su funcionamiento en redes neuronales artificiales (ANN, Artificial Neural Networks) y puede procesar datos en bruto sin ningún tipo de información previa. Es capaz automáticamente de determinar cuáles son las características que diferencian las categorías de datos analizados entre sí, evitando así la intervención humana. Las ANN tienen un funcionamiento similar a las neuronas del cerebro humano, en las que un núcleo transmite información al siguiente, y este al siguiente, creando una red de transmisión de datos con valores umbrales de entrada y de salida de datos. Una red neuronal de 3 o más capas ya puede considerarse *Deep Learning* [175].

El ML basa sus decisiones en algoritmos de aprendizaje que pueden ser supervisados o no supervisados.

5.1 Modelos supervisados

El aprendizaje supervisado o *supervised learning* requiere un conocimiento a priori de la muestra. Esta información -datos de entrenamiento- se utiliza como aprendizaje para construir el modelo y generar una función de clasificación que posteriormente se utilizará para predecir conjuntos de datos desconocidos -datos de validación- [146, 177]. Se emplean datos etiquetados para entrenar los algoritmos de clasificación y, a medida que los datos se introducen en el modelo, éste va calculando mediante una función de error la precisión del modelo, para adaptar el peso de cada función y ajustar el modelo correctamente.

Una vez que el modelo está entrenado basándose en la información existente, se pueden introducir nuevos datos desconocidos y obtener una nueva respuesta. Los modelos supervisados por lo general funcionan a base de *feedback*, es decir, una vez se revisan y corrigen los resultados arrojados por el modelo, el sistema los recuerda para la próxima operación. El funcionamiento de un modelo supervisado se resume en la Figura 9.

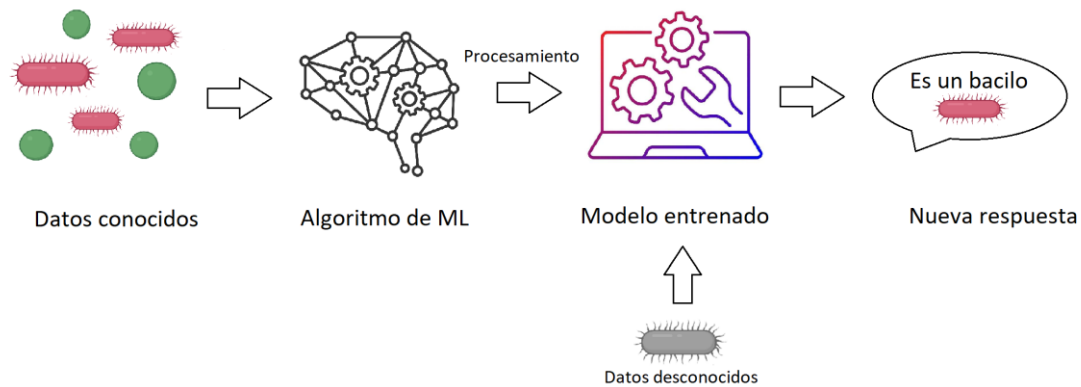


Figura 9. Funcionamiento de un modelo supervisado. Adaptado de Priyadharshini, 2023 [177]

Algunos de los algoritmos empleados en el aprendizaje supervisado son Random Forest (RF), Support Vector Machine (SVM), Light Gradient Boosting Machine (LightGBM), Partial Least Squares Discriminant Analysis (PLS-DA), o K-Nearest Neighbour (KNN) [175].

5.2 Modelos no supervisados

Los modelos no supervisados se utilizan cuando la diferenciación entre categorías se realiza sin una asignación previa a ningún grupo, es decir, sin información sobre los datos a analizar. Se emplean conjuntos de datos sin etiquetar y las relaciones se establecen de acuerdo con su similitud o diferencia empleando técnicas de reconocimiento de patrones para la clasificación en grupos, sin necesidad de la intervención humana [146, 175].

El análisis de componentes principales (PCA) y el agrupamiento jerárquico o dendrogramas (hierarchical clustering) son dos de los enfoques más habituales para realizar este proceso. El funcionamiento de un modelo no supervisado se resume en la Figura 10.

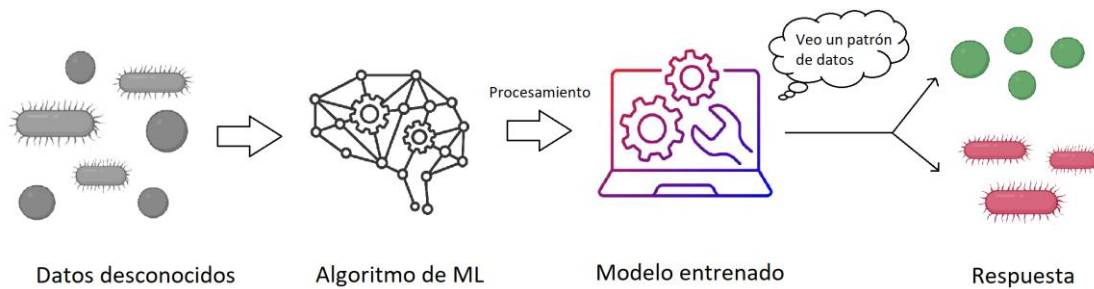


Figura 10. Funcionamiento de un modelo no supervisado. Adaptado de Priyadharshini, 2023 [177]

Por lo general, los modelos supervisados se emplean para predicciones de datos, mientras que los no supervisados se utilizan para encontrar patrones y estructuras ocultas entre los datos que permitan separarlos entre sí por sus similitudes o diferencias.

El proceso de creación de los modelos de *Machine Learning* sigue tres pasos que se conocen como “*Train-Test-Validation*” (Entrenamiento-Prueba-Validación). El primer paso consiste en dividir la muestra empleada en el entrenamiento en dos fracciones, la muestra de entrenamiento o ‘*Training set*’, y la muestra de Comprobación o ‘*Test set*’ [178].

El conjunto de entrenamiento se compone de datos caracterizados que se aportan al modelo para que los analice y aprenda. Es importante que incluya muestras representativas de todas las categorías y que tenga un tamaño suficiente para generar datos estadísticos significativos. Por lo general se acepta como representativo un conjunto de entrenamiento que contenga el 80-70% de los datos totales, y puede estar compuesto de datos etiquetados en los modelos supervisados, y no etiquetados en los modelos no supervisados. La muestra de comprobación o ‘*Test set*’ será el 20-30% restante y se empleará para probar el funcionamiento del modelo [179]. Estos datos no están etiquetados y se comprobarán a posteriori para observar el porcentaje de acierto. Los datos del ‘*Training set*’ y del ‘*Test set*’ deben ser datos independientes y no solaparse para evitar sesgos en el algoritmo [177, 180].

INTRODUCCIÓN

Una vez entrenado y comprobado el modelo con resultados adecuados, se puede comenzar a utilizar para obtener predicciones con nuevos datos independientes -el conjunto de validación o '*Validation set*'-, no empleados en el entrenamiento.

Cuando el conjunto total de datos no es lo suficientemente grande como para dividir los datos en dos conjuntos, uno de entrenamiento (*training + test*) y otro de validación, se divide la muestra en varios conjuntos de datos, y se entrena con algunos de ellas y con otros se valida, numerosas veces. Esto se conoce como validación cruzada o *cross-validation*, una técnica de aprendizaje cruzada donde internamente en el modelo de entrenamiento se separa un porcentaje de los datos y se emplea como muestra de validación.

El *K-fold cross validation* es el método más empleado de validación cruzada, donde la muestra se divide en 'k' conjuntos iguales más pequeños, que se emplearán para realizar la validación. Esta técnica es útil para evitar el *overfitting* o 'sobreajuste' que sucede si sólo se divide la muestra en dos partes y se entrena con una sola. En esta situación, el modelo se aprenderá el '*Training set*' demasiado bien, pero no funcionará correctamente con nuevos datos de validación [180]. El funcionamiento del proceso de validación cruzada se representa en la Figura 11. Otro método comúnmente empleado es el de "*leave one out*" o método LOO, en el que se entrena el modelo cada vez con el conjunto de todos los datos menos uno de ellos, que es el que se emplea para realizar la validación.



Figura 11. Validación cruzada o *Cross-validation* con 5-fold (20%)

Para analizar los resultados de los modelos de *Machine Learning* se emplea lo que se conoce como matriz de confusión, donde se representan en una tabla con las categorías analizadas los datos obtenidos (predicciones) frente a los datos reales. A partir de esta tabla se podrán calcular la exactitud del modelo (*accuracy*), los verdaderos positivos y negativos, los falsos positivos y negativos y el VPN y VPP [180].

5.3 Aplicaciones de Machine Learning en microbiología y espectrometría de masas

La tecnología de ML ya está presente en nuestro día a día, aunque no seamos conscientes de ello. Desde los anuncios personalizados en función del historial de búsqueda en ordenadores y teléfonos móviles, los coches con autonomía de conducción, hasta las recomendaciones de qué ver a continuación en Netflix, forman parte del sistema de aprendizaje de los ordenadores basado en nuestras decisiones previas. Las publicaciones relacionadas con *Machine Learning* se han multiplicado exponencialmente en los últimos años, y esto no ha hecho más que empezar. En la Figura 12 se puede observar el aumento en el número de publicaciones en PubMed buscando el término '*Machine Learning*'.

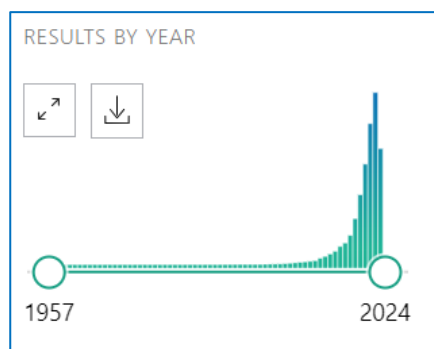


Figura 12. Búsqueda del término '*Machine Learning*' en Pubmed

INTRODUCCIÓN

En medicina también se está empezando a aplicar la inteligencia artificial, sobre todo en el campo de la onco-radiología, como predictores de evolución de la enfermedad. Se han publicado numerosos estudios en los últimos años en los que algoritmos de *Machine Learning* son capaces de analizar imágenes radiológicas y distinguir entre pacientes sanos y enfermos, e incluso detectar un tumor antes que el propio médico [181-184]. También se han desarrollado modelos predictivos, alimentados con datos extraídos de pruebas médicas, que son capaces de clasificar a los pacientes según el riesgo de evolución de la enfermedad [185, 186]. A medida que los modelos desarrollados se vayan mejorando con la experiencia acumulada, también mejorarán sus predicciones y sus aplicaciones aumentarán. El *Machine Learning* ya está aquí y cambiará la forma en la que se practica la medicina [176].

En el mundo microbiológico es muy probable que en los próximos años seamos testigos de una revolución diagnóstica, donde el uso de la inteligencia artificial sea cada vez más común para el diagnóstico microbiológico y de las enfermedades infecciosas. Actualmente ya estamos viviendo la automatización del laboratorio de microbiología mediante la instalación de sembradores automatizados y posterior lectura y procesamiento a través de un software. El siguiente paso será la lectura automatizada sin necesidad de personal, mediante el uso de la inteligencia artificial supervisada por humanos.

Ya se han descrito en artículos numerosas aplicaciones de la inteligencia artificial en el campo de la microbiología, principalmente basadas en distinción de patrones en imágenes. Algunos ejemplos son: Una aplicación de móvil que permite la lectura automática de antibiogramas de disco-placa para reducir la variabilidad interoperador [187], un lector automatizado de gota fina y gruesa para el diagnóstico de malaria [188], la detección automatizada de parásitos en heces mediante el análisis de los resultados de microscopía con inteligencia artificial [189] o la interpretación automatizada de la tinción de Gram [190].

En la espectrometría de masas MALDI-TOF concretamente, también es posible la aplicación de algoritmos de *Machine Learning* para la diferenciación entre espectros. Se han desarrollado numerosos software en los últimos años para el procesamiento de

espectros de masas (por ejemplo FlexAnalysis, ClinProTools, MassUp, CloverBiosoft o BioNumerics) que permiten realizar estudios con dichos espectros aplicando algoritmos de ML supervisados y no supervisados. Estos software incluyen también una etapa de preprocesado previo que suele comprender un suavizado de los espectros y una substracción de la línea base, así como una normalización y alineamiento previo con una determinada tolerancia en Da, para que los picos de interés coincidan en todos los espectros [191].

Numerosos estudios aplican ya estas tecnologías para el análisis de espectros de MALDI-TOF, para intentar diferenciar de manera rápida entre cepas de interés clínico que por otros métodos comúnmente aplicados o no es posible, o es costoso, o un proceso largo en el tiempo, [96, 192-195] e incluso para intentar predecir patrones de sensibilidad o resistencia analizando directamente los espectros obtenidos para la identificación de microorganismos [94, 95, 196]. En la misma tendencia, también se ha empleado la IA con la espectrometría de masas MALDI-TOF para el estudio de nuevos compuestos antimicrobianos, observando los cambios que se producen en el espectro de masas de las bacterias que se enfrentan a ellos [197].

Esta tesis se centra en 4 de estos microorganismos de interés: La diferenciación entre *Enterococcus faecium* VSE y VRE, la correcta identificación de las distintas especies del complejo de *Enterobacter cloacae*, la detección de un brote del clon de alto riesgo ST175 de *Pseudomonas aeruginosa* y un anexo de ribotipado de *Clostridioides difficile* mediante MALDI-TOF.

JUSTIFICACIÓN Y OBJETIVOS

JUSTIFICACIÓN

En Febrero del año 2017, la OMS emitió un informe en el que se detallaban los patógenos de prioridad internacional que suponían un mayor riesgo para la salud pública, por su incremento en el desarrollo de resistencias a los antibióticos y la urgente necesidad de investigación en nuevos antibióticos para tratarlos [198]. En esta lista, se establecieron 3 categorías según el grado de prioridad:

- Prioridad 1, o CRÍTICA: *A. baumannii* resistente a carbapenémicos (CRAB), *P. aeruginosa* resistente a carbapenémicos y *Enterobacteriaceae* resistentes a carbapenémicos y/o productoras de BLEE.
- Prioridad 2, o ELEVADA: *E. faecium* resistente a vancomicina, *S. aureus* con resistencia a la meticilina (SARM), *S. aureus* con resistencia o sensibilidad intermedia a la vancomicina (VRSA, VISA), *Helicobacter pylori* resistente a claritromicina, *Campylobacter* spp. y *Salmonella* spp. resistentes a fluoroquinolonas, *Neisseria gonorrhoeae* resistente a cefalosporinas de 3ª generación y fluoroquinolonas.
- Prioridad 3, o MEDIA: *S. pneumoniae* resistente a penicilina, *H. influenzae* resistente a ampicilina, *Shigella* spp. resistente a fluoroquinolonas.

El principal objetivo de este listado era el de promover la investigación y el desarrollo de nuevos antibióticos frente a estos patógenos, y concienciar a la población frente al problema emergente e incremental del desarrollo de resistencias a los antimicrobianos. Este enfoque se enmarcó dentro del ámbito del “*One Health*” o “Una sola salud”, destacando la necesidad de cooperación entre los sistemas de vigilancia del ámbito de salud humana, animal, alimentaria y medioambiental, ya que la multirresistencia es un problema global que se debe abordar desde todos estos ámbitos en conjunto.

Este listado de la OMS, concuerda con la lista de “Amenazas de resistencia a los antibióticos en Estados Unidos” publicada por el CDC americano, en el que destacan como “amenaza urgente” a las *Enterobacteriaceae* resistentes a los carbapenémicos y a *C. difficile* entre otros; y como “amenaza seria” los *Enterococcus* spp. resistentes a vancomicina (resistencia no intrínseca), *P. aeruginosa* MDR y *Enterobacteriaceae*

JUSTIFICACIÓN

productoras de BLEE junto con otros microorganismos. Estos listados de microorganismos relevantes y prioritarios para la OMS y el CDC se resumen en la Figura 13 [198, 199].

El término “ESKAPE” fue propuesto por primera vez en el año 2008 por Louis B. Rice [200], y en él ya incluía a los principales patógenos de prioridad crítica y elevada propuestos más adelante por la OMS. ESKAPE son las siglas que representan a los microorganismos *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* y *Enterobacter cloacae*. Tres de estos patógenos han sido objeto de estudio en esta tesis (*E. faecium*, *P. aeruginosa* y *E. cloacae*). El cuarto microorganismo que se ha estudiado, *C. difficile*, es un microorganismo que produce diarrea inhabilitante y el principal causante de infección relacionada con la asistencia sanitaria.

Recientemente, se ha propuesto en la literatura el cambio del término ESKAPE por ESCAPE, sustituyendo la letra K de *K. pneumoniae* por la C de *C. difficile*, y la última E, de *E. cloacae*, por el término más genérico *Enterobacteriaceae* que puedan expresar niveles crecientes de resistencia a los antibióticos [201-204].

Tras la instauración del MALDI-TOF en los laboratorios de microbiología para la identificación bacteriana, numerosos estudios han investigado sus posibles usos alternativos, intentando emplearlo para otros fines más allá de la identificación, como la detección de brotes, la identificación de microorganismos a partir de muestra directa, la determinación rápida de perfiles de resistencia, el tipado, o la identificación bacteriana intraespecie para la distinción de especies dentro de un mismo complejo [33].

MALDI-TOF se trata de una herramienta versátil, ya disponible en muchos laboratorios clínicos y con precios asequibles para cada medición, por lo que su optimización y aprovechamiento para usos alternativos a la identificación bacteriana puede suponer una mejora en el manejo de los pacientes y el control de las infecciones.

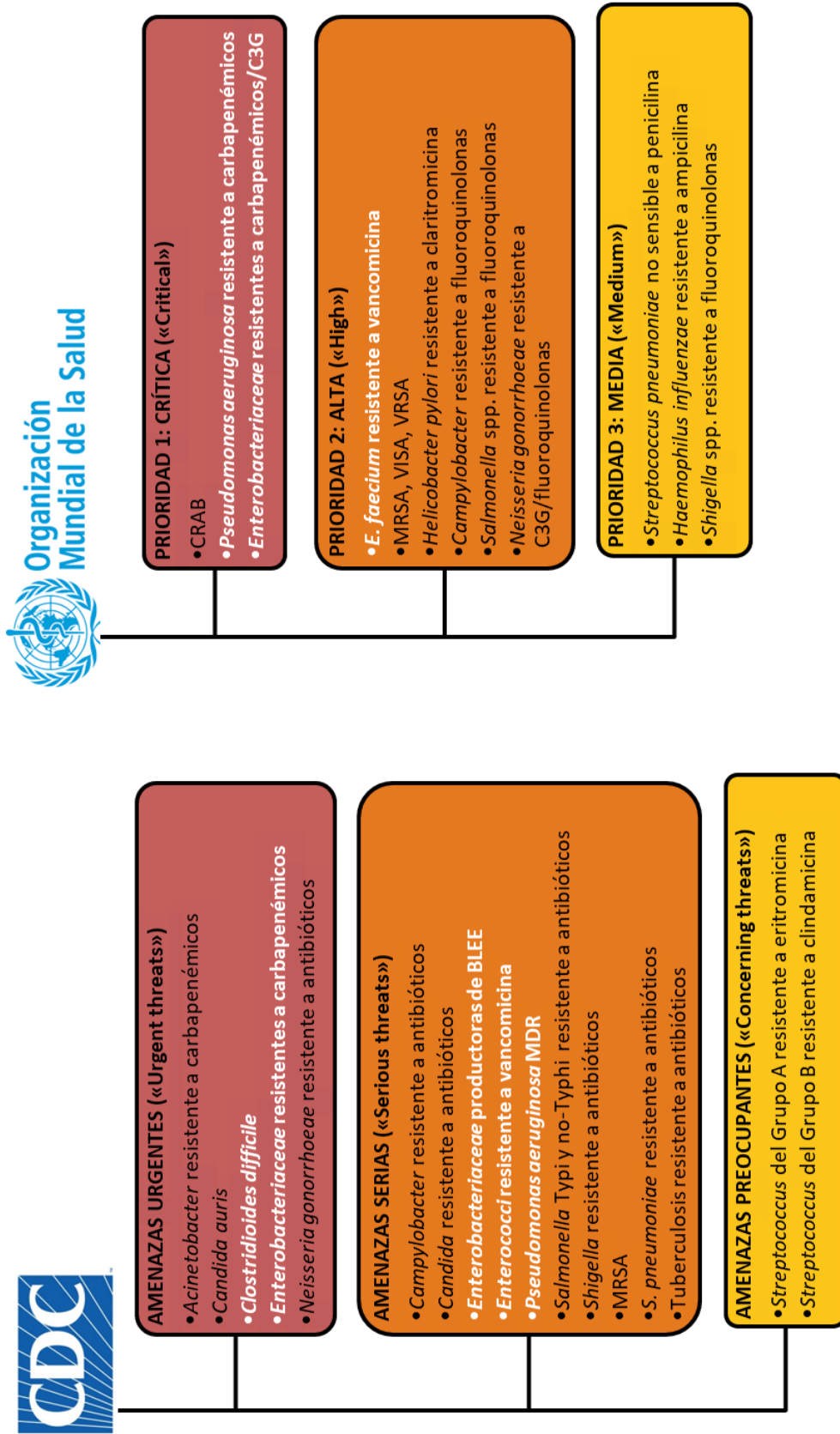


Figura 13. Especies de relevancia destacadas por la OMS y el CDC en 2017 y 2019, respectivamente

OBJETIVO PRINCIPAL

Por todo lo mencionado anteriormente, el objetivo principal de esta tesis ha sido el estudio de la aplicación clínica de la tecnología MALDI-TOF (en ocasiones combinada con FTIR) enfocada a estos cuatro microorganismos previamente mencionados, para mejorar su diagnóstico clínico y epidemiológico, y así poder vigilar de mejor manera su diseminación hospitalaria.

Esta tesis se centra en el asentamiento de unas bases generales de análisis de microorganismos mediante MALDI-TOF y el desarrollo de una metodología reproducible de obtención y procesamiento de espectros de masas, que permita una correcta diferenciación y clasificación de los microorganismos estudiados.

OBJETIVOS ESPECÍFICOS

- i. Desarrollar una metodología para el análisis de espectros de masas de MALDI-TOF, estudiando su reproducibilidad para la aplicación de algoritmos de *Machine Learning*.
- ii. Diferenciación entre aislados de *E. faecium* sensibles y resistentes a Vancomicina, portadores del gen *vanA* o *vanB*, empleando MALDI-TOF.
- iii. Identificación precisa y separación de las especies intracomplejo de *E. cloacae* complex combinando MALDI-TOF y algoritmos de *Machine Learning*.
- iv. Discriminación del clon de alto riesgo ST175 de *Pseudomonas aeruginosa*, el más prevalente en España, causante de un brote hospitalario en el Hospital General Universitario Gregorio Marañón, mediante MALDI-TOF y FTIR.
- v. Tipado de aislados clínicos de *Clostridioides difficile* mediante la tecnología MALDI-TOF, en comparación con la técnica de ribotipado clásico.

TRABAJOS CIENTÍFICOS

Capítulo 1. Método rápido y reproducible basado en MALDI-TOF para la detección de *Enterococcus faecium* resistentes a vancomicina empleando algoritmos de clasificación

Contexto

Enterococcus faecium es un microorganismo colonizador del tracto digestivo que se puede convertir en patógeno y causar infecciones graves. Especialmente, los *E. faecium* que presentan resistencia a la vancomicina representan una amenaza para la salud debido a su dificultad de tratamiento, ya que se trata de un antibiótico de primera línea frente a las infecciones causadas por este microorganismo. MALDI-TOF ha demostrado su utilidad para la identificación de *E. faecium*, pero su implementación para la detección de resistencias a vancomicina en este microorganismo aún se encuentra en evaluación.

Material y métodos

En este estudio se evalúa la reproducibilidad de MALDI-TOF para el análisis de picos de *E. faecium*, y su capacidad en la discriminación entre aislados sensibles (VSE) y resistentes a la vancomicina (VRE). El estudio se llevó a cabo con espectros de proteínas de 178 aislados clínicos únicos de *E. faecium* (92 VSE, 31 VanA VRE, 55 VanB VRE), procesados con el software Clover Mass Spectrometry Data Analysis Software (Clover MSDAS).

Se ensayó la reproducibilidad técnica y biológica del sistema, y se aplicaron algoritmos no supervisados (PCA) y supervisados (SVM, RF y PLS-DA) para la diferenciación entre las 3 categorías estudiadas. El ensayo de reproducibilidad se realizó con 18 picos comunes a todos los VSE y VRE, con intensidades superiores al 1,0% de la intensidad del pico de máxima intensidad.

Resultados

El estudio de reproducibilidad mostró una menor variabilidad con los datos normalizados y para los picos que se encontraban dentro del rango de 3000-9000 m/z . Se consiguió un 80,9%, 79,2% y 77,5% de discriminación entre VSE vs. VRE aplicando SVM, RF y PLS-DA, respectivamente. La mejor diferenciación interna entre los VRE VanA y los VRE VanB se obtuvo mediante PLS-DA, con un resultado correcto en el 86,6% de los casos.

Conclusiones

La implementación de MALDI-TOF y la técnica de análisis de picos puede representar una herramienta rápida y eficaz para la detección de VRE. Sin embargo, se necesitan más mejoras para aumentar la precisión de este enfoque.

Article

Rapid and Reproducible MALDI-TOF-Based Method for the Detection of Vancomycin-Resistant *Enterococcus faecium* Using Classifying Algorithms

Ana Candela ^{1,2,*}, Manuel J. Arroyo ^{3,†}, Ángela Sánchez-Molleda ³, Gema Méndez ³, Lidia Quiroga ^{1,2}, Adrián Ruiz ^{1,2}, Emilia Cercenado ^{1,2,4,5}, Mercedes Marín ^{1,2,4}, Patricia Muñoz ^{1,2,4,5}, Luis Mancera ³, David Rodríguez-Temporal ^{1,2,*} and Belén Rodríguez-Sánchez ^{1,2}

¹ Clinical Microbiology and Infectious Diseases Department, Hospital General Universitario Gregorio Marañón, 28007 Madrid, Spain; lidia.quirogam@gmail.com (L.Q.); arauko.bmcelta@gmail.com (A.R.); emilia.cercenado@salud.madrid.org (E.C.); mercedes.marin@salud.madrid.org (M.M.); patricia.munoz.garcia@salud.madrid.org (P.M.); mbelen.rodriguez@iisgm.com (B.R.-S.)

² Instituto de Investigación Sanitaria Gregorio Marañón, 28007 Madrid, Spain

³ Clover Bioanalytical Software, Av. del Conocimiento, 41, 18016 Granada, Spain; manuel.arroyo@cloverbiosoft.com (M.J.A.); angelamolleda97@gmail.com (Á.S.-M.); gema.mendez@cloverbiosoft.com (G.M.); luis.mancera@cloverbiosoft.com (L.M.)

⁴ CIBER de Enfermedades Respiratorias (CIBERES CB06/06/0058), 28029 Madrid, Spain

⁵ Medicine Department, School of Medicine, Universidad Complutense de Madrid, 28040 Madrid, Spain

* Correspondence: acandelagon@gmail.com (A.C.); david.rodriguez@iisgm.com (D.R.-T.); Tel.: +34-91-426-9595 (D.R.-T.)

† These authors contributed equally to this work.

Citation: Candela, A.; Arroyo, M.J.; Sánchez-Molleda, Á.; Méndez, G.; Quiroga, L.; Ruiz, A.; Cercenado, E.; Marín, M.; Muñoz, P.; Mancera, L.; et al. Rapid and Reproducible MALDI-TOF-Based Method for the Detection of Vancomycin-Resistant *Enterococcus faecium* Using Classifying Algorithms. *Diagnostics* **2022**, *12*, 328. <https://doi.org/10.3390/diagnostics12020328>

Academic Editors: Andreas Kjaer and Raul Colodner

Received: 29 December 2021

Accepted: 24 January 2022

Published: 27 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Vancomycin-resistant *Enterococcus faecium* represents a health threat due to its ability to spread and cause outbreaks. MALDI-TOF MS has demonstrated its usefulness for *E. faecium* identification, but its implementation for antimicrobial resistance detection is still under evaluation. This study assesses the repeatability of MALDI-TOF MS for peak analysis and its performance in the discrimination of vancomycin-susceptible (VSE) from vancomycin-resistant isolates (VRE). The study was carried out on protein spectra from 178 *E. faecium* unique clinical isolates—92 VSE, 31 VanA VRE, 55 VanB VRE—, processed with Clover MS Data Analysis software. Technical and biological repeatability were assayed. Unsupervised (principal component analysis, (PCA)) and supervised algorithms (support vector machine (SVM), random forest (RF) and partial least squares–discriminant analysis (PLS-DA)) were applied. The repeatability assay was performed with 18 peaks common to VSE and VRE with intensities above 1.0% of the maximum peak intensity. It showed lower variability for normalized data and for the peaks within the 3000–9000 *m/z* range. It was found that 80.9%, 79.2% and 77.5% VSE vs. VRE discrimination was achieved by applying SVM, RF and PLS-DA, respectively. Correct internal differentiation of VanA from VanB VRE isolates was obtained by SVM in 86.6% cases. The implementation of MALDI-TOF MS and peak analysis could represent a rapid and effective tool for VRE screening. However, further improvements are needed to increase the accuracy of this approach.

Keywords: enterococci; vancomycin resistance; MALDI-TOF; mass spectrometry; peak analysis

1. Introduction

Enterococcus faecium are Gram-positive, non-spore forming, facultative anaerobic cocci that can be found as part of the microbiota of the human gastrointestinal tract [1]. Because of its genomic plasticity and its adaptation to harsh conditions, *E. faecium* has become a health threat due to its ability to rapidly spread and cause outbreaks in hospital settings [2,3]. Adding up to the intrinsic antibiotic resistance displayed by *E. faecium*,

infections caused by strains with acquired resistance to certain antimicrobials are of special interest [4]. In the past years, vancomycin resistance in *E. faecium* has been a major concern owing to its rise and rapid spread of high-risk clones among hospitalized patients [5]. Therefore, *E. faecium* was included in the ESKAPE group of highly resistant microorganisms because of its ability to “escape” the action of conventional treatments [6]. Vancomycin resistance is by far the biggest threat regarding enterococci, mainly *E. faecium*, due to its position as first-line treatment for multidrug-resistant enterococcal infections [7,8].

Susceptibility to vancomycin can be routinely determined in the clinical microbiology laboratory using two approaches: i) phenotypically, by the standard broth antimicrobial susceptibility testing microdilution method or by gradient diffusion, or ii) genotypically, by amplification of the *vanA/vanB* genes and subsequent analysis of the specific amplicons [9,10]. The first approach has a turnaround time of approximately 2 days. Although the implementation of molecular methods provides final results in 1–3 hours after isolation in culture, its cost in laboratory reagents is high.

While the usefulness of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) in the microbiology laboratory for bacterial identification is settled, its implementation for antimicrobial resistance detection is not well standardized yet [11].

In the last years, MALDI-TOF MS has been proposed as a fast and cost-efficient method for the detection of some antimicrobial resistance mechanisms, such as β -lactamase activity [12], discrimination of methicillin-resistant *S. aureus* [13] or the detection of the *cfiA* gene in *B. fragilis* [14,15]. As for vancomycin resistance in *Enterococcus*, a few studies have been published recently but with variable results [16–18].

The use of MALDI-TOF MS in the clinical microbiology laboratory will be of interest as a rapid approach for the differentiation of VRE from VSE, based on their mass spectra protein profile.

The main objective of this study was the development of a MALDI-TOF-based classifying algorithm for the discrimination of vancomycin-resistant *E. faecium* (VRE) from vancomycin-susceptible *E. faecium* (VSE).

2. Materials and Methods

2.1. Bacterial Strains

A total of $n = 178$ *E. faecium* strains were included in the study and considered as the “classification set” (92 VSE, 31 VanA VRE, 55 VanB VRE). The isolates were collected consecutively throughout the years 2017 to 2019 from clinical samples of patients admitted at Hospital General Universitario Gregorio Marañón in Madrid (HGUGM) (Figure 1). Only one sample per patient was selected.

Strains were selected from blood cultures ($n = 95$) and rectal swabs ($n = 83$) (Table S1). Clonality was clinically and epidemiologically discarded by analyzing the date, patient location and department where the inpatients were located. No clinical outbreak was detected during the period of study. All strains were isolated from inpatient clinical samples, characterized and kept frozen at -80 °C for further analysis. For this study, samples were thawed and cultured overnight at 37 °C in Columbia blood agar.

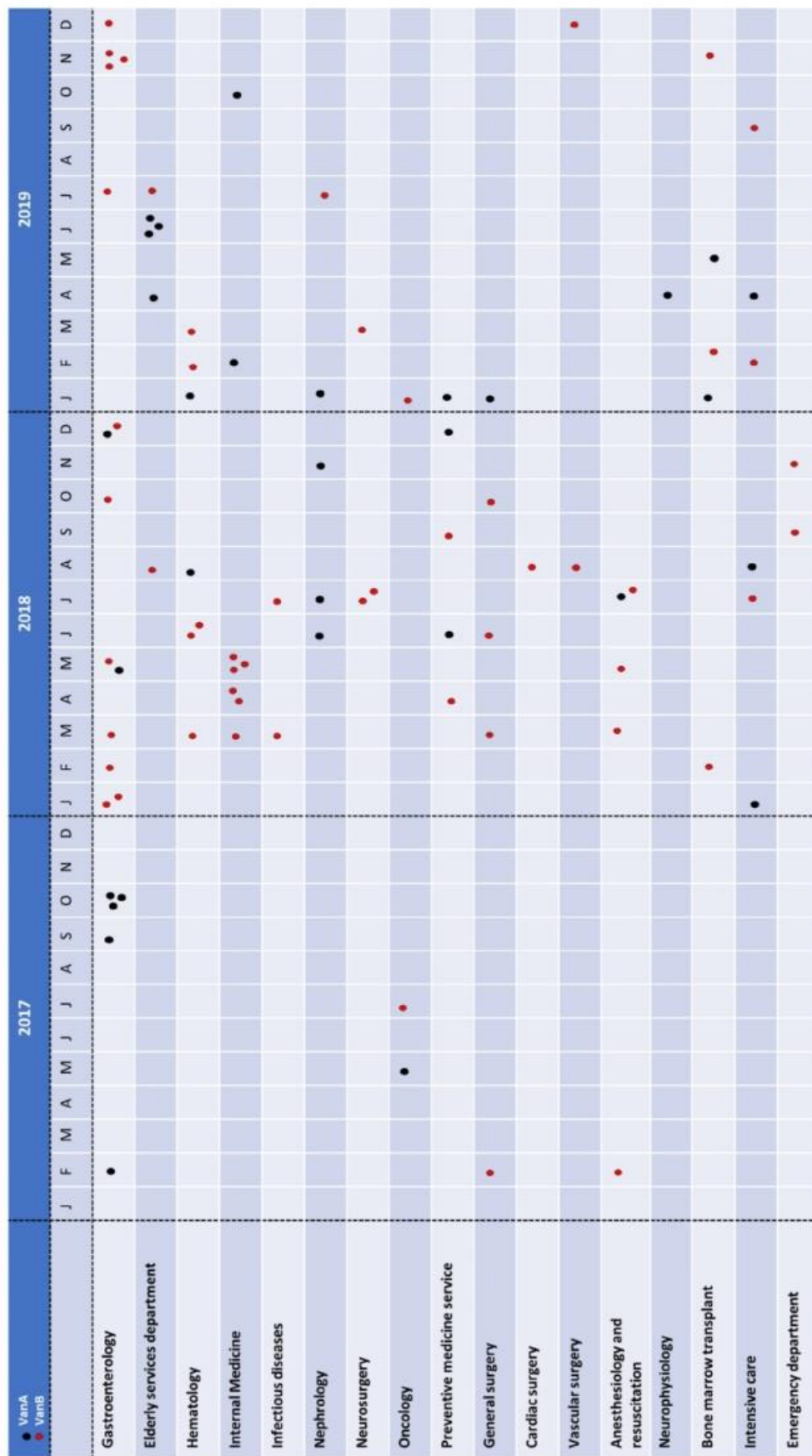


Figure 1. Distribution of the *E. faecium* isolates analyzed in this study by hospital departments and date. VanA *E. faecium* isolates are shown with black dots and VanB isolates with red dots.

2.2. Antibiotic Susceptibility Testing for Vancomycin

Antimicrobial susceptibility testing was performed with the automated microdilution method Microscan® System (Beckman-Coulter, Brea, CA, USA) using PM33 panels following the manufacturer's guidelines. Vancomycin and teicoplanin breakpoints were established as indicated by the EUCAST (2021) v. 11. The results obtained were confirmed by real-time PCR for the amplification of the *vanA* and *vanB* genes [9]. In addition, the presence of the vancomycin resistance genes was confirmed a second time by the implementation of the commercial Xpert® vanA/vanB cartridges (Cepheid, Sunnyvale, CA, USA).

2.3. Identification of the Isolates by MALDI-TOF MS

Bacterial strains were analyzed by MALDI-TOF MS in an MBT Smart MALDI Biotyper (Bruker Daltonics, Bremen, Germany) using the updated database containing 9957 mass spectra profiles (MSPs). A few bacterial colonies of each enterococcal isolate were spotted onto the MALDI target plate. On-plate protein extraction was performed by applying 1 µL of formic acid on each target spot and letting it dry at room temperature before adding 1 µL of HCCA matrix solution (Bruker Daltonics, Bremen, Germany), following the manufacturer's instructions. Spectra were acquired in positive mode in the range of 2000 to 20,000 Da, applying default settings [19].

2.4. Spectra Acquisition and Pre-Processing

Each isolate was analyzed in two different spots from the MALDI target plate, and each spot was read twice, obtaining four spectra per strain [20]. Protein spectra were visually inspected with FlexAnalysis (Bruker Daltonics) and aligned with the genus-specific peak at 4428 *m/z*, present in all isolates [16,21]. Outlier spectra and zero lines were discarded.

After the previous screening, protein spectra were processed with Clover MS Data Analysis software (<https://platform.clovermsdataanalysis.com>, accessed 30 November 2021, Clover Biosoft, Granada, Spain). For classification purposes, peak matrices were generated in the range of 2000 to 20,000 *m/z*. For this goal, pre-processing was performed as follows: a Savitzky–Golay filter (window length 11; polynomial order 3) was applied for smoothing spectra, and then the baseline was removed by the top-hat filter method (factor 0.02).

Processed spectra were aligned using the following method: replicates within the same spot were aligned to create an average spectrum per spot (shift medium; linear mass tolerance 2000 ppm). Then, average spectra from each replicated spot were aligned, and thus one average spectrum per sample was obtained. Finally, average spectra from different isolates were aligned together.

2.5. Repeatability Test

A repeatability study was performed to determine the robustness of MALDI-TOF-MS-based bacterial classification. For this purpose, 20 different *E. faecium* isolates (VSE, *n* = 10; and VRE strains, *n* = 10, 5 VanA VRE and 5 VanB VRE) were randomly selected and considered as the “repeatability set” (Table S2). For intra-spot repeatability, two spectra per spot were acquired, whilst for inter-spot repeatability (or technical repeatability), an average spectrum was built from each spot. Finally, each isolate was subcultured during three consecutive days, and their average spectra were compared in order to evaluate the biological (inter-day) repeatability of the method (Figure 2). This methodology has been described by Oberle et al., 2016 [20]. A final number of 12 spectra per isolate was obtained.

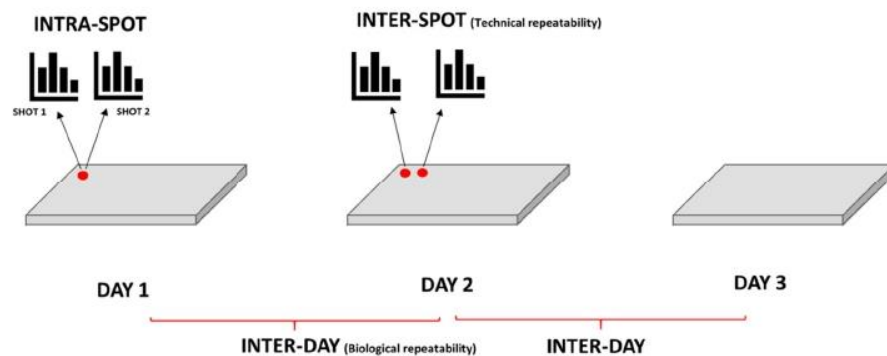


Figure 2. Graphic representation of 12 replicates spectra for each isolate: 2 spectra are acquired per spot; 2 spots analyzed each day and compared for 3 consecutive days.

All spectra were pre-processed using Clover MS Data Analysis Software following the pipeline described in Section 2.4. The post-processed spectra (smoothed and baseline removed) were aligned (shift medium; linear mass tolerance 2000 ppm) to obtain an average spectrum for each spot. This process was repeated to obtain a single average spectrum per day for each isolate.

Once the average Day 1 spectra from all isolates were aligned, a first assay was performed to identify all common peaks and establish their characteristic peak profiles. Peak finding was carried out by applying a threshold filter (0.01), so peaks with less than 1.0% of the maximal intensity recorded were discarded. Group-specific peaks were then searched in the repeatability set by the mass position method (constant mass tolerance 0.2 Da; linear mass tolerance 500 ppm) at the shot, spot and day levels.

The coefficient of variation (% CV) of the intensities registered for each of these common peaks was calculated from raw spectra and from spectra normalized with the TIC method in two different ways: i) normalizing by TIC the peaks previously found in raw spectra (pTIC) and ii) normalizing first the entire spectra by TIC and then finding the common peaks (TICp). Results from both methods and from raw data were compared for intra-spot, technical and biological repeatability.

In addition, Pearson correlation coefficient (p) was applied to protein spectra from VRE and VSE isolates to measure how biological variation could affect the discrimination of these two groups and how reproducibly the discrimination could be performed.

This peak study also included the calculation of arithmetic and post-alignment means and the subsequent comparison between them. The arithmetic means for spots and days were calculated directly from the intensity peak values of each shot. On the other hand, the post-alignment means were calculated automatically with Clover MS Data Analysis software after replicated average spectra were aligned. Both mean values were compared at spot and day levels. The assay was repeated without normalization and for both TIC methods. Thereby, the automatic alignment and replicate software process could be compared with the ideal arithmetic model in the three normalization cases. Once the Shapiro–Wilk and Levene test was applied to verify the normal distribution and the homoscedasticity of the data, a Student’s t -test was performed to verify whether the null hypothesis of equality of means was fulfilled for the three normalization methods.

2.6. Classification of *E. faecium* Isolates Based on Their MALDI-TOF MS Protein Spectra

Protein spectra from *E. faecium* isolates were acquired as described in Section 2.4. Clover MS Data Analysis Software was applied to differentiate (1) susceptible from resistant *E. faecium* strains and (2) isolates hosting *vanA* and *vanB* resistance genes. For this purpose, three different methods were evaluated: (i) a “full-spectrum method”, in which the matrix obtained included all peak intensities from the spectrum separated by 0.5 Da

regardless of their intensity. Their intensities were then normalized by total ion current (TIC) normalization. The other two methods used a peak matrix generated by a “threshold method”, in which only peaks with intensities above 1.0% of the maximum peak intensity (0.01 factor) were chosen. The difference in the last two methods was the order in which TIC normalization was applied: (ii) before (TICp) or (iii) after (pTIC) searching peaks by the threshold method.

The three peak matrices described above were used as input data for three different supervised machine learning algorithms: partial least squares–discriminant analysis (PLS-DA), support vector machine (SVM) and random forest (RF). These algorithms were first tested for the discrimination VSE and VRE isolates and, secondly, for the differentiation of VRE isolates hosting *vanA* and *vanB* genes.

Internal validation of the results provided by each algorithm was assayed using *k*-fold cross validation ($k = 10$) as previously described [22]. Briefly, data were randomly split into 10 data subsets of the same size. The algorithms were trained with nine of them and the remaining subset was used as a test set for internal validation. This process was iterated 10 times (once for each of the 10 subsets) and the accuracy rate of the classification was recorded [23].

Furthermore, a search for potential biomarkers was performed to use group-specific protein peaks as markers for the correct classification of VanA and VanB VRE and VSE strains. For this purpose, 178 average spectra (one per sample) from Day 1 were considered as input data for the Biomarker Analysis application within Clover MS Data Analysis software. The threshold method was applied as explained above, and peaks were merged with 0.5 Da and 300 ppm as constant and linear mass tolerance, respectively.

3. Results

3.1. Identification by MALDI-TOF MS

All isolates ($n = 178$) included in this study were correctly identified as *E. faecium* by MALDI-TOF MS with score ≥ 2.0 . Identifications for this microorganism were consistent along the top 10 identifications provided by MALDI-TOF MS since this is a common pathogen well represented in the updated commercial library.

3.2. Repeatability Study

When the different methods for peak finding described before were applied to the Day 1 spectra from the “classification set”, a total of 18 common peaks were found in all *E. faecium* (VSE and VRE) protein spectra with intensities above 1.0% of the maximum peak intensity (Table 1). The CV means of the intensities from these 18 common peaks were compared with the peaks present in the spectra from the repeatability set at intra-spot, inter-spot and inter-day levels.

Comparing the means of the CVs in the repeatability set showed lower CV values at the intra-spot level than at the inter-spot and inter-day levels. For non-normalized raw data, the average CVs were $CV_{\text{intra-spot}} = 15.35$, $CV_{\text{inter-spot}} = 29.29$ and $CV_{\text{inter-day}} = 31.25$. The same pattern was shown for the TICp ($CV_{\text{intra-spot}} = 8.46$, $CV_{\text{inter-spot}} = 20.88$ and $CV_{\text{inter-day}} = 20.66$) and pTIC ($CV_{\text{intra-spot}} = 7.99$, $CV_{\text{inter-spot}} = 19.91$ and $CV_{\text{inter-day}} = 19.30$) methods (Figure 3). Additionally, data normalization allowed the reduction of CV values at the three levels ($p < 0.0001$). Differences between both normalization methods (TICp and pTIC) varied between 0.47 for $CV_{\text{intra-spot}}$ and 1.36 for $CV_{\text{inter-day}}$, demonstrating that both allowed for reduced CV values at intra-spot (44.80% and 47.94% for TICp and pTIC, respectively), inter-spot (28.71% and 32.0%) and inter-day (33.89% and 38.24%) levels. Reduced CV values were consistently recorded for VSE and VRE isolates alike (Table S3).

Table 1. List of common peaks ($n = 18$) found in the average spectra of the vancomycin-resistant *E. faecium* and vancomycin-susceptible *E. faecium* isolates included in the classification set. CV = coefficient of variation of intensity. Mean in intensity units.

Mass (m/z)	Measurements		
	Appearance (%)	CV (%)	Mean
2059.79	178/178 (100)	53.42	4361.20
2188.53	178/178 (100)	56.61	5269.55
2408.35	178/178 (100)	49.04	3316.45
2537.05	178/178 (100)	53.06	3527.45
2676.89	178/178 (100)	51.89	2170.47
3026.71	178/178 (100)	47.65	1270.16
3517.85	178/178 (100)	56.14	1327.19
3673.77	178/178 (100)	50.43	3344.71
4430.01	178/178 (100)	46.26	17,437.72
4532.78	178/178 (100)	57.94	2296.61
5356.72	178/178 (100)	56.15	8719.83
5974.65	178/178 (100)	68.28	2531.38
6055.07	178/178 (100)	50.30	4187.74
6830.35	178/178 (100)	71.56	3020.35
6891.33	178/178 (100)	59.00	5252.80
7035.47	178/178 (100)	65.96	2656.59
7347.39	178/178 (100)	55.31	10,326.29
9062.75	178/178 (100)	66.94	2820.45

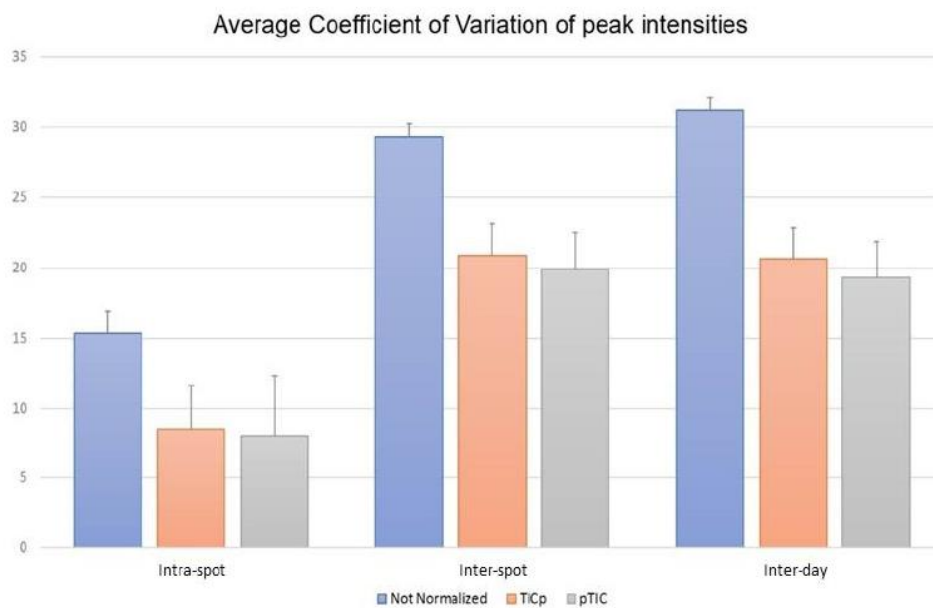


Figure 3. Coefficient of variation (CV) means for the intensity of the 18 common protein peaks of *E. faecium* analyzed at the intra-spot, inter-spot and inter-day levels using raw data (non-normalized) and both normalizations methods—before (TICp) and after (pTIC) finding peaks.

The variability of the 18 common peaks found in all *E. faecium* isolates showed that CV values were lower for peaks between 3000 and 9000 m/z at intra-spot (Figure 4A), inter-spot (Figure 4B) and inter-day levels (Figure 4C), especially when data were normalized, showing different means ($p < 0.05$) versus the 2000 and 3000 m/z range (Table S4). The 5974.6 m/z peak showed higher CV values at the three levels but the region between 6000

and 9000 m/z showed lower variation again, although the CVs were higher for inter-spot and inter-day repeatability (Figures 4B, C). These results support the fact that the central m/z region of the spectrum is the most reliable for peak analysis [24].

The Pearson correlation coefficient was applied to inter-day repeatability, showing a mean of 0.94 factor for all samples (Table S2). This value was higher for VRE ($p = 0.95$) than for VSE isolates (0.93). Within the VRE group, the p value for isolates hosting the *vanA* mechanism was 0.98 versus 0.92 for the isolates with the *vanB* resistance gene.

Arithmetic mean, as a representative value of peak intensity, and post-alignment mean were calculated and compared (Table 2). For non-normalized data, the arithmetic mean was higher than the post-alignment mean for inter-spot and inter-day levels, unlike the case for the TICp normalization method. Regarding pTIC normalization, the arithmetic and post-alignment means were identical. Furthermore, the homoscedasticity of the data and their normal distribution were checked by the implementation of the Levene and Shapiro–Wilk tests respectively. The t-Student test was then performed to check whether the null hypothesis of equality of means could be accepted among arithmetic and post-alignment means. In all possible comparisons, p -values obtained were >0.05 (Table S5), showing that, even when a post-alignment was performed, the intensity values were not affected. The fact that both means in all methods did not show statistical differences, proved the high repeatability of the assays based on protein spectra analysis.

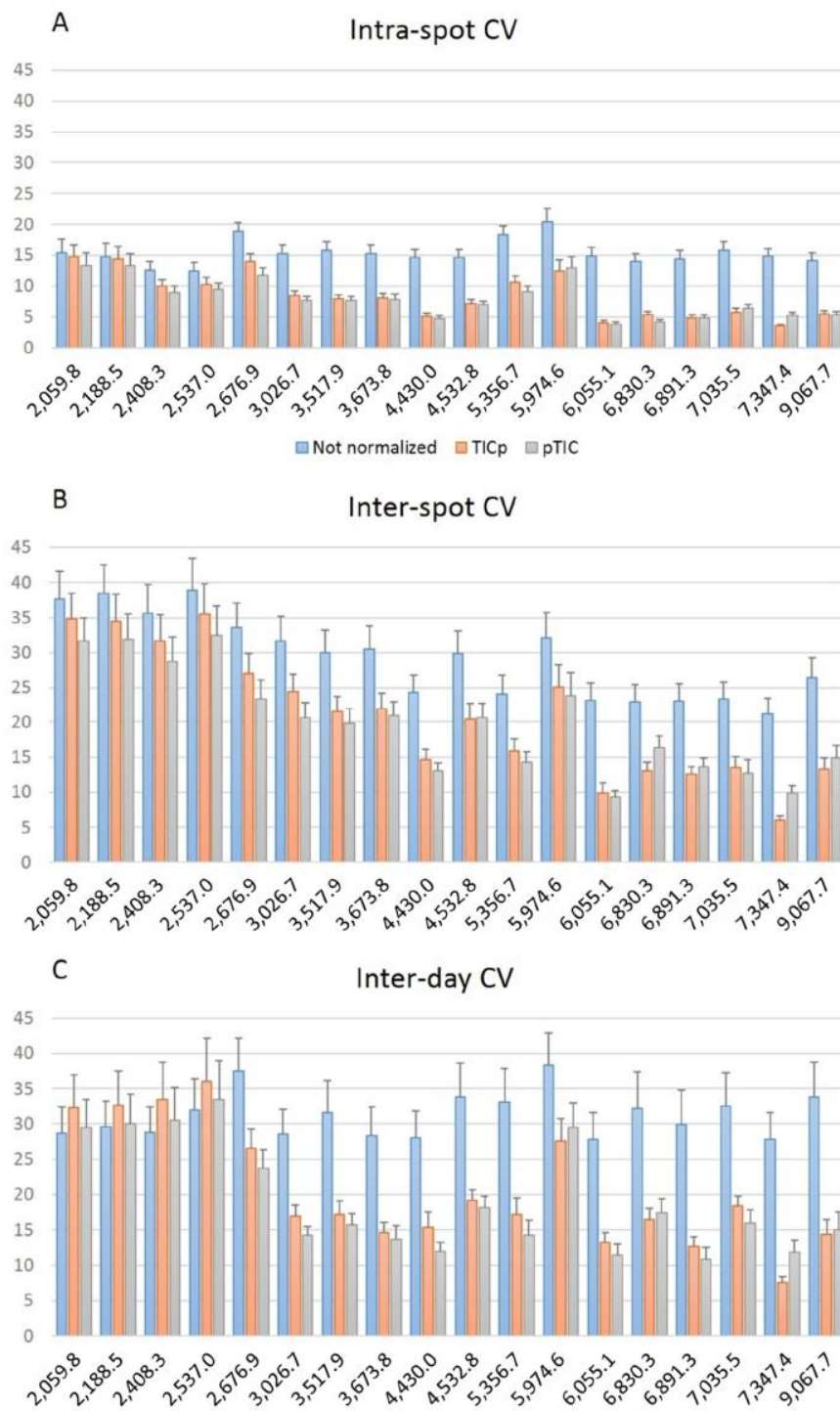


Figure 4. Comparison of the average coefficient of variation (CV) for the peak intensity of the 18 common peaks analyzed at intra-spot (A), inter-spot (B) and inter-day (C) levels.

Table 2. Arithmetic and post-alignment means at spot- and day-level without normalization and with the two different normalization methods applied—before (TICp) and after (pTIC) finding peaks.

	Not Normalized		TICp		pTIC	
	Arithmetic	Post-Alignment	Arithmetic	Post-Alignment	Arithmetic	Post-Alignment
Inter-spot mean	3825.44	3792.14	0.0005	0.0005	0.056	0.056
Inter-day mean	3974.07	3955.49	0.0005	0.0005	0.056	0.056

3.3. Classification of the Isolates Using Machine Learning

Three peak matrices generated (full-spectrum method and threshold methods TICp and pTIC) with data from the “classification set” were used as input data to test the capacity of the algorithms to discriminate VRE (VanA + VanB) from VSE isolates (Table S6). The 10-fold cross validation results for the SVM algorithm showed the best accuracy with 80.9% and an F1 score (the harmonic mean of the sensitivity and the accuracy of the model) of 80.5% for the full-spectrum-TIC method (Table 3A).

Table 3. Discrimination of vancomycin-susceptible *E. faecium* (VSE) from vancomycin-resistant *E. faecium* (VRE) isolates and, within the latter group, differentiation between the strains hosting the *vanA* and *vanB* resistance genes. The actual classification is shown in columns and the predictive classification in rows. (A) Results from the support vector machine (SVM) algorithm using the full-spectrum-TIC method. Accuracy: 80.9%; F1 score: 80.46%; sensitivity: 81.4% specificity: 80.43%; positive predictive value (PPV) or precision: 79.55%; negative predictive value (NPV): 82.22%. (B) Scores from the partial least squares–discrimination analysis (PLS-DA) algorithm using the total ion current normalization after peak finding (TICp) method. Accuracy: 86.05%; predictive value for VanA VRE: 80.65%; predictive value for VanB VRE: 89.09%.

Actual Classification	Predicted Classification	
(A) SVM Full Spectrum	VRE	VSE
VRE	70	16
VSE	18	74
(B) PLS Threshold TICp	VanA	VanB
VanA	25	6
VanB	6	49

The same procedures were also applied for the discrimination of VanA from VanB VRE strains (Table S7). In this approach, PLS-DA algorithm with TICp method provided 86.65% correct classification (Figure 5). This algorithm achieved 89.09% of predictive value for identifying VanB VRE strains in a 10-fold cross validation (Table 3B).

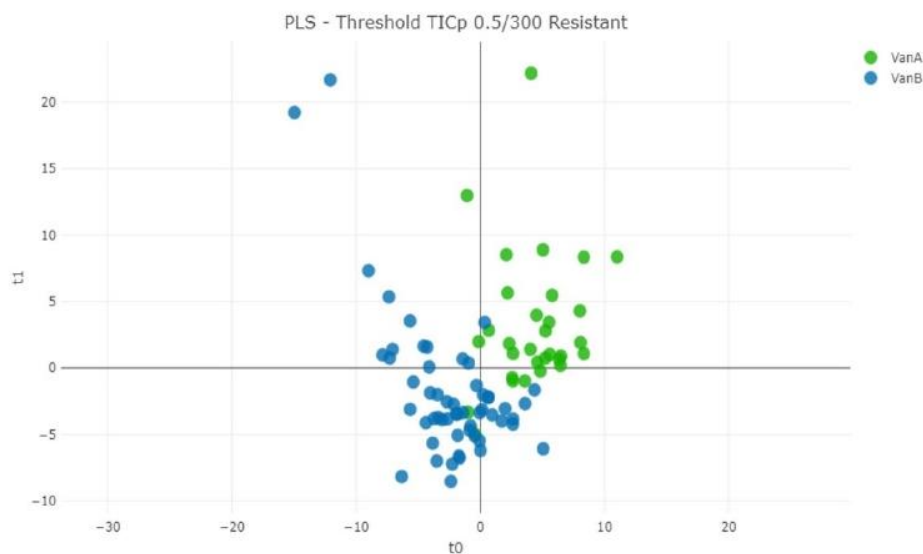


Figure 5. Distance plot of partial least squares–discrimination analysis (PLS-DA) machine learning algorithm for the discrimination of VanA from VanB VRE isolates using the TIC_p method.

The biomarker analysis revealed the presence of two potential resistance biomarkers at 6891.33 and 5095.01 m/z (Table 4). These two peaks showed AUC values greater than 0.8 in its receiver operating characteristic (ROC) curve (Figure S1). The 5095.01 m/z peak allowed the discrimination of the VRE strains from the VSE strains (Figure S2). This peak was present in 82 of the 86 VRE strains in this study. Furthermore, the AUC for peak at 6891.33 m/z allowed the discrimination of VRE isolates hosting the *vanA* resistance mechanism from those carrying the *vanB* one. Its intensity was higher in all VanA *E. faecium* isolates tested with an AUC of 0.831 and a CV value of 22.99% (Figure S2). These results were obtained by applying a threshold (0.01 factor) after a TIC normalization using the 178 pre-processed samples of the classification set.

Table 4. Biomarker peaks for the detection of the vancomycin-resistant *E. faecium* (VRE) isolates and the discrimination of strains hosting the VanA mechanism.

Peak m/z	AUC (≥ 0.8)	Appearance (Total Samples)	Positive Category	Coefficient of Variation (CV)
5095.01	0.814	123/178	82/86 (resistant)	61.63%

4. Discussion

The application of MALDI-TOF MS coupled with data analysis was shown to be a reproducible methodology—CV values ≤ 20.88 for normalized data—that allowed the discrimination of VRE (VanA + VanB) from VSE in 80.9% of the cases using SVM and the correct differentiation of 86.6% of the *E. faecium* VanB isolates from the VanA VRE isolates by PLS-DA. Specific peaks for the discrimination of the studied isolates were found: the 5095.01 m/z peak was present in 82/86 VREs. Although this peak has already been described as a biomarker for VRE isolates [25], its real meaning is currently under debate. Brackmann et al. (2020) recently reported the sequencing of the 5095.01 m/z protein and identified the protein hiracin, a secretory protein encoded by the *hirJ*M79 gene, whose role in the vancomycin resistance mechanism remains unknown [26]. Although our study supports the value of the 5095.01 m/z peak as a marker for VRE isolates, caution should be exercised when using this peak for VRE presence until further studies unravel its correlation with vancomycin resistance.

The 6891.33 m/z peak was found as a biomarker for VanA VRE isolates. Both the AUC (0.831) and the CV values (22.99%) indicate the uniqueness and repeatability of this marker. Although this peak had been related before with different clonal complexes and sequence types [17], its correlation with VanA VRE isolates had not been reported so far [27].

The repeatability study carried out with 20 *E. faecium* isolates (10 VSE, 5 VanA VRE and 5 VanB VRE) demonstrated that the lowest CV values for peak intensities were obtained for normalized data ($p < 0.0001$), regardless of the order in which normalization and peak finding is performed. Intra-spot repeatability showed the highest rate of repeatability (CV values ranging from 7.99 to 8.46), although the CV values for inter-spot and inter-day repeatability ranged between 19.91 and 20.88 for inter-spot variability and between 19.30 to 20.66 for inter-day repeatability. Similar CV values (6.5–17%) have been reported in a study that evaluated the technical repeatability of MALDI-TOF MS for quantitative protein profiling [28]. Therefore, this methodology is considered reproducible and feasible for peak analysis, especially in the range of 3000–9000 m/z . In addition, 13 of the 18 most representative common peaks for all *E. faecium* and also both biomarker peaks for VRE vs. VSE and VanA VRE vs. VanB VRE differentiation are located within this spectrum range. The implementation of the methodology described in this study could provide standardization for data comparison with other studies analyzing antimicrobial resistance with MALDI-TOF MS.

Previous studies have shown the ability of MALDI-TOF MS to differentiate among *E. faecium* vancomycin-resistant high-risk clones, clonal complexes and sequence types with different success rates [16–18]. Differentiation between VRE and VSE isolates has also been reported. Griffin et al. reported 88.45% correct discrimination using SVM and 88.24% with the implementation of the genetic algorithm [25]. This algorithm also allowed a discrimination rate of 92.4% for VanA VRE from VSE isolates in the study developed by Nakano et al. [27]. In our case, the application of SVM, RF and PLS-DA algorithms provided 80.9%, 79.2% and 77.5% correct classification of VRE and VSE isolates using the full-spectrum method. Moreover, discrimination between VanA and VanB VRE isolates was achieved in 86.6% of the cases by applying the TICp method. Discrepancies in the peaks used for discrimination of the different *E. faecium* groups were detected with the two previous studies: none of the peaks included in the study by Nakano et al. were found relevant in our models; besides, only the 5095.01 m/z peak (5094.7 m/z in Griffin et al.) was common to both studies for the discrimination of VRE from VSE. In our predictive models, the 6603 m/z peak proposed by Griffin et al. for the discrimination of VanA from VanB VRE was not considered discriminative. Instead, the 6891.78 m/z peak served this purpose in our study.

Although further studies are requested in order to clarify the role of the 5095.01 m/z peak for the routine detection of VanB VRE isolates, we propose the detection of this peak in combination with the 6891.78 m/z peak for the detection of suspected VanA VRE, for the differentiation of VRE using MALDI-TOF MS. Despite VanB VRE isolates being more prevalent in our setting, VanA VRE isolates have been reportedly correlated with hospital infections in different European countries [8]. Thus, both biomarker peaks could be used for rapid screening of VRE isolates with MALDI-TOF MS.

One of the limitations of this study is the lack of genomic background for the analyzed isolates. Only the genes encoding vancomycin resistance were targeted, and its presence or absence was confirmed by molecular methods. Although this information was useful for the development of predictive models, more comprehensive information about our VRE isolates could help obtain higher discrimination power from the applied algorithms and also explain the misclassifications from the current models.

Another limitation of the study is that clonality of the isolates was only analyzed within a clinical and epidemiological approach. HGUGM is a tertiary hospital where clinical departments are in different wards, far from each other, and contact among them is sporadic—as each department has its own medical staff—so an outbreak affecting

different departments is unlikely. The theory of non-clonality is also supported by Griffin et al., since the 5095.01 m/z peak was also found in their study, performed in a different continent [25]. We acknowledge that clonality cannot be fully discarded as we lack a deep whole-genome sequencing approach. More accurate molecular methods are needed to rule out that the *E. faecium* isolates analyzed in this study belong to the same clone. However, our results support the importance of a previously described peak for differentiating VRE from VSE isolates and add a new specific biomarker for the discrimination of VanA VRE strains.

The classification accuracy of the applied algorithms has shown to be <90% in all cases. Therefore, further studies with well-characterized isolates sourced from different geographic origins are needed to confirm the results obtained in this work and improve them if possible.

5. Conclusions

In conclusion, MALDI-TOF MS has demonstrated acceptable discrimination of *E. faecium* isolates beyond species assignment. Although further refining is requested and isolates from different clones and origins have to be included in predictive models in order to understand how they are discriminated by MALDI-TOF MS, protein profiling could become a suitable tool for the rapid detection of VRE in clinical microbiology laboratories. Its implementation could be key for the control of VRE isolates in hospital settings.

Supplementary Materials: The following supporting information can be downloaded at www.mdpi.com/article/10.3390/diagnostics12020328/s1, Table S1: Description of the 178 *E. faecium* isolates included in this study; Table S2: *E. faecium* isolates included in the repeatability assay. Their Pearson correlation coefficient for inter-day (biological) repeatability is shown; Table S3: Coefficient of variation of the intensity means per level, normalization after (pTIC) and before (TICp) and antibiotic susceptibility; Table S4: Comparison of the p -values between average coefficient of variation of intensity means for peaks from 2000 to 3000 m/z and 3000 to 9000 m/z range; Table S5: p -Values for arithmetic and post-alignment means at the spot and day level with and without normalization; Table S6: Accuracy and F1-factor of the internal validation for the classification of vancomycin-resistant *E. faecium* (VRE) and vancomycin-susceptible *E. faecium* (VSE) strains as a metric for algorithms scores. Data obtained with the k -fold cross validation method, where $k = 10$. Scaled data; Table S7: Accuracy of internal validation for the classification of VanA and VanB vancomycin-resistant *E. faecium* (VRE) strains as metric for algorithms scores. Data obtained with the k -fold cross validation method, where $k = 10$. Scaled data; Figure S1: Area under the curve (AUC) values for the two potential biomarkers found for (A) vancomycin-resistant *E. faecium* (VRE) isolates (5095.01 m/z) and (B) for VanA VRE isolates (6891.33 m/z); Figure S2: Visual discrimination of protein spectra from vancomycin-resistant *E. faecium* (VRE) (in red) and vancomycin-susceptible *E. faecium* (VSE) (green) according to the presence/absence of the peak at 5095.01 m/z (A). Differentiation of Van A (light blue) and VanB VRE isolates (purple) was based on the intensity of the peak at 6891.33 m/z (B).

Author Contributions: A.C.: experimental part, formal analysis, data collection, validation, visualization, writing—original draft preparation and review/editing; M.J.A., Á.S.-M. and G.M., data analysis, validation, writing—original draft preparation and review/editing; D.R.-T., E.C. and M.M., formal analysis, data analysis and review/editing; L.Q. and A.R., experimental part and data collection; P.M., manuscript review; L.M., conceptualization, formal analysis, validation, original draft preparation and review/editing; B.R.-S., conceptualization, project administration, formal analysis, supervision, validation, visualization, original draft preparation and review/editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by project PII5/01073 from the Health Research Fund (Instituto de Salud Carlos III. Plan Nacional de I + D + I 2013–2016) of the Carlos III Health Institute (ISCIII, Madrid, Spain) partially financed by the European Regional Development Fund (FEDER) “A way of making Europe”. B.R.-S. (CPII19/00002) is a recipient of a Miguel Servet contract supported by the ISCIII.

Institutional Review Board Statement: The research presented in this study was conducted in accordance with the Declaration of Helsinki. It did not involve humans or human data. The Ethics Committee for Research with Medicinal Products from the Hospital Gregorio Marañón—CEIm

(code no. MICRO.HGUGM.2020-002, 15 September 2020) approved this study. This committee determined that, since only bacterial isolates (not human products) were analyzed, all the conditions to waive the informed consent have been met.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated or analyzed during this study are included in this published article (and its Supplementary Information files). Classification algorithms are available at <https://platform.dovermsdataanalysis.com/>.

Acknowledgments: The authors are grateful to Luis Alcalá for his assistance with statistical analysis.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Vu, J.; Carvalho, J. Enterococcus: Review of its physiology, pathogenesis, diseases and the challenges it poses for clinical microbiology. *Front. Biol.* **2011**, *6*, 357–366.
2. Arias, C.A.; Murray, B.E. The rise of the Enterococcus: Beyond vancomycin resistance. *Nat. Rev. Microbiol.* **2012**, *10*, 266–278.
3. Cattoir, V.; Leclercq, R. Twenty-five years of shared life with vancomycin-resistant enterococci: Is it time to divorce? *J. Antimicrob. Chemother.* **2013**, *68*, 731–742.
4. Miller, W.R.; Murray, B.E.; Rice, L.B.; Arias, C.A. Resistance in Vancomycin-Resistant Enterococci. *Infect. Dis. Clin. N. Am.* **2020**, *34*, 751–771.
5. Cattoir, V.; Giard, J.C. Antibiotic resistance in Enterococcus faecium clinical isolates. *Expert Rev. Anti-Infect. Ther.* **2014**, *12*, 239–248.
6. Rice, L.B. Federal funding for the study of antimicrobial resistance in nosocomial pathogens: No ESKAPE. *J. Infect. Dis.* **2008**, *197*, 1079–1081.
7. Chiang, H.Y.; Perencevich, E.N.; Nair, R.; Nelson, R.E.; Samore, M.; Khader, K.; Chorazy, M.L.; Herwaldt, L.A.; Blevins, A.; Ward, M.A.; et al. Incidence and Outcomes Associated with Infections Caused by Vancomycin-Resistant Enterococci in the United States: Systematic Literature Review and Meta-Analysis. *Infect. Control. Hosp. Epidemiol.* **2017**, *38*, 203–215.
8. Lopez, M.; Cercenado, E.; Tenorio, C.; Ruiz-Larrea, F.; Torres, C. Diversity of clones and genotypes among vancomycin-resistant clinical Enterococcus isolates recovered in a Spanish hospital. *Microb. Drug Resist.* **2012**, *18*, 484–491.
9. Mirzaei, B.; Farivar, T.N.; Juhari, P.; Mehr, M.A.; Babaei, R. Investigation of the Prevalence of vanA and vanB genes in vancomycin resistant enterococcus (VRE) by Taq Man real time PCR Assay. *J. Microbiol. Infect. Dis.* **2013**, *3*, 192.
10. Hricová, K.; Štosová, T.; Kučová, P.; Fišerová, K.; Bardoň, J.; Kolář, M. Analysis of Vancomycin-Resistant Enterococci in Hemato-Oncological Patients. *Antibiotics* **2020**, *9*, 785.
11. Seng, P.; Drancourt, M.; Gouriet, F.; La Scola, B.; Fournier, P.E.; Rolain, J.M.; Raoult, D. Ongoing revolution in bacteriology: Routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **2009**, *49*, 543–551.
12. Oviaño, M.; Gómara, M.; Barba, M.J.; Revillo, M.J.; Barbeyto, L.P.; Bou, G. Towards the early detection of beta-lactamase-producing Enterobacteriaceae by MALDI-TOF MS analysis. *J. Antimicrob. Chemother.* **2017**, *72*, 2259–2262.
13. Schuster, D.; Josten, M.; Janssen, K.; Bodenstein, I.; Albert, C.; Schallenberg, A.; Gajdiss, M.; Sib, E.; Szekat, C.; Kehl, K.; et al. Detection of methicillin-resistant coagulase-negative staphylococci harboring the class A mec complex by MALDI-TOF mass spectrometry. *Int. J. Med. Microbiol.* **2018**, *308*, 522–526.
14. Oviano, M.; Bou, G. Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry for the Rapid Detection of Antimicrobial Resistance Mechanisms and Beyond. *Clin. Microbiol. Rev.* **2019**, *32*, 927–937.
15. Rodriguez-Sanchez, B.; Cercenado, E.; Coste, A.T.; Greub, G. Review of the impact of MALDI-TOF MS in public health and hospital hygiene, 2018. *Eurosurveillance* **2019**, *24*, 1800193.
16. Lasch, P.; Fleige, C.; Stänmler, M.; Layer, F.; Nübel, U.; Witte, W.; Werner, G. Insufficient discriminatory power of MALDI-TOF mass spectrometry for typing of Enterococcus faecium and Staphylococcus aureus isolates. *J. Microbiol. Methods* **2014**, *100*, 58–69.
17. Freitas, A.R.; Sousa, C.; Novais, C.; Silva, L.; Ramos, H.; Coque, T.M.; Lopes, J.; Peixe, L. Rapid detection of high-risk Enterococcus faecium clones by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Diagn. Microbiol. Infect. Dis.* **2017**, *87*, 299–307.
18. Holzknacht, B.J.; Dargis, R.; Pedersen, M.; Pinholt, M.; Christensen, J.J.; Hammerum, A.M.; Littauer, P.; Worning, P.; Westh, H.; Moser, C.; et al. Typing of vancomycin-resistant enterococci with MALDI-TOF mass spectrometry in a nosocomial outbreak setting. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **2018**, *24*, 1104 e1101–1104 e1104.

19. Rodríguez-Sánchez, B.; Marín, M.; Sánchez-Carrillo, C.; Cercenado, E.; Ruiz, A.; Rodríguez-Créixems, M.; Bouza, E.. Improvement of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry identification of difficult-to-identify bacteria and its impact in the workflow of a clinical microbiology laboratory. *Diagn Microbiol. Infect. Dis.* **2014**, *79*, 1–6.
20. Oberle, M.; Wohlwend, N.; Jonas, D.; Maurer, F.P.; Jost, G.; Tschudin-Sutter, S.; Vranckx, K.; Egli, A. The Technical and Biological Reproducibility of Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry (MALDI-TOF MS) Based Typing: Employment of Bioinformatics in a Multicenter Study. *PLoS ONE* **2016**, *11*, e0164260.
21. Quintela-Baluja, M.; Böhme, K.; Fernández-No, I.C.; Morandi, S.; Alnakip, M.E.; Caamaño-Antelo, S.; Barros-Velázquez, J.; Calomata, P. Characterization of different food-isolated Enterococcus strains by MALDI-TOF mass fingerprinting. *Electrophoresis* **2013**, *34*, 2240–2250.
22. Zvezdanova, M.E.; Arroyo, M.J.; Méndez, G.; Guinea, J.; Mancera, L.; Muñoz, P.; Rodríguez-Sánchez, B.; Escribano, P. Implementation of MALDI-TOF Mass Spectrometry and Peak Analysis: Application to the Discrimination of Cryptococcus neoformans Species Complex and Their Interspecies Hybrids. *J. Fungi* **2020**, *6*, 330.
23. Weis, C.V.; Jutzeler, C.R.; Borgwardt, K. Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: A systematic review. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **2020**, *26*, 1310–1317.
24. Timm, W.; Scherbar, A.; Bocker, S.; Kohlbacher, O.; Nattkemper, T.W. Peak intensity prediction in MALDI-TOF mass spectrometry: A machine learning study to support quantitative proteomics. *BMC Bioinform.* **2008**, *9*, 443.
25. Griffin, P.M.; Price, G.R.; Schooneveldt, J.M.; Schlebusch, S.; Tilse, M.H.; Urbanski, T.; Hamilton, B.; Venter, D. Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak. *J. Clin. Microbiol.* **2012**, *50*, 2918–2931.
26. Brackmann, M.; Leib, S.L.; Tonolla, M.; Schurch, N.; Wittwer, M. Antimicrobial resistance classification using MALDI-TOF-MS is not that easy: Lessons from vancomycin-resistant Enterococcus faecium. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **2020**, *26*, 391–393.
27. Nakano, S.; Matsumura, Y.; Kato, K.; Yunoki, T.; Hotta, G.; Noguchi, T.; Yamamoto, M.; Nagao, M.; Ito, Y.; Takakura, S.; et al. Differentiation of vanA-positive Enterococcus faecium from vanA-negative E. faecium by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Int. J. Antimicrob. Agents* **2014**, *44*, 256–259.
28. Pang, R.T.; Johnson, P.J.; Chan, C.M.; Kong, E.K.; Chan, A.T.; Sung, J.J.; Poon, T.C. Technical evaluation of MALDI-TOF mass spectrometry for quantitative proteomic profiling matrix formulation and application. *Clin. Proteom. J.* **2004**, *1*, 12.

MATERIAL SUPLEMENTARIO CAPÍTULO 1

Table S1. Description of the 178 *E. faecium* isolates included in this study.

Strain ID	Source	Resistance mechanism
1	2-17	Blood culture
2	2690-17	Rectal swab
3	2691-17	Rectal swab
4	2799-17	Rectal swab
5	2800-17	Rectal swab
6	1281-17	Rectal swab
7	250-18	Rectal swab
8	1290-18	Rectal swab
9	1552-18	Rectal swab
10	1785-18	Rectal swab
11	2051-18	Rectal swab
12	2082-18	Rectal swab
13	2271-18	Rectal swab
14	2272-18	Rectal swab
15	3307-18	Rectal swab
16	3412-18	Rectal swab
17	3549-18	Rectal swab
18	31-19	Rectal swab
19	40-19	Rectal swab
20	87-19	Rectal swab
21	92-19	Rectal swab
22	262-19	Rectal swab
23	457-19	Rectal swab
24	942-19	Blood culture
25	947-19	Blood culture
26	968-19	Rectal swab
27	1355-19	Rectal swab
28	1538-19	Rectal swab
29	1756-19	Rectal swab
30	19138416-19	Rectal swab
31	3024-19	Rectal swab
32	313-17	Blood culture
33	342-17	Blood culture
34	2038-17	Blood culture
35	107-18	Rectal swab
36	242-18	Rectal swab
37	620-18	Rectal swab
38	684-18	Rectal swab
39	742-18	Rectal swab
40	764-18	Rectal swab
41	804-18	Rectal swab
42	1111-18	Rectal swab
43	1112-18	Rectal swab
44	1124-18	Rectal swab
45	1287-18	Rectal swab
46	1289-18	Rectal swab
47	1403-18	Rectal swab
48	1451-18	Rectal swab
49	1472-18	Rectal swab
50	1553-18	Rectal swab
51	1590-18	Rectal swab
52	1786-18	Rectal swab
53	1876-18	Rectal swab

MATERIAL SUPLEMENTARIO CAPÍTULO 1

Rapid and reproducible MALDI-TOF-based method for detection Vancomycin-resistant *Enterococcus faecium* using classifying algorithms

54	1887-18	Rectal swab	vanB
55	1905-18	Rectal swab	vanB
56	1954-18	Rectal swab	vanB
57	2015-18	Rectal swab	vanB
58	2269-18	Rectal swab	vanB
59	1376-18	Rectal swab	vanB
60	2422-18	Rectal swab	vanB
61	3305-18	Blood culture	vanB
62	427-18	Rectal swab	vanB
63	428-18	Rectal swab	vanB
64	581-18	Rectal swab	vanB
65	3495-18	Rectal swab	vanB
66	2335-18	Rectal swab	vanB
67	2477-18	Blood culture	vanB
68	2905-18	Blood culture	vanB
69	3000-18	Blood culture	vanB
70	2063-19	Blood culture	vanB
71	2098-19	Blood culture	vanB
72	2168-19	Rectal swab	vanB
73	2608-19	Blood culture	vanB
74	3267-19	Rectal swab	vanB
75	2830-19	Rectal swab	vanB
76	3212-19	Blood culture	vanB
77	3217-19	Rectal swab	vanB
78	3388-19	Blood culture	vanB
79	3550-19	Rectal swab	vanB
80	3593-19	Blood culture	vanB
81	437-19	Rectal swab	vanB
82	655-19	Blood culture	vanB
83	241-19	Blood culture	vanB
84	458-19	Rectal swab	vanB
85	504-19	Rectal swab	vanB
86	640-19	Rectal swab	vanB
87	60-18	Blood culture	Susceptible
88	144-18	Blood culture	Susceptible
89	299-18	Blood culture	Susceptible
90	314-18	Blood culture	Susceptible
91	369-18	Blood culture	Susceptible
92	415-18	Blood culture	Susceptible
93	432-18	Blood culture	Susceptible
94	468-18	Blood culture	Susceptible
95	613-18	Blood culture	Susceptible
96	615-18	Blood culture	Susceptible
97	698-18	Blood culture	Susceptible
98	825-18	Blood culture	Susceptible
99	871-18	Blood culture	Susceptible
100	884-18	Blood culture	Susceptible
101	908-18	Blood culture	Susceptible
102	989-18	Blood culture	Susceptible
103	992-18	Blood culture	Susceptible
104	1002-18	Blood culture	Susceptible
105	1039-18	Blood culture	Susceptible
106	1053-18	Blood culture	Susceptible
107	1060-18	Blood culture	Susceptible
108	1116-18	Blood culture	Susceptible
109	1147-18	Blood culture	Susceptible
110	1162-18	Blood culture	Susceptible

MATERIAL SUPLEMENTARIO CAPÍTULO 1

Rapid and reproducible MALDI-TOF-based method for detection Vancomycin-resistant *Enterococcus faecium* using classifying algorithms

111	1170-18	Blood culture	Susceptible
112	1293-18	Blood culture	Susceptible
113	1337-18	Blood culture	Susceptible
114	1401-18	Blood culture	Susceptible
115	1426-18	Blood culture	Susceptible
116	1498-18	Blood culture	Susceptible
117	1545-18	Blood culture	Susceptible
118	1615-18	Blood culture	Susceptible
119	1641-18	Blood culture	Susceptible
120	1717-18	Blood culture	Susceptible
121	1771-18	Blood culture	Susceptible
122	1791-18	Blood culture	Susceptible
123	2061-18	Blood culture	Susceptible
124	2095-18	Blood culture	Susceptible
125	2287-18	Blood culture	Susceptible
126	2439-18	Blood culture	Susceptible
127	3277-18	Blood culture	Susceptible
128	3312-18	Blood culture	Susceptible
129	37-19	Blood culture	Susceptible
130	131-19	Blood culture	Susceptible
131	310-19	Blood culture	Susceptible
132	328-19	Blood culture	Susceptible
133	345-19	Blood culture	Susceptible
134	368-19	Blood culture	Susceptible
135	378-19	Blood culture	Susceptible
136	390-19	Blood culture	Susceptible
137	395-19	Blood culture	Susceptible
138	430-19	Blood culture	Susceptible
139	460-19	Blood culture	Susceptible
140	464-19	Blood culture	Susceptible
141	552-19	Blood culture	Susceptible
142	559-19	Blood culture	Susceptible
143	575-19	Blood culture	Susceptible
144	595-19	Blood culture	Susceptible
145	598-19	Blood culture	Susceptible
146	604-19	Blood culture	Susceptible
147	676-19	Blood culture	Susceptible
148	950-19	Blood culture	Susceptible
149	965-19	Blood culture	Susceptible
150	994-19	Blood culture	Susceptible
151	1021-19	Blood culture	Susceptible
152	1033-19	Blood culture	Susceptible
153	1039-19	Blood culture	Susceptible
154	1241-19	Blood culture	Susceptible
155	1366-19	Blood culture	Susceptible
156	1468-19	Blood culture	Susceptible
157	1554-19	Blood culture	Susceptible
158	1590-19	Blood culture	Susceptible
159	1593-19	Blood culture	Susceptible
160	1625-19	Blood culture	Susceptible
161	1684-19	Blood culture	Susceptible
162	1703-19	Blood culture	Susceptible
163	1721-19	Blood culture	Susceptible
164	19152313	Rectal swab	Susceptible
165	19152964	Rectal swab	Susceptible
166	19152967	Rectal swab	Susceptible
167	19154177	Rectal swab	Susceptible

MATERIAL SUPLEMENTARIO CAPÍTULO 1

Rapid and reproducible MALDI-TOF-based method for detection Vancomycin-resistant *Enterococcus faecium* using classifying algorithms

168	19154199	Rectal swab	Susceptible
169	19154200	Rectal swab	Susceptible
170	19154757	Rectal swab	Susceptible
171	19156397	Rectal swab	Susceptible
172	19156441	Rectal swab	Susceptible
173	19156448	Rectal swab	Susceptible
174	19157126	Rectal swab	Susceptible
175	19159082	Rectal swab	Susceptible
176	19159630	Rectal swab	Susceptible
177	19159642	Rectal swab	Susceptible
178	19159645	Rectal swab	Susceptible

Table S2. *E. faecium* isolates included in the repeatability assay. Their Pearson correlation coefficient for inter-days (biological) repeatability is shown.

Samples	P Pearson
"Enterococcus_faecium_S_1441"	0,856796493
"Enterococcus_faecium_S_41518"	0,968867739
"Enterococcus_faecium_S_82518"	0,988280503
"Enterococcus_faecium_S_96519"	0,968823737
"Enterococcus_faecium_S_98918"	0,965538082
"Enterococcus_faecium_S_100218"	0,989609441
"Enterococcus_faecium_S_124119"	0,945246999
"Enterococcus_faecium_S_171718"	0,888020473
"Enterococcus_faecium_S_327718"	0,969714629
"Enterococcus_faecium_S_19159642"	0,765205711
"Enterococcus_faecium_VanA_217"	0,988879506
"Enterococcus_faecium_VanA_2690"	0,990673724
"Enterococcus_faecium_VanA_2799"	0,984994445
"Enterococcus_faecium_VanA_2800"	0,990077871
"Enterococcus_faecium_VanA_3024"	0,953617436
"Enterococcus_faecium_VanB_2063"	0,908970353
"Enterococcus_faecium_VanB_2168"	0,91930336
"Enterococcus_faecium_VanB_3212"	0,966220338
"Enterococcus_faecium_VanB_3217"	0,908212247
"Enterococcus_faecium_VanB_3388"	0,898020244

Table S3. Coefficient of variation of the intensity means per level, normalization after (pTIC) and before (TICp) finding peaks and antibiotic susceptibility.

	Intra-spots			Inter-Spots			Inter-Days		
	Not Normalized	TICp	pTIC	Not Normalized	TICp	pTIC	Not Normalized	TICp	pTIC
TOTAL	15.35442	8.460269	7.999143	29.29323	20.88475	19.90829	31.24830	20.65608	19.29570
SUSCEPTIBLE	17.66087	9.102237	8.446666	23.68497	17.3779	16.92136	33.55983	22.47998	20.9593
RESISTANT	12.07580	7.538624	7.356173	33.96677	23.80711	22.39739	28.93677	18.83218	17.63204

Table S4. Comparison of the p-values between average coefficient of variation of intensity means for peaks from 2000 to 3000 *m/z* and 3000 to 9000 *m/z* range.

	Not normalized	TICp	pTIC
Days	0.9549	2.68E-06	0.000934
Spots	0.5663	0.000528	0.001684
Shoots	3.76E-05	2.77E-05	2.73E-05

Table S5. p-values for arithmetic and post-alignment means at spot- and day level with and without normalization.

p-Value	Not Normalized	TICp
Inter-Spots Mean	0.8366	0.09755
Inter-Days Mean	0.9298	0.1283

Table S6. Accuracy and F1-factor of the internal validation for the classification of Vancomycin-resistant *E. faecium* (VRE) and Vancomycin-susceptible *E. faecium* (VSE) strains as metric for algorithms scores. Data obtained with k-fold cross validation method where k=10. Scaled data.

	Full spectrum-TIC		Peaks-TIC		TIC-Peaks	
	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Partial Least Squares (4 Components)	77.53%	75.31%	72.47%	70.66%	76.4%	74.7%
Support Vector Machine Linear SVC (liblinear)	80.9%	80.46%	70.22%	70.72%	76.97%	75.45%
Random Forest	79.21%	77.85%	75.84%	74.56%	72.47%	70.3%

Table S7. Accuracy of internal validation for the classification of vanA and vanB Vancomycin-resistant *E. faecium* (VRE) strains as metric for algorithms scores. Data obtained with k-fold cross validation method where k=10. Scaled data.

	Full spectrum-TIC	Peaks-TIC	TIC-Peaks
Partial Least Squares (4 Components)	83.72%	83.72%	86.65%
Support Vector Machine - LinearSVC (liblinear)	76.74%	80.23	73.26
Random Forest	79.07%	80.23%	79.07%

MATERIAL SUPLEMENTARIO CAPÍTULO 1

Rapid and reproducible MALDI-TOF-based method for detection Vancomycin-resistant *Enterococcus faecium* using classifying algorithms

Figure S1. Area Under the Curve (AUC) values for the two potential biomarkers found for A) Vancomycin-resistant *E. faecium* (VRE) isolates -5095.01 *m/z*- and B) for vanA VRE isolates -6891.33 *m/z*-.

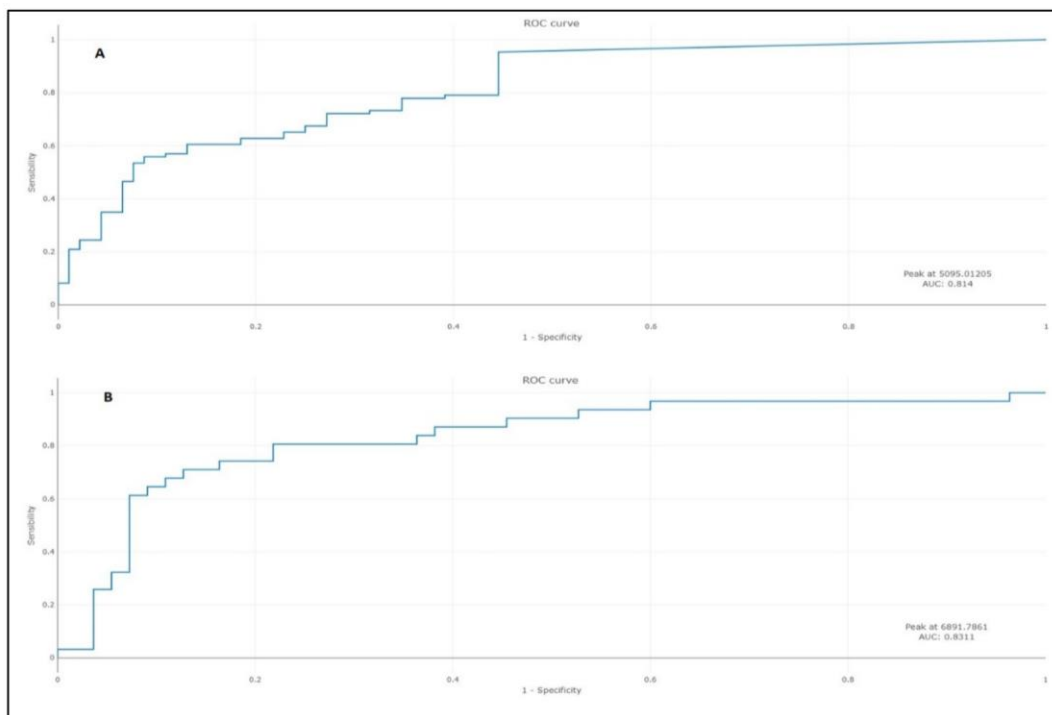
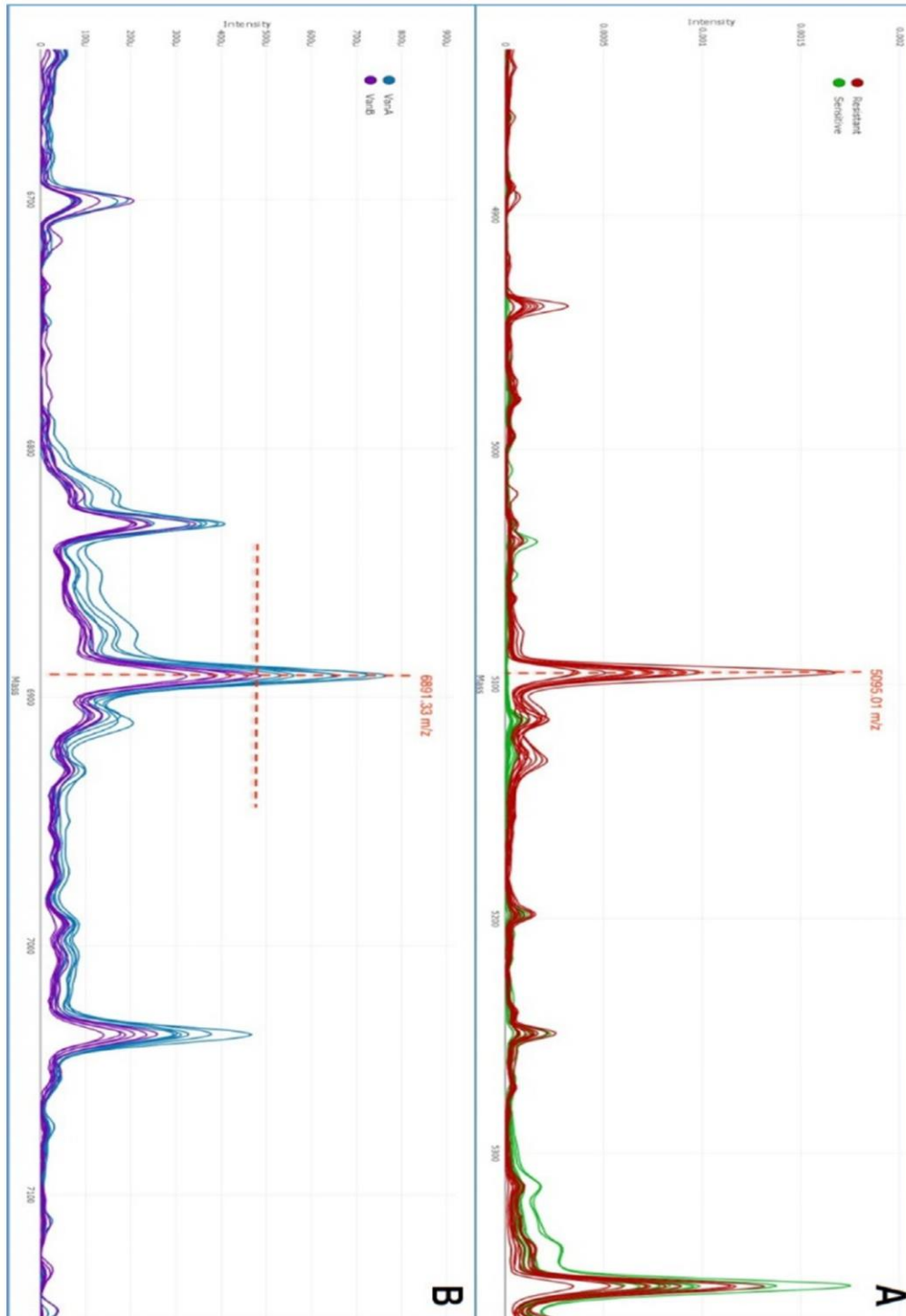


Figure S2. Visual discrimination of protein spectra from Vancomycin-resistant *E. faecium* (VRE) –in red- and Vancomycin-susceptible *E. faecium* (VSE) –green-according to the presence/absence of the peak at 5095.01 *m/z* (A). Differentiation of van A –blue- and vanB VRE isolates –purple- was based on the intensity of the peak at 6891.33 *m/z* (B).



Capítulo 2. Discriminación automática de las especies dentro del complejo *Enterobacter cloacae* empleando espectrometría de masas MALDI-TOF y algoritmos supervisados

Contexto

El complejo *E. cloacae* (ECC) abarca un grupo heterogéneo de especies que se asocian con brotes nosocomiales. Estas especies pueden tener diferentes mecanismos de virulencia y resistencia a los antimicrobianos, y su identificación a nivel de especie supone un desafío. Este estudio tiene como objetivo el desarrollo de modelos predictivos basados en el análisis de perfiles proteicos obtenidos con MALDI-TOF y la aplicación de *Machine Learning* para la identificación a nivel de especie.

Material y métodos

En este estudio se incluyeron un total de 219 ECC y 118 aislados clínicos de *Klebsiella aerogenes* pertenecientes a tres hospitales (Hospital General Universitario Gregorio Marañón -HGM-, Hospital Universitario Ramón y Cajal -HURC- y Hospital Universitario de Basel -UHB-). Se estudió la capacidad del método para diferenciar las especies de ECC más comunes (*Enterobacter asburiae*, *Enterobacter kobei*, *Enterobacter hormaechei*, *Enterobacter roggenkampii*, *Enterobacter ludwigii* y *Enterobacter bugandensis*) y *K. aerogenes* mediante la aplicación de algoritmos no supervisados de agrupamiento jerárquico y PCA, y algoritmos supervisados (PLS-DA, SVM y RF).

Resultados

Inicialmente y mediante algoritmos no supervisados se pudo observar una agrupación distintiva de *E. hormaechei* y *K. aerogenes* y una tendencia a la diferenciación entre el resto de las especies de ECC en el conjunto de datos empleados para el desarrollo del modelo.

Posteriormente se estudiaron modelos predictivos supervisados. La validación de estos modelos se realizó con espectros de proteínas de dos de los hospitales participantes (HURC y UHB) y arrojó una asignación de nivel de especie 100% correcta para *E. asburiae* y *E. kobei*, y entre 91,2% y 98,0% para las especies ECC restantes. Con espectros de nuevas cepas externas pertenecientes al HGUGM y analizados en los tres centros participantes, la precisión fue cercana al 100%.

Resultados similares se observaron con el análisis con la base de datos de identificación espectrométrica de masas (MSI) desarrollada recientemente (<https://msi.happy-dev.fr>), excepto en el caso de *E. hormaechei*, que se identificó con mayor precisión con nuestro algoritmo de RF.

Conclusiones

En este estudio se pudo demostrar que la aplicación de MALDI-TOF en combinación con algoritmos de *Machine Learning* puede emplearse como un método rápido y preciso para la diferenciación de especies de ECC.



Automatic Discrimination of Species within the *Enterobacter cloacae* Complex Using Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry and Supervised Algorithms

Ana Candela,^{a,b} Alejandro Guerrero-López,^c Miriam Mateos,^{d,l} Alicia Gómez-Asenjo,^{a,b} Manuel J. Arroyo,^e  Marta Hernández-García,^{d,f,i}  Rosa del Campo,^{d,f,i} Emilia Cercenado,^{a,b,g,h} Aline Cuénod,^{i,j} Gema Méndez,^e Luis Mancera,^e Juan de Dios Caballero,^{d,f,i} Laura Martínez-García,^{d,k,l} Desirée Gijón,^{d,f,i} María Isabel Morosini,^{d,f,i}  Patricia Ruiz-Garbajosa,^{d,f,i} Adrian Egli,^{i,j}  Rafael Cantón,^{d,f,i}  Patricia Muñoz,^{a,b,g,h}  David Rodríguez-Temporal,^{a,b}  Belén Rodríguez-Sánchez^{a,b}

^aClinical Microbiology and Infectious Diseases Department, Hospital General Universitario Gregorio Marañón, Madrid, Spain

[†]Institute of Health Research Gregorio Marañón, Madrid, Spain

[‡]Department of Signal Theory and Communication, University Carlos III of Madrid, Madrid, Spain

[§]Servicio de Microbiología, Hospital Universitario Ramón y Cajal, Madrid, Spain

[¶]Clover Bioanalytical Software, Granada, Spain

^{||}CIBER en Enfermedades Infecciosas, Madrid, Spain

^{¶¶}CIBER de Enfermedades Respiratorias, CIBERES CB06/06/0058, Madrid, Spain

^{***}Medicine Department, Faculty of Medicine, Universidad Complutense de Madrid, Madrid, Spain

^{††}Applied Microbiology Research, Department of Biomedicine, University of Basel, Basel, Switzerland

^{†††}Division of Clinical Bacteriology and Mycology, University Hospital Basel, Basel, Switzerland

^{††††}Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública, Madrid, Spain

^{†††††}Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain

Ana Candela and Alejandro Guerrero-López contributed equally to this study. Author order was determined alphabetically.

ABSTRACT The *Enterobacter cloacae* complex (ECC) encompasses heterogeneous clusters of species that have been associated with nosocomial outbreaks. These species may have different acquired antimicrobial resistance and virulence mechanisms, and their identification is challenging. This study aims to develop predictive models based on matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) profiles and machine learning for species-level identification. A total of 219 ECC and 118 *Klebsiella aerogenes* clinical isolates from three hospitals were included. The capability of the proposed method to differentiate the most common ECC species (*Enterobacter asburiae*, *Enterobacter kobei*, *Enterobacter hormaechei*, *Enterobacter roggenkampii*, *Enterobacter ludwigii*, and *Enterobacter bugandensis*) and *K. aerogenes* was demonstrated by applying unsupervised hierarchical clustering with principal-component analysis (PCA) preprocessing. We observed a distinctive clustering of *E. hormaechei* and *K. aerogenes* and a clear trend for the rest of the ECC species to be differentiated over the development data set. Thus, we developed supervised, nonlinear predictive models (support vector machine with radial basis function and random forest). The external validation of these models with protein spectra from two participating hospitals yielded 100% correct species-level assignment for *E. asburiae*, *E. kobei*, and *E. roggenkampii* and between 91.2% and 98.0% for the remaining ECC species; with data analyzed in the three participating centers, the accuracy was close to 100%. Similar results were obtained with the Mass Spectrometric Identification (MSI) database developed recently (<https://msi.happy-dev.fr>) except in the case of *E. hormaechei*, which was more accurately identified with the random forest algorithm. In short, MALDI-TOF MS combined with machine learning was demonstrated to be a rapid and accurate method for the differentiation of ECC species.

Editor Patricia J. Simner, Johns Hopkins University

Copyright © 2023 American Society for Microbiology. All Rights Reserved.

Address correspondence to David Rodríguez-Temporal, david.rodriguez@iisgm.com.

The authors declare no conflict of interest.

Received 20 July 2022

Returned for modification 14 August 2022

Accepted 28 February 2023

KEYWORDS *Enterobacter* species, mass spectrometry, MALDI-TOF MS, peak analysis, machine learning

E*nterobacter* is a genus of facultative anaerobic Gram-negative organisms that can be found as natural commensals in the gut microbiome of mammals (1). Several species have been associated with nosocomial outbreaks, causing urinary tract infections, skin and soft tissue infections, pneumonia, and bacteremia (2, 3). *Enterobacter cloacae* complex (ECC) is of particular clinical interest. This group is composed of 13 heterogenic genetic clusters according to *hsp60* gene sequencing, i.e., *Enterobacter asburiae* (cluster I), *Enterobacter kobei* (cluster II), *Enterobacter hormaechei* subsp. *hoffmannii* (cluster III), *Enterobacter roggkampii* (cluster IV), *Enterobacter ludwigii* (cluster V), *E. hormaechei* subsp. *oharae* and *E. hormaechei* subsp. *xiangfangensis* (cluster VI), *E. hormaechei* subsp. *hormaechei* (cluster VII), *E. hormaechei* subsp. *steigerwaltii* (cluster VIII), *Enterobacter bugandensis* (cluster IX), *Enterobacter nimipressuralis* (cluster X), *Enterobacter cloacae* subsp. *cloacae* (cluster XI), *E. cloacae* subsp. *dissolvens* (cluster XII), and a heterogeneous group of *E. cloacae* sequences are considered cluster XIII. However, the taxonomy of this genus is still under debate (4, 5). In fact, *Enterobacter aerogenes* has been recently reclassified into the *Klebsiella* genus as *Klebsiella aerogenes* (6). A more comprehensive study based on whole-genome sequencing (WGS) data from ECC isolates yielded a redistribution of the species defined by *hsp60* sequencing (5) into different clades (7) and allowed the characterization of new ECC species (8).

Discrimination of the ECC at the species level is usually performed by sequence-based methods. The most commonly targeted gene is *hsp60*, although multilocus sequence typing (MLST) and WGS have also been applied (5, 9, 10). Sequence-based diagnostic methods are laborious and require specific equipment. Therefore, new emerging techniques such as matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) have been proposed as alternatives to sequence-based methods. MALDI-TOF MS has been shown to be an excellent methodology for bacterial identification. It can easily identify ECC isolates, but it showed low discrimination power for the species in this group when using standard analyses and commercial databases with low resolution (11, 12).

This study aimed to develop and validate prediction models for automatic species differentiation within the ECC using MALDI-TOF MS and supervised learning algorithms. This task is important because of the diverse implications of ECC species in human pathologies and their involvement in nosocomial outbreaks (4). In addition, *E. hormaechei*, the ECC species most commonly encountered in clinical settings, has been correlated with the enhanced acquisition of antimicrobial resistance mechanisms and the expression of virulence factors (13, 14). To achieve this goal, three steps were conducted in this study. First, we performed an unsupervised clustering to determine the feasibility of using MALDI-TOF MS data for ECC species identification. Second, we applied a supervised machine learning algorithm with isolates from University Hospital Ramón y Cajal (UHRC) (Madrid, Spain) and validated our findings with different ECC isolates from the same hospital and from the University Hospital Basel (UHB) (Basel, Switzerland). Finally, the developed model was validated in the participating centers by the analysis of a batch of 23 ECC isolates.

MATERIALS AND METHODS

Bacterial isolates. Overall, we analyzed 219 clinical isolates belonging to the ECC and 118 *K. aerogenes* (formerly *Enterobacter aerogenes*) isolates. Among them, 164 ECC isolates and 9 *K. aerogenes* isolates were collected in a surveillance study of antimicrobial resistance in the UHRC (Madrid, Spain) between 2005 and 2018 and were identified by partial sequencing of the *hsp60* gene (15). A second set of 141 isolates (34 ECC isolates and 107 *K. aerogenes* isolates) were collected at the UHB (Basel, Switzerland) between 2016 and 2021 and were identified by WGS using KmerFinder v3.2 (16–18). MALDI-TOF MS profiles of these isolates were obtained at the UHB and submitted to the Hospital General Universitario Gregorio Marañón (HGM) (Madrid, Spain) for further analysis. Finally, a batch of 23 isolates (21 ECC isolates and 2 *K. aerogenes* isolates) were collected at the HGM in 2022 for validation purposes.

All isolates from UHRC and HGM were incubated overnight at 37°C and metabolically activated after three subcultures on Columbia blood agar (bioMérieux, Marcy l’Etoile, France) before their analysis with MALDI-TOF MS at the HGM.

TABLE 1 Number of ECC isolates used for the unsupervised feasibility study and the supervised model development

Species	No. of balanced samples in unsupervised study	No. of isolates in supervised study			
		Development dataset (UHRC)	Validation dataset (UHRC)	External validation dataset (UHB)	Validation dataset (HGM)
<i>K. aerogenes</i>	18	18	3	107	2
<i>E. asburiae</i>	18	18	1	0	3
<i>E. bugandensis</i>	18	0	0	0	0
<i>E. hormaechei</i>	18	18	51	33	15
<i>E. kobei</i>	18	18	9	0	2
<i>E. ludwigii</i>	18	0	0	0	0
<i>E. roggkampii</i>	18	18	62	1	1
Total	126	90	126	141	23

Spectrum acquisition using MALDI-TOF MS. We identified the isolates using the MALDI Biotyper^{smart} (Bruker Daltonics, Bremen, Germany). We spotted all strains from UHRC in duplicate onto the MALDI target plate and overlaid 1 μ L of 70% formic acid. After drying at room temperature, we covered and dried the spots with 1 μ L α -cyano-4-hydroxycinnamic acid (HCCA) matrix, according to the manufacturer's instructions (Bruker Daltonics). Each spot was read twice in the range of 2,000 to 20,000 Da, resulting in 4 composite spectra per isolate. The isolates from UHB were analyzed in the daily routine; therefore, 1 spot per strain was analyzed, and 1 composite spectrum from the spot was obtained.

Data processing of MALDI-TOF MS protein spectra and development of predictive models. For both feasibility and supervised studies, we processed all MALDI-TOF MS profiles with the Clover MS data analysis software (Clover BioSoft, Granada, Spain). We applied a preprocessing pipeline, which consisted of (i) smoothing (Savitzky-Golay filter: window length=11, polynomial order=3, and baseline subtraction; top-hat filter method: factor=0.02); (ii) creation of an average spectrum per isolate; (iii) alignment of the average spectra from different isolates (shift: medium; constant tolerance: 2 Da; linear mass tolerance: 600 ppm); and (iv) normalization by total ion current (TIC), to all protein spectra.

(i) Unsupervised feasibility study. To study the feasibility of using MALDI-TOF MS for differentiation of ECC species, we proposed an unsupervised study based on principal-component analysis (PCA) and *t*-distributed stochastic neighbor embedding (*t*-SNE). For this purpose, an oversampled balanced data set for each ECC species was used. We included a total of 126 spectra from the 7 ECC species analyzed in this study (sourced from UHRC and UHB), as indicated in Table 1.

(ii) Supervised model development. Once the feasibility of the study was determined, we proposed the supervised model development. In this case, three different data sets were created, i.e., a training validation set, an internal validation data set, and an external validation data set. The details of these data sets are shown in Table 1.

Due to the lack of validation samples for *E. ludwigii* and *E. bugandensis*, these species were not included in the development of the supervised model. Therefore, our supervised model was developed to predict four ECC species, namely, *E. asburiae* (cluster I), *E. kobei* (cluster II), *E. hormaechei* (clusters III, VI, and VIII considered together), and *E. roggkampii* (cluster IV). We applied four different supervised models, i.e., partial least-squares discriminant analysis (PLS-DA), support vector machine (SVM) with linear (SVM-L) kernel and SVM with radial basis function (SVM-R) kernel, and random forest (RF). The hyperparameter selection was performed with a 5-fold cross-validation technique.

Finally, we performed two external validations of the predictive models. First, 126 isolates from UHRC and 141 isolates from UHB were blindly classified by the same predictive models using Clover BioSoft software v0.6.1. Later, a batch of 23 isolates from HGM were sent to UHRC and UHB for external validation and study of the reproducibility of the developed models. The software applied uses the *scikit-learn* v0.23.2 Python library to implement all statistical methods used in this study. For reproducibility purposes under findability, accessibility, interoperability, and reusability (FAIR) principles, free access to all spectra to reproduce the analyses developed in this study is provided.

(iii) Evaluation of the MSI database. Recently, an online database was developed for the rapid differentiation of ECC species based on their MALDI-TOF MS protein profiles (19). This database has free access (<https://msi.happy-dev.fr>) and has been built using protein spectra from 42 ECC isolates characterized by sequencing of the *hsp60* gene. This identification method is considered the state-of-the-art method for the identification of ECC isolates at the species level. Therefore, both external validation data sets were also identified using the Mass Spectrometric Identification (MSI) database as a comparison to the methods proposed in this article. As stated above, MALDI-TOF MS spectra associated with this study have been made publicly available.

Ethics statement. The Ethics Committee of the HGM evaluated this project and considered that all of the conditions for waiving informed consent were met since the study was conducted with microbiological samples and not with human products. At the UHB, only anonymized data were used for the purpose of quality control and assay validation; according to the Swiss Human Research Act, no specific consent is required in such cases. Data either were acquired in routine microbiological diagnostics (excluding cases with a rejected general consent) or were used from a previously published data set (Database of Resistance Information on Antimicrobials and MALDI-TOF Mass Spectra [DRIAMS]).

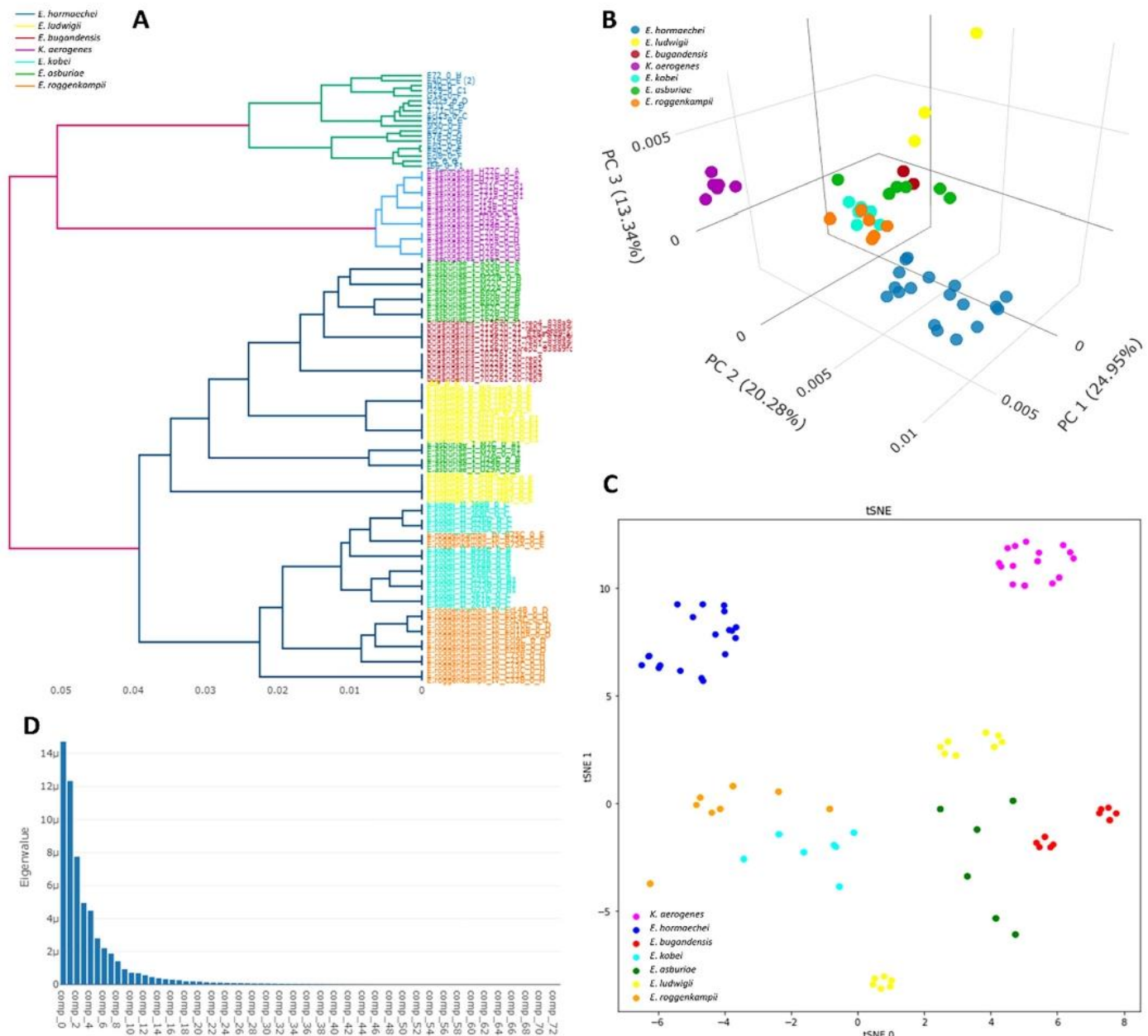


FIG 1 Comparative analysis of 126 MALDI-TOF MS spectra from *K. aerogenes* and different ECC species using unsupervised methods. (A) Dendrogram built with 126 spectra using Euclidean distance and Ward metric. Spectra from *K. aerogenes* (purple) and *E. hormaechei* (dark blue) are completely separated from the other species. (B) PCA of feasibility study spectra. *K. aerogenes* and *E. hormaechei* are completely separated. Visualization in three dimensions was needed for separation of *E. ludwigii* (yellow). (C) t-SNE analysis of study spectra. (D) PCA eigenvalues showing the variance of each component. A total of 14 components were needed to reach 95% of the total data variance.

Data availability. For reproducibility purposes under FAIR principles, all spectra to reproduce the analyses developed in this study are available in the Clover Repository (<https://platform.clovermsdataanalysis.com/repository/collection/EHGM001>).

RESULTS

Feasibility study. To prove the feasibility of using MALDI-TOF MS to differentiate ECC species, an unsupervised hierarchical clustering with PCA and t-SNE preprocessing was performed (Fig. 1). The protein spectra of the seven species (*E. asburiae*, *E. kobei*, *E. hormaechei*, *E. roggenkampii*, *E. ludwigii*, *E. bugandensis*, and *K. aerogenes*), which are equally represented in the model, were compared. The dendrogram built with these data showed three main clusters, one cluster containing *E. hormaechei*, a second

TABLE 2 Accuracy results for internal 5-fold cross-validation over development data set (90 spectral profiles)

Algorithm	No. identified/total no. (%)				
	<i>E. asburiae</i>	<i>E. hormaechei</i>	<i>E. kobei</i>	<i>E. roggenkampii</i>	<i>K. aerogenes</i>
PLS-DA	9/18 (50)	18/18 (100)	5/18 (27.8)	4/18 (22.2)	18/18 (100)
SVM-L	6/18 (33.3)	18/18 (100)	3/18 (16.7)	6/18 (33.3)	14/18 (77.8)
SVM-R	18/18 (100)	18/18 (100)	18/18 (100)	18/18 (100)	18/18 (100)
RF	18/18 (100)	18/18 (100)	18/18 (100)	18/18 (100)	18/18 (100)

cluster with *K. aerogenes*, and a third cluster with the rest of the species. Inside the latter cluster, *E. bugandensis* strains were clustered together and so were *E. asburiae*, *E. ludwigii*, *E. kobei*, and *E. roggenkampii*, although in those four cases some of the spectra were clustered with the wrong species (Fig. 1A to C).

The implementation of PCA to reduce the dimensionality showed that 14 components were needed to explain 95% of the variance (Fig. 1D). This fact and the relatively accurate classification of ECC species using an unsupervised algorithm demonstrated the potentiality of MALDI-TOF MS to differentiate ECC species.

Supervised models based on MALDI-TOF MS. To address the limitations of unsupervised learning, we added the label knowledge to the training phase by using supervised algorithms such as PLS-DA, SVM, and RF. We trained these models using the development data set shown in Table 1 and selected their hyperparameters by a 5-fold cross-validation technique. This cross-validation process led to the next hyperparameter selection; 2 components were used for PLS-DA, the value of *C* was 10 for SVM-L, and the value of *C* was 10 and the value of γ was 1,000 for SVM-R. Table 2 shows the results obtained for the internal 5-fold cross-validation, which have been further detailed in Table S1 in the supplemental material.

E. hormaechei and *K. aerogenes* presented the same trend as in the feasibility study, and their differentiation was 100% using nonlinear approaches (SVM-R and RF). Only the implementation of a linear approach (SVM-L) yielded lower results for *K. aerogenes* (Table 2). For the rest of the analyzed ECC species, we also obtained 100% correct classification by the application of nonlinear approaches (Table 2).

In Fig. 2, the distance between samples calculated by the RF classifier is shown. We detected a unique cluster for each species. Due to the results presented in Table 2, only SVM-R and RF were considered for further analysis. Table 3 shows the results of SVM-R and RF for the validation data set collected at UHRC and UHB.

Both algorithms, SVM-R and RF, yielded the same results in the external validation performed on the MALDI-TOF MS spectra from the validation data set sourced from the same hospital (UHRC). In that case, all *K. aerogenes*, *E. asburiae*, and *E. kobei* isolates were correctly classified, whereas 1 *E. hormaechei* strain was misclassified as *E. kobei* with both algorithms. For *E. roggenkampii*, 2 isolates were misclassified as *E. hormaechei* and 1 as *E. kobei* (see Table S2). The accuracy of the model is shown in Fig. 3A and B.

Since the SVM-R and RF algorithms performed equally, both of them were considered for external validation with MALDI-TOF MS profiles obtained at the UHB. In this case, 91.2% of the *E. hormaechei* isolates ($n = 33$), 100% of the *E. roggenkampii* isolates ($n = 1$), and 98.1% of the *K. aerogenes* isolates were correctly classified by RF, as shown in Table 3. The application of SVM-R yielded lower results for *E. hormaechei* and *K. aerogenes*. Figure 3C and D show the accuracy of both SVM-R and RF for the external validation collection from UHB.

Finally, for the 23 isolates from HGM used for the reproducibility study, 100% correct identification was obtained in all centers using the RF algorithm (Table 4). In the case of the SVM-R algorithm, 100% correct classification was obtained for strains analyzed in UHRC and UHB, whereas 1 strain of *E. kobei* was misclassified as *E. roggenkampii* in HGM.

Identification of the ECC isolates using the MSI database. The MSI platform was also used as an identification tool for ECC species, to compare the automatic approach developed in this study with the current state-of-the-art method (19). Among the UHRC isolates, the identification rates for *E. asburiae*, *E. hormaechei*, and *E. kobei* were

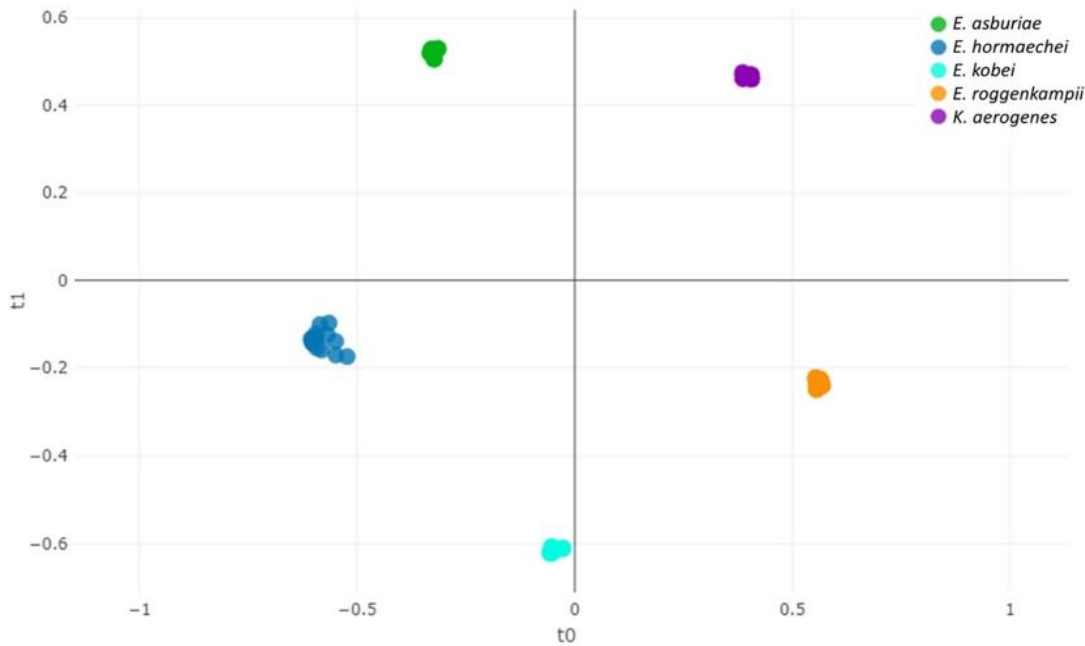


FIG 2 RF distance plot using Euclidean distance for MALDI-TOF MS spectra from the species included in the classification model.

similar to the rates yielded by the predictive models developed in this study (Table 3). The only difference detected between the two methods was that the MSI database correctly classified 1 more *E. roggenkampii* isolate. As for the protein spectra sourced from the UHB, 84.8% of the *E. hormaechei* isolates and 100% of the *E. roggenkampii* isolates ($n = 1$) were correctly identified with the MSI platform. In this case, the MSI platform provided a lower rate of correct identifications for *E. hormaechei* than the RF algorithm (Table 3).

DISCUSSION

In this study, the implementation of supervised, nonlinear algorithms (SVM-R and RF) with MALDI-TOF MS spectra allowed the correct species assignment for 100% of isolates belonging to two ECC species (*E. asburiae* and *E. kobei*) and between 91.2% and 98.1% for *E. hormaechei*, *E. roggenkampii*, and *K. aerogenes* (formerly *E. aerogenes*) sourced from three different hospitals.

Poor discrimination of ECC species by MALDI-TOF MS using either commercial (11, 15) or enriched, in-house (20) databases was reported previously. However, a recent study from a research group with broad experience in MALDI-TOF MS and the creation of expanded libraries reported 92.0% correct species-level identification by implementing a specific in-house library enriched with well-characterized ECC strains, with correct discrimination of

TABLE 3 Accuracy results for the validation data set from UHRC and UHB and the identification accuracy obtained with the MSI database^a

Center and algorithm	No. identified/total no. (%)				
	<i>E. asburiae</i>	<i>E. hormaechei</i>	<i>E. kobei</i>	<i>E. roggenkampii</i>	<i>K. aerogenes</i>
UHRC					
SVM-R	1/1 (100)	50/51 (98.0)	9/9 (100)	59/62 (95.2)	3/3 (100)
RF	1/1 (100)	50/51 (98.0)	9/9 (100)	59/62 (95.2)	3/3 (100)
MSI	1/1 (100)	50/51 (98.0)	9/9 (100)	60/62 (96.8)	NA
UHB					
SVM-R	NA	15/33 (44.1)	NA	1/1 (100)	102/107 (95.3)
RF	NA	31/33 (91.2)	NA	1/1 (100)	105/107 (98.1)
MSI	NA	28/33 (84.8)	NA	1/1 (100)	NA

^aNA, not applicable.

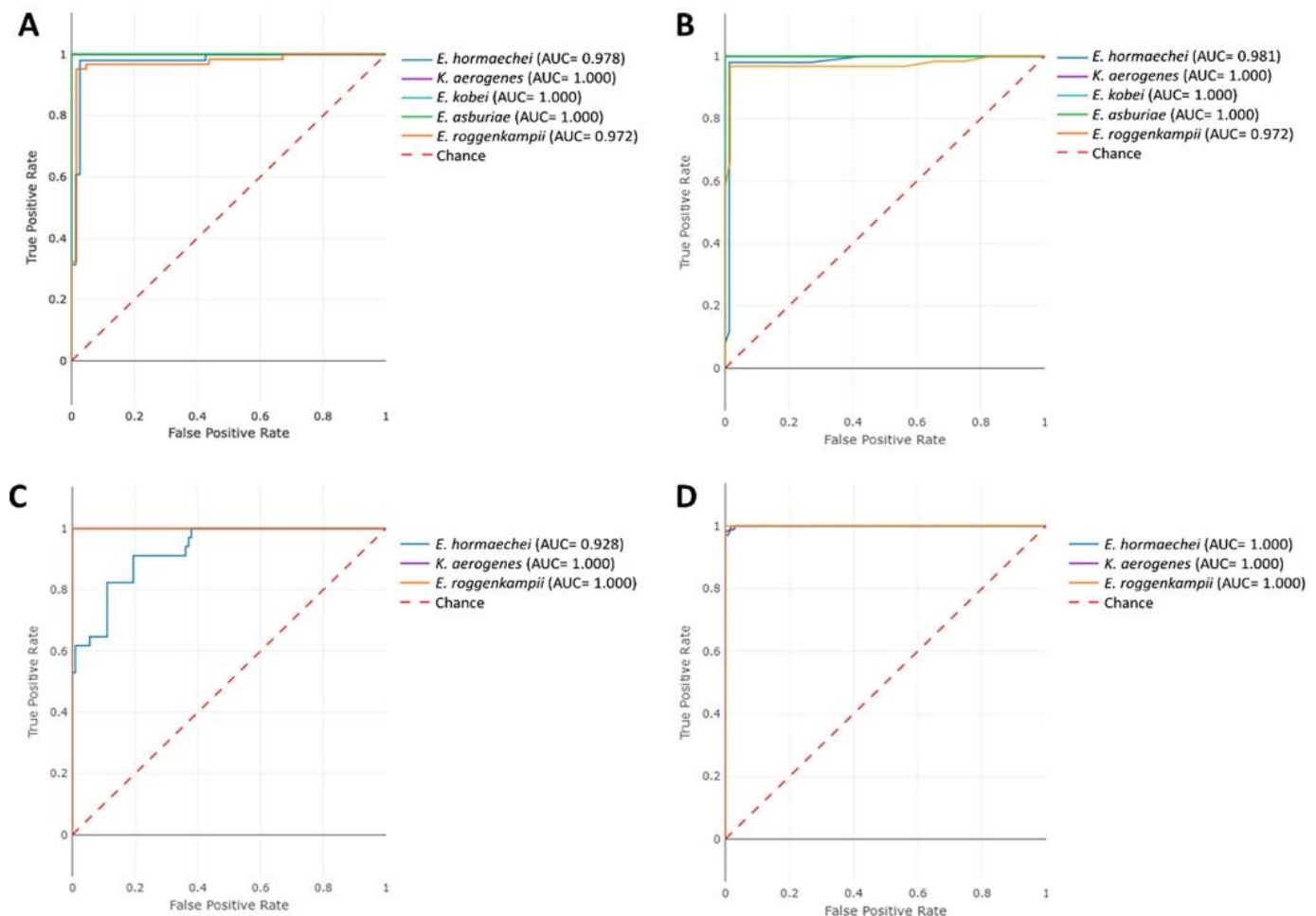


FIG 3 Receiver operating characteristic curve (ROC) and area under the curve (AUC) values obtained with different supervised algorithms for the two participating centers. (A) SVM-R model validated with UHRC strains. (B) RF model validated with UHRC strains. (C) SVM-R model validated with UHB strains. (D) RF model validated with UHB strains.

97.0% of *E. hormaechei* isolates (19). This approach can be useful for the discrimination of closely related species, but the construction of a database is cumbersome and requires highly trained personnel. The implementation of the MSI platform allowed 94.9% correct species-level identification of 155 ECC protein spectra in this study. This rate was slightly lower than the obtained with the nonlinear algorithms proposed by our approach.

In this study, we demonstrate the feasibility of using MALDI-TOF MS to identify species within the ECC. First, hierarchical clustering showed that it is possible to differentiate between species using the information contained in MALDI-TOF MS spectra, as reported previously (20). Second, a supervised study using machine learning algorithms yielded the correct classification of all ECC species. Therefore, different supervised classification algorithms were implemented to correctly provide species assignment of ECC species. The internal validation experiment demonstrated that nonlinear approaches, such as SVM-R or RF, were needed to correctly identify all species. Both models perfectly classified all samples in internal cross-validation.

To further demonstrate that the model can perform in different scenarios with data different from the spectral profiles used for model training, we performed three validation experiments. First, we carried out a validation with MALDI-TOF MS protein spectra sourced from UHRC. Of a total of 126 samples, both SVM-R and RF assigned only 4 isolates to species different than those assigned by molecular techniques, i.e., 96.5% accuracy in classifying species within the ECC was obtained. Second, we performed an external validation with MALDI-TOF MS spectra sourced from UHB to simulate a real-

TABLE 4 Reproducibility study of 23 isolates from HGM identified in parallel in the three participating centers, UHRC, UHB, and HGM

Algorithm and species	No. identified/total no. (%)		
	UHRC	UHB	HGM
SVM-R			
<i>E. asburiae</i>	3/3 (100)	3/3 (100)	3/3 (100)
<i>E. hormaechei</i>	15/15 (100)	15/15 (100)	15/15 (100)
<i>E. kobei</i>	2/2 (100)	2/2 (100)	1/2 (50)
<i>E. roggkampii</i>	1/1 (100)	1/1 (100)	1/1 (100)
<i>K. aerogenes</i>	2/2 (100)	2/2 (100)	2/2 (100)
SVM-R total	23/23 (100)	23/23 (100)	22/23 (95.6)
RF			
<i>E. asburiae</i>	3/3 (100)	3/3 (100)	3/3 (100)
<i>E. hormaechei</i>	15/15 (100)	15/15 (100)	15/15 (100)
<i>E. kobei</i>	2/2 (100)	2/2 (100)	2/2 (100)
<i>E. roggkampii</i>	1/1 (100)	1/1 (100)	1/1 (100)
<i>K. aerogenes</i>	2/2 (100)	2/2 (100)	2/2 (100)
RF total	23/23 (100)	23/23 (100)	23/23 (100)

world scenario. These MALDI-TOF MS protein spectra originated in a different epidemiological scenario and were processed by operators from the UHB. In this case, SVM-R was shown to be overfitted to the UHRC distribution, which was already pointed out by the γ value, with 83.7% accuracy. Finally, we validated the models by analyzing the same batch of 23 isolates in the three centers. In this case, 100% correct species-level classification was achieved using the RF algorithm. These results show that the models have not been overfitted to the spectra sourced from only one center. This phenomenon occurs when a machine learning model has learned the training data too well and does not recognize new data with small variations, yielding a lower accuracy than expected. The current state-of-the-art tool, i.e., the MSI database, performed better than SVM-R, with 94.9% accuracy, although it was not able to distinguish *K. aerogenes* (19). However, RF outperformed both approaches, with >96.0% accuracy in identifying the three species. Therefore, it is demonstrated that supervised machine learning algorithms are feasible, and indeed applicable, in microbiology laboratories.

One limitation of this study was the fact that all UHRC isolates were carbapenemase-producing isolates, because this was the source of the previously analyzed collection (15). However, the present study provides the first proof of concept for differentiating ECC species based on machine learning. For definitive validation, improvement, and implementation of these predictive models, future studies will involve strains with more diverse epidemiological and geographical origins and characteristics. In addition, not all analyzed ECC species could be represented in the external validation data set due to the lack of isolates from the species *E. ludwigii* and *E. bugandensis*.

This study shows promising results for differentiating ECC species based on machine learning and MALDI-TOF MS protein spectra. It also highlights the fact that MALDI-TOF MS data should be linked to WGS data in order to allow future work and to provide a reference standard. The MALDI-TOF MS and machine learning approach has been demonstrated to be a rapid and cost-effective method, suitable for correct species-level assignment of closely related species, as in the case of ECC species. The use of spectrum analysis tools is becoming user-friendly and easy to apply, and their use may provide species-level identification in a fast and inexpensive way. Once the model is validated with a comprehensive number of ECC species, an open Web application will be deployed to be used freely by the community.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.4 MB.

ACKNOWLEDGMENTS

This work was supported by projects PI15/01073 and PI18/00997 from the Health Research Fund (FIS) (Instituto de Salud Carlos III, Plan Nacional de I+D+I 2013 to 2016) of the Carlos III Health Institute (ISCIII) (Madrid, Spain), partially financed by the European Regional Development Fund (FEDER) program A Way of Making Europe. M. Mateos was funded by the Community of Madrid (Programa de Garantía Juvenil, grant PEJD-2017-PRE/BMD-5106) and D. Rodríguez-Temporal by a postdoctoral contract from the Intramural Funding Program of the Institute of Health Research Gregorio Marañón (IISGM). B. Rodríguez-Sánchez is a recipient of a Miguel Servet contract supported by the FIS (grant CPII19/00002). A. Guerrero-López was funded by a predoctoral contract from the Intramural Funding Program of the IISGM.

A. Candela: experimental part, formal analysis, data collection, validation, visualization, and writing (original draft preparation and review/editing); A. Guerrero-López: formal analysis, data collection, validation, visualization, and writing (original draft preparation and review/editing); M. Mateos and A. Gómez-Asenjo: experimental part, formal analysis, and data collection; M. J. Arroyo, G. Méndez, and L. Mancera: data analysis, validation, and writing (original draft preparation and review/editing); M. Hernández-García and R. del Campo: experimental part, formal analysis, submission of isolates, and writing (original draft preparation and review/editing); E. Cercenado: conceptualization, formal analysis, validation, and writing (review/editing); A. Cuénod and A. Egli: submission of isolates and writing (original draft preparation and review/editing); J. D. Caballero, L. Martínez-García, D. Gijón, M. I. Morosini, P. Ruiz-Garbajosa, R. Cantón, and P. Muñoz: validation and writing (review/editing); D. Rodríguez-Temporal: conceptualization, formal analysis, validation, and writing (original draft preparation and review/editing); B. Rodríguez-Sánchez: conceptualization, project administration, formal analysis, supervision, validation, visualization, and writing (original draft preparation and review/editing).

We have no conflict of interests.

REFERENCES

- Mezzatesta ML, Gona F, Stefani S. 2012. *Enterobacter cloacae* complex: clinical impact and emerging antibiotic resistance. *Future Microbiol* 7: 887–902. <https://doi.org/10.2217/fmb.12.61>.
- Akbari M, Bakhshi B, Najar Peerayah S. 2016. Particular distribution of *Enterobacter cloacae* strains isolated from urinary tract infection within clonal complexes. *Iran Biomed J* 20:49–55. <https://doi.org/10.7508/ibj.2016.01.007>.
- Kremer A, Hoffmann H. 2012. Prevalences of the *Enterobacter cloacae* complex and its phylogenetic derivatives in the nosocomial environment. *Eur J Clin Microbiol Infect Dis* 31:2951–2955. <https://doi.org/10.1007/s10096-012-1646-2>.
- Davin-Regli A, Lavigne JP, Pages JM. 2019. *Enterobacter* spp.: update on taxonomy, clinical aspects, and emerging antimicrobial resistance. *Clin Microbiol Rev* 32:e00002-19. <https://doi.org/10.1128/CMR.00002-19>.
- Hoffmann H, Roggenkamp A. 2003. Population genetics of the nomenclature *Enterobacter cloacae*. *Appl Environ Microbiol* 69:5306–5318. <https://doi.org/10.1128/AEM.69.9.5306-5318.2003>.
- Tindall BJ, Sutton G, Garrity GM. 2017. *Enterobacter aerogenes* Hormaeche and Edwards 1960 (Approved Lists 1980) and *Klebsiella mobilis* Bascomb et al. 1971 (Approved Lists 1980) share the same nomenclature type (ATCC 13048) on the Approved Lists and are homotypic synonyms, with consequences for the name *Klebsiella mobilis* Bascomb et al. 1971 (Approved Lists 1980). *Int J Syst Evol Microbiol* 67:502–504. <https://doi.org/10.1099/ijsem.0.001572>.
- Sutton GG, Brinkac LM, Clarke TH, Fouts DE. 2018. *Enterobacter hormaechei* subsp. *hoffmannii* subsp. nov., *Enterobacter hormaechei* subsp. *xiangfangensis* comb. nov., *Enterobacter roggkampii* sp. nov., and *Enterobacter muelleri* is a later heterotypic synonym of *Enterobacter asburiae* based on computational analysis of sequenced *Enterobacter* genomes. *F1000Res* 7:521. <https://doi.org/10.12688/f1000research.14566.2>.
- Wu W, Feng Y, Zong Z. 2019. Characterization of a strain representing a new *Enterobacter* species, *Enterobacter chengduensis* sp. nov. *Antonie Van Leeuwenhoek* 112:491–500. <https://doi.org/10.1007/s10482-018-1180-z>.
- Singh NK, Bezdan D, Checinska Sielaff A, Wheeler K, Mason CE, Venkateswaran K. 2018. Multi-drug resistant *Enterobacter bugandensis* species isolated from the International Space Station and comparative genomic analyses with human pathogenic strains. *BMC Microbiol* 18:175. <https://doi.org/10.1186/s12866-018-1325-2>.
- Hoffmann H, Stindl S, Ludwig W, Stumpf A, Mehlen A, Heesemann J, Monget D, Schleifer KH, Roggenkamp A. 2005. Reassignment of *Enterobacter dissolvens* to *Enterobacter cloacae* as *E. cloacae* subspecies *dissolvens* comb. nov. and emended description of *Enterobacter asburiae* and *Enterobacter kobei*. *Syst Appl Microbiol* 28:196–205. <https://doi.org/10.1016/j.syapm.2004.12.010>.
- Pavlovic M, Konrad R, Iwobi AN, Sing A, Busch U, Huber I. 2012. A dual approach employing MALDI-TOF MS and real-time PCR for fast species identification within the *Enterobacter cloacae* complex. *FEMS Microbiol Lett* 328:46–53. <https://doi.org/10.1111/j.1574-6968.2011.02479.x>.
- De Florio L, Riva E, Giona A, Dedej E, Fogolari M, Cella E, Spoto S, Lai A, Zehender G, Ciccozzi M, Angeletti S. 2018. MALDI-TOF MS identification and clustering applied to *Enterobacter* species in nosocomial setting. *Front Microbiol* 9:1885. <https://doi.org/10.3389/fmicb.2018.01885>.
- Barnes AI, Paraje MG, Battán PDC, Albesa I. 2001. Molecular properties and metabolic effect on blood cells produced by a new toxin of *Enterobacter cloacae*. *Cell Biol Toxicol* 17:409–418. <https://doi.org/10.1023/A:1013704801570>.
- Paaauw A, Caspers MPM, Leverstein-van Hall MA, Schuren FHJ, Montijn RC, Verhoef J, Fluit AC. 2009. Identification of resistance and virulence factors in an epidemic *Enterobacter hormaechei* outbreak strain. *Microbiology (Reading)* 155:1478–1488. <https://doi.org/10.1099/mic.0.024828-0>.
- Mateos M, Hernandez-Garcia M, Del Campo R, Martinez-Garcia L, Gijon D, Morosini MI, Ruiz-Garbajosa P, Canton R. 2020. Emergence and persistence over time of carbapenemase-producing *Enterobacter* isolates in a Spanish university hospital in Madrid, Spain (2005–2018). *Microb Drug Resist* 27:895–903. <https://doi.org/10.1089/mdr.2020.0265>.
- Hasman H, Saputra D, Sicheritz-Ponten T, Lund O, Svendsen CA, Frimodt-Moller N, Aarestrup FM. 2014. Rapid whole-genome sequencing for detection and characterization of microorganisms directly

- from clinical samples. *J Clin Microbiol* 52:139–146. <https://doi.org/10.1128/JCM.02452-13>.
17. Larsen MV, Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H, Sicheritz-Ponten T, Aarestrup FM, Ussey DW, Lund O. 2014. Benchmarking of methods for genomic taxonomy. *J Clin Microbiol* 52:1529–1539. <https://doi.org/10.1128/JCM.02981-13>.
 18. Clausen P, Aarestrup FM, Lund O. 2018. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* 19:307. <https://doi.org/10.1186/s12859-018-2336-6>.
 19. Godmer A, Benzerara Y, Normand AC, Veziris N, Gallah S, Eckert C, Morand P, Piarroux R, Aubry A. 2021. Revisiting species identification within the *Enterobacter cloacae* complex by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Microbiol Spectr* 9:e00661-21. <https://doi.org/10.1128/Spectrum.00661-21>.
 20. Wang YQ, Xiao D, Li J, Zhang HF, Fu BQ, Wang XL, Ai XM, Xiong YW, Zhang JZ, Ye CY. 2018. Rapid identification and subtyping of *Enterobacter cloacae* clinical isolates using peptide mass fingerprinting. *Biomed Environ Sci* 31:48–56. <https://doi.org/10.3967/bes2018.005>.

MATERIAL SUPLEMENTARIO CAPÍTULO 2

MATERIAL SUPLEMENTARIO CAPÍTULO 2

Table S1. Accuracy of internal 5-fold cross-validation over development collection for all algorithms tested.

Actual/Predicted	<i>E. asburiae</i>	<i>E. hormaechei</i>	<i>E. kobei</i>	<i>E. roggenkampii</i>	<i>K. aerogenes</i>	% correct
PLS-DA						
<i>E. asburiae</i>	9	0	4	2	3	50%
<i>E. hormaechei</i>	0	18	0	0	0	100%
<i>E. kobei</i>	6	0	5	7	0	27.8%
<i>E. roggenkampii</i>	7	0	7	4	0	22.2%
<i>K. aerogenes</i>	0	0	0	0	18	100%
Total PLS-DA						60.4%
SVM-L						
<i>E. asburiae</i>	6	4	2	3	3	33.3%
<i>E. hormaechei</i>	0	18	0	0	0	100%
<i>E. kobei</i>	6	3	3	6	0	16.7%
<i>E. roggenkampii</i>	4	4	4	6	0	33.3%
<i>K. aerogenes</i>	0	4	0	0	14	77.8%
Total SVM-L						52.7%
SVM-R						
<i>E. asburiae</i>	18	0	0	0	0	100%
<i>E. hormaechei</i>	0	18	0	0	0	100%
<i>E. kobei</i>	0	0	18	0	0	100%
<i>E. roggenkampii</i>	0	0	0	18	0	100%
<i>K. aerogenes</i>	0	0	0	0	18	100%
Total SVM-R						100%
RF						
<i>E. asburiae</i>	18	0	0	0	0	100%
<i>E. hormaechei</i>	0	18	0	0	0	100%
<i>E. kobei</i>	0	0	18	0	0	100%
<i>E. roggenkampii</i>	0	0	0	18	0	100%
<i>K. aerogenes</i>	0	0	0	0	18	100%
Total RF						100%

MATERIAL SUPLEMENTARIO CAPÍTULO 2

Table S2. Accuracy of validation collection from UHRC, UHB strains and MSI database.

Actual/Predicted	<i>E. asburiae</i>	<i>E. hormaechei</i>	<i>E. kobei</i>	<i>E. roggenkampii</i>	<i>K. aerogenes</i>	<i>E. bugandensis</i>	% correct
SVM-R UHRC							
<i>E. asburiae</i>	1	0	0	0	0	-	100%
<i>E. hormaechei</i>	0	50	1	0	0	-	98%
<i>E. kobei</i>	0	0	9	0	0	-	100%
<i>E. roggenkampii</i>	0	2	1	59	0	-	95.2%
<i>K. aerogenes</i>	0	0	0	0	3	-	100%
Total SVM-R							96.8%
RF UHRC							
<i>E. asburiae</i>	1	0	0	0	0	-	100%
<i>E. hormaechei</i>	0	50	1	0	0	-	98%
<i>E. kobei</i>	0	0	9	0	0	-	100%
<i>E. roggenkampii</i>	0	2	1	59	0	-	95.2%
<i>K. aerogenes</i>	0	0	0	0	3	-	100%
Total RF							96.8%
SVM-R UHB							
<i>E. hormaechei</i>	5	15	14	5	0	-	44.1%
<i>E. roggenkampii</i>	0	0	0	1	0	-	100%
<i>K. aerogenes</i>	5	0	0	0	102	-	95.3%
Total SVM-R							83.1%
RF UHB							
<i>E. hormaechei</i>	1	31	2	0	0	-	91.2%
<i>E. roggenkampii</i>	0	0	0	1	0	-	100%
<i>K. aerogenes</i>	0	0	2	0	105	-	98.1%
Total RF							96.5%
MSI							
<i>E. asburiae</i>	1	0	0	0	0	0	100%
<i>E. hormaechei</i>	0	78	0	4	0	2	92.9%
<i>E. kobei</i>	0	0	9	0	0	0	100%
<i>E. roggenkampii</i>	0	2	0	61	0	0	96.8%
Total MSI							94.9%

Capítulo 3. Discriminación rápida de aislados de *Pseudomonas aeruginosa* ST175 involucrados en un brote nosocomial empleando espectrometría de masas MALDI-TOF y espectroscopía FTIR combinados con *Machine Learning*

Contexto

El objetivo de este estudio fue evaluar la espectrometría de masas MALDI-TOF y la espectroscopia infrarroja de FTIR como alternativas de diagnóstico a los métodos basados en ADN para la detección de aislados *P. aeruginosa* ST175 involucrados en un brote hospitalario en el Hospital General Universitario Gregorio Marañón.

Material y métodos

Para ello, se analizaron mediante las metodologías mencionadas 27 aislados de *P. aeruginosa* procedentes de un brote detectado en el servicio de Hematología de nuestro hospital durante los años 2014-2015. Previamente, estos aislados se habían caracterizado mediante electroforesis en gel de campo pulsado (PFGE) y secuenciación de genoma completo (WGS). Además, también se analizaron 8 aislados de *P. aeruginosa* como controles no relacionados con el brote.

Los espectros de masas se adquirieron transfiriendo las colonias al pocillo de MALDI y cubriéndolas con 1µl de ácido fórmico y 1µl de matriz de ácido α -ciano-3,4-hidroxicinámico (HCCA). Para el análisis con FTIR, las colonias se resuspendieron en etanol al 70% y agua estéril siguiendo las instrucciones del fabricante.

Los espectros de ambas metodologías se analizaron empleando el software Clover MSDAS, que permitió el análisis de los datos mediante diferentes algoritmos y la validación de los modelos de clasificación.

Resultados

Se encontraron tres picos biomarcadores específicos de brote (5169, 6915 y 7236 m/z) en los espectros analizados. Los modelos de clasificación basados en estos tres biomarcadores mostraron el mismo poder de discriminación que muestra el PFGE. Por otro lado, la tecnología FTIR mostró un poder de discriminación similar al PFGE y también alcanzó una correcta discriminación de los diferentes ST analizados.

Conclusiones

La combinación de ambas tecnologías evaluadas, junto con herramientas de *Machine Learning*, puede suponer una poderosa herramienta para el seguimiento en tiempo real de clones y aislados de alto riesgo involucrados en brotes nosocomiales.

Research Article

Rapid Discrimination of *Pseudomonas aeruginosa* ST175 Isolates Involved in a Nosocomial Outbreak Using MALDI-TOF Mass Spectrometry and FTIR Spectroscopy Coupled with Machine Learning

Ana Candela ^{1,2}, Manuel J. Arroyo ³, María Sánchez-Cueto ^{1,2}, Mercedes Marín ^{1,2,4,5}, Emilia Cercenado ^{1,2,4,5}, Gema Méndez ³, Patricia Muñoz ^{1,2,4,5}, Luis Mancera ³, David Rodríguez-Temporal ^{1,2} and Belén Rodríguez-Sánchez ^{1,2}

¹Clinical Microbiology and Infectious Diseases Department, Hospital General Universitario Gregorio Marañón, Madrid, Spain

²Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain

³Clover Bioanalytical Software, Av. del Conocimiento, 41, Granada 18016, Spain

⁴CIBER de Enfermedades Respiratorias (CIBERES CB06/06/0058), Madrid, Spain

⁵Medicine Department, Faculty of Medicine, Universidad Complutense de Madrid, Madrid, Spain

Correspondence should be addressed to Ana Candela; acandelagon@gmail.com and David Rodríguez-Temporal; david.rodriguez@iisgm.com

Received 1 December 2022; Revised 23 March 2023; Accepted 19 July 2023; Published 7 September 2023

Academic Editor: Mohsin Khurshid

Copyright © 2023 Ana Candela et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The goal of this study was to evaluate matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) and Fourier-transform infrared spectroscopy (FTIR-S) as diagnostic alternatives to DNA-based methods for the detection of *Pseudomonas aeruginosa* sequence type (ST) 175 isolates involved in a hospital outbreak. For this purpose, 27 *P. aeruginosa* isolates from an outbreak detected in the Hematology department of our hospital were analyzed by the above-mentioned methodologies. Previously, these isolates had been characterized by pulse-field gel electrophoresis (PFGE) and whole-genome sequencing (WGS). Besides, eight *P. aeruginosa* isolates were analyzed as unrelated controls. MALDI-TOF MS spectra were acquired by transferring several colonies onto the MALDI target and covering them with 1 μ l of formic acid 100% and 1 μ l of α -ciano-3,4-hidroxicinamic acid matrix. For the analysis with FTIR-S, colonies were resuspended in 70% ethanol and sterile water according to the manufacturer's instructions. Spectra from both methodologies were analyzed using Clover Biosoft Software, which allowed data modeling using different algorithms and validation of the classifying models. Three outbreak-specific biomarkers were found at 5,169, 6,915, and 7,236 m/z in MALDI-TOF MS spectra. Classification models based on these three biomarkers showed the same discrimination power displayed by PFGE. Besides, K -nearest neighbor algorithm allowed the discrimination of the same clusters provided by WGS and the validation of this model achieved 97.0% correct classification. On the other hand, FTIR-S showed a discrimination power similar to PFGE and reached correct discrimination of the different STs analyzed. In conclusion, the combination of both technologies evaluated, paired with machine learning tools, may represent a powerful tool for real-time monitoring of high-risk clones and isolates involved in nosocomial outbreaks.

1. Introduction

Healthcare-associated infections are becoming one of the major health concerns of the 21st century. This term groups together infections developed during and/or resulting from a hospital or nursing home stay that were not detected at the

time of admission [1]. They represent the most frequent adverse event in healthcare settings (6.5% in acute care hospitals in the European Union and 3.2% in hospitalized patients in the United States) [2, 3]. Pathogens that cause nosocomial infections can spread and cause outbreaks among inpatients and staff members, requiring control and

treatment measures and ultimately increasing resource costs in hospital settings [4]. The emergence of multidrug resistant microorganisms is another concern that arises in nosocomial outbreaks, as they pose an added complication in relation to the correct choice of antimicrobial treatment [5]. The control of nosocomial infection caused by multiresistant bacteria and outbreak surveillance programs should be implemented in hospitals in order to reduce mortality/morbidity, length of stay, and hospital costs [6].

Pseudomonas aeruginosa is one of the most frequent pathogens involved in outbreaks in hospitals and long-term care facilities [7]. This microorganism is an environmental, gram-negative, nonfermenting bacterium that can easily become a multidrug-resistant (MDR) and extensively drug-resistant pathogen through mutations in chromosomal genes in addition to its intrinsic resistance mechanisms [8]. Its wide range of virulence factors, such as its ability to produce biofilm and also its capacity to persist in moist environments, such as sinks and shower plates in hospitals, coupled with its clonality and fast spread of high-risk clones, make it an ideal candidate for being the cause of nosocomial outbreaks [9, 10].

The reference method for outbreak characterization remains pulse-field gel electrophoresis (PFGE) but it may be insufficient for clone discrimination in some cases [11]. Multilocus sequence typing (MLST) provides complementary information to PFGE but it is still laborious and time consuming [12]. Although the implementation of novel approaches such as whole-genome sequencing (WGS) would improve the follow-up of clinical outbreaks by increasing the quantity and quality of the information obtained, it requires expensive and sophisticated equipment and highly skilled personnel, making it unaffordable in most clinical laboratories nowadays.

Matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) is currently implemented in most clinical microbiology laboratories for bacterial species identification [13]. This technology has also been evaluated for purposes beyond identification, such as antimicrobial resistance detection or bacterial typing [14–16]. Therefore, MALDI-TOF MS could be a rapid and available alternative for outbreak characterization [17, 18]. In addition, Fourier-transform infrared spectroscopy (FTIR-S) has emerged as a promising new tool for bacterial typing [19]. This methodology has been recently applied to *Streptococcus pneumoniae* typing [20].

In this study, we evaluated MALDI-TOF MS and FTIR-S coupled with machine learning classification methods for the rapid detection and follow-up of a nosocomial outbreak caused by a *P. aeruginosa* high-risk clone with the same level of accuracy provided by advanced molecular methods.

2. Materials and Methods

2.1. Outbreak Description and Bacterial Isolates. Between October 2013 and December 2014, a nosocomial outbreak caused by *P. aeruginosa* showing an MDR phenotype was detected in the Hematology ward of a Spanish tertiary

hospital (Hospital General Universitario Gregorio Marañón (HGUGM), Madrid, Spain; 1,350 beds) [21]. Patients hospitalized at this ward showed an increased incidence of *P. aeruginosa* bacteremia, while this was not reflected at the overall hospital level. During this 15 month period, at least one isolate of *P. aeruginosa* showing the same MDR phenotype was detected in 14 patients, while this microorganism was not found in the same ward during the previous 8 months. The isolates were resistant to carbapenems (imipenem and meropenem), antipseudomonal fluoroquinolones (ciprofloxacin and levofloxacin), and aminoglycosides (gentamycin and tobramycin), fulfilling the criteria established by Magiorakos et al. [22] for MDR microorganisms.

As a first approach, PFGE was performed on 23 *P. aeruginosa* isolates: 12 available isolates out of the 14 outbreak-suspected strains from 2013 to 2014, 2 of them sourcing from the same patient; 5 controls strains related to the Hematology ward and 6 unrelated strains from the same time period. Moreover, a reference *P. aeruginosa* strain (ATCC 27853) was also included in the analysis in duplicate [21]. The methodology applied for PFGE has been described before [23]. Briefly, the strains were digested with SpeI overnight at 37°C and the DNA fragments were separated by electrophoresis in a Chef DRII instrument (Bio-Rad Laboratories, Inc., Hercules, CA) using a 1% agarose gel in 0.5X Tris-borate-EDTA buffer. The fingerprints obtained were analyzed with Bionumerics Software 4.0 (bioMérieux, Marcy-l'Étoile, France). Strains were considered to be identical when 99.9% similarity was achieved.

The strains that grouped in the outbreak-pulsotype (P1) were further characterized by WGS (Table S1). This methodology differentiated three clusters within pulsotype P1: Group 1, which contained the *P. aeruginosa* strains correlated to the outbreak, and Groups 2 and 3, where isolates detected during the same period as the outbreak strains but with enough single nucleotide polymorphisms (SNPs) to be considered as nonoutbreak strains clustered (Figure 1) [21].

With the outbreak-specific SNPs detected by WGS analysis, a multiplex allele-specific oligonucleotide polymerase chain reaction (ASO-PCR) was designed for the rapid differentiation of the outbreak-related *P. aeruginosa* isolates (Group 1). The ASO-PCR was tested with a collection of isolates ($n = 32$) from a broader period of time (2010–2018), allowing for the detection of new outbreak-related strains which were not considered part of the outbreak initially (Figure 1) [21].

A total of $n = 67$ available *P. aeruginosa* strains from this outbreak were included in this study (Table S1): 35 *P. aeruginosa* isolates analyzed by WGS (27 outbreak strains—all of them ST175–, and 8 control strains; Figure 1) and 32 isolates (16 outbreak and 16 nonoutbreak strains) analyzed by ASO-PCR. All strains sourced from inpatients from HGUGM and were kept frozen at -80°C until further analysis. Isolates were cultured on 5% Columbia sheep blood agar plates at 37°C in aerobic conditions, metabolically activated after three passages and analyzed after a 24 hr overnight incubation period.

2.2. MALDI-TOF MS Identification. A small amount of biomass from each isolate was spotted onto the MALDI target plate, covered with $1\ \mu\text{l}$ of formic acid 100% for on-plate

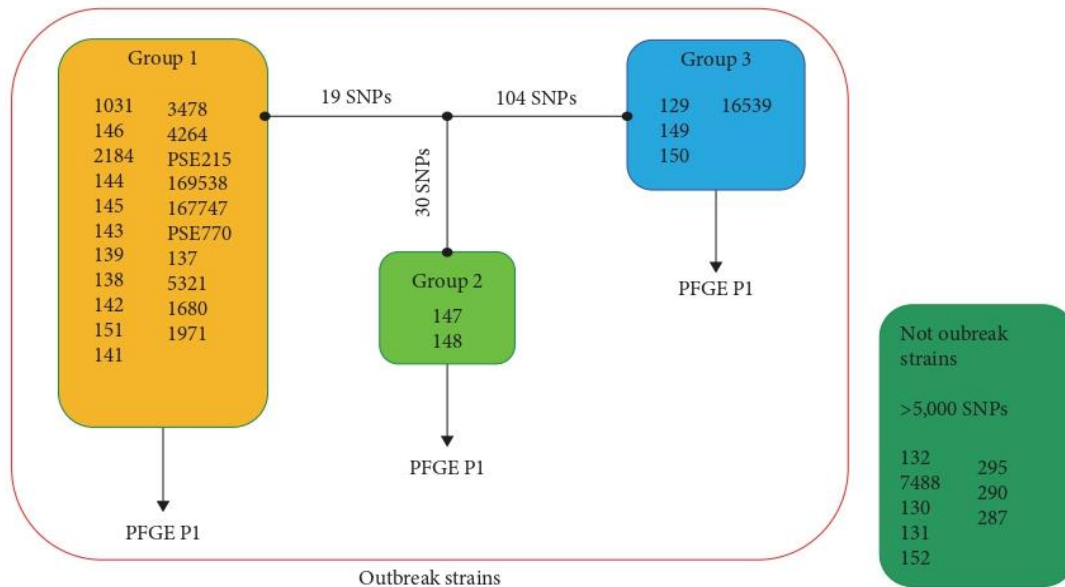


FIGURE 1: Classification of *Pseudomonas aeruginosa* strains as “outbreak” by pulse-field gel electrophoresis (PFGE) and “not outbreak” (in dark green). The three different groups obtained by whole-genome sequencing (WGS) are shown: Group 1, where the outbreak isolates are clustered; Group 2, containing two isolates separated by 30 single nucleotide polymorphisms (SNPs) from Group 1; and Group 3, where 4 isolates separated by 104 SNPs from Group 1 clustered. According to WGS, Groups 2 and 3 are not part of the outbreak.

protein extraction and allowed to dry. Then, $1\mu\text{l}$ of organic matrix (α -ciano-3,4-hidroxicinamic acid) was added and allowed to dry again before MALDI-TOF MS analysis. Strains were identified with an MBT Smart MALDI Biotyper (Bruker Daltonics, Bremen, Germany) using the updated database containing 9,957 reference spectra profiles.

2.3. MALDI-TOF MS Spectra Processing and Data Modeling. Spectra were acquired using default settings and visualized with FlexAnalysis Software (Bruker Daltonics), where outliers and zero lines were removed. MALDI-TOF MS spectra were exported to and further processed with Clover MS Data Analysis Software (Clover Biosoft, Granada, Spain) as follows: (a) variance stabilization, (b) smoothing by Savitzky–Golay filter (window length: 11, polynomial order: 3), (c) baseline subtraction using top-hat filter (0.02), and (d) TIC (total ion current)-normalized (Figure S1). Replicated peaks were aligned in the 2,000–20,000 Daltons region of the spectra—were most bacterial proteins can be found—and then merged in an average spectrum for each isolate according to the information compiled in a previous study [24].

As a first approach, the mass spectra from the 35 strains characterized by WGS were used as the training set for data modeling. Two peak matrices were built: (a) using the threshold method, which consisted of applying a 0.01 threshold value to average spectra, which selected only the peaks above 1.0% of the maximal intensity (Prominence: 0.01; Distance: 1) within 2,000–20,000 Daltons range and (b) using the biomarker selection method, that searches for specific peaks for each category (“outbreak” or “control” *P. aeruginosa* isolates). Peaks within the 2,000–10,000 Daltons range and with an area under the received operating characteristic curve (AUROC) higher than 0.85 were evaluated and selected for

the construction of the biomarker peak matrix (Figure S1). The AUROC values were obtained by matching all samples of both categories with peaks above 0.01 threshold.

Matrices built using both the threshold and biomarker selection methods were used as input data for training machine learning: first, two unsupervised algorithms—principal component analysis (PCA) and hierarchical clustering—were applied to check the clustering between samples and categories and settle their distances in a dendrogram and second, different supervised algorithms—partial least squares discriminant analysis (PLS-DA); linear support vector machine (SVM); random forest (RF) and neighborhood components analysis with K -nearest neighbors (NCA-KNN)—were used as classifiers to build predictive models. Supervised algorithms training was carried out by systematically testing different combinations of values for their respective hyperparameters. Each hyperparameter is given a list of possible values, and all combinations are tested for training. The combination which achieved the highest balanced accuracy was chosen. For SVM, we automatically tuned the hyperparameter C , which controls the trade-off between the training error and the classification error on unseen data. Thus, a large hyperparameter C value indicates a highly strict and potentially overfitting model whilst a lower value leads to a more heavily regularized model which will allow more misclassifications in the training set in exchange for a better generalization to unobserved samples. For NCA-KNN, this optimization is performed on the number of neighbors used (K). On the other hand, for RF we automatically tested a range of potential values in the following hyperparameters: number of trees in the forest, number of features to consider when looking for the best split, maximum depth of each tree, minimum number of samples required to split an internal node, and the minimum number of samples required

to be a leaf node. Details on the values explored can be found in Table S2. Cross-validation was performed for each model by k -fold method ($k = 10$) as described previously [16] and leave-one-out (LOO) method (Figure S1). The k -fold method divides the dataset into k -stratified folds and tests all folds as the validation set using the remaining folds as the training set. With the LOO method, each sample is used as a single set for validation and the rest as the training set. The final configuration of each supervised algorithm was saved as a prediction model to be validated.

For the validation of the predictive models described in the previous paragraph, a total of 32 isolates, 15 “outbreak” and 17 “control,” analyzed by ASO-PCR were included in the validation set. Average spectra obtained from four replicates analyzed twice for each isolate were preprocessed as described in the creation of predictive models and used as input data to be automatically classified by the prediction models.

2.4. Reproducibility Assessment. A reproducibility assay was carried out with the 35 *P. aeruginosa* strains characterized by WGS. Each isolate was spotted in quadruplicate (technical replicates) to assess the interspot or the technical reproducibility of the method. Besides, two spectra were acquired per spot to make an average spot spectrum. The biological reproducibility was analyzed by subculturing overnight under the same conditions as indicated above and acquiring protein spectra again in quadruplicates [24, 25]. Data analysis was carried out with Clover MS Data Analysis Software. The pipeline described above was applied to the acquired spectra, which were subsequently aligned to obtain an average spectrum per spot from two spectra per spot (shift medium; linear mass tolerance 200 ppm). The same procedure was applied to obtain a single average spectrum per day from four spots for each isolate. This approach allowed the detection of common peaks present in all protein spectra. The coefficient of variation (% CV) of the peak intensities registered for each common peak after a 0.1 threshold applied was calculated from TIC-normalized spectra.

2.5. FTIR-S Spectra Acquisition and Processing. FTIR-S was performed only on the strains which had combined PFGE plus WGS information in the initial study by Acosta et al. [21] ($n = 20$). FTIR-S spectra acquisition was performed using IR Biotyper (Bruker Daltonics) following the manufacturer’s instructions. Briefly, a $1\ \mu\text{l}$ loopful of biomass was resuspended in $50\ \mu\text{l}$ of 70% ethanol and homogenized with metal rods (Bruker Daltonics). Then, $50\ \mu\text{l}$ of sterile water were added and $15\ \mu\text{l}$ of the suspension were spotted on a silicon plate. Samples were analyzed in triplicates in three independent experiments along with two standards (Bruker Infrared Test Standard 1 and 2, Bruker Daltonics). FTIR-S spectra were visualized and processed using Clover MS Data Analysis Software. The processing consisted on standard normal variation followed by Savitzky–Golay filter with nine smoothing points, polynomial order = 2 and second derivative using the entire wavenumbers range. Moreover, spectra with valid target quality controls were further analyzed using hierarchical cluster analysis (HCA) with Euclidean distance and ward metric and PCA to evaluate

clustering of isolates according to molecular techniques (Figure S1).

3. Results

A preliminary version of the results shown in this article have been previously published as a preprint [26].

3.1. MALDI-TOF MS Analysis. All strains analyzed in this study ($n = 67$) were identified by MALDI-TOF MS as *P. aeruginosa* with a score ≥ 2.0 , showing, therefore, correct species-level identification according to the manufacturer’s criteria [27]. Besides, this identification was the only one provided by MALDI-TOF MS along the top 10 identifications given for each isolate.

3.1.1. Analysis of Outbreak Strains Characterized by WGS. Firstly, a peak matrix was constructed using MALDI-TOF MS spectra from the strains previously analyzed by WGS ($n = 35$) by applying the threshold method with a total of 413 peaks as features. The isolates were initially classified according to PFGE results, where the strains were grouped as P1 (outbreak) and other pulsotypes considered as unrelated strains. The cross-validation of this approach ($k = 10$) yielded 97.1% isolates correctly classified using PLS-DA, RF, and NCA-KNN algorithms and 94.3% with SVM (Table S3). Besides, using the biomarker selection method three potential biomarkers were found at 5,169, 6,915, and 7,236 m/z . The peak matrix constructed with these three peaks correctly classified all strains (100%) in all prediction models tested (PLS-DA, SVM, RF, and NCA-KNN) by k -fold validation ($k = 10$). The implementation of unsupervised algorithms also achieved optimal separation of the two main categories (“outbreak” and “control” strains) displaying two well-defined clusters in the PCA plot and HCA dendrogram (Figure 2).

In a second step, MALDI-TOF MS spectra were further compared according to WGS clustering, where the outbreak strains clustered by PFGE in the pulsotype 1 (P1) were divided into three outbreak groups: Group 1, considered the main outbreak strains, Group 2 and 3 (separated by <125 SNPs from Group 1) and Controls ($>5,000$ SNPs difference) (Figure 1). Differentiation of what WGS considered the main outbreak (Group 1) from the rest of the strains (“Controls,” “Group 2,” and “Group 3”) was attempted in this step. For this purpose, a peak matrix was created by applying the threshold method and used as input data to PLS-DA, SVM, RF, and NCA-KNN algorithms. They obtained a correct classification of 94.3% by SVM (C optimized hyperparameter: 0.01), 85.7% by PLS-DA and NCA-KNN (NCA component reduced to 3, neighbors optimized hyperparameter: 3), and 88.6% by RF (number of estimators optimized hyperparameter: 100) (Table S4; Figure S2). Group 2 strains ($n = 2$) appeared closer to the outbreak strains than Group 3 and control strains (Figure 2(c)), as it is closer to the Group 1 strains in a number of SNPs (50 SNPs).

3.1.2. External Validation of the Initial Model. Lastly, the biomarker classification model—SVM, three biomarker

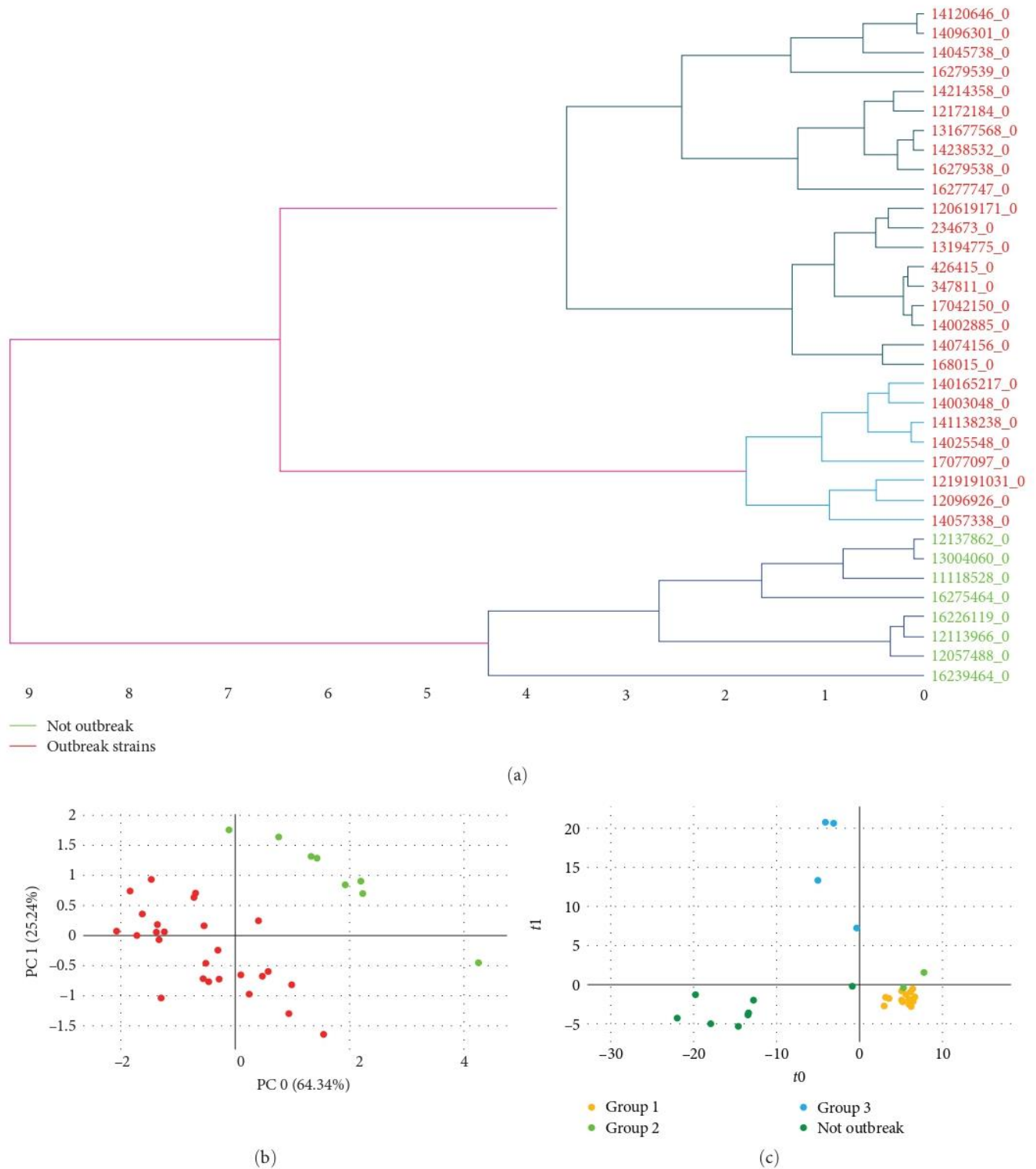


FIGURE 2: Classification of MALDI-TOF MS spectra between “outbreak” and “not outbreak” isolates. (a) Hierarchical clustering analysis with Euclidean distance and ward metric, (b) principal component analysis, and (c) differentiation of WGS-groups by PLS-DA (partial least squares discriminant analysis) algorithm. MALDI-TOF MS, matrix-assisted laser desorption ionization–time of flight mass spectrometry; WGS, whole-genome sequencing.

peaks—(C optimized hyperparameter: 0.19) developed with the strains characterized by WGS was blindly validated with the isolates characterized by ASO-PCR ($n=32$: 15 “outbreak” and 17 “control”). The validation of the model yielded 90.0% correct classification of the strains, misclassifying only 3/32 ASO-PCR strains (two outbreak isolates as nonoutbreak, and one nonoutbreak as outbreak strain).

3.1.3. Reproducibility of the Method. The comparison of the 13 common peaks from protein spectra obtained from the different spots in which the same isolate was smeared (interspot variability) showed 11.0% and 14.5% of mean of their CVs values for day 1 and day 2, respectively. Besides, the analysis of the variance yielded a maximum of 20.5% CV for the most common peaks detected within the 7,000–10,000 m/z range, although most peaks between 2,000 and 7,000 m/z interval showed CV values between 10.0% and 15.0% (Figure S3(a)). The mean and median interday CV values were 9.2% and 5.8%, respectively. CV values varied between 0.02% and 18.0%, a range considered as acceptable by Pang et al. [28] (Figure S3(b)). The only exception here was the 5,740 m/z peak, whose interday CV was almost 30.0%.

The reproducibility of the specific biomarkers that allowed the differentiation of the *P. aeruginosa* isolates belonging to the outbreak was further analyzed. In this case, three peaks were identified as potential biomarkers for differentiation between outbreak strains and those not related with the outbreak (peaks 5,169, 6,915, and 7,236 m/z). The average intraday CV for these peaks was 13.6% and this value varied between 7.0% and 20.5% (Figure S3(c)). However, the biomarker peak at 6,915 m/z showed 32.5% interday CV variability (Figure S3(d)).

3.2. FTIR-S Analysis. After analyzing different spectral ranges, infrared absorbance in the lipid region (1,400–1,500 and 2,800–3,000 cm^{-1}) showed the highest discriminatory power for the correct classification of *P. aeruginosa* outbreak strains. HCA using Euclidean distance and Ward linkage differentiated the outbreak strains (ST175) from control strains, which belonged to different sequence types (STs) such as ST227, ST253, ST381, ST557, and ST885 (Figure 3), with a discriminatory power similar to PFGE classification. In this case, the cut-off dendrogram distance for outbreak and controls differentiation was 0.350. For isolate differentiation, the automated cut-off distance assigned by IR Biotyper was 0.062. The two strains belonging to the same patient were clustered together with lower distance between them. Using PCA analysis in both IR Biotyper Software and Clover MS Data Analysis Software, ST175 strains were clustered together, distinctly from other STs (Figures 4(a) and 4(b)). The specific differences in infrared spectra absorbances among ST175 and other STs were found in 1,445 cm^{-1} (Figure 5(a)), 2,925 and 2,955 cm^{-1} (Figure 5(b)). The three subgroups inside the outbreak observed by WGS were not discriminated by HCA (Figure 3). In the PCA scatter plot (Figures 4(c) and 4(d)), although WGS Group 2 was partially separated from other groups, Groups 1 and 3 overlapped. Using PLS-DA and RF algorithm (Figure S4), it was also shown that Group 2 formed a distinct cluster, and the k -fold cross-validation obtained for differentiation of the three groups were 91.1% and 97.8%, respectively (Table S5).

4. Discussion

The implementation of MALDI-TOF MS and FTIR-S technology, combined with machine learning algorithms, allowed the correct classification of the MDR *P. aeruginosa* isolates causing a nosocomial outbreak in 2013–2014: 97.0% and 100% of the isolates were correctly classified by the algorithms applied to the protein peaks by MALDI-TOF MS with intensities above the established threshold (0.1) or to the specific biomarker peaks found with this technology, respectively. These results showed that MALDI-TOF MS yields a discrimination power similar to PFGE, the reference method for bacterial typing. Besides, the same protein spectra were further classified according to the information provided by WGS analysis. In this case, PLS-DA and SVM algorithms allowed a good classification of the *P. aeruginosa* isolates specifically correlated with the outbreak (Group 1) and showed that Group 2 was closely related with the outbreak strains, as the genomic analysis pointed out, and Group 3 and the control group were clearly unrelated to the outbreak. The validation of the classification models, carried out with 32 *P. aeruginosa* isolates characterized by ASO-PCR, yielded 90.0% correct classification of their protein spectra. Only three isolates were misclassified using the developed models, which indicates that the methodology described in this study may be applied as a rapid screening method when an outbreak is suspected. The implementation of FTIR-S technology showed the same discrimination power as PFGE to differentiate *P. aeruginosa* outbreak isolates. However, when the same level of classification provided by WGS technology was attempted with FTIR-S, the control group and outbreak Group 2 were clearly defined but Groups 1 and 3 overlapped. However, FTIR-S provided complementary information to the classification obtained with MALDI-TOF MS spectra by showing the correct classification of the different STs analyzed.

Previous studies have demonstrated that specific biomarker peaks present in the protein spectra obtained by MALDI-TOF MS could be used for the monitoring of *P. aeruginosa* STs. Cabrolier et al. [14] described a specific peak at 7,359 m/z that, combined with the absence of peaks at 7,329 and 12,154 m/z was specific of *P. aeruginosa* ST175. These results were further confirmed by Mulet et al. [29], who also found the peak at 7,359 m/z as a biomarker for *P. aeruginosa* ST175 and described another peak at 6,911 m/z as specific for this ST. In the present study, a peak at 6,915 m/z has been described as a biomarker of the *P. aeruginosa* outbreak strains belonging to the same ST. Although the difference of only 4 m/z between both peaks fall within the accepted margin of error of the applied pipeline ($\pm 4.5 m/z$) and no other marker peak has been found in the area, further analysis is needed to confirm that both studies refer the same peak. Finally, the other two potential biomarker peaks described for the *P. aeruginosa* outbreak strains (5,169 and 7,236 m/z) have never been described before and, therefore, they might be outbreak-specific markers.

The reproducibility of this MALDI-TOF MS-based method for the rapid detection of a *P. aeruginosa* outbreak

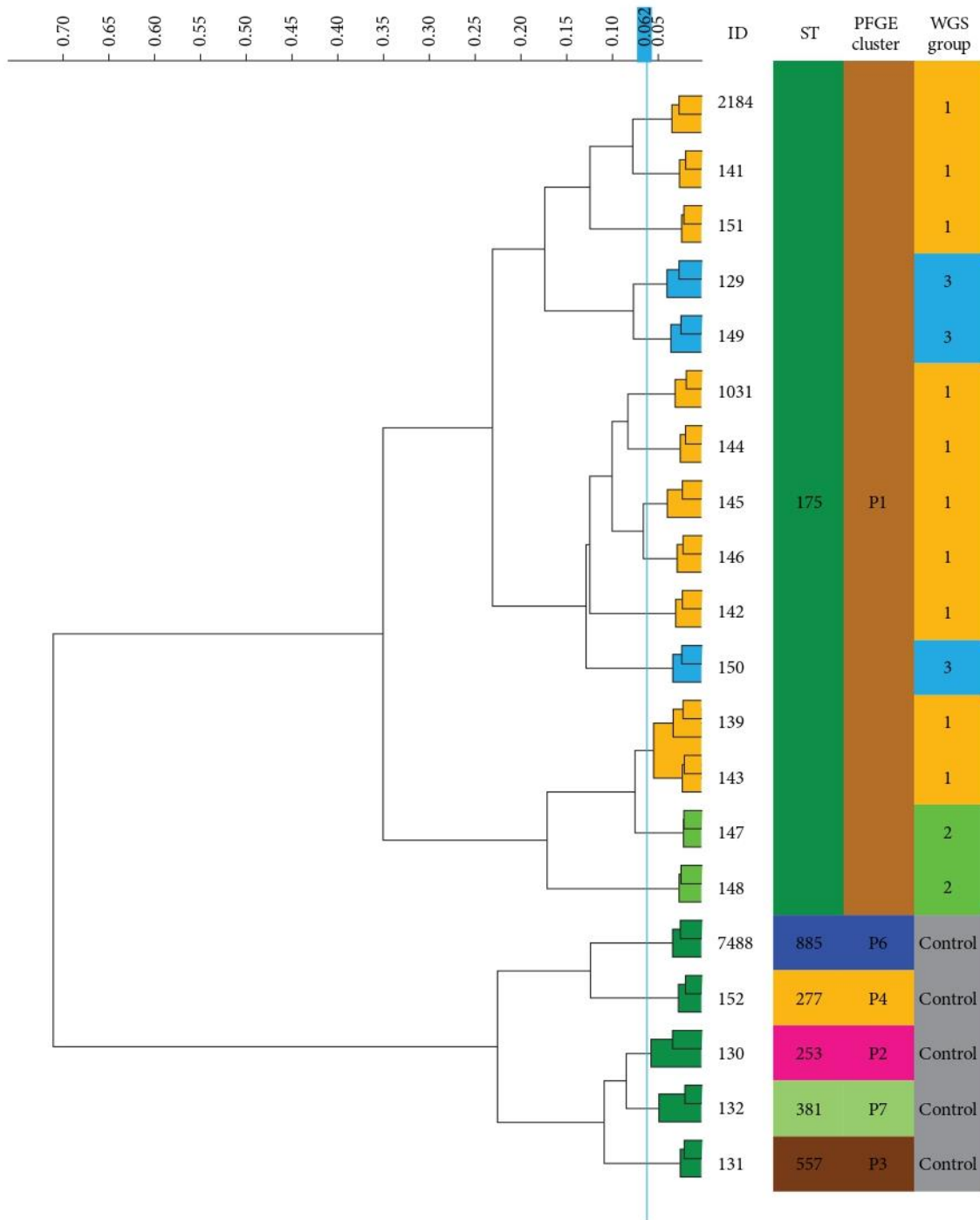


FIGURE 3: Dendrogram obtained by IR Biotyper for *Pseudomonas aeruginosa* isolates analyzed in the lipids region (1,400–1,500 and 2,800–3,000 cm^{-1}). Hierarchical cluster analysis with Euclidean distance and ward metric was performed. ST, sequence type; PFGE, pulsed-field gel electrophoresis; WGS, whole-genome sequencing.

was evaluated. With the exception of the peak at 5,740 m/z , whose interday CV was just below 30.0%, the peaks located between 2,000 and 7,000 m/z —where the most common bacterial proteins locate—showed CV values ranging between 10.0% and 15.0% and have been considered as acceptable in previous studies [28]. The specific biomarker peaks that

allowed the differentiation of the outbreak isolates showed an average intraday CV of 13.6% (range between 7.0% and 20.5%). Again, these values can be considered as acceptable. Only the biomarker peak at 6,915 m/z showed an interday CV variability CV of 32.5% (Figure S4(d)). Although this value is above the established limit (20.0%), the peak was

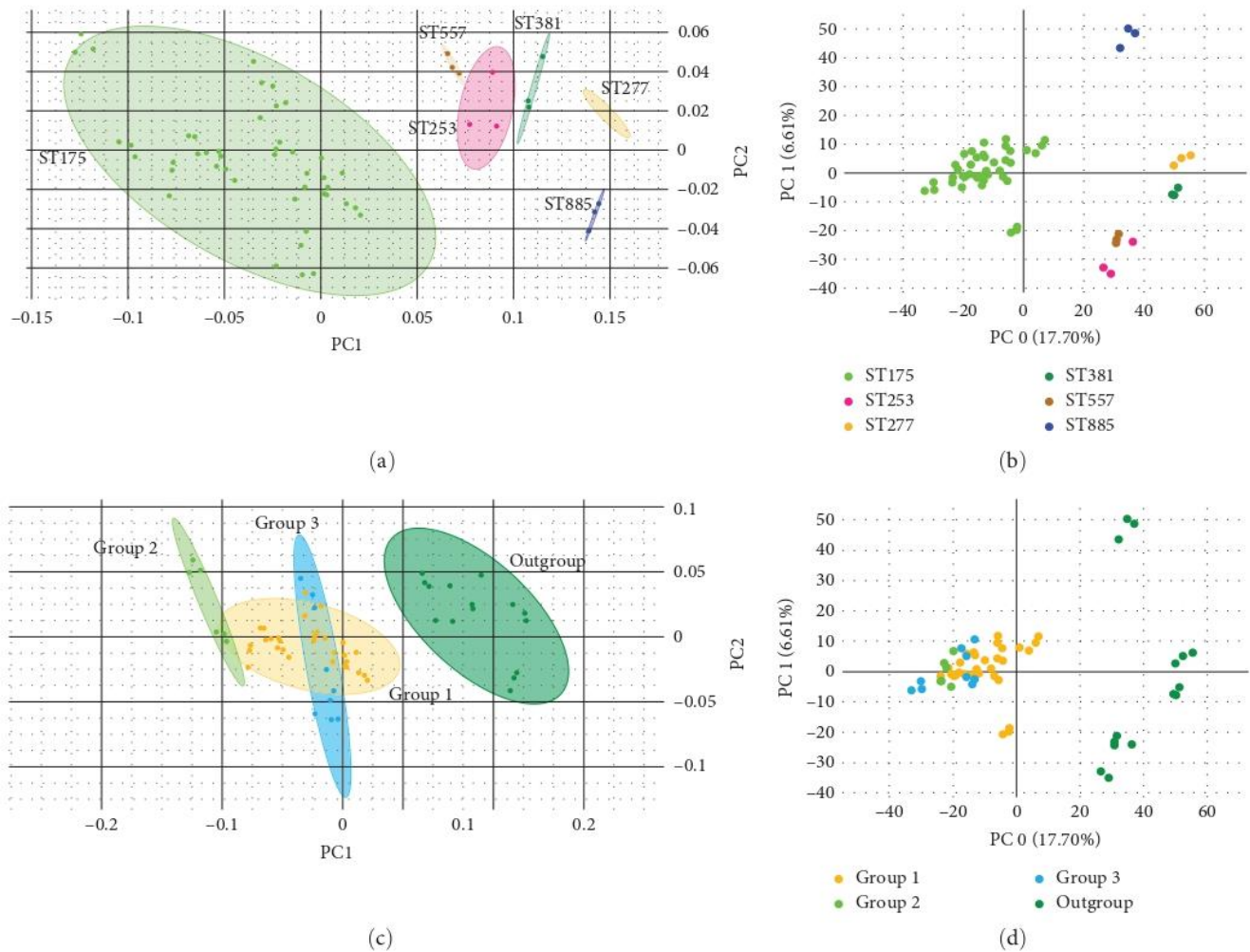


FIGURE 4: Principal component analysis scatter plot obtained by Fourier-transform infrared spectroscopy. (a) Plot according to sequence type (ST) of isolates obtained by IR Biotyper Software. (b) Plot according to ST obtained by Clover MS Data Analysis Software. (c) Plot according to whole-genome sequencing (WGS) outbreak groups obtained by IR Biotyper Software. (d) Plot according to WGS groups obtained by Clover MS Data Analysis Software.

always detected in *P. aeruginosa* outbreak strains regardless its intensity. Therefore, the presence of the 6,915 m/z peak can be reliably correlated with the outbreak along with the detection of the other two biomarker peaks at 5,169 and 7,236 m/z .

In recent years, FTIR-S has emerged as a reliable technology for outbreak analysis in clinical microbiology laboratories [30]. The simplicity and low costs of the sample preparation procedure and interpretation of results allows the follow-up of nosocomial outbreaks in real time since the turnaround time for the analysis of 30 isolates with this technology is approximately 3 hr. Although the most studied microorganism so far is *Klebsiella pneumoniae* [31], FTIR-S has been evaluated for typing other bacterial species such as *Salmonella* [19] or *Streptococcus pneumoniae* [20].

In our study, FTIR-S was able to discriminate *P. aeruginosa* outbreak isolates from nonoutbreak isolates at the same level than PFGE, either by HCA with a cut-off score of 0.350 or PCA (Figures 3 and 4). In addition, the two strains

isolated from the same patient (numbers 139 and 143) were clustered together with very low distance, which indicates that this method recognizes them as the same strain and is reproducible (Figure 3). For ST differentiation, the lipid region (1,400–1,500 and 2,800–3,000 cm^{-1}) showed the highest discriminatory power and allowed the correct classification of *P. aeruginosa* outbreak strains according to PFGE results. However, when this classification was compared with WGS information, the three outbreak subgroups were not clearly differentiated by FTIR-S by applying HCA. Similarly, the implementation of the PCA algorithm did not cluster the three groups of isolates separately, although Group 2 isolates were almost grouped apart (Figures 4(c) and 4(d)). When using PLS-DA and RF analyses for differentiation of these groups, it was also observed that Group 2 was clearly separated from the other outbreak groups according to WGS and the biggest differences among them were found in the 800–1,600 cm^{-1} region (Figure S5). It is important to note that WGS groups are based on SNP distance, and

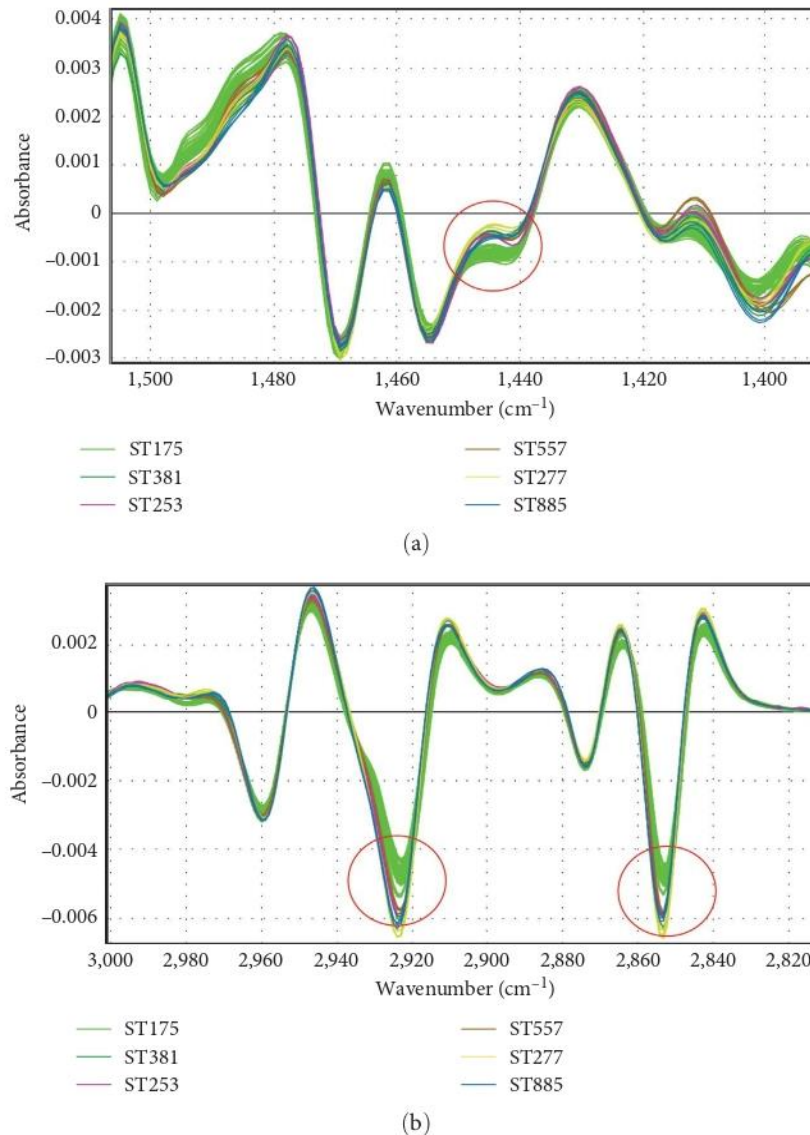


FIGURE 5: Fourier-transform infrared spectroscopy spectra of *Pseudomonas aeruginosa* isolates in the lipids region after applying second derivative. (a) 1,400–1,500 cm^{-1} region with different absorbance among outbreak and not-outbreak isolates (1,445 cm^{-1}). (b) 2,800–3,000 cm^{-1} region with different absorbance among outbreak and not-outbreak isolates (2,925 and 2,955 cm^{-1}).

maybe these differences are not expressed phenotypically, and thus, FTIR-S clusters may not reflect differences detected by WGS. At the moment of writing, only one previous study has evaluated *P. aeruginosa* typing by FTIR-S in comparison to PFGE and MLST results [32]. The authors showed the reliability of the method for differentiation of STs with an optimal cut-off distance between 0.184 and 0.374. Besides, these results can be obtained in a turnaround time of 3 hr, a great advantage over PFGE.

One of the limitations of this study is the small number of strains available, but it is important to acknowledge that bacterial outbreaks usually involve a limited number of patients if they are well contained, thus making a larger number of samples unavailable for research purposes. The other limitation would be the limited number of isolates characterized by

WGS. But, due to the costs of this technique, a more affordable approach (ASO-PCR) was carried out to classify the rest of the isolates.

However, despite these drawbacks, the results of this study showed that rapid diagnostic methods such as MALDI-TOF MS and FTIR-S may represent fast alternatives to conventional strategies—based on DNA sequencing—for real-time monitoring of nosocomial outbreaks, providing complementary information for the prompt characterization of the suspected isolates in a cost-efficient way. Although confirmation of the outbreak strains may request further analysis by WGS, the implementation and further validation of these rapid typing methods could help to reduce the number of isolates that require confirmation by expensive tests available to a limited number of microbiology laboratories.

Data Availability

Pseudomonas aeruginosa sequences were deposited in the Pseudomonas Genome Database (<http://www.pseudomonas.com>). Proteomic data sourcing from MALDI-TOF and FTIR are available upon request.

Ethical Approval

The Ethics Committee of the Gregorio Marañón Hospital (CEIm) evaluated this project and considered that all the conditions for waiving informed consent were met, since the study was conducted with microbiological samples and not with human products.

Conflicts of Interest

The authors declare that they have no conflicts of interest. GM, MA, and LM are employees of Clover Bioanalytical Software, S.L.

Acknowledgments

This work was supported by the projects PI15/01073 and PI18/00997 from the Health Research Fund (FIS. Instituto de Salud Carlos III. Plan Nacional de I+D+I 2013-2016) of the Carlos III Health Institute (ISCIII, Madrid, Spain) partially financed by the European Regional Development Fund (FEDER) “A way of making Europe.” BRS (CPII19/00002) is the recipient of a Miguel Servet contract supported by the FIS. DRT has been funded by the IISGM through its intramural program.

Supplementary Materials

Table S1: Description of the *P. aeruginosa* isolates included in this study and the techniques performed on them. Figure S1: Workflow of the *P. aeruginosa* isolates analysis with MALDI-TOF MS and FTIR S and the subsequent analysis of the spectra obtained by both methods. Table S2: Range of values explored for the automatic optimization of the balanced accuracy of model predictions obtained by the different methods of supervised training used. Table S3: Leave-one-out method for differentiation of outbreak and not-outbreak MALDI-TOF MS spectra by using the threshold matrix for PLS-DA, RF, NCA-KNN, and SVM algorithms. Table S4: Leave-one-out method for differentiation of WGS-groups in MALDI-TOF MS spectra by PLS-DA, SVM, and RF algorithms. Figure S2: Differentiation of outbreak strains according to whole-genome sequencing groups, using Clover MS Data Analysis Software and MALDI-TOF MS spectra. Figure S3: Reproducibility assay. Figure S4: Differentiation of outbreak strains according to whole-genome sequencing groups, using Clover MS Data Analysis Software and FTIR-S spectra. Table S5: Leave-one-out method for differentiation of whole-genome sequencing groups using PLS-DA and RF algorithms for FTIR-S spectra. Figure S5: Important regions in the FTIR spectra for strain discrimination. (*Supplementary Materials*)

References

- [1] J.-Y. Liu and J. K. Dickter, “Nosocomial infections: a history of hospital-acquired infections,” *Gastrointestinal Endoscopy Clinics of North America*, vol. 30, no. 4, pp. 637–652, 2020.
- [2] A. Sikora and F. Zahra, “Nosocomial infections,” in *StatPearls*, Treasure Island (FL), 2021.
- [3] C. Suetens, K. Latour, T. Kärki et al., “Prevalence of healthcare-associated infections, estimated incidence and composite antimicrobial resistance index in acute care hospitals and long-term care facilities: results from two European point prevalence surveys, 2016 to 2017,” *Eurosurveillance*, vol. 23, no. 46, 2018.
- [4] J.-L. Vincent, Y. Sakr, M. Singer et al., “Prevalence and outcomes of infection among patients in intensive care units in 2017,” *JAMA*, vol. 323, no. 15, pp. 1478–1487, 2020.
- [5] P. Chhatwal, E. Ebadi, F. Schwab et al., “Epidemiology and infection control of carbapenem resistant *Acinetobacter baumannii* and *Klebsiella pneumoniae* at a German University Hospital: a retrospective study of 5 years (2015–2019),” *BMC Infectious Diseases*, vol. 21, no. 1, Article ID 1196, 2021.
- [6] M. Sanchez, A. Garcia-de-Lorenzo, R. Herruzo, M. J. Asensio, and F. Leyva, “Clinical and economic benefits associated with the management of a nosocomial outbreak resulting from multidrug-resistant *Klebsiella pneumoniae*,” *Critical Care Medicine*, vol. 40, no. 6, 2007; author reply 2007-2008, 2012.
- [7] D. Reynolds and M. Kollef, “The epidemiology and pathogenesis and treatment of *Pseudomonas aeruginosa* infections: an update,” *Drugs*, vol. 81, no. 18, pp. 2117–2131, 2021.
- [8] Z. Pang, R. Raudonis, B. R. Glick, T.-J. Lin, and Z. Cheng, “Antibiotic resistance in *Pseudomonas aeruginosa*: mechanisms and alternative therapeutic strategies,” *Biotechnology Advances*, vol. 37, no. 1, pp. 177–192, 2019.
- [9] E. del Barrio-Tofiño, C. López-Causapé, and A. Oliver, “*Pseudomonas aeruginosa* epidemic high-risk clones and their association with horizontally-acquired β -lactamases: 2020 update,” *International Journal of Antimicrobial Agents*, vol. 56, no. 6, Article ID 106196, 2020.
- [10] A. Oliver, X. Mulet, C. López-Causapé, and C. Juan, “The increasing threat of *Pseudomonas aeruginosa* high-risk clones,” *Drug Resistance Updates*, vol. 21-22, pp. 41–59, 2015.
- [11] A. J. Sabat, A. Budimir, D. Nashev et al., “Overview of molecular typing methods for outbreak detection and epidemiological surveillance,” *Eurosurveillance*, vol. 18, no. 4, Article ID 20380, 2013.
- [12] S. J. Peacock, G. D. I. de Silva, A. Justice et al., “Comparison of multilocus sequence typing and pulsed-field gel electrophoresis as tools for typing *Staphylococcus aureus* isolates in a microepidemiological setting,” *Journal of Clinical Microbiology*, vol. 40, no. 10, pp. 3764–3770, 2002.
- [13] B. Rodríguez-Sánchez, E. Cercenado, A. T. Coste, and G. Greub, “Review of the impact of MALDI-TOF MS in public health and hospital hygiene, 2018,” *Eurosurveillance*, vol. 24, no. 4, 2019.
- [14] N. Cabroler, M. Sauget, X. Bertrand, and D. Hocquet, “Matrix-assisted laser desorption ionization–time of flight mass spectrometry identifies *Pseudomonas aeruginosa* high-risk clones,” *Journal of Clinical Microbiology*, vol. 53, no. 4, pp. 1395–1398, 2015.
- [15] E. Gato, A. Anantharajah, M. J. Arroyo et al., “Multicenter performance evaluation of MALDI-TOF MS for rapid detection of carbapenemase activity in enterobacterales: the future of networking data analysis with online software,” *Frontiers in Microbiology*, vol. 12, Article ID 789731, 2022.

- [16] M. E. Zvezdanova, M. J. Arroyo, G. Méndez et al., “Detection of azole resistance in *Aspergillus fumigatus* complex isolates using MALDI-TOF mass spectrometry,” *Clinical Microbiology and Infection*, vol. 28, no. 2, pp. 260–266, 2022.
- [17] M. Feucherolles, M. Nennig, S. L. Becker et al., “Investigation of MALDI-TOF mass spectrometry for assessing the molecular diversity of *Campylobacter jejuni* and comparison with MLST and cgMLST: a Luxembourg one-health study,” *Diagnostics*, vol. 11, no. 11, Article ID 1949, 2021.
- [18] M. Oviaño and G. Bou, “Matrix-assisted laser desorption ionization–time of flight mass spectrometry for the rapid detection of antimicrobial resistance mechanisms and beyond,” *Clinical Microbiology Reviews*, vol. 32, no. 1, 2018.
- [19] M. Cordovana, N. Mauder, M. Kostrzewa et al., “Classification of *Salmonella enterica* of the (Para-)typhoid fever group by Fourier-transform infrared (FTIR) spectroscopy,” *Microorganisms*, vol. 9, no. 4, Article ID 853, 2021.
- [20] I. Burckhardt, K. Sebastian, N. Mauder, M. Kostrzewa, F. Burckhardt, and S. Zimmermann, “Analysis of *Streptococcus pneumoniae* using Fourier-transformed infrared spectroscopy allows prediction of capsular serotype,” *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 38, no. 10, pp. 1883–1890, 2019.
- [21] F. Acosta, A. Fernández-Cruz, S. R. Maus et al., “In-depth study of a nosocomial outbreak caused by extensively drug-resistant *Pseudomonas aeruginosa* using whole genome sequencing coupled with a polymerase chain reaction targeting strain-specific single nucleotide polymorphisms,” *American Journal of Epidemiology*, vol. 189, no. 8, pp. 841–849, 2020.
- [22] A.-P. Magiorakos, A. Srinivasan, R. B. Carey et al., “Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance,” *Clinical Microbiology and Infection*, vol. 18, no. 3, pp. 268–281, 2012.
- [23] C. Sánchez-Carrillo, B. Padilla, M. Marín et al., “Contaminated feeding bottles: the source of an outbreak of *Pseudomonas aeruginosa* infections in a neonatal intensive care unit,” *American Journal of Infection Control*, vol. 37, no. 2, pp. 150–154, 2009.
- [24] A. Candela, M. J. Arroyo, A. Sánchez-Molleda et al., “Rapid and reproducible MALDI-TOF-based method for the detection of vancomycin-resistant *Enterococcus faecium* using classifying algorithms,” *Diagnostics*, vol. 12, no. 2, Article ID 328, 2022.
- [25] M. Oberle, N. Wohlwend, D. Jonas et al., “The technical and biological reproducibility of matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) based typing: employment of bioinformatics in a multicenter study,” *PLoS One*, vol. 11, no. 10, Article ID e0164260, 2016.
- [26] B. Rodríguez-Sánchez, A. Candela, M. J. Arroyo et al., “Rapid discrimination of *Pseudomonas aeruginosa* ST175 isolates involved in a nosocomial outbreak using MALDI-TOF mass spectrometry and FTIR spectroscopy coupled with machine learning,” *Authorea*, <https://www.authorea.com/users/498267/articles/579081-rapid-discrimination-of-pseudomonas-aeruginosa-st175-isolates-involved-in-a-nosocomial-outbreak-using-maldi-tof-mass-spectrometry-and-ftir-spectroscopy-coupled-with-machine-learning>, 2022.
- [27] M. Oviaño and B. Rodríguez-Sánchez, “MALDI-TOF mass spectrometry in the 21st century clinical microbiology laboratory,” *Enfermedades Infecciosas y Microbiología Clínica (English ed.)*, vol. 39, no. 4, pp. 192–200, 2021.
- [28] R. T. K. Pang, P. J. Johnson, C. M. L. Chan et al., “Technical evaluation of MALDI-TOF mass spectrometry for quantitative proteomic profiling matrix formulation and application,” *Clinical Proteomics*, vol. 1, pp. 259–270, 2004.
- [29] X. Mulet, M. Fernández-Esgueva, C. Norte, L. Zamorano, E. Del Barrio-Tofiño, and A. Oliver, “Validation of MALDI-TOF for the early detection of the ST175 high-risk clone of *Pseudomonas aeruginosa* in clinical isolates belonging to a Spanish nationwide multicenter study,” *Enfermedades Infecciosas y Microbiología Clínica (English ed.)*, vol. 39, no. 6, pp. 279–282, 2021.
- [30] C. Quintelas, E. C. Ferreira, J. A. Lopes, and C. Sousa, “An overview of the evolution of infrared spectroscopy applied to bacterial typing,” *Biotechnology Journal*, vol. 13, no. 1, Article ID 1700449, 2018.
- [31] N. Rakovitsky, S. Frenk, H. Kon et al., “Fourier transform infrared spectroscopy is a new option for outbreak investigation: a retrospective analysis of an extended-spectrum-beta-lactamase-producing *Klebsiella pneumoniae* outbreak in a neonatal intensive care unit,” *Journal of Clinical Microbiology*, vol. 58, no. 5, 2020.
- [32] D. Martak, B. Valot, M. Sauget et al., “Fourier-transform infrared spectroscopy can quickly type gram-negative bacilli responsible for hospital outbreaks,” *Frontiers in Microbiology*, vol. 10, Article ID 1440, 2019.

MATERIAL SUPLEMENTARIO CAPÍTULO 3

Table S1. Description of the *P. aeruginosa* isolates included in this study and the techniques performed on them.

AST	STRAIN ID	MALDI-TOF n=67	FTIR-S n=20	PFGE n=20	WGS n=35	ASO-PCR n=32	AST n=67
CONTROLS	132	Yes	Yes	Yes (Control)	Yes (Control)	No	non-MDR
Not-outbreak	7488	Yes	Yes	Yes (Control)	Yes (Control)	No	MDR
	130	Yes	Yes	Yes (Control)	Yes (Control)	No	non-MDR
	131	Yes	Yes	Yes (Control)	Yes (Control)	No	non-MDR
	152	Yes	Yes	Yes (Control)	Yes (Control)	No	non-MDR
	295	Yes	No	No	Yes (Control)	No	MDR
	290	Yes	No	No	Yes (Control)	No	MDR
	287	Yes	No	No	Yes (Control)	No	MDR
GROUP 1	151	Yes	Yes	Yes (Outbreak)	Yes (Outbreak)	No	MDR
	2184	Yes	Yes	Yes (Outbreak)	Yes (Outbreak)	No	MDR
	145	Yes	Yes	Yes (Outbreak)	Yes (Outbreak)	No	MDR
	139	Yes	Yes	Yes (Outbreak)	Yes (Outbreak)	No	MDR
	143	Yes	Yes	Yes (Outbreak)	Yes (Outbreak)	No	MDR
	144	Yes	Yes	Yes (Outbreak)	Yes (Outbreak)	No	MDR
	138	Yes	No	No	Yes (Outbreak)	No	MDR
	146	Yes	Yes	Yes (Outbreak)	Yes (Outbreak)	No	MDR
	142	Yes	Yes	Yes (Outbreak)	Yes (Outbreak)	No	MDR
	141	Yes	Yes	Yes (Outbreak)	Yes (Outbreak)	No	MDR
1031	Yes	Yes	Yes (Outbreak)	Yes (Outbreak)	No	MDR	

MATERIAL SUPLEMENTARIO CAPÍTULO 3

	1680	Yes	No	No	Yes (Outbreak)	No	MDR
	137	Yes	No	No	Yes (Outbreak)	No	MDR
	3478	Yes	No	No	Yes (Outbreak)	No	MDR
	4264	Yes	No	No	Yes (Outbreak)	No	MDR
	1971	Yes	No	No	Yes (Outbreak)	No	MDR
	5321	Yes	No	No	Yes (Outbreak)	No	MDR
	296	Yes	No	No	Yes (Outbreak)	No	MDR
	288	Yes	No	No	Yes (Outbreak)	No	MDR
	291	Yes	No	No	Yes (Outbreak)	No	MDR
	292	Yes	No	No	Yes (Outbreak)	No	MDR
GROUP 2	148	Yes	Yes	Yes (Outbreak)	Yes (Control)	No	MDR
	147	Yes	Yes	Yes (Outbreak)	Yes (Control)	No	MDR
GROUP 3	149	Yes	Yes	Yes (Outbreak)	Yes (Control)	No	MDR
	129	Yes	Yes	Yes (Outbreak)	Yes (Control)	No	MDR
	150	Yes	Yes	Yes (Outbreak)	Yes (Control)	No	MDR
	169539	Yes	No	No	Yes (Control)	No	MDR
ASO-PCR	790-12	Yes	No	No	No	Yes (Outbreak)	non-MDR
Not-outbreak	565-13	Yes	No	No	No	Yes (Outbreak)	non-MDR
	882-13	Yes	No	No	No	Yes (Outbreak)	non-MDR
	4138-15	Yes	No	No	No	Yes (Outbreak)	non-MDR
	4292-15	Yes	No	No	No	Yes (Outbreak)	MDR

MATERIAL SUPLEMENTARIO CAPÍTULO 3

	919-16	Yes	No	No	No	Yes (Outbreak)	non-MDR
	117-16	Yes	No	No	No	Yes (Outbreak)	non-MDR
	1,7E+07	Yes	No	No	No	Yes (Outbreak)	MDR
	1054-17	Yes	No	No	No	Yes (Outbreak)	non-MDR
	1070-17	Yes	No	No	No	Yes (Outbreak)	non-MDR
	1455-17	Yes	No	No	No	Yes (Outbreak)	non-MDR
	1505-17	Yes	No	No	No	Yes (Outbreak)	non-MDR
	57-17	Yes	No	No	No	Yes (Outbreak)	non-MDR
	772-17	Yes	No	No	No	Yes (Outbreak)	MDR
	921-17	Yes	No	No	No	Yes (Outbreak)	MDR
	1503-17	Yes	No	No	No	Yes (Outbreak)	non-MDR
ASO-PCR	1,2E+07	Yes	No	No	No	Yes (Control)	MDR
Outbreak	1983-13	Yes	No	No	No	Yes (Control)	MDR
	2645-15	Yes	No	No	No	Yes (Control)	MDR
	836-16	Yes	No	No	No	Yes (Control)	MDR
	1626-17	Yes	No	No	No	Yes (Control)	MDR
	1,7E+07	Yes	No	No	No	Yes (Control)	MDR
	1093-17	Yes	No	No	No	Yes (Control)	MDR
	1453-17	Yes	No	No	No	Yes (Control)	MDR
	1512-17	Yes	No	No	No	Yes (Control)	MDR
	1565-17	Yes	No	No	No	Yes (Control)	MDR
	1677-17	Yes	No	No	No	Yes (Control)	MDR

MATERIAL SUPLEMENTARIO CAPÍTULO 3

1862-17	Yes	No	No	No	Yes (Control)	MDR
1928-17	Yes	No	No	No	Yes (Control)	MDR
1964-17	Yes	No	No	No	Yes (Control)	MDR
229-17	Yes	No	No	No	Yes (Control)	MDR
378-17	Yes	No	No	No	Yes (Control)	MDR

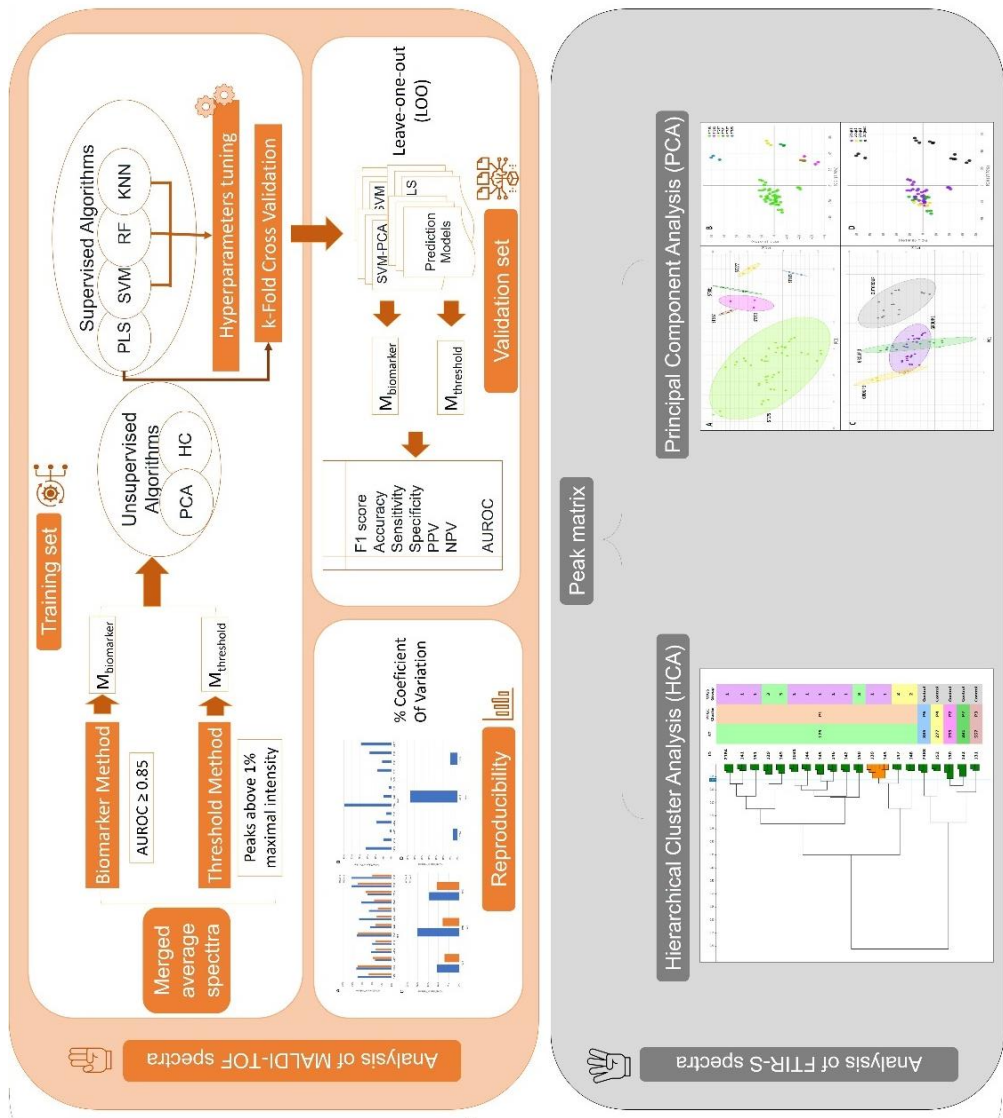
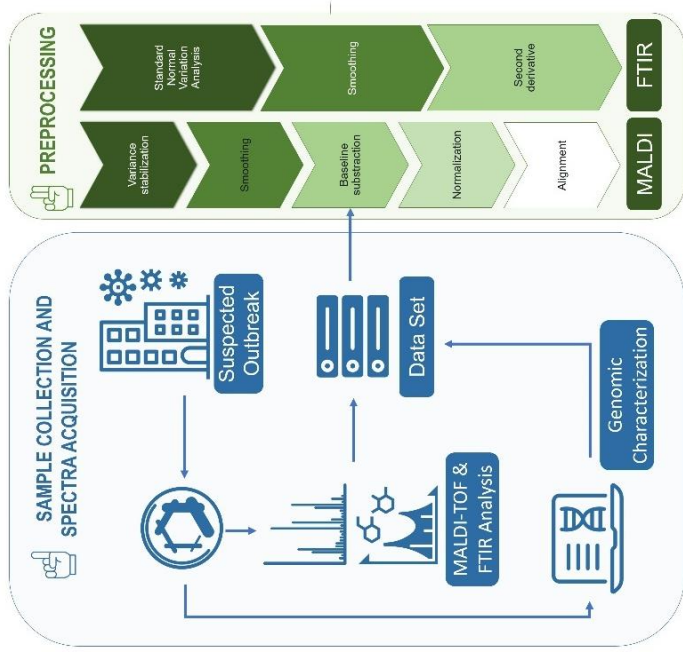


Figure S1. Workflow of the *P. aeruginosa* isolates analysis with MALDI-TOF MS and FTIR S and the subsequent analysis of the spectra obtained by both methods



MATERIAL SUPLEMENTARIO CAPÍTULO 3

Table S2. Range of values explored for the automatic optimization of the balanced accuracy of model predictions obtained by the different methods of supervised training used.

Supervised method	Hyperparameter	Range of values
SVM	C	1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 10
NCA-KNN	K	Odd values in [3, 21]
RF	Number of estimators	50, 100, 200, 400
	Number of features to look for the best split	Square root or logarithm in base 2 of the total number of features.
	Maximum depth of each tree	10, 30, 50, 70, 110
	Minimum number of samples required to split an internal node	2, 5, 10
	Minimum number of samples required to be a leaf node	1, 2, 4

Table S3. Leave-one-out method for differentiation of outbreak and not-outbreak MALDI-TOF MS spectra by using the threshold matrix for PLS-DA, RF, NCA-KNN and SVM algorithms. RF hyperparameters: 50 Estimators, 20 Max Features, 10 Max Depth, 2 Min split size and 1 Min Samples per leaf. PLS-DA components: 3. SVM hyperparameter C: 0.00001 (liblinear kernel). Hyperparameter C is the trade-off value of misclassified samples. PCA-SVM hyperparameter: 10 (liblinear kernel). NCA-KNN number of neighbors: 3

Actual\Predicted	Outbreak	Not outbreak	% Correct
PLS-DA, PCA-SVM, RF and NCA-KNN			
Outbreak	27	0	100%
Not Outbreak	1	7	87.5%
PLS-DA, PCA-SVM, RF and NCA-KNN Accuracy / Balanced Accuracy			97.1%/93.75%
SVM			
Outbreak	26	1	96.3%
Not Outbreak	1	7	87.5%
SVM Accuracy / Balanced Accuracy			94.3%/91.9%

Table S4. Leave-one-out method for differentiation of WGS-groups in MALDI-TOF MS spectra by PLS-DA, SVM and RF algorithms. RF hyperparameters: 50 Estimators, 20 Max Features, 10 Max Depth, 2 Min split size and 1 Min Samples per leaf. PLS-DA components: 3. SVM hyperparameter C: 0.01 (liblinear kernel). Hyperparameter C is the trade-off value of misclassified samples. NCA-KNN number of neighbors: 3

Actual\Predicted	Group 1	Group 2	Group 3	Not Outbreak	% Correct
PLS-DA					
Group 1	21	0	0	0	100%
Group 2	2	0	0	0	0%
Group 3	1	0	3	0	75%
Not Outbreak	2	0	0	6	75%
PLS-DA Accuracy / Balanced Accuracy					85.7%/62.5%
RF					
Group 1	21	0	0	0	100%
Group 2	2	0	0	0	0%
Group 3	1	0	3	0	75%
Not Outbreak	1	0	0	7	87.5%
RF Accuracy / Balanced Accuracy					88.57%/65.63%
SVM					
Group 1	21	0	0	0	100%
Group 2	1	1	0	0	50%
Group 3	0	0	4	0	100%
Not Outbreak	0	0	1	7	87.5%
SVM Accuracy / Balanced Accuracy					94.3%/84.38%
NCA- KNN					
Group 1	21	0	0	0	100%
Group 2	2	0	0	0	0%
Group 3	1	0	3	0	75%
Not Outbreak	1	0	1	6	75%
NCA-KNN Accuracy / Balanced Accuracy					85.7%/62.5%

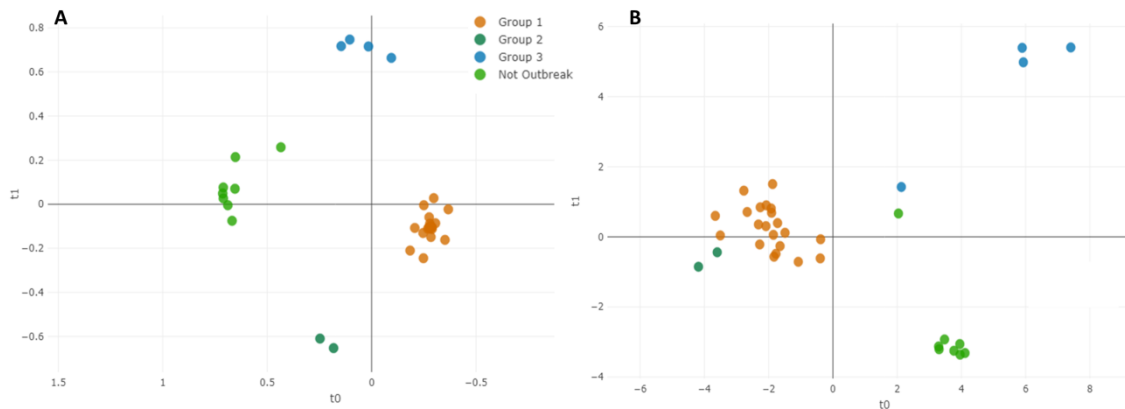


Figure S2. Differentiation of outbreak strains according to Whole-genome sequencing groups, using Clover MS Data Analysis software and MALDI-TOF MS spectra. **A:** Random Forest scatter plot. **B:** Neighborhood Component Analysis-K Nearest Neighbors scatter plot.

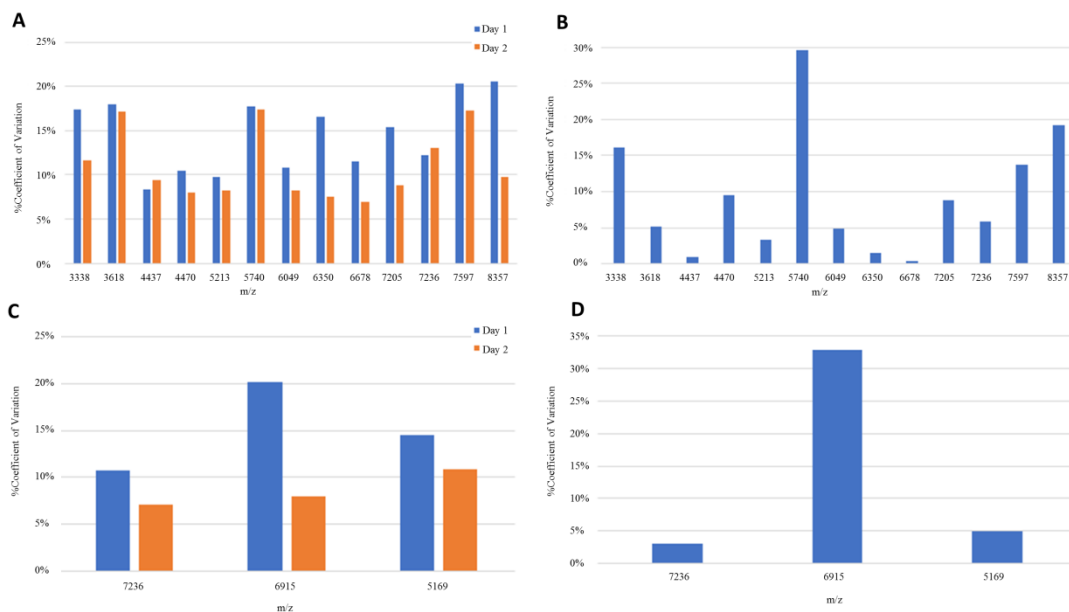


Figure S3. Reproducibility assay. **A:** Coefficient of Variation (CV) of most common peaks each day (intra-day reproducibility); **B:** CV of most common peaks between both days (inter-day reproducibility); **C:** CV of biomarker peaks each day (intra-day reproducibility); **D:** CV of biomarker peaks between both days (inter-day reproducibility).

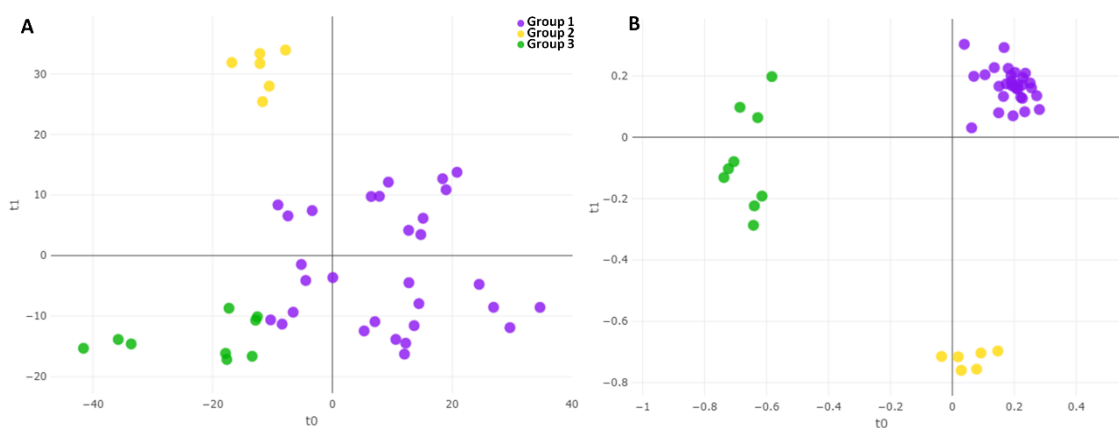


Figure S4. Differentiation of outbreak strains according to Whole-genome sequencing groups, using Clover MS Data Analysis software and FTIR-S spectra. **A:** Partial Least Squares-Discriminant Analysis scatter plot. **B:** Random Forest scatter plot.

Table S5 Leave-one-out method for differentiation of Whole-genome sequencing groups using PLS-DA and RF algorithms for FTIR-S spectra. PLS-DA Components: 3, RF hyperparameters: 50 estimators, 59 max. features, 10 max. depth, 5 min. split size and 1 min. samples per leaf.

Actual\Predicted	Group 1	Group 2	Group 3	% Correct
PLS-DA				
Group 1	30	0	0	100%
Group 2	1	5	0	83.3%
Group 3	4	0	5	55.56%
PLS-DA Accuracy / Balanced Accuracy				91.11%/79.63%
RF				
Group 1	30	0	0	100%
Group 2	0	6	0	100%
Group 3	1	0	8	88.89%
RF Accuracy / Balanced Accuracy				97.78% / 96.29%



Figure S5. Important regions in the FTIR spectra for strain discrimination. **Up:** Second derivate of FTIR-S spectra according to Whole-genome sequencing groups (Group 1: purple; Group 2: yellow; Group 3: green). **Down:** Important peaks for the differentiation of *P. aeruginosa* isolates according to Whole-genome sequencing using Random Forest for FTIR-S spectra.

Capítulo 4. Ribotipado de *Clostridioides difficile* empleando MALDI-TOF y algoritmos de *Machine Learning*

En el último capítulo de esta tesis se plantea un estudio basado en el análisis mediante MALDI-TOF y algoritmos de *Machine Learning* que permita el ribotipado rápido de *Clostridioides difficile* y la diferenciación de ribotipos hipervirulentos: ribotipo 027 (RT027), ribotipo 181 (RT181) y otros cercanos a ellos y clínicamente significativos.

Para ello se analizaron 363 aislados clínicos de *C. difficile* recogidos en España y representativos de los ribotipos más comunes causantes de infección clínica en España y Europa. De estas cepas, 348 pertenecían a una colección multicéntrica de 10 hospitales españoles, y aisladas entre los años 2010 y 2022, mientras que las 15 restantes se recogieron prospectivamente tras la detección de dos brotes nosocomiales que tuvieron lugar en Madrid entre mayo y junio de 2023 en el HCDGU (n=12) y el HGM (n=3).

La identificación de las cepas se llevó a cabo mediante MALDI-TOF y la PCR comercial Xpert® BT *C. difficile* (Cepheid, USA), y su ribotipado mediante amplificación por PCR de la región espaciadora intergénica entre el ARNr 16S y el ARNr 23S y electroforesis capilar [205-207]. Los resultados de esta secuenciación se interpretaron con el software Bionumerics 5.0 (bioMérieux, Marcy l'Etoile, France) [208].

Para la obtención de los espectros de masas mediante MALDI-TOF, los aislados se incubaron en atmósfera anaerobia en agar Brucella (Beckton Dickinson®, Franklin Lakes, NJ, US) a 37°C durante 48h para asegurar el crecimiento suficiente de las colonias. Las colonias se depositaron en los pocillos de la placa de MALDI-TOF y se cubrieron con 1µl ácido fórmico, seguido de 1µl de matriz HCCA. Los aislados se analizaron por duplicado en dos pocillos de la placa, de los que se realizaron 2 lecturas de cada uno, obteniendo un total de 4 espectros por cada cepa. La lectura se llevó a cabo en la región de análisis por defecto del MALDI-TOF, de 2000 a 20000 Da.

El análisis de los espectros se realizó mediante el software comercial Clover MSDAS y AutoCdiff, un software propio desarrollado en este estudio. Con el software Clover MSDAS se aplicaron diferentes algoritmos de ML (RF, KNN, PLS-DA, SVM y Light-GBM) tras un estudio previo de detección de picos biomarcadores en el que se incluyeron todos los picos con intensidades con AUC >0,7. Por otra parte, AutoCdiff es

un novedoso software desarrollado por nuestros colaboradores específicamente para este estudio, en abierto para uso público y construido mediante lenguaje Python, que emplea algoritmos bayesianos de ML como FA-VAE, DBLR-FS, RF o árboles de decisión para el ribotipado de cepas de *C. difficile*. Este software estudia el espectro completo subido por el usuario sin restricciones de picos biomarcadores, y analiza toda la región entre 2 y 20 kDa, aunque también permite la opción de seleccionar picos biomarcadores.

El estudio que aquí se presenta se desarrolló en tres experimentos sucesivos. El primero de ellos centró el foco en diferenciar los verdaderos RT027 de aquellos que el test Xpert® *C. difficile* BT clasifica como “presuntivos RT027”, empleando un conjunto de 100 aislados con ese resultado, de los que sólo 42 eran verdaderos RT027. Este modelo se validó con 31 aislados independientes pertenecientes a estas dos categorías.

En el segundo experimento el objetivo fue diferenciar los principales ribotipos toxigénicos del HGM (RT027 y RT181) de otros ribotipos directamente a partir del espectro de MALDI-TOF. Para ello se empleó la colección completa de 269 aislados formada por los ribotipos más prevalentes en España y Europa (n=131 aislados RT027 y presuntivos RT027 y n=138 de la categoría “otros ribotipos”). Estos aislados se clasificaron en tres categorías (RT027, RT181 y “Otros RTs”) y este modelo se validó con una colección de 79 aislados independientes pertenecientes a estas tres categorías.

Por último, se llevó a cabo un tercer experimento a tiempo real con aislados pertenecientes a dos brotes nosocomiales en la comunidad de Madrid, para evaluar la eficacia de estos modelos en un escenario real. Para ello se entrenaron los modelos previos con todos los aislados caracterizados (n=348) y se validaron con estas cepas prospectivas (n=15). El diseño completo del estudio y la colección de cepas empleada se resumen en la Tabla 3 a continuación.

Tabla 3. Relación de cepas empleadas en el estudio

	Total	Exp 1		Exp 2		Exp 3	
		Training	Test	Training	Test	Training	Test
RTs Oficiales:							
RT181	67	45	10	44	11	55	12
RT027	53	42	11	42	11	53	0
RT106	31	0	0	25	6	31	0
RT001	30	0	0	24	6	30	0
RT017	30	0	0	24	6	30	0
RT078	30	0	0	24	6	30	0
RT207	30	0	0	24	6	30	0
RT014	29	0	0	23	6	29	0
RT023	19	0	0	15	4	19	0
RT002	18	0	0	14	4	18	0
RT651	3	0	0	0	0	0	3
RT250	2	1	1	1	1	2	0
RT578	1	1	0	0	1	1	0
HGM RTs:							
165	6	4	2	4	2	6	0
166	4	1	3	3	1	4	0
170	1	0	1	0	1	1	0
171	1	0	1	0	1	1	0
173	1	1	0	0	1	1	0
174	1	1	0	0	1	1	0
175	1	0	1	0	1	1	0
176	1	1	0	0	1	1	0
187	1	1	0	0	1	1	0
UNK	3	2	1	2	1	3	0
Total	363	100	31	269	79	348	15

Exp: Experimento. **Training:** Entrenamiento. **Test:** Validación.

UNK: Ribotipo desconocido.

Los experimentos con Clover MSDAS confirmaron la presencia de siete picos de biomarcadores de interés con un AUC >0,7. Los picos específicos que diferenciaban el RT027 eran el 2463 *m/z* y el 4933 *m/z* mientras que el pico 4993 *m/z* estaba ausente en todos los RT027 y presente en el resto de RT estudiados incluido el RT181. Los picos que permitían diferenciar los RT027 y RT181 de los otros ribotipos fueron: 3353 *m/z*, 6187 *m/z*, 6651 *m/z* y 6710 *m/z*.

CAPÍTULO 4

Estos dos últimos picos (6651 m/z y 6710 m/z) correspondían a dos isoformas de la proteína L28 según la base de datos de proteínas ribosomales Ribopeaks (<https://lcad.deinfo.uepg.br/~ppgca/ribopeaks>), Uniprot (<https://www.uniprot.org>) y un artículo previo en la literatura [209].

Para diferenciar el RT027 del resto de presuntivos RT027 (entre ellos el RT181) mediante Clover MSDAS, en el primer experimento se incluyeron los picos 2463 m/z , 4933 m/z , 4993 m/z y 6651 m/z . El resto de picos no se incluyeron ya que eran comunes en los ribotipos RT027 y RT181. Todos los algoritmos estudiados presentaron resultados superiores al 88,0%, siendo Light-GBM el que ofrecía mejores resultados con un acierto del 96,6%. Con AutoCdiff se obtuvieron resultados >97,0% de acierto para todos los algoritmos. En la Figura 14 a continuación se resumen los resultados obtenidos en este experimento.

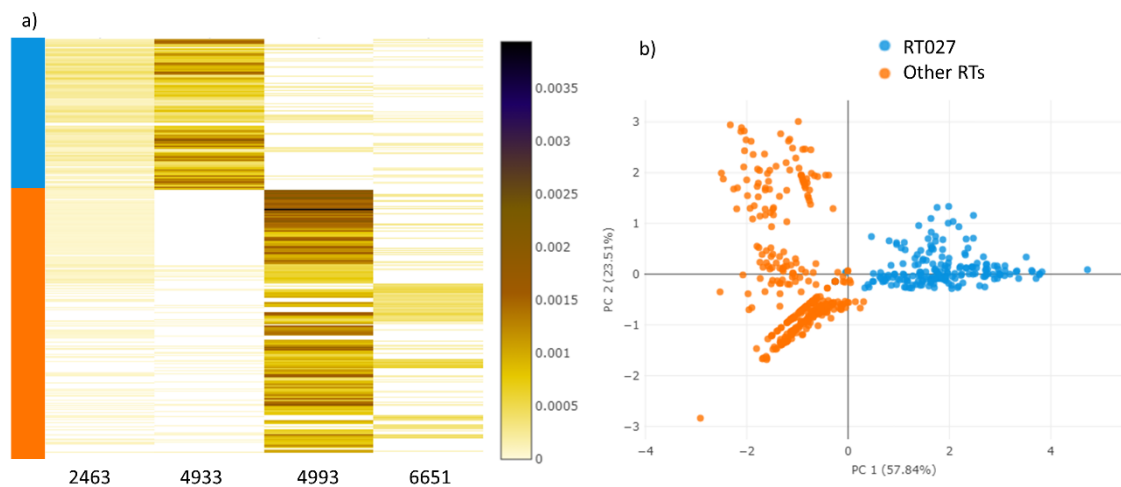


Figura 14. a) Mapa térmico de distribución de los cuatro picos seleccionados para el Experimento 1. b) Análisis de PCA con los picos seleccionados para la diferenciación de aislados RT027 de aislados pertenecientes a otros ribotipos

Para la diferenciación entre RT027, RT181 y el resto de ribotipos (presuntivos y no presuntivos 027) se añadieron a los modelos los tres picos restantes (3353 m/z , 6187 m/z y 6710 m/z) y se obtuvieron resultados >95,0% con Clover y >92,0% con AutoCdiff. En la Figura 15 se resumen los resultados obtenidos para este experimento.

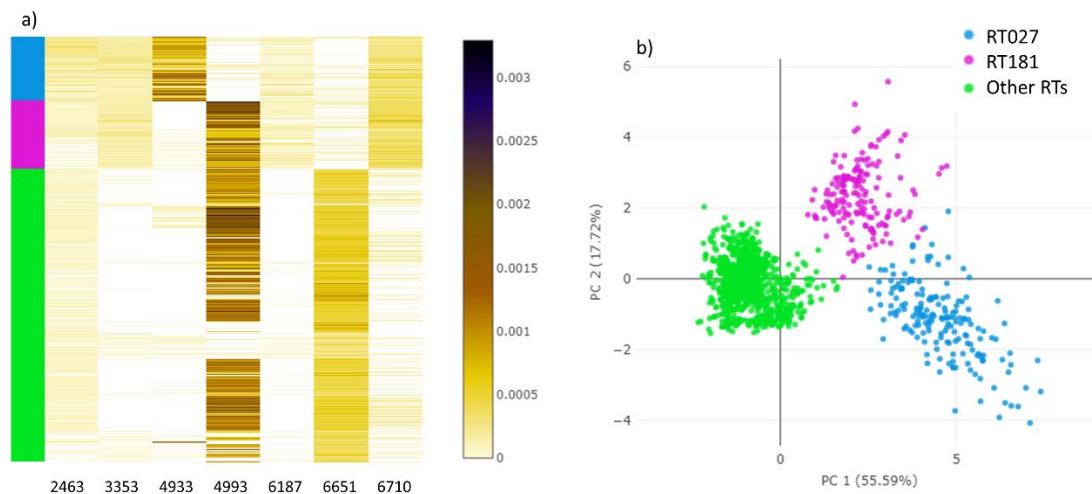


Figura 15. a) Mapa térmico de distribución de los siete picos seleccionados para el Experimento 2. b) Análisis de PCA con los 7 picos para la diferenciación de los aislados RT027 y RT181 de los aislados pertenecientes a Otros ribotipos.

Para el ribotipado de los dos brotes a tiempo real se ampliaron los modelos de ML estudiados previamente, añadiendo al conjunto de entrenamiento todas las cepas caracterizadas de los experimentos anteriores. Con estos modelos se estudiaron aplicando Clover MSDAS y AutoCdiff las cepas del brote del HCDGU, que se clasificaron como RT181 y las del HGM como “Otros ribotipos” (pertenecían al RT651). Los resultados de ribotipado confirmaron la correcta clasificación de estas cepas con un 100% de acierto con ambos software.

CAPÍTULO 4

Los resultados de todos los experimentos y modelos de ML estudiados se resumen en la Tabla 4 a continuación.

Tabla 4. Resumen de resultados de los algoritmos estudiados con Clover MSDAS y AutoCdiff en los tres experimentos llevado a cabo

Método	Algoritmo	Exp 1	Exp 2	Exp 3	Tiempo
Ribotipado por PCR	<i>Gold Estándar</i>	100%	100%	100%	48-96h
Clover MSDAS	RF	95,1%	99,7%	100%	1 h
	Light-GBM	96,6%	95,6%	100%	
	KNN	93,7%	96,7%	100%	
	SVM	92,9%	95,6%	100%	
	PLS-DA	88,4%	32,1%	100%	
AutoCdiff	DBLR-FS*	100%	99,9%	100%	1 h
	DBLR-FS	100%	99,5%	100%	
	RF	100%	99,9%	100%	
	DT	97,2%	94,5%	100%	
	FA-VAE	99,1%	92,0%	100%	

Exp: Experimento **DBLR-FS*:** Empleando picos biomarcadores

**RESUMEN GLOBAL DE RESULTADOS Y
DISCUSIÓN**

El informe O'Neill realizado por el gobierno de Reino Unido en 2016, estimó que para el año 2050 las muertes por infecciones causadas por bacterias resistentes a antibióticos superarán a las provocadas por el cáncer, convirtiéndose en la primera causa de mortalidad en el mundo [210]. Este informe también resalta que las muertes asociadas a infecciones por bacterias multirresistentes eran en el año 2014 aproximadamente 700000, sin embargo, más recientemente se ha estimado que en el año 2019 este número había crecido hasta el 1,27 millones de muertes atribuibles a la resistencia antibiótica [211]. Se ha calculado que esta cifra aumentará hasta los 10 millones en 2050.

La práctica clínica de la medicina actual (incluyendo cirugías, quimioterapia, trasplante de órganos, y otros procedimientos invasivos) sería inconcebible sin la existencia de antibióticos eficaces, por lo que se están promoviendo numerosas acciones enmarcadas en el enfoque de "Una Sola Salud" o "*One Health*" para preservar la actividad de los antimicrobianos activos disponibles [212]. Los planes de acción para combatir la resistencia antimicrobiana hacen especial hincapié en la importancia de la vigilancia, monitorización y control de los microorganismos multirresistentes a nivel ambiental, animal, comunitario y hospitalario. Esto es fundamental para evitar su diseminación y prevenir un aumento de las infecciones. Para ello, las pruebas de diagnóstico rápido basadas en proteómica y genómica para el diagnóstico de enfermedades infecciosas juegan un papel fundamental.

El rol del MALDI-TOF en el laboratorio de microbiología clínica ya se encuentra claramente instaurado para la identificación rutinaria de microorganismos a partir de cultivo convencional, habiendo desplazado a la mayoría de métodos fenotípicos y bioquímicos de identificación. Es una tecnología rápida y coste-efectiva que ha permitido adelantar de varios días a minutos el tiempo de espera para la identificación de microorganismos en el laboratorio, haciendo posible una mejora en la instauración del tratamiento antibiótico y en el control de la infección [213]. Además, también ha permitido que nuevos géneros y especies de bacterias sean ahora identificables, cuando esto no era posible empleando técnicas rutinarias bioquímicas [214].

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

La versatilidad de MALDI-TOF, junto con su rapidez y bajo coste, ha potenciado la exploración de nuevos campos de aplicación más allá de la identificación de microorganismos. Con el rápido desarrollo de las tecnologías bioinformáticas y la inteligencia artificial, estas nuevas aplicaciones parecen ir en aumento. El foco actual de la investigación en MALDI-TOF no se centra tanto en mejorar el funcionamiento de la técnica en sí, que ya está lo suficientemente optimizada si se siguen los criterios del fabricante, sino en la aplicación de algoritmos de *Machine Learning* para analizar los patrones de picos y encontrar similitudes o diferencias entre ellos según lo que se pretenda estudiar [215].

Esta tesis doctoral se compone de cuatro capítulos orientados a la aplicación de MALDI-TOF y FTIR para fines diferentes a la identificación rutinaria en el laboratorio de microbiología, pero de alta importancia para el control de la infección y prevención de brotes por microorganismos multirresistentes. Se centra en cuatro microorganismos de especial relevancia clínica destacados por la OMS y el CDC como patógenos prioritarios para la investigación: *E. faecium*, *E. cloacae* complex, *P. aeruginosa* y *C. difficile*.

En el primer capítulo, se plantea un estudio inicial de reproducibilidad para sentar las bases de la metodología que se debe emplear en estudios de tipado o análisis de resistencias mediante MALDI-TOF. Además, también se centra en la diferenciación mediante análisis de espectros de masas de cepas de *E. faecium* sensibles y resistentes a la vancomicina, y dentro de estas últimas, entre las portadoras del gen *vanA* y *vanB*.

Enterococcus faecium es un coco Gram-positivo comensal del tracto gastrointestinal, que bajo las condiciones adecuadas se puede volver patógeno y causar infecciones graves como bacteriemia, endocarditis, meningitis, pielonefritis e infecciones intraabdominales, entre otras. Está asociado en mayor medida con infecciones nosocomiales, especialmente en pacientes críticos e inmunodeprimidos [6, 7, 216].

En los últimos años, se ha observado en Europa un incremento de la resistencia transmisible a vancomicina en *E. faecium*, habiendo aumentado de un 13,4% en 2017 a un 17,2% en el año 2021 [217]. A la par, en Estados Unidos, se ha estimado que aproximadamente el 80% de los *E. faecium* son resistentes a vancomicina [216, 218]. En el año 2017, el VRE causó un total estimado de 54500 infecciones hospitalarias y 5400 muertes, siendo de todos los diagnósticos de infección nosocomial por *Enterococcus* spp., el 30% causadas por VRE, y en su mayoría, *E. faecium* resistente a vancomicina [199]. Esto genera preocupación a nivel mundial, ya que la vancomicina es uno de los tratamientos de elección frente a las infecciones causadas por este microorganismo [216, 218]. Además, su capacidad de colonización del tracto gastrointestinal, lo hace un patógeno que necesita una especial vigilancia en pacientes inmunodeprimidos, que puedan llegar a desarrollar una infección por este microorganismo si las condiciones adecuadas tienen lugar.

La resistencia a vancomicina en *Enterococcus* está mediada por 9 operones diferentes que se denominan *van* (*vanA*, *vanB*, *vanC*, *vanD*, *vanE*, *vanG*, *vanL*, *vanM* y *vanN*) [216, 219]. En *E. faecium*, los más prevalentes son *vanA* y *vanB* (96% y 4% respectivamente), y son los que se detectan en el laboratorio cuando se estudia la resistencia a vancomicina en *Enterococcus*. El genotipo VanA -portador del gen *vanA*- confiere resistencia simultánea a vancomicina y teicoplanina, mientras que el genotipo VanB -portador del gen *vanB*- únicamente presenta resistencia a vancomicina pero es sensible a teicoplanina [216, 220].

La detección de la resistencia a vancomicina en *Enterococcus* spp., se puede estudiar en el laboratorio de dos maneras. De manera convencional y rutinaria mediante cultivo y estudio de sensibilidad por microdilución, tiras de gradiente o difusión de discos con antibiótico en agar, o bien mediante detección genómica y amplificación de los genes *vanA* y *vanB* codificantes del mecanismo de resistencia a vancomicina (o secuenciando el genoma completo por secuenciación masiva) [6]. Ambas aproximaciones tienen sus ventajas y desventajas. El cultivo convencional aporta resultados fenotípicos de viabilidad bacteriana, pero se demora en el tiempo aproximadamente unas 48 horas, mientras que la detección genotípica de genes de resistencia, es una técnica rápida pero costosa en el caso de PCRs comerciales, y

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

laboriosa con necesidad de personal entrenado si se realiza una PCR casera o secuenciación masiva.

Es por ello que la detección de resistencia a vancomicina mediante el uso de MALDI-TOF puede suponer una mejora en el tiempo de espera, así como en el gasto económico, frente a los otros dos enfoques. El MALDI-TOF es una tecnología fácil de implementar para su uso rutinario, que actualmente ya se encuentra instaurada en la mayoría de laboratorios clínicos, aporta resultados en forma de un espectro de masas en cuestión de minutos y no necesita una formación extensa del personal que lo lleva a cabo.

En este primer artículo se estudian 178 aislados únicos de *E. faecium* recolectados durante los años 2017 a 2019 y pertenecientes a pacientes ingresados en el Hospital General Universitario Gregorio Marañón (HGM). Estos aislados se identificaron de manera rutinaria mediante MALDI-TOF a partir de muestras clínicas de infección y de colonización, todos ellos con un score >2,0. Su sensibilidad a vancomicina se estudió mediante microdilución en caldo con el sistema MicroScan (Beckman-Coulter) siguiendo las recomendaciones de EUCAST (v.11, 2021). Estos resultados se comprobaron con una PCR casera a tiempo real y esto se confirmó una segunda vez con la PCR comercial GeneXpert (Cepheid) para observar la presencia o ausencia de los genes *vanA* y *vanB*.

En total se incluyeron en el estudio 92 aislados VSE, 31 VRE VanA y 55 VRE VanB. Para obtener reproducibilidad técnica entre los espectros, cada cepa se analizó en dos pocillos diferentes (reproducibilidad interpocillo) y cada pocillo se leyó dos veces (reproducibilidad intrapocillo), lo que aportó un total de 4 espectros por cada aislado y día. Para estudiar la reproducibilidad biológica, esta misma metodología se realizó durante 3 días consecutivos.

Todos los espectros obtenidos se analizaron en el rango 2000-20000 Da, el rango por defecto de lectura de MALDI-TOF, y se calibraron frente al pico 4428 *m/z*, presente en todas las especies del género *Enterococcus* spp. [90, 219, 221]. Posteriormente y tras realizar un preprocesado inicial de suavizado (mediante el filtro Savitzky-Golay 11:3) y sustracción de la línea base (mediante el filtro Top-Hat, factor

0,02), se alinearon las réplicas de cada aislado y se crearon espectros media para cada pocillo y día de análisis.

Inicialmente se realizó un estudio de reproducibilidad siguiendo la metodología descrita por Oberle et al en 2016 [118]. Hasta el momento, no se han realizado más estudios acerca de la reproducibilidad de la técnica en las mismas condiciones y en un mismo laboratorio. Sí se ha estudiado la reproducibilidad de MALDI-TOF con diferentes medios de cultivo o condiciones de extracción e incubación, y en distintos centros, para diferentes objetivos como la identificación o la detección de mecanismos de resistencia, con resultados variables [80, 116, 222, 223].

Para este estudio inicial, se calculó el espectro media de las réplicas de cada pocillo, y el espectro media de ambos pocillos entre sí (espectro único de cada día: Día 1, Día 2 y Día 3). Con el espectro media del Día 1 de cada aislado se realizó una búsqueda de picos comunes que presentaran una intensidad superior al 1,0% del pico de máxima intensidad. Con estas condiciones se encontraron 18 picos biomarcadores comunes a los 178 aislados, todos ellos en la región de 2000 a 10000 m/z . El ensayo de reproducibilidad se realizó empleando estos 18 picos en 20 aislados seleccionados al azar (10 VSE, 5 VRE VanA y 5 VRE VanB) y estudiando el coeficiente de variación (% CV) de las intensidades de estos picos intrapocillo, interpocillo e interdía a partir de los espectros en bruto, y normalizados antes (TICp) y después (pTIC) de la búsqueda de estos picos.

Los menores valores de CV se observaron entre los espectros intrapocillo en comparación con los interpocillo e interdía. También se observó que la normalización de los espectros, ya fuera antes o después de la búsqueda de picos, reducía todos los CV (intra e interpocillo, e interdia) en comparación con los espectros en crudo. La menor variabilidad entre picos se observó en la región entre 3000 y 9000 m/z , apoyando la hipótesis descrita con anterioridad de que la región media del espectro es la más estable para el análisis de picos biomarcadores [224, 225].

Posteriormente, para el estudio de diferenciación entre aislados VSE y VRE VanA y VRE VanB mediante *Machine Learning*, se aplicaron tres algoritmos supervisados (SVM, RF y PLS-DA) para la separación de las categorías. Esto se realizó

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

con tres matrices de picos diferentes: la primera, empleando el espectro completo obtenido de la lectura mediante MALDI-TOF, que incluía todos los picos sin tener en cuenta las intensidades, y las otras dos matrices en las que se tuvo en cuenta solo los picos con intensidades superiores al 1% de la intensidad del pico más intenso del espectro, normalizando antes (TICp) y después (pTIC) de la búsqueda de picos (al igual que en el estudio de reproducibilidad).

Primeramente, se intentó una diferenciación inicial entre aislados sensibles y resistentes a vancomicina, seguida de una más específica que pretendía diferenciar entre aislados portadores del gen *vanA* de los portadores de *vanB* entre los aislados resistentes. Estos resultados se validaron siguiendo una validación k-fold con k=10. También se llevó a cabo una búsqueda de picos biomarcadores que pudieran potencialmente separar las 3 categorías estudiadas (sensible, VanA y VanB). Los mejores resultados se obtuvieron con SVM empleando el espectro completo, con el que se obtuvo un 80,9% de clasificaciones correctas en VSE y VRE. Para la diferenciación entre aislados VanA y VanB, el mejor resultado se obtuvo con PLS-DA empleando la matriz previamente normalizada antes de la búsqueda de picos (TICp), que clasificó los aislados con un acierto del 86,6%.

En el análisis de picos biomarcadores se encontraron dos picos potencialmente diferenciadores. El pico 5095m/z presente en 82 de las 86 las cepas resistentes de este estudio (tanto VanA como VanB), y el pico 6891m/z, cuya intensidad era marcadamente mayor en las cepas VanA que en las VanB y permitió diferenciarlas entre sí.

Varios estudios han analizado la resistencia a vancomicina en *E. faecium* hasta el momento, con resultados variables. El pionero de todos fue Griffin en 2012 [85], que diferenció el genotipo VanB (el más prevalente en Australia, lugar de realización del estudio) de *E. faecium* sensibles con una sensibilidad y especificidad de 92,4% y 85,2% respectivamente. Posteriormente, en 2014 dos estudios independientes demostraron que *E. faecium* VanA podía diferenciarse de cepas sensibles mediante MALDI-TOF con una sensibilidad mayor del 90% en ambos casos [226, 227]. Estos estudios iniciales

supusieron una prueba de concepto, ya que el número de cepas que emplean es limitado, y no todos los mecanismos de resistencia se encontraban representados.

Varios estudios se han centrado en la detección de brotes causados por *E. faecium* resistente a vancomicina, y en la separación de complejos clonales. En 2016, Schlebusch y colaboradores tiparon un brote de *E. faecium* VanB mediante MALDI-TOF con unos resultados similares a los obtenidos por secuenciación masiva, en los que se pudo observar que los aislados pertenecientes a un brote se agrupan en un mismo clúster [228]. Freitas et al y Lasch et al fueron capaces de diferenciar clones según su pertenencia a brotes hospitalarios (de alto riesgo) o comunitarios (aislados a partir de humanos sanos o animales) [90, 225]. Un estudio en Turquía en 2018 [229] y otro en Noruega en 2019 [230], comparó el poder discriminativo de MALDI-TOF frente a PFGE para el tipado de brotes de *E. faecium* resistente a vancomicina, obteniendo resultados comparables entre ambas técnicas y presentando la ventaja de la rapidez de MALDI-TOF frente al campo pulsado.

Por último, se han publicado estudios más recientes que emplean algoritmos de *Machine Learning* con un elevado número de espectros (>3000) que incluyen tanto cepas resistentes como sensibles y que presentan resultados similares a los obtenidos en nuestro estudio, con sensibilidades cercanas al 80% [231, 232]. También se ha publicado un novedoso estudio que emplea la tecnología FTIR para el tipado de un brote por *E. faecium* resistente a vancomicina, con resultados superiores al tipado mediante MLST [233].

El estudio que se presenta en esta tesis supone una primera aproximación para el empleo de MALDI-TOF en el estudio de mecanismos de resistencia mediante algoritmos de *Machine Learning*. La ventaja que presenta este estudio frente a otros previos es una mayor precisión a la hora del análisis de picos, ya que no se realiza de manera visual sino mediante algoritmos matemáticos. Además, el estudio de reproducibilidad que se lleva a cabo en este artículo ha servido de base a la plataforma bioinformática CloverBiosoft para el desarrollo de un módulo automatizado de reproducibilidad incluido en el software Clover MSDAS para estudios con espectros de masas, que ahora ya se realiza de manera automatizada.

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

Al tratarse de un estudio preliminar, presenta una serie de desventajas que en estudios posteriores de nuestro grupo se han intentado corregir. La primera de todas es que la caracterización genómica se realizó únicamente mediante una PCR a tiempo real, pero sin el abordaje más completo que aportaría la secuenciación masiva. Se descartó que las cepas incluidas en el estudio se trataran de cepas pertenecientes a un mismo brote analizando la procedencia de todos los aislados, la unidad de aislamiento en el hospital y la fecha de aislamiento. Además, durante ese período no se detectó un aumento considerable de infecciones causadas por *E. faecium* resistente a vancomicina en comparación con otros períodos en los seguimientos realizados por los servicios de Medicina Preventiva y Microbiología, lo que excluía que fuera sugestivo del inicio de un brote. Sin embargo, no se realizó un estudio genómico de SNPs al respecto por lo que estos resultados no se pueden asegurar al 100%.

En cuanto al análisis de biomarcadores, todos los estudios presentados con anterioridad seleccionan una serie de picos que para su propio análisis son válidos y permiten la diferenciación entre cepas sensibles y resistentes, pero que en nuestro estudio no hemos encontrado como potenciales biomarcadores. Por este motivo, los picos biomarcadores encontrados en nuestro estudio para diferenciación de VRE y VSE (5095 *m/z* y 6891 *m/z*) deben emplearse con precaución para la diferenciación entre cepas, ya que puede tratarse de picos presente solo en un complejo clonal y no en todos los VRE.

El pico 5095 *m/z* es un pico ya descrito con anterioridad como marcador de resistencia a vancomicina en otro artículo anterior [85]. Sin embargo, este pico se estudió con posterioridad en otro trabajo en el que se caracterizó y se observó que pertenecía a la proteína secretora hiracina, no relacionada con la resistencia a vancomicina, por lo que su validez como marcador de resistencia se encuentra actualmente bajo debate [89]. El otro pico encontrado, el 6891 *m/z*, se había descrito con anterioridad como pico exclusivo de la especie *E. faecium* [234], algo que se cumple también en este trabajo, ya que el pico se encuentra presente en todos los aislados, y es su diferencia en intensidad lo que permite la diferenciación entre aislados VanA y VanB.

Además, al tratarse de un estudio modelo que emplea diferentes algoritmos de *Machine Learning* para la separación de categorías, todos los aislados incluidos se han estudiado como conjunto de entrenamiento, y no se ha realizado una validación externa de los datos con un número superior de aislados, únicamente una validación cruzada con los propios espectros del conjunto de entrenamiento. El segundo estudio de validación externa se está llevando a cabo actualmente en el HGM con nuevas cepas aisladas en años posteriores.

Este estudio nos ha permitido sentar las bases del análisis de espectros de MALDI-TOF en cuanto a estudios de reproducibilidad, y automatizarlos para posteriores estudios, así como llevar a cabo un primer abordaje empleando ML para el análisis de espectros y diferenciación entre aislados sensibles y resistentes.

En el segundo capítulo el objetivo es caracterizar y diferenciar las distintas especies incluidas dentro del complejo de *Enterobacter cloacae* complex (ECC) empleando espectrometría de masas junto con *Machine Learning*; un problema que no es posible resolver mediante el uso de técnicas rutinarias de identificación en el laboratorio, como el MALDI-TOF o la secuenciación de ARNr 16S.

El género *Enterobacter* spp. pertenece a la familia *Enterobacteriaceae*. Está compuesto por un grupo de patógenos que presentan especial relevancia en infección nosocomial y diseminación de clones a nivel hospitalario, lo que ha hecho que se incluyan dentro del grupo de bacterias ESKAPE prioritario de la OMS [198, 200].

Este género lo componen bacterias presentes en el medio ambiente (agua, suelo, plantas...) que colonizan el tracto gastrointestinal y pueden causar una gran variedad de cuadros infecciosos entre los que destacan neumonía, infección de piel y partes blandas, bacteriemia e infecciones vasculares, y meningitis postquirúrgica [7].

Estas bacterias además son capaces de colonizar numerosas superficies y diseminarse por el ambiente hospitalario. Se han descrito brotes por contaminación de soluciones intravenosas, nutriciones parenterales, instrumental médico, respiradores e incubadoras entre otros [235]. Puede afectar a adultos inmunodeprimidos y con estancias prolongadas en unidades de cuidados intensivos (UCI) o tratados

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

extensamente con antibióticos, y también a neonatos y bebés prematuros de bajo peso [216, 236].

La mayoría de los brotes descritos en la literatura, especialmente los provocados por el complejo *Enterobacter cloacae*, son en unidades neonatales [236, 237]. Se cree que esto es debido a que la colonización temprana del tracto gastrointestinal de los neonatos por esta bacteria puede hacerles predisponentes a la infección, aunque el papel de las UCI neonatales también jugaría un rol importante como reservorio, y la administración de antibióticos de amplio espectro podría ser un factor desencadenante [238].

Las bacterias del género *Enterobacter* spp. presentan resistencia intrínseca a ampicilina, amoxicilina, amoxicilina-clavulánico y cefalosporinas de primera y segunda generación. También presentan un AmpC constitutivo inducible que se puede “hiperexpresar”, lo que hace que no sea recomendado su tratamiento con cefalosporinas de tercera generación, aunque in vitro se observe sensibilidad. Además, pueden desarrollar resistencia a otros β -lactámicos por adquisición de BLEE y carbapenemasas [216]. Recientemente se ha publicado un estudio español, en el que se alerta del aumento de la producción de carbapenemasas en ECC, que ya suponen el 25% de todas las carbapenemasas detectadas en *Enterobacterales* en Andalucía [239]. Debido a que su tratamiento puede verse limitado por todos estos factores, su correcta identificación y control en el ambiente hospitalario son esenciales para evitar su diseminación.

El género *Enterobacter* spp. lo conforman varias especies, pero su taxonomía está en constante evolución y ha sufrido múltiples reclasificaciones en los últimos años. El *Gold Estándar* actual de identificación intracomplejo puede considerarse el MLSA, en el que se secuencian de 4 a 7 genes conservados que permiten la diferenciación de las especies entre sí (entre ellos *atpD*, *fusA*, *infB*, *gyrB*, *leuS*, *dnaA*, *pyrG*, *rplB*, *rpoB*, o *hsp60*), y que es el método que más se correlaciona con WGS y el análisis de SNPs [236]. Uno de estos genes es el *hsp60*, cuya secuenciación es la más extendida y aceptada para la diferenciación de las especies de ECC y es la metodología que emplearon Hoffmann y Roggenkamp para la clasificación más comúnmente

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

utilizada de 13 clústeres [240]. La secuenciación convencional de ARNr 16S no permite diferenciar las especies de ECC entre sí [236, 241]. Recientemente se ha propuesto un nuevo gen (*dnaJ*) para la diferenciación de las especies de ECC que codifica una proteína chaperona específica de la especie *E. cloacae* [242, 243].

Con el desarrollo de la secuenciación masiva, se ha hecho una redistribución de especies que antes no era posible empleando solo la secuenciación del gen *hsp60* o el ARNr 16S. Así, por ejemplo, *Enterobacter aerogenes* ahora se ha incluido dentro del género *Klebsiella* y se le denomina *Klebsiella aerogenes* [244], al igual que *Enterobacter sakazakii* que ahora se conoce como *Cronobacter sakazakii* [245]. Otras especies dentro del género *Enterobacter* spp. se han reclasificado también en nuevos géneros que se resumen en la Tabla 5 [216, 246, 247].

Tabla 5. Nueva clasificación propuesta de especies de *Enterobacter* spp.

Antigua clasificación	Nueva clasificación propuesta
<i>E. amnigenus</i> y <i>E. nimipressuralis</i>	Género <i>Lelliottia</i> spp.
<i>E. agglomerans</i>	<i>Pantoea agglomerans</i>
<i>E. intermedium</i>	<i>Kluyvera intermedia</i>
<i>E. sakazakii</i>	<i>Cronobacter sakazakii</i>
<i>E. aerogenes</i>	<i>Klebsiella aerogenes</i>
<i>E. cowanii</i> , <i>E. arachidis</i> , <i>E. oryzae</i> , <i>E. oryzendophyticus</i> , <i>E. oryziphilus</i> , <i>E. radicincitas</i> , <i>E. sacchari</i>	Género <i>Kosakonia</i> spp.
<i>E. helveticus</i> y <i>E. pulveris</i>	Género <i>Franconibacter</i> spp.
<i>E. gergoviae</i> y <i>E. pyrinus</i>	Género <i>Pluralibacter</i> spp.
<i>E. massiliensis</i>	<i>Metakosakonia massiliensis</i>
<i>E. turicensis</i>	<i>Siccibacter turicensis</i>
<i>E. timonensis</i>	Género <i>Pseudenterobacter</i> spp.

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

Dentro del género *Enterobacter* spp., el ECC está formado por 13 clústeres según la secuenciación del gen *hsp60*. Este complejo también se ha ido actualizando en los últimos años. Inicialmente y tras la publicación de la clasificación de ECC en 13 clústeres, se añadieron nuevas especies al complejo según estudios de hibridación de DNA y pruebas bioquímicas, como las subespecies de *E. cloacae*, *E. cloacae* subsp. *cloacae* y subsp. *dissolvens* (incluidas en los clústeres XI y XII de Hoffmann) [248], las subespecies de *Enterobacter hormaechei*, *E. hormaechei* subsp. *oharae*, *E. hormaechei* subsp. *hormaechei* y *E. hormaechei* subsp. *steigerwaltii* (clústeres VI, VII y VIII respectivamente) [249] y *Enterobacter ludwigii*, incluido en el clúster V [250].

Con la irrupción de las nuevas técnicas de secuenciación masiva en el laboratorio, también se han descubierto nuevas especies pertenecientes al complejo en los últimos años. Entre ellas destaca el *Enterobacter bugandensis* (clasificado como clúster IX) [251], *Enterobacter roggenkampii* (clúster IV) y nuevas subespecies de *Enterobacter hormaechei* (*E. hormaechei* subsp. *hoffmannii* y subsp. *xiangfangensis*, clústeres III y VI respectivamente) [252].

Otras especies, sin embargo, se han eliminado del complejo e incluso cambiado de género según los resultados obtenidos por secuenciación masiva. En la última actualización de la taxonomía en 2020 [247], se han descrito dos nuevas especies, *E. quasiroggenkampii* y *E. quasimori*, y las subespecies *E. cloacae* subsp. *dissolvens* y *E. hormaechei* subsp. *hoffmannii* han pasado a considerarse nuevas especies (*E. dissolvens* y *E. hoffmannii*). También se han reclasificado las subespecies *E. hormaechei* subsp. *oharae*, *xiangfangensis* y *steigerwaltii* en una nueva especie, *E. xiangfangensis*.

Dentro del complejo ECC, la tecnología MALDI-TOF no es capaz de diferenciar correctamente las distintas especies que lo componen. El segundo artículo que se presenta en esta tesis se centra en la discriminación automática de las especies dentro del complejo ECC mediante el uso de la tecnología MALDI-TOF y algoritmos supervisados de *Machine Learning*.

La importancia de la separación entre las diferentes especies del complejo ECC radica en que algunas de ellas están implicadas de manera más relevante que otras en determinados tipos de infección [253], como por ejemplo *E. hormaechei*, la especie más identificada en hospitales y que presenta una gran capacidad para adquirir mecanismos de resistencia y factores de virulencia [254], y *E. bugandensis*, que se ha descrito como una importante entidad productora de infecciones graves y brotes en unidades neonatales [235, 251].

Una identificación errónea de las especies incluidas dentro del ECC no supondría un problema grave en cuanto a la elección de antibioterapia, ya que todas las especies dentro del complejo presentan los mismos patrones de sensibilidad intrínsecos; sin embargo, sí difieren en la producción de factores de virulencia o su capacidad de provocar brotes, y de ahí la importancia de su diferenciación [236, 238]. Para la realización de este estudio se incluyeron un total de 219 aislados de ECC: 164 procedentes del Hospital Ramón y Cajal de Madrid (UHRC), 34 del Hospital de Basel en Suiza (UHB), y 21 del Hospital Universitario Gregorio Marañón de Madrid (HGM) y 118 aislados de *Klebsiella aerogenes* (9 aislados del UHRC, 107 del UHB y 2 del HGM).

La caracterización genómica de los aislados se realizó de dos maneras diferentes según la procedencia de las cepas, las cepas pertenecientes al UHB se analizaron mediante secuenciación masiva, mientras que los aislados del UHRC y HGM por secuenciación del gen *hsp60*, por lo que para la clasificación de las especies dentro del complejo ECC se siguió el modelo propuesto por Hoffmann y Roggenkamp de 13 clústeres, pero actualizado hasta la fecha del estudio con las nuevas subespecies. La clasificación empleada en este estudio se resume en la Tabla 6. Actualmente esta distribución se ha modificado y ya no se encuentra actualizada debido a nuevos estudios con secuenciación masiva.

Tabla 6. Clasificación de las especies de ECC empleada en el estudio

Clúster	Especies
I	<i>E. asburiae</i>
II	<i>E. kobei</i>
III	<i>E. hormaechei</i> subsp. <i>hoffmannii</i>
IV	<i>E. roggenkampii</i>
V	<i>E. ludwigii</i>
VI	<i>E. hormaechei</i> subsp. <i>xiangfangensis</i> y subsp. <i>oharae</i>
VII	<i>E. hormaechei</i> subsp. <i>hormaechei</i>
VIII	<i>E. hormaechei</i> subsp. <i>steigerwaltii</i>
IX	<i>E. bugandensis</i>
X	<i>E. nimipressuralis</i>
XI	<i>E. cloacae</i> subsp. <i>cloacae</i>
XII	<i>E. cloacae</i> subsp. <i>dissolvens</i>
XIII	Grupo heterogéneo de aislados

Para el análisis mediante MALDI-TOF, los aislados del UHRC y del HGM se procesaron en dos pocillos de la placa de MALDI que se leyeron por duplicado, obteniendo un total de 4 espectros por cada cepa, mientras que para los aislados del UHB se dispuso solo del espectro obtenido en la identificación rutinaria del laboratorio, por lo que solo se incluyó un espectro por cada cepa. El preprocesado de los espectros se realizó con el software Clover MSDAS al igual que en el artículo previo de *Enterococcus*, con el *pipeline* habitual de suavizado (con el filtro Savitzky-Golay 11:3) y sustracción de la línea base (con el filtro Top-Hat 0,02), seguido de la creación de un espectro medio, un alineamiento y normalización TIC.

Inicialmente, para estudiar la viabilidad de la aplicación de MALDI-TOF para la separación de especies dentro del complejo ECC se realizó un estudio no supervisado en el que se construyó un dendrograma mediante análisis de componentes principales (PCA) con una muestra representativa hasta alcanzar n=18 de cada una de las 6 especies más comunes dentro del complejo (*E. asburiae*, *E. kobei*, *E. hormaechei*, *E.*

roggenkampii, *E. ludwigii* y *E. bugandensis*) y *K. aerogenes*. En las situaciones en las que esto no fue posible por falta de aislados, se realizó un proceso de *oversampling* o sobremuestreo, en el que se replicaron espectros para evitar un posible sobreajuste posterior del modelo. En el dendrograma construido se pudieron observar 3 clústeres diferenciados, el de *E. hormaechei*, el de *K. aerogenes* y un tercero que agrupaba el resto de las especies. Estos resultados iniciales demostraron que era posible una diferenciación de las especies dentro del complejo, con la ayuda de algoritmos supervisados.

Para ello, posteriormente se realizó un estudio supervisado en el que se aplicaron diferentes algoritmos para la diferenciación de las especies más comunes. El total de aislados se repartió en cuatro grupos: el conjunto de entrenamiento (únicamente con cepas del UHRC), el conjunto de validación interna (cepas de UHRC no empleadas en el entrenamiento) y dos conjuntos de validación externa, uno con cepas del UHB y otro con cepas del HGM.

Debido a la falta de aislados para la validación de *E. ludwigii* y *E. bugandensis*, estas dos especies no se incluyeron en el desarrollo y validación de los modelos supervisados, por lo que esta parte del estudio se centró en diferenciar 4 especies dentro del complejo ECC (*E. asburiae*, *E. kobei*, *E. hormaechei* y *E. roggkampii*) y *K. aerogenes* mediante la aplicación de 4 modelos diferentes de *Machine Learning* (PLS-DA, SVM lineal y radial, y RF). Estos resultados se comprobaron internamente con una validación cruzada (*cross-validation*) de $k=5$, cuyos resultados fueron: 60,4% para PLS-DA, 52,7% para SVM-L, 100% para SVM-R y 100% para RF. De acuerdo con estos resultados, se decidió solo considerar los algoritmos SVM-R y RF (algoritmos no lineales) para el análisis posterior de validación interna y externa.

Se realizaron tres validaciones de los algoritmos SVM-R y RF. La primera de ellas, una validación interna con los aislados del UHRC que no se emplearon para el entrenamiento del modelo, seguida de una validación externa con aislados independientes del UHB, y por último, una segunda validación externa con nuevos aislados procedentes del HGM que se enviaron al UHB y al UHRC para validación y estudio de reproducibilidad del empleo de estos algoritmos.

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

En el primer estudio de validación, los resultados con las cepas del UHRC que no se emplearon en el entrenamiento fueron idénticos con los dos algoritmos estudiados, RF y SVM-R, con un porcentaje global de acierto del 96,8%. Todos los aislados de *K. aerogenes* (n=3), *E. asburiae* (n=1) y *E. kobei* (n=9) fueron correctamente clasificados en esta validación, mientras que de los aislados de *E. hormaechei* (n=51), y *E. roggenkampii* (n=62), 1 y 3 aislados fueron mal clasificados respectivamente, dando un total de acierto del 98,0% y 95,2% para estas especies.

Dado que ambos algoritmos obtuvieron iguales resultados, se decidió validarlos externamente con cepas del UHB (33 *E. hormaechei*, 1 *E. roggenkampii* y 107 *K. aerogenes*). El algoritmo que obtuvo mejores resultados de validación externa fue el RF con un porcentaje de acierto del 96,5%, mientras que el SVM-R obtuvo un 83,1%, ya que no clasificó con precisión los aislados de *E. hormaechei*, acertando sólo en 15 aislados.

Finalmente se realizó una última validación externa por triplicado para estudiar la reproducibilidad con nuevas cepas procedentes del HGM (2 *K. aerogenes*, 3 *E. asburiae*, 15 *E. hormaechei*, 2 *E. kobei* y 1 *E. roggenkampii*), que fueron analizadas en los tres centros participantes. Con estas cepas se obtuvieron resultados del 100% de clasificación correcta con RF para todas las especies testadas y en todos los centros, y para el algoritmo SVM-R, se obtuvo un 100% de clasificación correcta en todos los centros, excepto en el HGM donde se clasificó erróneamente un único aislado de *E. kobei* como *E. roggenkampii* (95,6% de clasificación correcta).

Por último, los resultados de estos algoritmos de clasificación se compararon con la nueva base de datos online MSI (<https://msi.happy-dev.fr>) que clasifica los aislados del ECC también según la secuenciación del gen *hsp60* (no incluye *K. aerogenes* ya que actualmente ya no forma parte del ECC). Esta base de datos es una herramienta totalmente libre y online desarrollada por un grupo de investigación francés con 42 aislados de ECC [255]. En ella se pueden cargar los espectros obtenidos en la lectura de MALDI-TOF, y por comparación con la base de datos, emite unos resultados de identificación de la cepa que se ha subido a la plataforma.

Para la identificación con esta base de datos se emplearon todos los aislados de la validación interna del UHRC y externa del UHB excepto los aislados identificados como *K. aerogenes* (1 *E. asburiae*, 84 *E. hormaechei*, 9 *E. kobei* y 63 *E. roggenkampii*). Se obtuvieron resultados similares a nuestros algoritmos, con un 94,9% de acierto en la clasificación de los aislados.

Muy pocos artículos científicos publicados han intentado abordar el problema de la separación de las especies del ECC mediante espectrometría de masas hasta ahora. El primer artículo en el que se intentó la separación de las especies con MALDI-TOF, se publicó en 2011 por Pavlovic et al, con resultados no satisfactorios en comparación con la secuenciación del gen *dnaJ* [242]. En este estudio pudieron diferenciar el ECC de otras especies de *Enterobacter* spp., pero no identificar correctamente las especies dentro del complejo.

Posteriormente, en el año 2015, en un estudio con cepas de Argelia y Francia se intentó tipar mediante MALDI-TOF cepas de ECC [256]. Empleando los espectros de MALDI-TOF construyeron un dendrograma en el que pudieron observar 3 clústeres diferenciados: C1 que incluía todos los aislados de *E. aerogenes* (actual *K. aerogenes*) procedentes de Francia, C2 que incluía 29 aislados de ECC procedentes también de Francia, y C3 en el que se mezclaban 12 ECC de Francia y el total de los 27 aislados de ECC de Argelia. Lo resaltable de este estudio es que todas las cepas de *K. aerogenes* se agruparon en un mismo clúster (lo que a su vez justifica su posterior separación taxonómica del complejo ECC y del género *Enterobacter* spp.), y que los ECC se distribuyeron según su procedencia. Sin embargo, no es extrapolable a la diferenciación de especies de ECC ya que del total de 77 aislados incluidos, 63 de ellos eran *Enterobacter cloacae* y 9 *K. aerogenes*, por lo que el resto de especies dentro del complejo se encontraba infrarrepresentado (solo 5 aislados en total).

En los últimos años, al estudiarse más a fondo la importancia de ECC en la infección nosocomial y tras la publicación de varios brotes por determinadas especies en el entorno hospitalario, ha crecido el interés por caracterizar el complejo y diferenciar las especies que lo componen entre sí.

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

En 2018, de Florio y colaboradores [257] intentaron la separación de especies mediante MALDI-TOF, pero nuevamente, no todos los aislados de ECC estaban representados, ya que solo se incluyeron cepas de *E. cloacae* (el 63% de los aislados), *E. asburiae* y *E. aerogenes* (*K. aerogenes*). Además, estas cepas se identificaron únicamente con MALDI-TOF (de dos casas comerciales: MALDI Biotyper de Bruker y VITEK de Biomerieux), pero no se compararon estos resultados de identificación con otras técnicas como el MLSA o la secuenciación de *hsp60*, considerando los resultados correctos únicamente por el resultado del score de identificación de MALDI-TOF. En este estudio y en concordancia con el artículo publicado en 2015 de Argelia y Francia, también son capaces de separar *K. aerogenes* de ECC, y a los aislados en diferentes clústeres por su procedencia de distintas áreas del hospital [256].

Además, se ha estudiado la tecnología de espectroscopía de infrarrojos en un artículo publicado recientemente, en el que el objetivo era el uso del FTIR para realizar un estudio epidemiológico de las cepas circulantes de ECC en una unidad de UCI neonatal. Para ello los autores compararon la capacidad de tipado de FTIR frente a WGS, con resultados moderados de concordancia entre ambos métodos, y agrupación conjunta en el dendrograma de los aislados de un mismo paciente, pero con variabilidad en el tiempo. Para mejorar esto y conseguir una mejor agrupación de los aislados, deciden entrenar redes neuronales artificiales (ANN) con la información sobre el ST y obtienen una clasificación correcta del 90%. Esta metodología la aplican a un brote posterior por ECC, en el que obtienen los mismos resultados por FTIR que por análisis de SNPs por WGS [166]. Otro estudio coetáneo analiza 23 aislados de ECC mediante FTIR y obtiene una correcta clasificación en clústeres de los STs, salvo por cuatro aislados pertenecientes a dos STs diferentes clasificados erróneamente [159].

A la vez, otros estudios han intentado actualizar y mejorar las bases de datos comerciales mediante la inclusión de nuevas cepas de ECC, como es el caso de la base de datos de MSI con la que comparamos nuestros resultados [255], o la desarrollada por Wang y colaboradores con 87 nuevos aislados pertenecientes a diferentes clústeres de ECC [258]. Las bases de datos comerciales de Bruker también se han ido actualizando periódicamente, añadiendo espectros de más cepas de ECC, y las

versiones más recientes han mejorado significativamente en la distinción de las especies dentro del complejo.

El artículo presentado en esta tesis supone una novedad en el campo ya que incluye una representación amplia y balanceada de los aislados más significativos del complejo, y emplea nuevas tecnologías de *Machine Learning* fáciles de utilizar que mejoran la diferenciación rutinaria de las especies del ECC mediante MALDI-TOF.

El tercer capítulo se centra en la diferenciación y tipado mediante espectrometría de masas MALDI-TOF y FTIR de un brote provocado por el clon de alto riesgo ST175 de *Pseudomonas aeruginosa*, que tuvo lugar en la unidad de hematología del hospital HGM.

Pseudomonas aeruginosa es un bacilo Gram-negativo no fermentador, que por su capacidad de desarrollar numerosos mecanismos de resistencia se puede convertir con facilidad en una bacteria MDR/XDR e incluso PDR según la definición de Magiorakos [259]. Estas definiciones de clasificación de bacterias según su resistencia a día de hoy se han actualizado debido a la comercialización de nuevos antibióticos, y se acepta mejor el término DTR o microorganismos difíciles de tratar, cuando son resistentes a todos los β -lactámicos incluyendo carbapenémicos y a fluoroquinolonas; reservando para último recurso antibióticos con mayor toxicidad, como la colistina o los aminoglucósidos [260, 261]. En esta nueva categoría también fácilmente puede encontrarse *P. aeruginosa*. Su capacidad de producir β -lactamasas y carbapenemasas (sobre todo metalo- β -lactamasas), la pérdida de porinas en su pared, la sobreexpresión de bombas de eflujo que expulsan los antibióticos, la hiperexpresión de su AmpC cromosómica, la producción de biofilm o la mutación de sus PBP, la hacen un microorganismo que necesita una especial vigilancia y control epidemiológico, con medidas de aislamiento de contacto en los pacientes colonizados o infectados por cepas multirresistentes [262].

Sus distintivas características de sensibilidad, propiciadas sobre todo por la baja permeabilidad de su membrana externa, hacen que en numerosas ocasiones, si se trata con antibióticos β -lactámicos, tenga que ser a dosis altas, en perfusión continua o

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

prolongada y tras una dosis inicial de carga (lo que se representaría con la nueva categoría 'I' de EUCAST) [263].

A pesar de ser un patógeno ambiental, es de especial relevancia por causar infecciones graves en pacientes inmunodeprimidos (especialmente hematológicos y trasplantados), y por su gran capacidad de mutar en clones XDR/DTR de alto riesgo y diseminarse causando brotes hospitalarios [262]. Puede provocar bacteriemia, neumonía, infección urinaria, abscesos e infecciones de piel y partes blandas, sobre todo en pacientes hospitalizados o institucionalizados. También es un patógeno frecuente en otitis e infecciones oculares como queratitis o úlceras corneales. Los pacientes con fibrosis quística y EPOC, por sus especiales características pulmonares, suelen estar colonizados por cepas de *P. aeruginosa* que en la mayoría de los casos presentan un fenotipo de multirresistencia [216].

P. aeruginosa es el principal microorganismo causante de infección nosocomial y neumonía asociada a ventilación mecánica, por su prevalencia en el hábitat hospitalario y posibilidad de diseminación a través de fómites como los respiradores, catéteres, sumideros y demás material médico presente en el hospital [264, 265]. La capacidad de desarrollo y adquisición de nuevos mecanismos de resistencia de *P. aeruginosa* han hecho que sea incluida también en el grupo ESKAPE, y que se necesite una especial vigilancia de este patógeno en el entorno hospitalario [198, 200].

En el artículo que se presenta en esta tesis el objetivo es tipar, empleando la tecnología MALDI-TOF en combinación con FTIR, un brote de *P. aeruginosa* MDR causado por el clon ST175 en la unidad de hematología del hospital HGM durante un período de 15 meses, entre los años 2013 y 2014.

Durante este período de tiempo se observó un incremento de incidencia de *P. aeruginosa* en esta unidad, que no se detectó a nivel global en el hospital. En este tiempo, se detectó al menos 1 aislado de *P. aeruginosa* con el mismo fenotipo MDR (resistencia a carbapenémicos, fluoroquinolonas y aminoglucósidos) en 14 pacientes, mientras que en los meses previos no se había aislado ninguna *P. aeruginosa* con este perfil de sensibilidad en el servicio de hematología.

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

Los aislados se identificaron mediante MALDI-TOF de manera rutinaria, y para la caracterización molecular de las cepas en un estudio preliminar se realizó PFGE a las cepas sospechosas de pertenecer al brote y a controles externos [266]. En el análisis de estos aislados se observó que se repartían en varios pulsotipos, siendo el pulsotipo P1 el que agrupaba las cepas pertenecientes al brote.

Las cepas de este pulsotipo se secuenciaron mediante WGS, y se pudo observar 3 clústeres: El grupo 1 que incluía todos los aislados pertenecientes al auténtico brote, y los grupos 2 y 3, que se consideraron aislados externos al brote por distancia de SNPs. Los controles externos incluidos se agruparon a una distancia superior a 5000 SNPs del resto (Figura 16). Con los SNPs que separaban el grupo 1 (considerado el auténtico brote) de los demás, se diseñó una ASO-PCR para poder diferenciar aislados del brote del resto aislados. Esta ASO-PCR se testó con cepas aisladas durante un período de tiempo más amplio (de 2010 a 2018), permitiendo la detección de nuevos aislados pertenecientes al brote que no se habían detectado previamente [266].

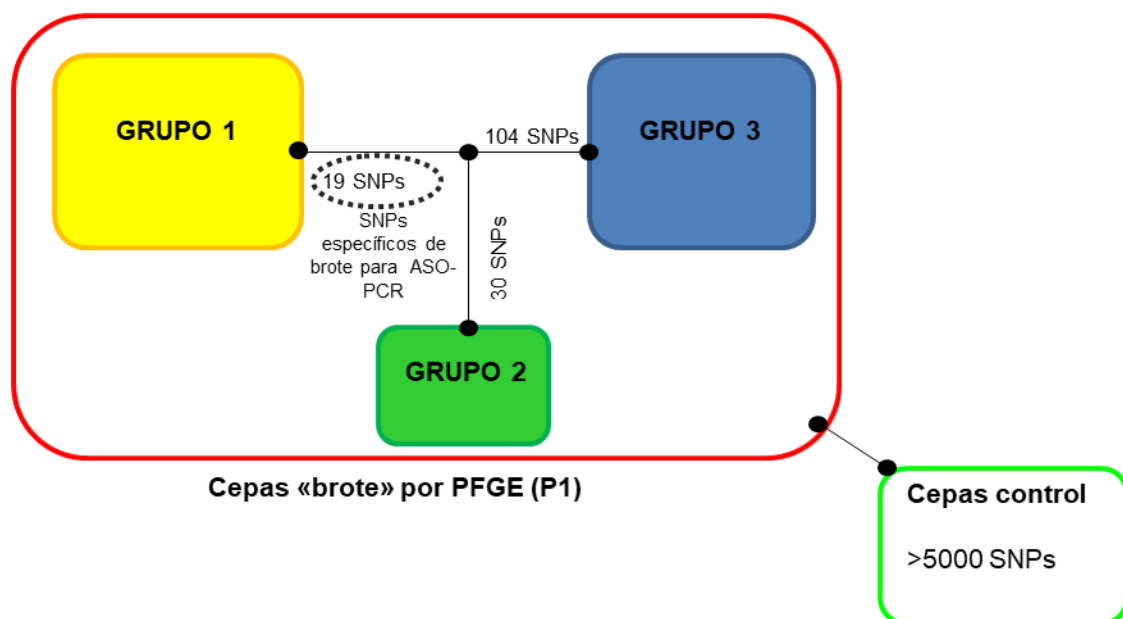


Figura 16. Distribución de las cepas incluidas en el estudio de *P. aeruginosa*

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

En total, en el estudio con MALDI-TOF se incluyeron 67 aislados identificados como *P. aeruginosa* con una puntuación $>2,0$; 35 aislados analizados por WGS (27 aislados del brote y 8 controles) y 32 aislados analizados por ASO-PCR (16 pertenecientes al brote y 16 controles). Los espectros obtenidos por MALDI-TOF se procesaron con la plataforma Clover MSDAS con el preprocesado habitual de suavizado (Savitzky-Golay 11:3), sustracción de la línea base (filtro Top-Hat 0,02) y normalización TIC.

Se empleó como conjunto de entrenamiento los 35 aislados caracterizados por WGS y se construyeron dos matrices de picos, el primero empleando el “método *threshold*”, incluyendo solo aquellos picos con una intensidad superior a 1% de la intensidad del pico máximo (413 picos en total), y otra seleccionando picos biomarcadores que permitieran la diferenciación entre la categoría de brote y control. Con esta técnica se encontraron 3 picos biomarcadores que se emplearon para distinguir las cepas procedentes del brote (5169 *m/z*, 6915 *m/z*, 7236 *m/z*).

Ambas matrices de picos se estudiaron mediante *Machine Learning* con dos algoritmos no supervisados, PCA y agrupamiento jerárquico, y cuatro algoritmos supervisados (PLS-DA, RF, SVM y KNN). Para comprobar estos modelos se realizaron dos validaciones internas con los datos de entrenamiento, una validación cruzada de $k=10$ y una validación “Leave One Out” o LOO. Estos modelos se validaron externamente con las 32 cepas procedentes de la ASO-PCR.

Inicialmente se compararon los resultados de la separación mediante MALDI-TOF con los obtenidos por PFGE, considerando el pulsotipo P1 como brote, y el resto de aislados como controles. En esta aproximación, los modelos creados por *threshold* obtuvieron validaciones cruzadas $k=10$ de 97,1% para el caso de PLS-DA, RF y KNN, y de 94,3% para el SVM. Con la matriz construida con los tres picos biomarcadores, todas las validaciones cruzadas fueron del 100% para todos los modelos. En los modelos no supervisados también se pudo observar una separación en dos clústeres claramente diferenciados de brote y controles.

Posteriormente, se compararon los resultados con los obtenidos mediante WGS, donde el pulsotipo P1 se dividió en tres grupos diferentes según la distancia de

SNPs. Para conseguir la separación de estos tres grupos mediante MALDI-TOF se aplicaron los mismos 4 algoritmos de *Machine Learning* a la matriz de picos construida por treshold. Se realizó una validación cruzada LOO (ya que las categorías tenían menos de 10 aislados y no se pudo realizar una validación por k-fold) y se obtuvo un porcentaje de acierto del 94,3% con SVM, del 85,7% con PLS-DA y KNN, y del 88,6% aplicando RF. A pesar de que no se obtuvieron resultados satisfactorios para la separación de los grupos 2 y 3 del resto de muestras, en la representación gráfica de los algoritmos se pudo observar que el grupo 2 estaba más cercano al grupo 1 que el grupo 3 y las cepas control, lo que coincide con los resultados de WGS (Figura 16), ya que el grupo 2 está separado del 1 por 49 SNPs, y el grupo 3 por 123 SNPs.

Por último se realizó una validación externa con las cepas ASO-PCR con el modelo que rindió mejores resultados en este último ensayo, el SVM y se obtuvo una clasificación correcta del 90%, clasificando erróneamente 3 aislados (dos aislados pertenecientes al brote como controles, y un aislado control como perteneciente al brote).

Además, también se realizó un estudio de reproducibilidad de MALDI-TOF con los 35 aislados de *P. aeruginosa* caracterizados mediante WGS. Cada aislado se analizó por cuadruplicado en 4 pocillos de la placa de MALDI, para estudiar la reproducibilidad técnica o interpocillo, y a su vez cada pocillo se leyó dos veces, dando un total de 8 espectros por cada cepa y día. Se realizó la media de cada pocillo obteniendo solo un espectro por pocillo, con el que se realizó el estudio de reproducibilidad interpocillo. La reproducibilidad biológica se estudió en dos días diferentes, volviendo a realizar la lectura de igual manera en el segundo día. Aplicando el método treshold de búsqueda de picos con intensidades superiores al 1% del pico de máxima intensidad a los espectros normalizados, se encontraron 13 picos comunes a todos los aislados con los que se realizó el estudio. Los CV interpocillo fueron del 11,0% y del 14,5% para el día 1 y día 2 respectivamente. Se observó que la región más estable (con menor CV) se encontraba entre los 2000 y 7000 Da. La media de CV interdía fue de 9,2% y todos los CV interpocillo e interdía se encontraban por debajo de 18,0%, un rango que se considera adecuado (<20-30%) [267], a excepción de uno de los picos que tenía un CV interdía cercano al 30%.

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

Para el estudio con FTIR se analizaron las cepas que disponían de análisis por PFGE y WGS, es decir, 20 aislados. El estudio con FTIR se realizó siguiendo las instrucciones del fabricante, resuspendiendo un asa de 1µl de colonia en 50µl de etanol al 70%, homogeneizando y añadiendo 50µl de agua estéril. De esta suspensión se depositó en la placa de FTIR 15µl por triplicado y se procedió a la lectura. Se analizó la segunda derivada de los espectros con el software de Clover MSDAS, mediante algoritmos no supervisados (agrupamiento jerárquico y PCA) y supervisados (PLS-DA y RF). Se analizaron varias regiones del espectro infrarrojo, y se observó que la región de los lípidos (de 1400 a 1500 cm^{-1} y de 2800 a 3000 cm^{-1}) fue la que tenía mayor poder discriminatorio entre los aislados, a pesar de ser la región de los carbohidratos la más comúnmente empleada para el tipado mediante FTIR por la variabilidad de los carbohidratos de la pared celular entre bacterias.

En el agrupamiento jerárquico se observó que las cepas pertenecientes al brote (P1), y por tanto al mismo ST (ST175), se agruparon juntas, al igual que los dos aislados del mismo paciente, que se agruparon una al lado de la otra y que el FTIR consideró que se trataba del mismo clon (cepas 139 y 143). El punto de corte para separar los aislados de brote de los controles fue de 0,35. También se observó que dentro del pulsotipo P1, el FTIR no fue capaz de diferenciar los tres grupos resultantes de WGS separados por pocos SNPs. En cuanto a los algoritmos supervisados, se realizó una validación cruzada LOO con un 91,1% de acierto en PLS-DA y 97,8% en RF en la clasificación de los aislados.

En este artículo se ha demostrado que, si bien el tipado mediante MALDI-TOF no es comparable al obtenido mediante técnicas de secuenciación masiva, sí que se equipara a la capacidad de tipado mediante PFGE, que se considera el actual *Gold Estándar* de tipado [114]. La aplicación de MALDI-TOF para el tipado de brotes permite una reducción en costes y material en comparación con el desarrollo de esta técnica. Pese a que los resultados obtenidos con MALDI-TOF no se correlacionan completamente con la técnica de WGS, sí se pudo observar una clasificación adecuada de los aislados de brote (pulsotipo P1), y una clasificación relativamente correcta dentro de éstos, donde el grupo 2 estaba más cercano al brote que el grupo 3 y los controles, al igual que se observó en la secuenciación de los aislados.

Lo mismo sucede con el tipado mediante FTIR, se observa una correlación con los datos de PFGE en los que se diferencia el pulsotipo P1 causante del brote, pero sin clasificar correctamente los tres grupos de aislados dentro de P1 caracterizados por WGS.

Es importante destacar que los grupos de WGS se originan en función a la distancia de SNPs, y estas diferencias pueden traducirse en cambios que no se expresen fenotípicamente, que es lo que se detecta mediante el MALDI-TOF y el FTIR. A pesar de no poder distinguir entre los grupos de P1, la clasificación y tipado por STs fue correcta, diferenciando claramente el ST175 de los demás.

Pocos estudios han intentado tipar clones de alto riesgo de *P. aeruginosa* mediante MALDI-TOF. El primero de ellos fue el publicado por Cabroler en 2015 [124], en el que intentan clasificar 5 clones de alto riesgo de *P. aeruginosa* empleando MALDI-TOF (ST111, ST175, ST235, ST253 y ST395) con 46 aislados para el conjunto de entrenamiento, que luego validan externamente con 295 aislados independientes. Desarrollan varios modelos enfrentando el ST que quieren diferenciar del resto de cepas de otros STs y consiguen clasificar los aislados correctamente con un porcentaje de acierto >90% en el conjunto de entrenamiento y también en la validación externa (excepto para el ST111 que tiene un 82,0% de acierto en la validación). En el caso del ST175, que es el que se diferencia en esta tesis, observan que la presencia de un pico biomarcador en 7359 *m/z* junto con la ausencia de picos en 7329 y 12154 *m/z* es lo que les permite diferenciar este ST de los demás.

Estos resultados fueron validados por otro grupo de investigación en Mallorca, que llevó a cabo primeramente un estudio con cepas de su hospital para el tipado del ST175 [122], y posteriormente validó sus resultados con un nuevo estudio con 506 aislados recogidos en 51 hospitales españoles [123]. En estos aislados confirman la presencia del pico 7359 *m/z* descrito por Cabroler en todos los ST175 (n=35) excepto en uno, pero también en otros aislados no pertenecientes a este ST, por lo que la especificidad y el valor predictivo positivo que obtienen es bajo (85,0% y 33,0%). Por eso actualizan el modelo y descubren un nuevo pico, el 6911 *m/z* presente en todos los aislados ST175 y ausente en el resto de aislados. Con estos dos picos, obtienen una

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

sensibilidad en la clasificación del 97,0% (34 de 35 aislados, ya que uno no presenta el pico 7359 m/z). Este estudio es significativo ya que se lleva a cabo con aislados de toda España y ese hecho no disminuye la sensibilidad de la técnica (también puede significar que el clon ST175 se encuentra bastante conservado en todo el país).

En nuestro estudio, encontramos el mismo pico 6915 m/z con un desplazamiento dentro de los límites aceptados ($\pm 4,5$ m/z) y sin otros picos alrededor, pero no encontramos el 7359 m/z . Los otros dos picos biomarcadores encontrados en nuestro estudio (5169 m/z y 7236 m/z) no han sido descritos con anterioridad, por lo que pueden ser picos específicos del propio brote.

Otros artículos han analizado *P. aeruginosa* mediante MALDI-TOF pero no para el tipado o la detección de clones, sino para determinar de manera rápida su sensibilidad o resistencia a antibióticos [268-270], principalmente para la detección de resistencia a carbapenémicos. Nuevas publicaciones han estudiado la detección de resistencia a colistina empleando el nuevo MALDI-TOF SIRIUS de Bruker en polaridad negativa para el análisis de modificaciones en el lípido A [271]. Otros autores han estudiado otras especies de *Pseudomonas* en brotes (*Pseudomonas stutzeri* y *Pseudomonas putida*) y empleado MALDI-TOF y algoritmos no supervisados para su tipado y estudio de relaciones filogenéticas [272-274].

Hasta el momento, no se ha publicado ningún artículo científico que emplee el FTIR para el tipado de cepas de *P. aeruginosa* exclusivamente. Uno reciente de Martak en 2019 intentó tipar bacilos Gram-negativos en general [159]. En ese estudio analizan 100 aislados de *P. aeruginosa* procedentes de un brote, tipados mediante MLST y PFGE, y todos se clasifican correctamente según su ST, excepto un aislado ST395 que había sufrido una delección de 131kb en comparación con el resto de aislados de ese ST procedentes del brote.

Por último, en el cuarto capítulo de esta tesis se compara el uso de MALDI-TOF con una de las técnicas de referencia de tipado de *Clostridioides difficile*, el ribotipado mediante electroforesis capilar, para la diferenciación entre los ribotipos RT027, RT181 (027-like) y otros ribotipos clínicamente significativos aislados de muestras de heces.

C. difficile es un bacilo Gram-positivo anaerobio que puede persistir en el medio ambiente y es resistente a la mayoría de desinfectantes convencionales como el alcohol o la clorhexidina por su capacidad de formar esporas. Esto lo vuelve altamente transmisible si no se llevan a cabo buenas medidas de higiene y control de la infección, como una correcta higiene de manos entre el personal sanitario [275].

Es el principal patógeno causante de diarrea nosocomial y asociada al consumo de antibióticos, y supone un problema de salud pública por sus altas tasas de morbimortalidad y su capacidad de causar brotes nosocomiales. Las infecciones por *C. difficile* suelen ser intestinales, y rara vez se dan fuera de esta localización. Las infecciones extraintestinales causadas por esta bacteria generalmente se deben a contaminaciones con materia fecal en pacientes sometidos a intervenciones quirúrgicas o portadores de catéteres [276].

Los síntomas más comunes de la infección por *C. difficile* son los procesos diarreicos, que pueden ser desde leves a muy graves e incapacitantes, pero la infección también puede manifestarse mediante otros síntomas más graves como la colitis pseudomembranosa, el íleo paralítico o el megacolon tóxico [275-277]. El diagnóstico de *C. difficile* en un paciente implica su aislamiento, preferiblemente en habitaciones individuales, y también medidas de aislamiento de contacto del personal sanitario para evitar su transmisión [216, 275].

Dentro de la especie *C. difficile*, algunos ribotipos presentan un locus de patogenicidad (PaLoc) que codifica la producción de toxinas que dañan la barrera intestinal provocando muerte celular, como la toxina A (enterotoxina) y la toxina B (citotoxina). Otras cepas presentan además una tercera toxina, la toxina binaria, que no se encuentra en PaLoc y cuyo papel en el desarrollo de la infección no está del todo claro aún. Además, el ribotipo hipervirulento NAP1/B1/027 presenta una delección de un nucleótido en la posición 117 del gen regulador negativo *tcdC*, lo que da lugar a un *frameshift* o cambio en la pauta de lectura que se traduce en una pérdida de actividad de este gen regulador. Esto implica una hiperproducción de toxinas y por tanto un aumento en la virulencia de estas cepas [276-278]. El ribotipo RT181, que ha sido descrito recientemente en la literatura, pertenece al mismo clado de MLST que el

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

RT027 (Clado 2) y también presenta esta delección en el gen regulador *tcdC* [279-281]. Estas cepas con delección en *tcdC* se consideran hipervirulentas por su mayor capacidad de producción de toxinas.

Los ribotipos hipervirulentos presentan mayor patogenicidad que el resto, y una peor respuesta a tratamiento con mayores tasas de recaídas y morbi-mortalidad, por lo que resulta necesario identificarlos correctamente al momento del diagnóstico clínico. La detección rápida y precisa de estos ribotipos es crucial para la instauración de un tratamiento eficaz y el control de la diseminación de la infección [282, 283].

El diagnóstico de la infección por *C. difficile* en el laboratorio de microbiología se lleva a cabo a partir de muestras de heces no formes, mayoritariamente mediante técnicas de inmunoensayo en las que se detecta la enzima glutamato deshidrogenasa (GDH) y la producción de toxinas (generalmente la toxina B por ser más prevalente), o bien mediante técnicas moleculares donde se amplifican los genes que codifican estas toxinas [276, 277]. La detección mediante inmunoensayo presenta la ventaja de la rapidez, pero unos valores de sensibilidad bajos en comparación con el cultivo toxigénico, mientras que los estudios moleculares son más sensibles pero también más costosos.

Dentro de las técnicas moleculares comerciales, una de las más empleadas es el Xpert® *C. difficile* BT, que permite la detección de la toxina B y toxina binaria, y a la vez es capaz de detectar la delección en la posición 117 del *tcdC*. Sin embargo, no distingue entre los diferentes ribotipos toxigénicos que presentan esta delección, impidiendo la detección de brotes. Métodos de tipado como el PFGE, el MLST, MLVA o la secuenciación masiva se emplean para la diferenciación de clones de *C. difficile*, pero la falta de estandarización entre laboratorios dificulta su comparación interlaboratorio y además, se trata de técnicas de alto coste, laboriosas y que requieren de personal especializado en biología molecular. Estas técnicas son lentas y sus resultados se obtienen al cabo de varios días [277, 281].

Es por ello que el objetivo de este último estudio fue el desarrollo de una metodología de tipado empleando la espectrometría de masas para la

diferenciación de las cepas hipervirulentas RT027 y RT181 de otros ribotipos de interés clínico utilizando MALDI-TOF y algoritmos de ML.

Para ello se emplearon dos herramientas bioinformáticas independientes: Clover MSDAS (<https://cloverbiosoft.com>) y AutoCdiff (<https://www.bacteria.id>), un software en abierto de acceso libre desarrollado específicamente para este estudio en colaboración con ingenieros bioinformáticos.

Todas las cepas empleadas en este estudio se aislaron en cultivo anaerobio y se ribotiparon mediante la técnica *Gold Estándard* de PCR de la región intergénica ARNr 16S y ARNr 23S seguida de electroforesis capilar. Se incluyeron aislados representativos de los ribotipos más frecuentemente aislados en Europa y España, incluyendo el RT027 y aislados identificados por Xpert® *C.difficile* BT como “presuntivos 027” [283].

En el primer experimento el objetivo fue diferenciar el RT027 de *C. difficile* de los demás ribotipos que el Xpert® *C. difficile* BT clasifica como “presuntivos 027”. Para esto se encontraron 4 picos biomarcadores (2463 *m/z*, 4933 *m/z*, 4993 *m/z* y 6651 *m/z*) con los que se realizó el estudio en Clover MSDAS. En el segundo experimento, el objetivo fue tipar directamente empleando MALDI-TOF los ribotipos RT027, RT181 y otros ribotipos clínicamente significativos y frecuentemente aislados en España y Europa. Para ello, además de los 4 picos empleados en el experimento anterior, se añadieron otros tres que presentaban la capacidad de diferenciación entre ribotipos (3353 *m/z*, 6187 *m/z* y 6710 *m/z*). El software AutoCdiff fue capaz de clasificar correctamente los ribotipos estudiados sin la búsqueda de picos biomarcadores.

Dos de los picos descubiertos ya habían sido descritos con anterioridad en otros estudios en las regiones 6647-6654 *m/z* (en nuestro estudio el pico 6651 *m/z*) y 6707-6712 *m/z* (en nuestro estudio el pico 6710 *m/z*) [126, 204, 209]. Estos picos corresponden a las dos isoformas de la proteína L28, la isoforma con masa 6710 *m/z* presente en los ribotipos RT027 y RT181 y la isoforma con masa 6651 *m/z* presente en los demás ribotipos, incluyendo el resto de “presuntivos 027”. Otro de los picos específicos al RT027 y RT181 que se estudió, el 3353 *m/z*, puede representar la misma isoforma de L28 en 6710 *m/z*, presentando ésta doble carga iónica. Sin embargo, con

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

el pico 6651 m/z no se encontró la misma correlación con otro pico de una masa inferior.

El pico 4933 m/z también había sido descrito con anterioridad en otros estudios, con masa de 4928 m/z en el estudio de Corver, 4930 m/z en el estudio de Reil y de 4932 m/z en el de Flores-Treviño [126, 204, 284]. Este pico está descrito como específico de los RT027 (MLST-1) y las variaciones en el rango de masas pueden deberse a desviaciones por errores en la calibración. También se encontró el pico 2463 m/z en todos los aislados RT027, lo que sugiere que podría representar la misma proteína con doble carga iónica. Sin embargo, el pico 4933 m/z presente en las categorías RT181 y en "Otros ribotipos", que permitió diferenciarlos del RT027, no ha sido reportado previamente, probablemente por la falta de bibliografía acerca del ribotipo RT181, recientemente descrito.

En nuestro estudio, los algoritmos de aprendizaje automático mostraron una precisión de hasta el 100%, incluso para aislados de brotes en tiempo real. En Clover MSDAS, Light-GBM y RF fueron los algoritmos que mostraron un rendimiento superior en la diferenciación del RT027 de otros aislados "presuntivos RT027". Este patrón se repitió con todos los algoritmos, excepto PLS-DA, para la diferenciación de los RT027 y RT181 del resto de los RT analizados en el experimento 2. Además, todos los algoritmos de ML aplicados lograron una precisión del 100% en la clasificación a tiempo real de los aislados de *C. difficile* del brote. En AutoCdiff, el DBLR-FS y RF mostraron la mayor precisión para diferenciar RT027 de otros aislados "presuntivos RT027". En el segundo experimento, solo un aislado, HGM-RT173 del grupo "Otros RT", se clasificó erróneamente como RT027, probablemente debido a la presencia del pico de 4933 m/z , comúnmente asociado con RT027. Este aislado se clasificó erróneamente también por Xpert® *C. difficile* BT como "presuntivo 027". Todos los algoritmos de AutoCdiff clasificaron correctamente los aislados pertenecientes a los brotes de *C. difficile*.

En 2022, Calderaro y colaboradores publicaron un artículo en el que aplicaban MALDI-TOF para el ribotipado de cepas clínicas de *C. difficile*. La técnica molecular de referencia que emplean para el ribotipado de las cepas clasificaba los aislados en diez

ribotipos aleatoriamente nombrados PR1 a PR10, y no la correlacionaron con la nomenclatura internacional de ribotipado. Demostraron que MALDI-TOF presentaba poder discriminatorio para tipar esos ribotipos propios, pero su técnica no es comparable con otros trabajos ni extrapolable fuera de ese hospital [285]. Posteriormente en otra publicación, fueron capaces de tipar un brote de *C. difficile* mediante MALDI-TOF. En sus aislados, ven como se agrupan juntos mediante PCA aquellos pertenecientes a la misma unidad hospitalaria, con el mismo perfil toxigénico y sensibilidad antibiótica, pero tampoco los pueden asociar a un ribotipo internacional en concreto ni extrapolar sus resultados por el mismo motivo que en el estudio anterior [286].

Lo mismo sucede en otros dos estudios realizados en Suecia en los que analizan regiones de masas de mayor peso molecular a 20000 Da [125, 287]. En este nuevo enfoque se estudia la región entre 30 y 50 kDa, que es la región donde se encuentra la denominada capa S proteica de la superficie de *C. difficile*, con una matriz orgánica de ácido transferúlico. Son capaces también de tipar sus aislados creando una base de datos interna de ribotipos según su espectro de masas altas (lo que denominan *High Molecular Weight Profile*), que luego comparan con el ribotipado convencional.

Otro estudio realizado en Brasil consigue diferenciar mediante MALDI-TOF los ribotipos de *C. difficile* aislados únicamente en Brasil de otros ribotipos de circulación global. Sin embargo, de los picos biomarcadores que describen no identificamos ninguno en nuestro estudio que nos permitiera la diferenciación de nuestros ribotipos [288]. Por otra parte, dos estudios en China consiguen la diferenciación mediante MALDI-TOF de las cepas del Clado IV de MLST de *C. difficile*, que se corresponde con el ST37 o RT017, el más prevalente en ese país [289, 290].

La metodología desarrollada en nuestro estudio supone una valiosa herramienta para la determinación rápida del ribotipo de cepas clínicas de *C. difficile*, con el valor añadido de que es comparable a la técnica *Gold Estándar* de ribotipado directamente. Este enfoque reduce significativamente el tiempo y los costes para el ribotipado de *C. difficile*, permitiendo una mejor optimización de los recursos

RESUMEN GLOBAL DE RESULTADOS Y DISCUSIÓN

hospitalarios y una instauración más temprana del tratamiento según el ribotipo caracterizado, así como un control más eficiente y coste-efectivo de la infección.

En este estudio se demostró que es posible el ribotipado mediante MALDI-TOF, y que se trata una herramienta automatizada que reduce el tiempo para la obtención de resultados de varios días a pocas horas desde el aislamiento del microorganismo en cultivo. Además, la aplicación de esta técnica es relativamente sencilla, siendo la preparación de la muestra idéntica a la preparación para la identificación rutinaria mediante MALDI-TOF.

Una de las principales limitaciones de este estudio es que, a pesar que se realizó con numerosos aislados pertenecientes a hospitales de toda la geografía española, solo se incluyeron cepas de este país. En un futuro, se intentará seguir desarrollando tanto los modelos de ML de Clover MSDAS como el software AutoCdiff mediante colaboraciones con centros externos pertenecientes a España, así como a otros países. Por otro lado, la falta de estandarización en el ribotipado de *C. difficile*, con diversas técnicas aplicadas en cada país, como PFGE, ribotipado por electroforesis capilar, MLVA, MLST, WGS..., y la falta de un consenso general y bases de datos internacionales con las que comparar dichos ribotipos, hacen difícil la estandarización universal de una metodología desarrollada a partir de MALDI-TOF.

Por último, tanto para el ribotipado convencional como para la realización de MALDI-TOF es necesario el crecimiento de la bacteria en cultivo cuando se trata de muestras de heces debido a su compleja composición. Esta siembra clínica de la muestra no se realiza en todos los países y hospitales, ya que muchos de ellos se limitan solo a la realización de una PCR para la detección de *C. difficile*. La realización de PCRs comerciales como el Xpert® *C. difficile* BT supone una herramienta más rápida para la detección de *C. difficile*, ya que puede realizarse directamente a partir de la muestra clínica sin necesidad de esperar al cultivo. Sin embargo, el coste de esta PCR frente al cultivo convencional supone una desventaja, por lo que no todos los laboratorios clínicos disponen de estas PCR comerciales debido a su precio, mientras que el equipo de MALDI-TOF está cada vez más presente en muchos de ellos para la

identificación bacteriana, lo que permitiría también emplearlo para el ribotipado de muestras de *C. difficile*.

El tipado de *C. difficile* mediante MALDI-TOF ha demostrado ser una herramienta con un gran potencial que supone una alternativa real a los métodos moleculares, y que mejora el flujo de trabajo en el laboratorio de microbiología. A la vez, se ha demostrado que se trata de una tecnología reproducible ya que se probó con diferentes algoritmos de ML así como en dos plataformas de software distintas. Actualmente, el software AutoCdiff desarrollado en este estudio se encuentra en proceso de obtención de patente, y se espera su crecimiento y desarrollo en colaboración con grupos de investigación internacionales que aporten nuevas cepas de ribotipos diversos, para poder aumentar nuestra base de datos y expandirla globalmente.

Todos los trabajos incluidos en esta tesis presentan un denominador común, que es la investigación en espectrometría de masas MALDI-TOF para usos alternativos a la identificación bacteriana. En adición a los objetivos específicos de cada artículo, estos trabajos sientan las bases generales del procesamiento de espectros de masas y la aplicación de algoritmos de *Machine Learning* para la diferenciación entre espectros. Además, en uno de los artículos se incluye el tipado mediante FTIR, una novedosa herramienta en los laboratorios de microbiología clínica que ganará importancia en los próximos años por su precisión en el tipado pero a la vez su rapidez y fácil ejecución.

La implementación de MALDI-TOF en los laboratorios de microbiología ha supuesto un beneficio sustancial en el manejo de los pacientes, permitiendo instaurar de forma temprana un tratamiento antibiótico dirigido, lo que reduce la duración de los tratamientos y la estancia hospitalaria. Esto ha conseguido a su vez una mejora en la distribución de recursos económicos y humanos. La metodología de realización de MALDI-TOF en el laboratorio clínico es rápida y sencilla, facilitando el flujo de trabajo [22, 291].

El estudio alternativo para aplicaciones más allá de la identificación podrá proporcionar mejoras adicionales en estos campos, permitiendo siempre una optimización asistencial y un beneficio de cara al paciente.

CONCLUSIONES

- I. En esta tesis se desarrolló una metodología reproducible para el análisis de espectros de masas de MALDI-TOF, que permitió el análisis preciso mediante la aplicación de algoritmos de *Machine Learning*. Se sentaron las bases en el procesamiento de espectros de masas para aplicaciones alternativas a la identificación de microorganismos, y se definieron los parámetros a aplicar en el análisis de espectros de masas.
- II. En el estudio de reproducibilidad en *E. faecium* se observó que la región más estable y que presenta menores coeficientes de variación para el análisis de picos se encuentra entre 3000 Da y 9000 Da, obteniendo mejores resultados con los datos normalizados.
- III. Se consiguió la diferenciación entre aislados de *E. faecium* sensibles y resistentes a vancomicina mediante el uso de MALDI-TOF y algoritmos supervisados con un porcentaje de acierto superior al 80% y una separación entre aislados VanA y VanB con un acierto superior al 85%. Se describieron dos potenciales picos biomarcadores, el 6891 *m/z* que permitía separar aislados sensibles de resistentes, y el 5095 *m/z* que permitía separar los aislados VanA de VanB.
- IV. La identificación precisa y separación de las especies intracomplejo de *E. cloacae* complex se consiguió combinando MALDI-TOF y algoritmos supervisados de *Machine Learning* con un porcentaje de acierto del 100% con las cepas empleadas para el desarrollo del modelo, y un acierto en la validación superior al 96% empleando aislados externos.
- V. La discriminación del clon de alto riesgo ST175 de *Pseudomonas aeruginosa* causante de un brote hospitalario en el Hospital General Universitario Gregorio Marañón fue posible mediante la aplicación de MALDI-TOF y FTIR. Se descubrieron tres picos biomarcadores específicos de brote, con masas 5169 *m/z*, 6915 *m/z* y 7236 *m/z*. Estos picos permitieron la correcta clasificación del 100% de los aislados en comparación con PFGE con todos los algoritmos

CONCLUSIONES

estudiados. La tecnología FTIR mostró un poder discriminatorio similar al del PFGE analizando la región de los lípidos, permitiendo también la diferenciación de los distintos STs estudiados.

- VI.** Se desarrolló una metodología rápida para el ribotipado de aislados clínicos de *Clostridioides difficile* mediante la tecnología MALDI-TOF, en comparación con la técnica de ribotipado clásico. Se describieron 7 picos biomarcadores (2463 *m/z*, 4933 *m/z*, 4993 *m/z*, 3353 *m/z*, 6187 *m/z*, 6651 y 6710 *m/z*) que permitieron la separación del RT027 y RT181 del resto de ribotipos clínicos analizados. Además, se desarrolló un software propio que permitía la clasificación automática de estos ribotipos.

- VII.** Todo esto permitió la mejora en el diagnóstico de las infecciones causadas por estos 4 microorganismos, suponiendo una ayuda en la vigilancia epidemiológica hospitalaria y el control de la infección.

BIBLIOGRAFÍA

1. van Belkum, A., et al., *Rapid clinical bacteriology and its future impact*. Annals of laboratory medicine, 2013. **33**(1): p. 14-27.
2. Khan, Z.A., M.F. Siddiqui, and S. Park, *Current and Emerging Methods of Antibiotic Susceptibility Testing*. Diagnostics, 2019. **9**(2).
3. García-Lechuz Moya JM, G.L.J., Orta Mira N, Sánchez Romero MI., *1b. Recogida, transporte y procesamiento general de muestras en el laboratorio de microbiología*. Procedimientos en Microbiología Clínica. Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica (SEIMC), 2017.
4. Lagier, J.C., et al., *Current and past strategies for bacterial culture in clinical microbiology*. Clinical microbiology reviews, 2015. **28**(1): p. 208-36.
5. Bou, G., et al., *[Bacterial identification methods in the microbiology laboratory]*. Enfermedades infecciosas y microbiología clínica, 2011. **29**(8): p. 601-8.
6. Carroll, K.C., et al., *12th Manual of clinical microbiology*. 12th edition. ed2019. 2 volumes.
7. Bennett, J.E., R. Dolin, and M.J. Blaser, *Mandell, douglas, and bennett's principles and practice of infectious diseases*. 9 ed2019.
8. Gajic, I., et al., *Antimicrobial Susceptibility Testing: A Comprehensive Review of Currently Used Methods*. Antibiotics, 2022. **11**(4).
9. Khardori, N., *Future of diagnostic microbiology*. Indian journal of medical microbiology, 2014. **32**(4): p. 371-7.
10. Ozenci, V. and G.M. Rossolini, *Rapid microbial identification and antimicrobial susceptibility testing to drive better patient care: an evolving scenario*. The Journal of antimicrobial chemotherapy, 2019. **74**(Suppl 1): p. i2-i5.
11. Li, Y., X. Yang, and W. Zhao, *Emerging Microtechnologies and Automated Systems for Rapid Bacterial Identification and Antibiotic Susceptibility Testing*. SLAS technology, 2017. **22**(6): p. 585-608.
12. Vasala, A., V.P. Hytonen, and O.H. Laitinen, *Modern Tools for Rapid Diagnostics of Antimicrobial Resistance*. Frontiers in cellular and infection microbiology, 2020. **10**: p. 308.
13. Kralik, P. and M. Ricchi, *A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything*. Frontiers in microbiology, 2017. **8**: p. 108.
14. Osei Sekyere, J., U. Govinden, and S.Y. Essack, *Review of established and innovative detection methods for carbapenemase-producing Gram-negative bacteria*. Journal of applied microbiology, 2015. **119**(5): p. 1219-33.
15. Buchan, B.W. and N.A. Ledebor, *Emerging technologies for the clinical microbiology laboratory*. Clinical microbiology reviews, 2014. **27**(4): p. 783-822.
16. Lopez-Causapé C., G.-C.F., Tomás Carmona M., Oliver-Palomo A., *71. Aplicaciones de las técnicas de secuenciación masiva en la Microbiología Clínica*. Procedimientos en Microbiología Clínica. Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica (SEIMC), 2021.
17. Gilchrist, C.A., et al., *Whole-genome sequencing in outbreak analysis*. Clinical microbiology reviews, 2015. **28**(3): p. 541-63.
18. Quainoo, S., et al., *Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis*. Clinical microbiology reviews, 2017. **30**(4): p. 1015-1063.
19. Clark, A.E., et al., *Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology*. Clinical microbiology reviews, 2013. **26**(3): p. 547-603.
20. Croxatto, A., G. Prod'hom, and G. Greub, *Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology*. FEMS microbiology reviews, 2012. **36**(2): p. 380-407.

BIBLIOGRAFÍA

21. Pliakos, E.E., et al., *The Cost-Effectiveness of Rapid Diagnostic Testing for the Diagnosis of Bloodstream Infections with or without Antimicrobial Stewardship*. *Clinical microbiology reviews*, 2018. **31**(3).
22. Campos, A.F., et al., *Impact of an Antimicrobial Stewardship Program Intervention Associated with the Rapid Identification of Microorganisms by MALDI-TOF and Detection of Resistance Genes in ICU Patients with Gram-Negative Bacteremia*. *Antibiotics*, 2022. **11**(9).
23. Morency-Potvin, P., D.N. Schwartz, and R.A. Weinstein, *Antimicrobial Stewardship: How the Microbiology Laboratory Can Right the Ship*. *Clinical microbiology reviews*, 2017. **30**(1): p. 381-407.
24. Herbert Beynon, J.B., L. *Mass Spectrometry*. *Encyclopaedia Britannica*. 2023 16 June 2023 June 19, 2023]; Available from: <https://www.britannica.com/science/mass-spectrometry>.
25. Han, X., A. Aslanian, and J.R. Yates, 3rd, *Mass spectrometry for proteomics*. *Current opinion in chemical biology*, 2008. **12**(5): p. 483-90.
26. Griffiths, J., *A brief history of mass spectrometry*. *Analytical chemistry*, 2008. **80**(15): p. 5678-83.
27. Centres, R., *Mass spectrometry at Johns Hopkins and the University of Maryland Baltimore County*. *Rapid communications in mass spectrometry*, 1994. **8**(9): p. 763-765.
28. Anhalt, J. and C. Fenselau, *Identification of bacteria using mass spectrometry*. *Analytical chemistry*, 1975. **47**(2): p. 219-225.
29. Tanaka, K., *The origin of macromolecule ionization by laser irradiation (Nobel lecture)*. *Angewandte Chemie*, 2003. **42**(33): p. 3860-70.
30. Tanaka, K.W., H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T. , *Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-flight Mass Spectrometry*. *Rapid communications in mass spectrometry*, 1988. **2**(8).
31. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons*. *Analytical chemistry*, 1988. **60**(20): p. 2299-301.
32. Hillenkamp, F. and M. Karas, *Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization*. *Methods in enzymology*, 1990. **193**: p. 280-95.
33. Oviano, M.R.-S., B.; Caballero-Pérez, JdD.; Muñoz-Bellido, JL., *65. Aplicaciones de la espectrometría de masas MALDI-TOF en Microbiología Clínica*. *Procedimientos en Microbiología Clínica*. Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica (SEIMC), 2019.
34. Fenselau, C. and P.A. Demirev, *Characterization of intact microorganisms by MALDI mass spectrometry*. *Mass spectrometry reviews*, 2001. **20**(4): p. 157-71.
35. Beavis, R.C. and B.T. Chait, *Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins*. *Rapid communications in mass spectrometry : RCM*, 1989. **3**(12): p. 432-5.
36. Oviano, M. and B. Rodriguez-Sanchez, *MALDI-TOF mass spectrometry in the 21st century clinical microbiology laboratory*. *Enfermedades infecciosas y microbiología clínica*, 2021. **39**(4): p. 192-200.
37. Rodriguez-Sanchez, B., et al., *Improvement of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry identification of difficult-to-identify bacteria and its impact in the workflow of a clinical microbiology laboratory*. *Diagnostic microbiology and infectious disease*, 2014. **79**(1): p. 1-6.
38. Oviano, M., et al., *Direct identification of clinical pathogens from liquid culture media by MALDI-TOF MS analysis*. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 2018. **24**(6): p. 624-629.

39. Rindi, L., et al., *Rapid and Accurate Identification of Nontuberculous Mycobacteria Directly from Positive Primary MGIT Cultures by MALDI-TOF MS*. *Microorganisms*, 2022. **10**(7).
40. Fernandez-Esgueva, M., et al., *Use of MALDI-TOF MS (Bruker Daltonics) for identification of Mycobacterium species isolated directly from liquid medium*. *Enfermedades infecciosas y microbiología clínica*, 2021. **39**(5): p. 241-243.
41. Almuhayawi, M.S., et al., *Identification of microorganisms directly from blood culture bottles with polymicrobial growth: comparison of FilmArray and direct MALDI-TOF MS*. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*, 2021. **129**(4): p. 178-185.
42. Scohy, A., et al., *Evaluation of the Bruker(R) MBT Sepsityper IVD module for the identification of polymicrobial blood cultures with MALDI-TOF MS*. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*, 2018. **37**(11): p. 2145-2152.
43. Zboromyrska, Y., et al., *A multicentre study investigating parameters which influence direct bacterial identification from urine*. *PLoS one*, 2018. **13**(12): p. e0207822.
44. Vila, J., et al., *[Future applications of mass spectrometry in microbiology]*. *Enfermedades infecciosas y microbiología clínica*, 2016. **34 Suppl 2**: p. 53-8.
45. Martiny, D., et al., *Impact of rapid microbial identification directly from positive blood cultures using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry on patient management*. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 2013. **19**(12): p. E568-81.
46. Burillo, A., et al., *Gram-stain plus MALDI-TOF MS (Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry) for a rapid diagnosis of urinary tract infection*. *PLoS one*, 2014. **9**(1): p. e86915.
47. Bishop, B., et al., *The use of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for rapid bacterial identification in patients with smear-positive bacterial meningitis*. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 2018. **24**(2): p. 171-174.
48. Lallemand, E., et al., *Use of MALDI-TOF mass spectrometry after liquid enrichment (BD Bactec) for rapid diagnosis of bone and joint infections*. *Research in microbiology*, 2017. **168**(2): p. 122-129.
49. Bhavsar, S.M., T.C. Dingle, and C.L. Hamula, *The impact of blood culture identification by MALDI-TOF MS on the antimicrobial management of pediatric patients*. *Diagnostic microbiology and infectious disease*, 2018. **92**(3): p. 220-225.
50. Verroken, A., et al., *Clinical Impact of MALDI-TOF MS Identification and Rapid Susceptibility Testing on Adequate Antimicrobial Treatment in Sepsis with Positive Blood Cultures*. *PLoS one*, 2016. **11**(5): p. e0156299.
51. Perse, G., et al., *Sepsityper((R)) Kit versus In-House Method in Rapid Identification of Bacteria from Positive Blood Cultures by MALDI-TOF Mass Spectrometry*. *Life*, 2022. **12**(11).
52. Oviano, M., et al., *Multicenter Evaluation of Rapid BACpro((R)) II for the Accurate Identification of Microorganisms Directly from Blood Cultures Using MALDI-TOF MS*. *Diagnostics*, 2021. **11**(12).
53. Rodriguez-Sanchez, B., et al., *Direct identification of pathogens from positive blood cultures using matrix-assisted laser desorption-ionization time-of-flight mass spectrometry*. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 2014. **20**(7): p. O421-7.

BIBLIOGRAFÍA

54. Verroken, A., et al., *Reducing time to identification of positive blood cultures with MALDI-TOF MS analysis after a 5-h subculture*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2015. **34**(2): p. 405-13.
55. Idelevich, E.A., et al., *Rapid identification of microorganisms from positive blood cultures by MALDI-TOF mass spectrometry subsequent to very short-term incubation on solid medium*. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 2014. **20**(10): p. 1001-6.
56. Azrad, M., et al., *Cheap and rapid in-house method for direct identification of positive blood cultures by MALDI-TOF MS technology*. BMC infectious diseases, 2019. **19**(1): p. 72.
57. Moussaoui, W., et al., *Matrix-assisted laser desorption ionization time-of-flight mass spectrometry identifies 90% of bacteria directly from blood culture vials*. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 2010. **16**(11): p. 1631-8.
58. Lopez-Pintor, J.M., et al., *Direct antimicrobial susceptibility testing from the blood culture pellet obtained for MALDI-TOF identification of Enterobacterales and Pseudomonas aeruginosa*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2019. **38**(6): p. 1095-1104.
59. Marin, M., et al., *Identification of Nocardia species from clinical isolates using MALDI-TOF mass spectrometry*. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 2018. **24**(12): p. 1342 e5-1342 e8.
60. Rodriguez-Temporal, D., et al., *Impact of updating the MALDI-TOF MS database on the identification of nontuberculous mycobacteria*. Journal of mass spectrometry : JMS, 2017. **52**(9): p. 597-602.
61. Zvezdanova, M.E., et al., *Validation of an expanded, in-house library and an optimized preparation method for the identification of fungal isolates using MALDI-TOF mass spectrometry*. Medical mycology, 2023. **61**(4).
62. Alcalá, L., et al., *Identifying Anaerobic Bacteria Using MALDI-TOF Mass Spectrometry: A Four-Year Experience*. Frontiers in cellular and infection microbiology, 2021. **11**: p. 521014.
63. Rodriguez-Sanchez, B., et al., *Evaluation of MALDI Biotyper Mycobacteria Library v3.0 for Identification of Nontuberculous Mycobacteria*. Journal of clinical microbiology, 2016. **54**(4): p. 1144-7.
64. Rodriguez-Temporal, D., B. Rodriguez-Sanchez, and F. Alcaide, *Evaluation of MALDI Biotyper Interpretation Criteria for Accurate Identification of Nontuberculous Mycobacteria*. Journal of clinical microbiology, 2020. **58**(10).
65. Garrigos, T., et al., *Distribution of Achromobacter Species in 12 French Cystic Fibrosis Centers in 2020 by a Retrospective MALDI-TOF MS Spectrum Analysis*. Journal of clinical microbiology, 2022. **60**(6): p. e0242221.
66. Dematheis, F., et al., *Machine Learning Algorithms for Classification of MALDI-TOF MS Spectra from Phylogenetically Closely Related Species Brucella melitensis, Brucella abortus and Brucella suis*. Microorganisms, 2022. **10**(8).
67. Pizzato, J., et al., *Discrimination of Escherichia coli, Shigella flexneri, and Shigella sonnei using lipid profiling by MALDI-TOF mass spectrometry paired with machine learning*. MicrobiologyOpen, 2022. **11**(4): p. e1313.
68. Jensen, C.S., et al., *Identification of Streptococcus pseudopneumoniae and other mitis group streptococci using matrix assisted laser desorption/ionization - time of flight*

- mass spectrometry*. Diagnostic microbiology and infectious disease, 2021. **101**(3): p. 115487.
69. Cuenod, A., et al., *Whole-genome sequence-informed MALDI-TOF MS diagnostics reveal importance of Klebsiella oxytoca group in invasive infections: a retrospective clinical study*. Genome medicine, 2021. **13**(1): p. 150.
 70. Christner, M., et al., *Rapid MALDI-TOF mass spectrometry strain typing during a large outbreak of Shiga-Toxigenic Escherichia coli*. PloS one, 2014. **9**(7): p. e101924.
 71. Cuenod, A., et al., *Factors Associated With MALDI-TOF Mass Spectral Quality of Species Identification in Clinical Routine Diagnostics*. Frontiers in cellular and infection microbiology, 2021. **11**: p. 646648.
 72. CLSI, *Methods for the Identification of Cultured Microorganisms Using Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry*. 1st ed. CLSI guideline M58, 2017.
 73. Cuenod, A., et al., *Quality of MALDI-TOF mass spectra in routine diagnostics: results from an international external quality assessment including 36 laboratories from 12 countries using 47 challenging bacterial strains*. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 2023. **29**(2): p. 190-199.
 74. Alatoon, A.A., et al., *Comparison of direct colony method versus extraction method for identification of gram-positive cocci by use of Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry*. Journal of clinical microbiology, 2011. **49**(8): p. 2868-73.
 75. Rhoads, D.D., et al., *Clinical microbiology informatics*. Clinical microbiology reviews, 2014. **27**(4): p. 1025-47.
 76. Mitchell, M., et al., *Enhancing MALDI time-of-flight mass spectrometer performance through spectrum averaging*. PloS one, 2015. **10**(3): p. e0120932.
 77. Jonasson, E., E. Matuschek, and G. Kahlmeter, *The EUCAST rapid disc diffusion method for antimicrobial susceptibility testing directly from positive blood culture bottles*. The Journal of antimicrobial chemotherapy, 2020. **75**(4): p. 968-978.
 78. Akerlund, A., et al., *EUCAST rapid antimicrobial susceptibility testing (RAST) in blood cultures: validation in 55 European laboratories*. The Journal of antimicrobial chemotherapy, 2020. **75**(11): p. 3230-3238.
 79. Rodriguez-Sanchez, B., et al., *Review of the impact of MALDI-TOF MS in public health and hospital hygiene, 2018*. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin, 2019. **24**(4).
 80. Gato, E., et al., *Direct Detection of Carbapenemase-Producing Klebsiella pneumoniae by MALDI-TOF Analysis of Full Spectra Applying Machine Learning*. Journal of clinical microbiology, 2023. **61**(6): p. e0175122.
 81. Schuster, D., et al., *Detection of methicillin-resistant coagulase-negative staphylococci harboring the class A mec complex by MALDI-TOF mass spectrometry*. International journal of medical microbiology : IJMM, 2018. **308**(5): p. 522-526.
 82. Josten, M., et al., *Identification of agr-positive methicillin-resistant Staphylococcus aureus harbouring the class A mec complex by MALDI-TOF mass spectrometry*. International journal of medical microbiology : IJMM, 2014. **304**(8): p. 1018-23.
 83. Nagy, E., et al., *Differentiation of division I (cfiA-negative) and division II (cfiA-positive) Bacteroides fragilis strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry*. Journal of medical microbiology, 2011. **60**(Pt 11): p. 1584-1590.
 84. Cordovana, M., et al., *Bacteroides fragilis: A whole MALDI-based workflow from identification to confirmation of carbapenemase production for routine laboratories*. Anaerobe, 2018. **54**: p. 246-253.

BIBLIOGRAFÍA

85. Griffin, P.M., et al., *Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant enterococci and investigate the epidemiology of an outbreak*. Journal of clinical microbiology, 2012. **50**(9): p. 2918-31.
86. Lau, A.F., et al., *A rapid matrix-assisted laser desorption ionization-time of flight mass spectrometry-based method for single-plasmid tracking in an outbreak of carbapenem-resistant Enterobacteriaceae*. Journal of clinical microbiology, 2014. **52**(8): p. 2804-12.
87. Cordovana, M., et al., *A Full MALDI-Based Approach to Detect Plasmid-Encoded KPC-Producing Klebsiella pneumoniae*. Frontiers in microbiology, 2018. **9**: p. 2854.
88. Gato, E., et al., *Occurrence of the p019 Gene in the bla(KPC)-Harboring Plasmids: Adverse Clinical Impact for Direct Tracking of KPC-Producing Klebsiella pneumoniae by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry*. Journal of clinical microbiology, 2021. **59**(8): p. e0023821.
89. Brackmann, M., et al., *Antimicrobial resistance classification using MALDI-TOF-MS is not that easy: lessons from vancomycin-resistant Enterococcus faecium*. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 2020. **26**(3): p. 391-393.
90. Lasch, P., et al., *Insufficient discriminatory power of MALDI-TOF mass spectrometry for typing of Enterococcus faecium and Staphylococcus aureus isolates*. Journal of microbiological methods, 2014. **100**: p. 58-69.
91. Walker, J., et al., *Intact cell mass spectrometry (ICMS) used to type methicillin-resistant Staphylococcus aureus: media effects and inter-laboratory reproducibility*. Journal of microbiological methods, 2002. **48**(2-3): p. 117-26.
92. GmbH, B.D. *MALDI Biotyper HT Subtyping Module*. [cited 2023 24 July]; Available from: <https://www.bruker.com/en/products-and-solutions/microbiology-and-diagnostics/microbial-identification/mbt-subtyping-ruo-module.html>.
93. Cordovana, M., et al., *MALDI-TOF bacterial subtyping to detect antibiotic resistance*. Clinical mass spectrometry, 2019. **14 Pt A**: p. 3-8.
94. Weis, C.V., C.R. Jutzeler, and K. Borgwardt, *Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: a systematic review*. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 2020. **26**(10): p. 1310-1317.
95. Weis, C., et al., *Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning*. Nature medicine, 2022. **28**(1): p. 164-174.
96. Feucherolles, M., et al., *Combination of MALDI-TOF Mass Spectrometry and Machine Learning for Rapid Antimicrobial Resistance Screening: The Case of Campylobacter spp*. Frontiers in microbiology, 2021. **12**: p. 804484.
97. Sparbier, K., et al., *Matrix-assisted laser desorption ionization-time of flight mass spectrometry-based functional assay for rapid detection of resistance against beta-lactam antibiotics*. Journal of clinical microbiology, 2012. **50**(3): p. 927-37.
98. Oviano, M. and G. Bou, *Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry for the Rapid Detection of Antimicrobial Resistance Mechanisms and Beyond*. Clinical microbiology reviews, 2019. **32**(1).
99. Oviano, M., et al., *Rapid detection of the plasmid-mediated quinolone resistance determinant AAC(6')-Ib-cr in Enterobacteriaceae by MALDI-TOF MS analysis*. The Journal of antimicrobial chemotherapy, 2017. **72**(4): p. 1074-1080.
100. Pardo, C.A., et al., *Rapid detection of AAC(6')-Ib-cr production using a MALDI-TOF MS strategy*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2016. **35**(12): p. 2047-2051.
101. Oviano, M., et al., *Universal protocol for the rapid automated detection of carbapenem-resistant Gram-negative bacilli directly from blood cultures by matrix-*

- assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF/MS)*. International journal of antimicrobial agents, 2016. **48**(6): p. 655-660.
102. Oviano, M., et al., *Rapid direct detection of carbapenemase-producing Enterobacteriaceae in clinical urine samples by MALDI-TOF MS analysis*. The Journal of antimicrobial chemotherapy, 2017. **72**(5): p. 1350-1354.
 103. Oviano, M., et al., *Towards the early detection of beta-lactamase-producing Enterobacteriaceae by MALDI-TOF MS analysis*. The Journal of antimicrobial chemotherapy, 2017. **72**(8): p. 2259-2262.
 104. Dortet, L., et al., *MALDI-TOF for the rapid detection of carbapenemase-producing Enterobacteriaceae: comparison of the commercialized MBT STAR(R)-Carba IVD Kit with two in-house MALDI-TOF techniques and the RAPIDEC(R) CARBA NP*. The Journal of antimicrobial chemotherapy, 2018. **73**(9): p. 2352-2359.
 105. Lange, C., et al., *Quantitative matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid resistance detection*. Journal of clinical microbiology, 2014. **52**(12): p. 4155-62.
 106. Sparbier, K., et al., *MALDI biotyper-based rapid resistance detection by stable-isotope labeling*. Journal of clinical microbiology, 2013. **51**(11): p. 3741-8.
 107. Jung, J.S., et al., *Rapid detection of antibiotic resistance based on mass spectrometry and stable isotopes*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2014. **33**(6): p. 949-55.
 108. Florio, W., et al., *Detection of Antibiotic-Resistance by MALDI-TOF Mass Spectrometry: An Expanding Area*. Frontiers in cellular and infection microbiology, 2020. **10**: p. 572909.
 109. Idelevich, E.A., et al., *Rapid detection of antibiotic resistance by MALDI-TOF mass spectrometry using a novel direct-on-target microdroplet growth assay*. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 2018. **24**(7): p. 738-743.
 110. Idelevich, E.A., et al., *Rapid Simultaneous Testing of Multiple Antibiotics by the MALDI-TOF MS Direct-on-Target Microdroplet Growth Assay*. Diagnostics, 2021. **11**(10).
 111. Nix, I.D., et al., *MALDI-TOF Mass Spectrometry-Based Optochin Susceptibility Testing for Differentiation of Streptococcus pneumoniae from other Streptococcus mitis Group Streptococci*. Microorganisms, 2021. **9**(10).
 112. Nix, I.D., et al., *Detection of Methicillin Resistance in Staphylococcus aureus From Agar Cultures and Directly From Positive Blood Cultures Using MALDI-TOF Mass Spectrometry-Based Direct-on-Target Microdroplet Growth Assay*. Frontiers in microbiology, 2020. **11**: p. 232.
 113. Sauget, M., et al., *Can MALDI-TOF Mass Spectrometry Reasonably Type Bacteria?* Trends in microbiology, 2017. **25**(6): p. 447-455.
 114. Simar, S.R., B.M. Hanson, and C.A. Arias, *Techniques in bacterial strain typing: past, present, and future*. Current opinion in infectious diseases, 2021. **34**(4): p. 339-345.
 115. Valentine, N., et al., *Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry*. Applied and environmental microbiology, 2005. **71**(1): p. 58-64.
 116. Goldstein, J.E., et al., *Culture conditions and sample preparation methods affect spectrum quality and reproducibility during profiling of Staphylococcus aureus with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry*. Letters in applied microbiology, 2013. **57**(2): p. 144-50.
 117. Spinali, S., et al., *Microbial typing by matrix-assisted laser desorption ionization-time of flight mass spectrometry: do we need guidance for data interpretation?* Journal of clinical microbiology, 2015. **53**(3): p. 760-5.

BIBLIOGRAFÍA

118. Oberle, M., et al., *The Technical and Biological Reproducibility of Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS) Based Typing: Employment of Bioinformatics in a Multicenter Study*. PloS one, 2016. **11**(10): p. e0164260.
119. Huber, C.A., S.J. Reed, and D.L. Paterson, *Bacterial Sub-Species Typing Using Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry: What Is Promising?* Current issues in molecular biology, 2021. **43**(2): p. 749-757.
120. Rodriguez-Temporal, D., et al., *Identification of Mycobacterium abscessus Subspecies by MALDI-TOF Mass Spectrometry and Machine Learning*. Journal of clinical microbiology, 2023. **61**(1): p. e0111022.
121. Kehrmann, J., et al., *Principal component analysis of MALDI TOF MS mass spectra separates M. abscessus (sensu stricto) from M. massiliense isolates*. BMC microbiology, 2016. **16**: p. 24.
122. Mulet, X., et al., *O-antigen serotyping and MALDI-TOF, potentially useful tools for optimizing semi-empiric antipseudomonal treatments through the early detection of high-risk clones*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2019. **38**(3): p. 541-544.
123. Mulet, X., et al., *Validation of MALDI-TOF for the early detection of the ST175 high-risk clone of Pseudomonas aeruginosa in clinical isolates belonging to a Spanish nationwide multicenter study*. Enfermedades infecciosas y microbiología clinica, 2021. **39**(6): p. 279-282.
124. Cabrolier, N., et al., *Matrix-assisted laser desorption ionization-time of flight mass spectrometry identifies Pseudomonas aeruginosa high-risk clones*. Journal of clinical microbiology, 2015. **53**(4): p. 1395-8.
125. Rizzardi, K. and T. Akerlund, *High Molecular Weight Typing with MALDI-TOF MS - A Novel Method for Rapid Typing of Clostridium difficile*. PloS one, 2015. **10**(4): p. e0122457.
126. Reil, M., et al., *Recognition of Clostridium difficile PCR-ribotypes 001, 027 and 126/078 using an extended MALDI-TOF MS system*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2011. **30**(11): p. 1431-6.
127. Camoez, M., et al., *Automated categorization of methicillin-resistant Staphylococcus aureus clinical isolates into different clonal complexes by MALDI-TOF mass spectrometry*. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 2016. **22**(2): p. 161 e1-161 e7.
128. Sauget, M., et al., *Matrix-assisted laser desorption ionization-time of flight Mass spectrometry can detect Staphylococcus aureus clonal complex 398*. Journal of microbiological methods, 2016. **127**: p. 20-23.
129. Ueda, O., et al., *Development of a novel matrix-assisted laser desorption/ionization time-of-flight mass spectrum (MALDI-TOF-MS)-based typing method to identify methicillin-resistant Staphylococcus aureus clones*. The Journal of hospital infection, 2015. **90**(2): p. 147-55.
130. Zhang, T., et al., *Analysis of methicillin-resistant Staphylococcus aureus major clonal lineages by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS)*. Journal of microbiological methods, 2015. **117**: p. 122-7.
131. Wang, H.Y., et al., *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, 2018. **13**(3): p. e0194289.

132. Bar-Meir, M., et al., *The utility of MALDI-TOF MS for outbreak investigation in the neonatal intensive care unit*. European journal of pediatrics, 2020. **179**(12): p. 1843-1849.
133. Huang, Y., et al., *Rapid Detection of K1 Hypervirulent Klebsiella pneumoniae by MALDI-TOF MS*. Frontiers in microbiology, 2015. **6**: p. 1435.
134. Mansson, V., et al., *Capsule Typing of Haemophilus influenzae by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry*. Emerging infectious diseases, 2018. **24**(3): p. 443-452.
135. Ojima-Kato, T., et al., *Application of proteotyping Strain Solution ver. 2 software and theoretically calculated mass database in MALDI-TOF MS typing of Salmonella serotype*. Applied microbiology and biotechnology, 2017. **101**(23-24): p. 8557-8569.
136. Pinto, T.C., et al., *Potential of MALDI-TOF MS as an alternative approach for capsular typing Streptococcus pneumoniae isolates*. Scientific reports, 2017. **7**: p. 45572.
137. Rothen, J., et al., *Subspecies Typing of Streptococcus agalactiae Based on Ribosomal Subunit Protein Mass Variation by MALDI-TOF MS*. Frontiers in microbiology, 2019. **10**: p. 471.
138. Huang, L., et al., *Rapid Classification of Multilocus Sequence Subtype for Group B Streptococcus Based on MALDI-TOF Mass Spectrometry and Statistical Models*. Frontiers in cellular and infection microbiology, 2020. **10**: p. 577031.
139. Sauget, M., et al., *Matrix-assisted laser desorption ionization-time of flight mass spectrometry assigns Escherichia coli to the phylogroups A, B1, B2 and D*. International journal of medical microbiology : IJMM, 2014. **304**(8): p. 977-83.
140. Chui, H., et al., *Rapid, Sensitive, and Specific Escherichia coli H Antigen Typing by Matrix-Assisted Laser Desorption Ionization-Time of Flight-Based Peptide Mass Fingerprinting*. Journal of clinical microbiology, 2015. **53**(8): p. 2480-5.
141. Lafolie, J., et al., *Detection of Escherichia coli sequence type 131 by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: implications for infection control policies?* The Journal of hospital infection, 2015. **90**(3): p. 208-12.
142. Cuénod, A. and A. Egli, *Advanced Applications of MALDI-TOF MS – Typing and Beyond*, in *Application and Integration of Omics-powered Diagnostics in Clinical and Public Health Microbiology*, J. Moran-Gilad and Y. Yagel, Editors. 2021, Springer International Publishing: Cham. p. 153-173.
143. Naumann, D., V.L. Fijala, H., and P. Giesbrecht, *The rapid differentiation and identification of pathogenic bacteria using Fourier transform infrared spectroscopic and multivariate statistical analysis*. Journal of Molecular Structure, 1988. **174**: p. 165-170.
144. Naumann, D., D. Helm, and H. Labischinski, *Microbiological characterizations by FT-IR spectroscopy*. Nature, 1991. **351**(6321): p. 81-2.
145. Helm, D., et al., *Classification and identification of bacteria by Fourier-transform infrared spectroscopy*. Journal of general microbiology, 1991. **137**(1): p. 69-79.
146. Novais, Â. and L. Peixe, *Fourier Transform Infrared Spectroscopy (FT-IR) for Food and Water Microbiology*, in *Application and Integration of Omics-powered Diagnostics in Clinical and Public Health Microbiology*, J. Moran-Gilad and Y. Yagel, Editors. 2021, Springer International Publishing: Cham. p. 191-217.
147. Naumann, D., *Infrared spectroscopy in microbiology*. Encyclopedia of analytical chemistry, 2000. **102**: p. 131.
148. Novais, A., et al., *Fourier transform infrared spectroscopy: unlocking fundamentals and prospects for bacterial strain typing*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2019. **38**(3): p. 427-448.
149. Cordovana, M., et al., *Classification of Salmonella enterica of the (Para-)Typhoid Fever Group by Fourier-Transform Infrared (FTIR) Spectroscopy*. Microorganisms, 2021. **9**(4).

BIBLIOGRAFÍA

150. Cordovana, M., et al., *Machine learning-based typing of Salmonella enterica O-serogroups by the Fourier-Transform Infrared (FTIR) Spectroscopy-based IR Biotyper system*. Journal of microbiological methods, 2022. **201**: p. 106564.
151. Davis, R., et al., *Detection of E. coli O157:H7 from ground beef using Fourier transform infrared (FT-IR) spectroscopy and chemometrics*. Journal of food science, 2010. **75**(6): p. M340-6.
152. Abu-Aqil, G., et al., *Fast identification and susceptibility determination of E. coli isolated directly from patients' urine using infrared-spectroscopy and machine learning*. Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy, 2023. **285**: p. 121909.
153. Bagcioglu, M., et al., *Detection and Identification of Bacillus cereus, Bacillus cytotoxicus, Bacillus thuringiensis, Bacillus mycoides and Bacillus weihenstephanensis via Machine Learning Based FTIR Spectroscopy*. Frontiers in microbiology, 2019. **10**: p. 902.
154. Mouwen, D.J., et al., *Artificial neural network based identification of Campylobacter species by Fourier transform infrared spectroscopy*. Journal of microbiological methods, 2006. **67**(1): p. 131-40.
155. Wortberg, F., et al., *Identification of Yersinia ruckeri from diseased salmonid fish by Fourier transform infrared spectroscopy*. Journal of fish diseases, 2012. **35**(1): p. 1-10.
156. Stamm, I., et al., *Yersinia enterocolitica in diagnostic fecal samples from European dogs and cats: identification by fourier transform infrared spectroscopy and matrix-assisted laser desorption ionization-time of flight mass spectrometry*. Journal of clinical microbiology, 2013. **51**(3): p. 887-93.
157. Davis, R. and L.J. Mauer, *Subtyping of Listeria monocytogenes at the haplotype level by Fourier transform infrared (FT-IR) spectroscopy and multivariate statistical analysis*. International journal of food microbiology, 2011. **150**(2-3): p. 140-9.
158. Janbu, A.O., et al., *FT-IR microspectroscopy: a promising method for the rapid identification of Listeria species*. FEMS microbiology letters, 2008. **278**(2): p. 164-70.
159. Martak, D., et al., *Fourier-Transform InfraRed Spectroscopy Can Quickly Type Gram-Negative Bacilli Responsible for Hospital Outbreaks*. Frontiers in microbiology, 2019. **10**: p. 1440.
160. Tata, A., et al., *A multi-center validation study on the discrimination of Legionella pneumophila sg.1, Legionella pneumophila sg. 2-15 and Legionella non-pneumophila isolates from water by FT-IR spectroscopy*. Frontiers in microbiology, 2023. **14**: p. 1150942.
161. Pascale, M.R., et al., *Use of Fourier-Transform Infrared Spectroscopy With IR Biotyper(R) System for Legionella pneumophila Serogroups Identification*. Frontiers in microbiology, 2022. **13**: p. 866426.
162. Dinkelacker, A.G., et al., *Typing and Species Identification of Clinical Klebsiella Isolates by Fourier Transform Infrared Spectroscopy and Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry*. Journal of clinical microbiology, 2018. **56**(11).
163. Silva, L., et al., *Fourier transform infrared (FT-IR) spectroscopy typing: a real-time analysis of an outbreak by carbapenem-resistant Klebsiella pneumoniae*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2020. **39**(12): p. 2471-2475.
164. Wang-Wang, J.H., et al., *Evaluation of Fourier Transform Infrared Spectroscopy as a First-Line Typing Tool for the Identification of Extended-Spectrum beta-Lactamase-Producing Klebsiella pneumoniae Outbreaks in the Hospital Setting*. Frontiers in microbiology, 2022. **13**: p. 897161.

165. Dieckmann, R., et al., *Rapid characterisation of Klebsiella oxytoca isolates from contaminated liquid hand soap using mass spectrometry, FTIR and Raman spectroscopy*. Faraday discussions, 2016. **187**: p. 353-75.
166. Vogt, S., et al., *Fourier-Transform Infrared (FTIR) Spectroscopy for Typing of Clinical Enterobacter cloacae Complex Isolates*. Frontiers in microbiology, 2019. **10**: p. 2582.
167. Sousa, C., et al., *Identification of carbapenem-resistant Acinetobacter baumannii clones using infrared spectroscopy*. Journal of biophotonics, 2014. **7**(5): p. 287-94.
168. Lombardo, D., et al., *Application of Fourier transform infrared spectroscopy for real-time typing of Acinetobacter baumannii outbreak in intensive care unit*. Future microbiology, 2021. **16**: p. 1239-1250.
169. Hu, Y., et al., *Evaluation of IR Biotyper for carbapenem-resistant Pseudomonas aeruginosa typing and its application potential for the investigation of nosocomial infection*. Frontiers in microbiology, 2023. **14**: p. 1068872.
170. Passaris, I., et al., *Validation of Fourier Transform Infrared Spectroscopy for Serotyping of Streptococcus pneumoniae*. Journal of clinical microbiology, 2022. **60**(7): p. e0032522.
171. Jöhler, S., et al., *High-resolution subtyping of Staphylococcus aureus strains by means of Fourier-transform infrared spectroscopy*. Systematic and applied microbiology, 2016. **39**(3): p. 189-194.
172. Grunert, T., et al., *Rapid and reliable identification of Staphylococcus aureus capsular serotypes by means of artificial neural network-assisted Fourier transform infrared spectroscopy*. Journal of clinical microbiology, 2013. **51**(7): p. 2261-6.
173. Amiali, N.M., et al., *Epidemiological typing of methicillin-resistant Staphylococcus aureus strains by Fourier transform infrared spectroscopy*. Journal of microbiological methods, 2007. **69**(1): p. 146-53.
174. Curtoni, A., et al., *Application of FT-IR Spectroscopy for Mycobacterium abscessus complex subspecies differentiation*. Journal of microbiological methods, 2023: p. 106792.
175. IBM. *¿Qué es Machine Learning?* 2023 [cited 2023 August 18]; Available from: <https://www.ibm.com/es-es/topics/machine-learning>.
176. Alvarez-Vega, M., L.M. Quiros-Mora, and M.V. Cortes-Badilla, *Artificial intelligence and machine learning in medicine*. Revista Médica Sinergia, 2020. **5**(8).
177. Priyadarshini. *What Is Machine Learning and How Does It Work?* 2023 March 10, 2023 [cited 2023 August 21]; Available from: <https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-machine-learning>.
178. Mora-Pineda, J., *Modelos predictivos en salud basados en aprendizaje de maquina (machine learning)* Revista Médica Clínica Las Condes, 2022. **33**(6): p. 583-590.
179. Géron, A., *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow : concepts, tools, and techniques to build intelligent systems*. Second edition. ed2019: O'Reilly Media, Inc. xxv, 819 pages.
180. Goodswen, S.J., et al., *Machine learning and applications in microbiology*. FEMS microbiology reviews, 2021. **45**(5).
181. Shoshan, Y., et al., *Artificial Intelligence for Reducing Workload in Breast Cancer Screening with Digital Breast Tomosynthesis*. Radiology, 2022. **303**(1): p. 69-77.
182. Lam, C.S.N., et al., *A machine learning model for colorectal liver metastasis post-hepatectomy prognostications*. Hepatobiliary surgery and nutrition, 2023. **12**(4): p. 495-506.
183. Vasilev, Y., et al., *Clinical application of radiological AI for pulmonary nodule evaluation: Replicability and susceptibility to the population shift caused by the COVID-19 pandemic*. International journal of medical informatics, 2023. **178**: p. 105190.

BIBLIOGRAFÍA

184. Ueda, D., et al., *Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study*. *The Lancet. Digital health*, 2023. **5**(8): p. e525-e533.
185. Boulitsakis Logothetis, S., et al., *Predicting acute clinical deterioration with interpretable machine learning to support emergency care decision making*. *Scientific reports*, 2023. **13**(1): p. 13563.
186. Liu, X., et al., *Illness severity assessment of older adults in critical illness using machine learning (ELDER-ICU): an international multicentre study with subgroup bias evaluation*. *The Lancet. Digital health*, 2023.
187. Pascucci, M., et al., *AI-based mobile application to fight antibiotic resistance*. *Nature communications*, 2021. **12**(1): p. 1173.
188. Rajaraman, S., et al., *Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images*. *PeerJ*, 2018. **6**: p. e4568.
189. Mathison, B.A., et al., *Detection of Intestinal Protozoa in Trichrome-Stained Stool Specimens by Use of a Deep Convolutional Neural Network*. *Journal of clinical microbiology*, 2020. **58**(6).
190. Smith, K.P., A.D. Kang, and J.E. Kirby, *Automated Interpretation of Blood Culture Gram Stains by Use of a Deep Convolutional Neural Network*. *Journal of clinical microbiology*, 2018. **56**(3).
191. Chung, C.R., et al., *Towards Accurate Identification of Antibiotic-Resistant Pathogens through the Ensemble of Multiple Preprocessing Methods Based on MALDI-TOF Spectra*. *International journal of molecular sciences*, 2023. **24**(2).
192. Zvezdanova, M.E., et al., *Detection of azole resistance in Aspergillus fumigatus complex isolates using MALDI-TOF mass spectrometry*. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 2022. **28**(2): p. 260-266.
193. Li, Y., et al., *Accurate classification of Listeria species by MALDI-TOF mass spectrometry incorporating denoising autoencoder and machine learning*. *Journal of microbiological methods*, 2022. **192**: p. 106378.
194. Jeon, K., et al., *Performance of a Machine Learning-Based Methicillin Resistance of Staphylococcus aureus Identification System Using MALDI-TOF MS and Comparison of the Accuracy according to SCCmec Types*. *Microorganisms*, 2022. **10**(10).
195. Zvezdanova, M.E., et al., *Implementation of MALDI-TOF Mass Spectrometry and Peak Analysis: Application to the Discrimination of Cryptococcus neoformans Species Complex and Their Interspecies Hybrids*. *Journal of fungi*, 2020. **6**(4).
196. Weis, C., et al., *Topological and kernel-based microbial phenotype prediction from MALDI-TOF mass spectra*. *Bioinformatics*, 2020. **36**(Suppl_1): p. i30-i38.
197. van Oosten, L.N. and C.D. Klein, *Machine Learning in Mass Spectrometry: A MALDI-TOF MS Approach to Phenotypic Antibacterial Screening*. *Journal of medicinal chemistry*, 2020. **63**(16): p. 8849-8856.
198. Tacconelli, E., et al., *Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis*. *The Lancet. Infectious diseases*, 2018. **18**(3): p. 318-327.
199. Centers for Disease Control and Prevention, A.G. *Centers for Disease Control and Prevention: Antibiotic Resistance Threats in the United States*. 2019; Available from: <https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf>.
200. Rice, L.B., *Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE*. *The Journal of infectious diseases*, 2008. **197**(8): p. 1079-81.

201. De Rosa, F.G., et al., *From ESKAPE to ESCAPE, from KPC to CCC*. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2015. **60**(8): p. 1289-90.
202. Peterson, L.R., *Bad bugs, no drugs: no ESCAPE revisited*. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2009. **49**(6): p. 992-3.
203. Eatemadi, A., et al., "*ESKAPE*"? No, "*ESCAPE*"! Journal of Emerging Technologies and Innovative Research (JETIR), 2021. **8**(1): p. 2.
204. Flores-Trevino, S., et al., *Screening of biomarkers of drug resistance or virulence in ESCAPE pathogens by MALDI-TOF mass spectrometry*. Scientific reports, 2019. **9**(1): p. 18945.
205. Marin, M., et al., *Clostridium difficile isolates with high linezolid MICs harbor the multiresistance gene cfr*. Antimicrobial agents and chemotherapy, 2015. **59**(1): p. 586-9.
206. Indra, A., et al., *Characterization of Clostridium difficile isolates using capillary gel electrophoresis-based PCR ribotyping*. Journal of medical microbiology, 2008. **57**(Pt 11): p. 1377-1382.
207. Stubbs, S.L., et al., *PCR targeted to the 16S-23S rRNA gene intergenic spacer region of Clostridium difficile and construction of a library consisting of 116 different PCR ribotypes*. Journal of clinical microbiology, 1999. **37**(2): p. 461-3.
208. Reigadas, E., et al., *Breakthrough Clostridium difficile Infection in Cirrhotic Patients Receiving Rifaximin*. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2018. **66**(7): p. 1086-1091.
209. Emele, M.F., et al., *Proteotyping of Clostridioides difficile as Alternate Typing Method to Ribotyping Is Able to Distinguish the Ribotypes RT027 and RT176 From Other Ribotypes*. Frontiers in microbiology, 2019. **10**: p. 2087.
210. O'Neill, J., *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations*. Review on Antimicrobial Resistance. Wellcome Trust and HM Government., 2016.
211. Collaborators*, A.R., *Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis*. Lancet, 2022. **399**(10325): p. 629-655.
212. Laxminarayan, R., *The overlooked pandemic of antimicrobial resistance*. Lancet, 2022. **399**(10325): p. 606-607.
213. Tsuchida, S., H. Umemura, and T. Nakayama, *Current Status of Matrix-Assisted Laser Desorption/Ionization-Time-of-Flight Mass Spectrometry (MALDI-TOF MS) in Clinical Diagnostic Microbiology*. Molecules, 2020. **25**(20).
214. Gupta, A., et al., *Matrix-Assisted Laser Desorption Ionization Time of Flight (MALDI-TOF) as an Indispensable Tool in Diagnostic Bacteriology: A Comparative Analysis With Conventional Technique*. Cureus, 2023. **15**(3): p. e36984.
215. Becker, K. and A. Lupetti, *Editorial: MALDI-TOF MS in microbiological diagnostics: future applications beyond identification*. Frontiers in microbiology, 2023. **14**: p. 1204452.
216. Mensa, J., et al., *Guía de Terapéutica Antimicrobiana 2023*2023: Editorial Antares. 1106.
217. Control., E.C.f.D.P.a., *Antimicrobial resistance in the EU/EEA (EARS-Net) - Annual Epidemiological Report 2021*. ECDC, 2022.
218. Garcia-Solache, M. and L.B. Rice, *The Enterococcus: a Model of Adaptability to Its Environment*. Clinical microbiology reviews, 2019. **32**(2).
219. Sabenca, C., et al., *Next-Generation Sequencing and MALDI Mass Spectrometry in the Study of Multiresistant Processed Meat Vancomycin-Resistant Enterococci (VRE)*. Biology, 2020. **9**(5).
220. Cercenado, E., *[Enterococcus: phenotype and genotype resistance and epidemiology in Spain]*. Enfermedades infecciosas y microbiología clínica, 2011. **29 Suppl 5**: p. 59-65.

BIBLIOGRAFÍA

221. Quintela-Baluja, M., et al., *Characterization of different food-isolated Enterococcus strains by MALDI-TOF mass fingerprinting*. Electrophoresis, 2013. **34**(15): p. 2240-50.
222. Rodriguez-Temporal, D., et al., *Multicentre study on the reproducibility of MALDI-TOF MS for nontuberculous mycobacteria identification*. Scientific reports, 2022. **12**(1): p. 1237.
223. Paskova, V., et al., *Insufficient repeatability and reproducibility of MALDI-TOF MS-based identification of MRSA*. Folia microbiologica, 2020. **65**(5): p. 895-900.
224. Timm, W., et al., *Peak intensity prediction in MALDI-TOF mass spectrometry: a machine learning study to support quantitative proteomics*. BMC bioinformatics, 2008. **9**: p. 443.
225. Freitas, A.R., et al., *Rapid detection of high-risk Enterococcus faecium clones by matrix-assisted laser desorption ionization time-of-flight mass spectrometry*. Diagnostic microbiology and infectious disease, 2017. **87**(4): p. 299-307.
226. Wang, L.J., et al., *Application of matrix-assisted laser desorption ionization time-of-flight mass spectrometry in the screening of vanA-positive Enterococcus faecium*. European journal of mass spectrometry, 2014. **20**(6): p. 461-5.
227. Nakano, S., et al., *Differentiation of vanA-positive Enterococcus faecium from vanA-negative E. faecium by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry*. International journal of antimicrobial agents, 2014. **44**(3): p. 256-9.
228. Schlebusch, S., et al., *MALDI-TOF MS meets WGS in a VRE outbreak investigation*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2017. **36**(3): p. 495-499.
229. Savas, S., et al., *From days to hours: Can MALDI-TOF MS system replace both conventional and molecular typing methods with new cut off level for Vancomycin Resistant Enterococcus faecium*. Journal of microbiological methods, 2019. **162**: p. 62-68.
230. Rangberg, A., et al., *Molecular analysis and epidemiological typing of Vancomycin-resistant Enterococcus outbreak strains*. Scientific reports, 2019. **9**(1): p. 11917.
231. Wang, H.Y., et al., *Clinically Applicable System for Rapidly Predicting Enterococcus faecium Susceptibility to Vancomycin*. Microbiology spectrum, 2021. **9**(3): p. e0091321.
232. Wang, H.Y., et al., *Efficiently Predicting Vancomycin Resistance of Enterococcus Faecium From MALDI-TOF MS Spectra Using a Deep Learning-Based Approach*. Frontiers in microbiology, 2022. **13**: p. 821233.
233. Scheier, T.C., et al., *Fourier-transform infrared spectroscopy for typing of vancomycin-resistant Enterococcus faecium: performance analysis and outbreak investigation*. Microbiology spectrum, 2023. **11**(5): p. e0098423.
234. Santos, T., et al., *Use of MALDI-TOF mass spectrometry fingerprinting to characterize Enterococcus spp. and Escherichia coli isolates*. Journal of proteomics, 2015. **127**(Pt B): p. 321-31.
235. Hernandez-Alonso, E., et al., *Contaminated Incubators: Source of a Multispecies Enterobacter Outbreak of Neonatal Sepsis*. Microbiology spectrum, 2022. **10**(4): p. e0096422.
236. Davin-Regli, A., J.P. Lavigne, and J.M. Pages, *Enterobacter spp.: Update on Taxonomy, Clinical Aspects, and Emerging Antimicrobial Resistance*. Clinical microbiology reviews, 2019. **32**(4).
237. Talon, D., et al., *Emergence of Enterobacter cloacae as a common pathogen in neonatal units: pulsed-field gel electrophoresis analysis*. The Journal of hospital infection, 2004. **57**(2): p. 119-25.
238. Girlich, D., et al., *Uncovering the novel Enterobacter cloacae complex species responsible for septic shock deaths in newborns: a cohort study*. The Lancet. Microbe, 2021. **2**(10): p. e536-e544.

239. Rivera-Izquierdo, M., *Alarming increase in hospital outbreaks of carbapenemase-producing Enterobacter cloacae in Southern Spain*. The Journal of hospital infection, 2023.
240. Hoffmann, H. and A. Roggenkamp, *Population genetics of the nomenclature species Enterobacter cloacae*. Applied and environmental microbiology, 2003. **69**(9): p. 5306-18.
241. Mezzatesta, M.L., F. Gona, and S. Stefani, *Enterobacter cloacae complex: clinical impact and emerging antibiotic resistance*. Future microbiology, 2012. **7**(7): p. 887-902.
242. Pavlovic, M., et al., *A dual approach employing MALDI-TOF MS and real-time PCR for fast species identification within the Enterobacter cloacae complex*. FEMS microbiology letters, 2012. **328**(1): p. 46-53.
243. Hernandez-Alonso, E., et al., *dnaJ: a New Approach to Identify Species within the Genus Enterobacter*. Microbiology spectrum, 2021. **9**(3): p. e0124221.
244. Tindall, B.J., G. Sutton, and G.M. Garrity, *Enterobacter aerogenes Hormaeche and Edwards 1960 (Approved Lists 1980) and Klebsiella mobilis Bascomb et al. 1971 (Approved Lists 1980) share the same nomenclature type (ATCC 13048) on the Approved Lists and are homotypic synonyms, with consequences for the name Klebsiella mobilis Bascomb et al. 1971 (Approved Lists 1980)*. International journal of systematic and evolutionary microbiology, 2017. **67**(2): p. 502-504.
245. Iversen, C., et al., *The taxonomy of Enterobacter sakazakii: proposal of a new genus Cronobacter gen. nov. and descriptions of Cronobacter sakazakii comb. nov., Cronobacter sakazakii subsp. sakazakii, comb. nov., Cronobacter sakazakii subsp. malonicus subsp. nov., Cronobacter turicensis sp. nov., Cronobacter muytjensii sp. nov., Cronobacter dublinensis sp. nov. and Cronobacter genomospecies 1*. BMC evolutionary biology, 2007. **7**: p. 64.
246. Brady, C., et al., *Taxonomic evaluation of the genus Enterobacter based on multilocus sequence analysis (MLSA): proposal to reclassify E. nimipressuralis and E. amnigenus into Lelliottia gen. nov. as Lelliottia nimipressuralis comb. nov. and Lelliottia amnigena comb. nov., respectively, E. gergoviae and E. pyrinus into Pluralibacter gen. nov. as Pluralibacter gergoviae comb. nov. and Pluralibacter pyrinus comb. nov., respectively, E. cowanii, E. radicincitans, E. oryzae and E. arachidis into Kosakonia gen. nov. as Kosakonia cowanii comb. nov., Kosakonia radicincitans comb. nov., Kosakonia oryzae comb. nov. and Kosakonia arachidis comb. nov., respectively, and E. turicensis, E. helveticus and E. pulveris into Cronobacter as Cronobacter zurichensis nom. nov., Cronobacter helveticus comb. nov. and Cronobacter pulveris comb. nov., respectively, and emended description of the genera Enterobacter and Cronobacter*. Systematic and applied microbiology, 2013. **36**(5): p. 309-19.
247. Wu, W., Y. Feng, and Z. Zong, *Precise Species Identification for Enterobacter: a Genome Sequence-Based Study with Reporting of Two Novel Species, Enterobacter quasiroggenkampii sp. nov. and Enterobacter quasimori sp. nov.* mSystems, 2020. **5**(4).
248. Hoffmann, H., et al., *Reassignment of enterobacter dissolvens to Enterobacter cloacae as E. cloacae subspecies dissolvens comb. nov. and emended description of Enterobacter asburiae and Enterobacter kobei*. Systematic and applied microbiology, 2005. **28**(3): p. 196-205.
249. Hoffmann, H., et al., *Enterobacter hormaechei subsp. oharae subsp. nov., E. hormaechei subsp. hormaechei comb. nov., and E. hormaechei subsp. steigerwaltii subsp. nov., three new subspecies of clinical importance*. Journal of clinical microbiology, 2005. **43**(7): p. 3297-303.
250. Hoffmann, H., et al., *Description of Enterobacter ludwigii sp. nov., a novel Enterobacter species of clinical relevance*. Systematic and applied microbiology, 2005. **28**(3): p. 206-12.

BIBLIOGRAFÍA

251. Pati, N.B., et al., *Enterobacter bugandensis: a novel enterobacterial species associated with severe clinical infection*. Scientific reports, 2018. **8**(1): p. 5392.
252. Sutton, G.G., et al., *Enterobacter hormaechei subsp. hoffmannii subsp. nov., Enterobacter hormaechei subsp. xiangfangensis comb. nov., Enterobacter roggenkampii sp. nov., and Enterobacter muelleri is a later heterotypic synonym of Enterobacter asburiae based on computational analysis of sequenced Enterobacter genomes*. F1000Research, 2018. **7**: p. 521.
253. Morand, P.C., et al., *Specific distribution within the Enterobacter cloacae complex of strains isolated from infected orthopedic implants*. Journal of clinical microbiology, 2009. **47**(8): p. 2489-95.
254. Yeh, T.K., et al., *Antibiotic resistance in Enterobacter hormaechei*. International journal of antimicrobial agents, 2022. **60**(4): p. 106650.
255. Godmer, A., et al., *Revisiting Species Identification within the Enterobacter cloacae Complex by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry*. Microbiology spectrum, 2021. **9**(1): p. e0066121.
256. Khennouchi, N.C., et al., *MALDI-TOF MS as a Tool To Detect a Nosocomial Outbreak of Extended-Spectrum-beta-Lactamase- and ArmA Methyltransferase-Producing Enterobacter cloacae Clinical Isolates in Algeria*. Antimicrobial agents and chemotherapy, 2015. **59**(10): p. 6477-83.
257. De Florio, L., et al., *MALDI-TOF MS Identification and Clustering Applied to Enterobacter Species in Nosocomial Setting*. Frontiers in microbiology, 2018. **9**: p. 1885.
258. Wang, Y.Q., et al., *Rapid Identification and Subtyping of Enterobacter cloacae Clinical Isolates Using Peptide Mass Fingerprinting*. Biomedical and environmental sciences : BES, 2018. **31**(1): p. 48-56.
259. Magiorakos, A.P., et al., *Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance*. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 2012. **18**(3): p. 268-81.
260. Kadri, S.S., et al., *Difficult-to-Treat Resistance in Gram-negative Bacteremia at 173 US Hospitals: Retrospective Cohort Analysis of Prevalence, Predictors, and Outcome of Resistance to All First-line Agents*. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2018. **67**(12): p. 1803-1814.
261. Strich, J.R. and S.S. Kadri, *Difficult-to-Treat Antibiotic-Resistant Gram-Negative Pathogens in the Intensive Care Unit: Epidemiology, Outcomes, and Treatment*. Seminars in respiratory and critical care medicine, 2019. **40**(4): p. 419-434.
262. Horcajada, J.P., et al., *Epidemiology and Treatment of Multidrug-Resistant and Extensively Drug-Resistant Pseudomonas aeruginosa Infections*. Clinical microbiology reviews, 2019. **32**(4).
263. Nabal Diaz, S.G., O. Algara Robles, and J.M. Garcia-Lechuz Moya, *New definitions of susceptibility categories EUCAST 2019: clinic application*. Revista espanola de quimioterapia : publicacion oficial de la Sociedad Espanola de Quimioterapia, 2022. **35 Suppl 3**(Suppl 3): p. 84-88.
264. Reynolds, D. and M. Kollef, *The Epidemiology and Pathogenesis and Treatment of Pseudomonas aeruginosa Infections: An Update*. Drugs, 2021. **81**(18): p. 2117-2131.
265. Pang, Z., et al., *Antibiotic resistance in Pseudomonas aeruginosa: mechanisms and alternative therapeutic strategies*. Biotechnology advances, 2019. **37**(1): p. 177-192.
266. Acosta, F., et al., *In-Depth Study of a Nosocomial Outbreak Caused by Extensively Drug-Resistant Pseudomonas aeruginosa Using Whole Genome Sequencing Coupled With a Polymerase Chain Reaction Targeting Strain-Specific Single Nucleotide Polymorphisms*. American journal of epidemiology, 2020. **189**(8): p. 841-849.

267. Pang, R.T.K., et al., *Technical evaluation of MALDI-TOF mass spectrometry for quantitative proteomic profiling matrix formulation and application*. Clinical Proteomics, 2004. **1**: p. 259-270.
268. Hoyos-Mallemcot, Y., et al., *MALDI-TOF MS, a useful instrument for differentiating metallo-beta-lactamases in Enterobacteriaceae and Pseudomonas spp.* Letters in applied microbiology, 2014. **58**(4): p. 325-9.
269. Miltgen, G., et al., *Detection of carbapenemase activity in Pseudomonas aeruginosa by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS)*. Journal of microbiological methods, 2018. **145**: p. 66-68.
270. Sanchez, D., et al., *A rapid MALDI-TOF mass spectrometry-based method for categorization of Pseudomonas aeruginosa from blood cultures as susceptible or resistant to meropenem*. Journal of microbiological methods, 2023. **212**: p. 106803.
271. Jeannot, K., et al., *Detection of Colistin Resistance in Pseudomonas aeruginosa Using the MALDIxin Test on the Routine MALDI Biotyper Sirius Mass Spectrometer*. Frontiers in microbiology, 2021. **12**: p. 725383.
272. Scotta, C., et al., *Whole-cell MALDI-TOF mass spectrometry and multilocus sequence analysis in the discrimination of Pseudomonas stutzeri populations: three novel genomovars*. Microbial ecology, 2013. **66**(3): p. 522-32.
273. Angrup, A., et al., *Role of MALDI-TOF mass spectrometry and molecular typing methods in an outbreak investigation of Pseudomonas stutzeri acute endophthalmitis post-phacoemulsification*. The Journal of hospital infection, 2023.
274. Teramoto, K., et al., *Phylogenetic classification of Pseudomonas putida strains by MALDI-MS using ribosomal subunit proteins as biomarkers*. Analytical chemistry, 2007. **79**(22): p. 8712-9.
275. Czepliel, J., et al., *Clostridium difficile infection: review*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2019. **38**(7): p. 1211-1221.
276. Alcalá-Hernández, L., et al., *53. Diagnóstico microbiológico de la infección por Clostridium difficile*. Procedimientos en Microbiología Clínica. Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica (SEIMC), 2015.
277. Burnham, C.A. and K.C. Carroll, *Diagnosis of Clostridium difficile infection: an ongoing conundrum for clinicians and for clinical laboratories*. Clinical microbiology reviews, 2013. **26**(3): p. 604-30.
278. Wolff, D., T. Bruning, and A. Gerritzen, *Rapid detection of the Clostridium difficile ribotype 027 tcdC gene frame shift mutation at position 117 by real-time PCR and melt curve analysis*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2009. **28**(8): p. 959-62.
279. Kachrimanidou, M., et al., *An outbreak of Clostridioides difficile infections due to a 027-like PCR ribotype 181 in a rehabilitation centre: Epidemiological and microbiological characteristics*. Anaerobe, 2020. **65**: p. 102252.
280. Kachrimanidou, M., et al., *Predominance of Clostridioides difficile PCR ribotype 181 in northern Greece, 2016-2019*. Anaerobe, 2022. **76**: p. 102601.
281. Baktash, A., et al., *Comparison of Whole-Genome Sequence-Based Methods and PCR Ribotyping for Subtyping of Clostridioides difficile*. Journal of clinical microbiology, 2022. **60**(2): p. e0173721.
282. Davies, K.A., et al., *Diversity of Clostridium difficile PCR ribotypes in Europe: results from the European, multicentre, prospective, biannual, point-prevalence study of Clostridium difficile infection in hospitalised patients with diarrhoea (EUCLID), 2012 and 2013*. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin, 2016. **21**(29).
283. (ECDC), E.C.f.D.P.a.C. *Clostridioides (Clostridium) difficile infection. Annual epidemiological report for 2016-2017*. 2022; Available from:

BIBLIOGRAFÍA

- <https://www.ecdc.europa.eu/en/publications-data/clostridioides-difficile-infections-annual-epidemiological-report-2016-2017>.
284. Corver, J., et al., *Identification and validation of two peptide markers for the recognition of Clostridioides difficile MLST-1 and MLST-11 by MALDI-MS*. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases, 2019. **25**(7): p. 904 e1-904 e7.
 285. Calderaro, A., et al., *Rapid Classification of Clostridioides difficile Strains Using MALDI-TOF MS Peak-Based Assay in Comparison with PCR-Ribotyping*. Microorganisms, 2021. **9**(3).
 286. Calderaro, A., et al., *Characterization of Clostridioides difficile Strains from an Outbreak Using MALDI-TOF Mass Spectrometry*. Microorganisms, 2022. **10**(7).
 287. Ortega, L., A. Ryberg, and A. Johansson, *HMW-profiling using MALDI-TOF MS: A screening method for outbreaks of Clostridioides difficile*. Anaerobe, 2018. **54**: p. 254-259.
 288. Carneiro, L.G., et al., *MALDI-TOF MS: An alternative approach for ribotyping Clostridioides difficile isolates in Brazil*. Anaerobe, 2021. **69**: p. 102351.
 289. Li, R., et al., *Identification and Characterization of Clostridium difficile Sequence Type 37 Genotype by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry*. Journal of clinical microbiology, 2018. **56**(5).
 290. Cheng, J.W., et al., *Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify MLST clade 4 Clostridium difficile isolates*. Diagnostic microbiology and infectious disease, 2018. **92**(1): p. 19-24.
 291. Uzuriaga, M., et al., *Clinical Impact of Rapid Bacterial Microbiological Identification with the MALDI-TOF MS*. Antibiotics, 2023. **12**(12).