



UNIVERSIDAD
NACIONAL
DE COLOMBIA

**Predicción del perfil de resistencia a
partir de las secuencias del genoma
de aislamientos colombianos de
*Acinetobacter baumannii***

Karen Jhovana Aguilar González

Universidad Nacional de Colombia
Facultad de Ciencias, Posgrado Interfacultades
Bogotá, Colombia

2021

Predicción del perfil de resistencia a partir de las secuencias del genoma de aislamientos colombianos de *Acinetobacter baumannii*

Karen Jhovana Aguilar González

Tesis de investigación presentada como requisito parcial para optar al título de:
Magister en Ciencias -Microbiología

Director:

PhD. Emiliano Barreto Hernández

Codirectora:

MSc. María Teresa Jesús Reguero Reza

Línea de Investigación:

Biología Molecular de Agentes Infecciosos

Grupos de Investigación:

Bioinformática y EPIMOL

Universidad Nacional de Colombia

Facultad de Ciencias, Posgrado Interfacultades

Bogotá, Colombia

2021

“Recuerda mirar a las estrellas y no tus pies. Intenta dar sentido a lo que ves y pregúntate por lo que hace al universo existir. Sé curioso. Aunque la vida puede parecer difícil, siempre hay algo que puedes hacer y tener éxito. Lo importante es que no te rindas.”

Stephen Hawking

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

Karen Aguilar González

Fecha 09/04/2021

Agradecimientos

Siempre que uno termina una etapa empieza a recordar el camino recorrido, y sin duda visualiza todas esas personas que estuvieron involucradas en este hermoso y a la vez difícil proceso. Hoy que tengo la oportunidad quiero mencionar a todas aquellas que hicieron posible estar en donde estoy. Agradezco profundamente a mis padres, quienes hicieron posible mi estadía en un país ajeno, quienes me han enseñado a perseverar y jamás rendirme, gracias porque a pesar de la distancia encontraron la manera de siempre hacerme sentir cerca, y mil gracias por confiar en mí y apoyarme en cada uno de los proyectos y metas que me propongo. A mis hermanas, quienes se caracterizan por llenar de alegría mis días, gracias por escucharme en momentos difíciles y celebrar buenos momentos. A mi novio David Ochoa, gracias por su infinita paciencia, por su amor, muchas gracias por correr junto a mi lado durante este proceso y empujarme cuando quería detenerme.

Igualmente, un grato agradecimiento al profesor Emiliano Barreto y la profesora Ma. Teresa Reguero, quienes compartieron conmigo sus enormes e invaluable conocimientos, gracias por su tiempo, por cada una de sus correcciones y observaciones que han permitido conseguir cada uno de los objetivos planteados en el presente estudio. A mis compañeras de este grupo de investigación, pero más que nada amigas Caridad Tenorio, Mishelle Coello y Laura Rojas quiero agradecerles por su apoyo constante y por todas las experiencias compartidas. También agradezco a Diego Talero quien me colaboró mucho para que sea posible el desarrollo de este trabajo.

Finalmente agradezco al Instituto de Biotecnología quien me brindó el espacio y los recursos necesarios que hicieron posible esta investigación, a la Universidad Nacional de

Colombia por permitirme ser parte de tan prestigiosa Institución y finalmente gracias Colombia por la acogida durante estos dos años de formación Mil gracias.

Resumen

Predicción del perfil de resistencia a partir de las secuencias del genoma de aislamientos colombianos de *Acinetobacter baumannii*

La creciente resistencia a los antibióticos y las pocas alternativas terapéuticas disponibles vuelven una urgencia la necesidad de optimizar los diagnósticos actuales que nos permitan prescripciones más rápidas y efectivas. Últimamente uno de los enfoques para predecir resistencia a partir de los datos de secuenciación de genoma consta en aplicar modelos basados en aprendizaje de máquina, los cuales han ido tomando credibilidad debido a la capacidad de realizar predicciones precisas. Además, gracias al creciente conocimiento acerca de mecanismos de resistencia asociados a *A. baumannii*, este patógeno nos brinda una alternativa para desarrollar estos modelos.

En este trabajo se utilizaron 343 genomas, 76 colombianos del Instituto Nacional de Salud y 267 recolectados de la base de datos *Biosample NCBI*, para la obtención de modelos basados en aprendizaje de máquina empleando regresión *lasso*, *random forest* y *gradient boosting*, para predecir la concentración mínima inhibitoria de 10 antibióticos. *Random forest* fue el algoritmo que mostró los mejores resultados, logrando una precisión promedio dentro de +/- una dilución doble de 91% (I.C 95, 85- 97), una tasa de *very major error* y *major error* de 1,71% y 0,7%, respectivamente. Como datos de entrada para los modelos se utilizaron genes de resistencia, los cuales fueron identificados utilizando el software *Resistance Gene Identifier*. Estos resultados demuestran que la predicción de la susceptibilidad de *A. baumannii* a los antibióticos, basada en la secuencia del genoma son prometedoras como posibles herramienta de diagnóstico en la clínica.

Palabras clave: *Acinetobacter baumannii*, aprendizaje de máquina, predicción fenotípica de resistencia, concentración mínima inhibitoria, regresión *lasso*, *random forest*, *gradient boosting*

Abstract

Prediction of the resistance profile from the genome sequences of Colombian isolates of *Acinetobacter baumannii*

The increasing resistance to antibiotics and the few therapeutic choices available turns into an urgency the need to optimize the current diagnoses that allow faster and more efficient prescriptions. Lately, an approach to predict resistance from genome sequencing data applies machine learning models, which has taken credibility due to its capability to make reliable predictions. Further, thanks to the increasing knowledge about resistance mechanisms associated with *A. baumannii*, there is enough accessible data for developing these models.

We used 343 genomes, 76 from the Colombian National Institute of Health, and 267 gathered from the Biosample NCBI database. We created models based on machine learning using lasso regression, random forest, and gradient boosting, to predict the minimum inhibitory concentration of 10 antibiotics. Random forest was the algorithm that show is better results, achieving an average accuracy within +/- a double dilution of 91% (I.C 95, 85- 97), a very major error, and a major error rate of 1.71% and 0.7% respectively. We employ known resistance genes as the model input, which were identified using the Resistance Gene Identifier software. These results show that the *A. baumannii* antibiotics susceptibility prediction, based on genome sequence, is promising as a possible diagnostic tool in the clinic.

Keywords *Acinetobacter baumannii*, machine learning, phenotypic resistance prediction, minimum inhibitory concentration, lasso regression, random forest, gradient boosting.

Contenido

	<u>Pág.</u>
Lista de figuras	XI
Lista de tablas	XII
Introducción	1
1. Capítulo 1: Generalidades	5
1.1 Enfermedades Infecciosas	5
1.2 Resistencia Bacteriana	7
1.3 Secuenciación de genoma completo como tecnología de diagnóstico microbiológico.....	9
1.4 Inteligencia artificial: predicción de resistencia.....	10
1.5 <i>Acinetobacter baumannii</i> como modelo para predicción de resistencia antimicrobiana	12
1.5.1 Mecanismos de resistencia en <i>A. baumannii</i>	14
2. Capítulo 2: Caracterización del perfil fenotípico de resistencia de <i>Acinetobacter baumannii</i>.....	17
2.1 Introducción.....	17
2.2 Metodología.....	18
2.2.1 Aislamientos bacterianos	18
2.2.2 Selección de antibióticos	18
2.2.3 Preparación de soluciones madre de antibióticos	20
2.2.4 Preparación de la suspensión bacteriana	21
2.2.5 Determinación de la concentración mínima inhibitoria	21
2.2.6 Recolección de datos públicos.....	21
2.3 Resultados y discusión	22
2.3.1 Obtención de la suspensión bacteriana.....	22
2.3.2 Determinación de la concentración mínima inhibitoria	23
2.3.3 Recolección de datos públicos.....	28
3. Capítulo 3: Identificación de las características genómicas asociadas a resistencia utilizadas como entrada para el desarrollo de modelos de predicción. 31	

3.1	Introducción.....	31
3.2	Metodología.....	33
3.2.1	Predicción de resistoma	33
3.2.2	Determinación de las variaciones aminoacídicas de los elementos asociados a resistencia	34
3.3	Resultados	36
3.3.1	Predicción de resistoma	36
3.3.2	Determinación de las variaciones de los elementos asociados a resistencia	48
4.	Capítulo 4: Desarrollo del modelo de predicción del perfil de resistencia a partir de las secuencias del genoma de <i>Acinetobacter baumannii</i>.....	51
4.1	Introducción.....	51
4.2	Metodología.....	57
4.2.1	Pre-procesamiento de datos	57
4.2.2	Características importantes (<i>feature importance</i>)	59
4.2.3	Construcción de modelos de aprendizaje de máquina	59
4.2.4	Ajuste de hiperparámetros.....	61
4.2.5	Evaluación del modelo.....	62
4.3	Resultados y discusión	64
4.3.1	Pre- procesamiento de datos	64
4.3.2	Características importantes (<i>feature importance</i>)	65
4.3.3	Construcción de modelos de aprendizaje de máquina	68
4.3.4	Acceso a los datos	82
5.	Conclusiones y recomendaciones	85
5.1	Conclusiones.....	85
5.2	Recomendaciones	87
A.	Anexo: Tablas	88
B.	Anexo: Procedimientos	127
C.	Anexo: Figuras.....	129
	Bibliografía	137

Lista de figuras

Figura 1-1: La secuenciación de patógenos durante enfermedades infecciosas puede informar intervenciones precisas (Ladner, 2019).....	9
Figura 2-1: Perfiles de resistencia de <i>A. baumannii</i>	26
Figura 3-1: Flujo de trabajo del programa <i>tablaAlelos.py</i> para la determinación de las variaciones a nivel de secuencia de proteína de cada uno de los elementos asociados a resistencia encontrados en los genomas de los 343 aislamientos de <i>A. baumannii</i> analizados.	34
Figura 3-2: Mecanismos de resistencia de los 343 genomas de <i>A. baumannii</i>	36
Figura 3-3: Resistoma de los 343 genomas de <i>A. baumannii</i> utilizados para el entrenamiento de los modelos de aprendizaje de máquina.	37
Figura 4-1: Esquema del estudio empleando selección de características	60
Figura 4-2: Principales genes seleccionados por la función “ <i>permutation feature importance</i> ”.....	66
Figura 4-3: Número de características seleccionadas por antibiótico	67
Figura 4-4: Distribución de los genomas por cada MIC y por antibiótico	71
Figura 4-5: <i>Very major error</i>	73
Figura 4-6: <i>Major Error</i>	74

Lista de tablas

Pág.

Tabla 1-1: Carbapenemasas tipo OXA asociadas a <i>A. baumannii</i> en Suramérica (Andrés Opazo et al., 2012)	14
Tabla 2-1: Antibióticos seleccionados	19
Tabla 2-2: Rango de concentraciones de soluciones madre de antibióticos	20
Tabla 2-3: Recuento Celular	22
Tabla 2-4: Susceptibilidad de los 76 aislamientos colombianos frente a los antibióticos probados menos trimetoprim-sulfametoxazol.....	23
Tabla 2-5: Resultados de la concentración mínima inhibitoria para el trimetoprim-sulfametoxazol.....	24
Tabla 2-6 : Información de los genomas de referencia de <i>A. baumannii</i>	28
Tabla 2-7: Perfiles de susceptibilidad correspondientes a los 267 genomas de referencia obtenidos de la base de datos <i>Biosample del NCBI</i>	29
Tabla 3-1: Ejemplo de la matriz de las variaciones de elementos asociados a resistencia generada utilizando el programa <i>tablaAlelos.py</i>	35
Tabla 3-2: Resistoma accesorio correspondiente a Inactivación del antibiótico de los 343 aislamientos de <i>A. baumannii</i>	40
Tabla 3-3: Resistoma accesorio correspondiente a Alteración de la permeabilidad y del sitio blanco de los 343 aislamientos de <i>A. baumannii</i>	44
Tabla 3-4: Resistoma accesorio correspondiente a protección y reemplazo del sitio blanco de los 343 aislamientos de <i>A. baumannii</i>	46
Tabla 3-5: Resistoma accesorio correspondiente a bombas de eflujo de los 343 aislamientos de <i>A. baumannii</i>	47
Tabla 3-6: Genes y alelos en el resistoma de los 343 aislamientos de <i>Acinetobacter baumannii</i> agrupados al 100%, 97% y 95% de identidad.	49
Tabla 4-1: Estructura de los datos de entrada del modelo de aprendizaje de máquina ..	58
Tabla 4-2: Hiperparámetros ajustados en <i>random forest</i> y <i>gradient boosting</i>	62
Tabla 4-3: Definición de los parámetros empleados para evaluación del modelo	63

Tabla 4-4: Resumen de las 343 muestras utilizadas para construir y evaluar modelos de predicción de resistencia y el número de genomas disponibles para cada antibiótico	65
Tabla 4-5: Precisión de los modelos desarrollados con <i>random forest</i> , regresión <i>lasso</i> y <i>gradient boosting</i>	70
Tabla 4-6: Precisiones dentro de +/- una dilución doble y tasas de error de los modelos desarrollados con " <i>permutation feature importance</i> ".....	77
Tabla 4-7: Modelos que cumplen con las limitaciones del FDA.	82

Introducción

Los antibióticos son un recurso esencial en el tratamiento de enfermedades infecciosas, ya que desde su descubrimiento han disminuido sustancialmente las tasas de mortalidad y morbilidad asociadas a infecciones (B. Li & Webster, 2018; Pataki et al., 2019). No obstante, el uso excesivo e indiscriminado de estos ha contribuido a la aparición y amplia propagación de resistencia bacteriana. (Jeukens et al., 2019; Organización Mundial de la Salud, 2018d). La resistencia a los antimicrobianos (RAM) aparece con el tiempo de forma natural, generalmente por la capacidad evolutiva de las bacterias para desarrollar día a día nuevos mecanismos de resistencia, los cuales se están diseminando en todo el planeta; y si a esto le sumamos la disminución en el descubrimiento y desarrollo de nuevos antibióticos en las últimas décadas, combatirlas se tornará en un verdadero desafío (Khaledi et al., 2019; B. Li & Webster, 2018; Organización Mundial de la Salud, 2018c). Las RAMs en el siglo XXI se han convertido en uno de los grandes problemas médicos (López, 2019), debido a que no solo eleva las tasas de morbilidad, mortalidad y costos, sino obliga al uso frecuente de medicamentos de último recurso que generalmente son tóxicos (Su et al., 2019).

Por lo tanto, tomar decisiones para frenar la creciente propagación de la resistencia antimicrobiana y así prolongar la efectividad de los antibióticos existentes, es urgente. (B. Li & Webster, 2018). La detección temprana y precisa de la resistencia antimicrobiana es crucial para guiar hacia un tratamiento óptimo y disminuir esta propagación (B. Li & Webster, 2018; Pataki et al., 2019; Su et al., 2019). Las pruebas de susceptibilidad antimicrobiana basadas en cultivo siguen siendo el principal método para la detección de resistencia, no obstante, en los últimos 5 años se han desarrollado nuevos enfoques de predicción de resistencia basados en la genómica. (B. Li & Webster, 2018; Pataki et al., 2019; Su et al., 2019). La secuenciación de última generación es potencialmente capaz de mejorar la microbiología clínica (Gwinn et al., 2019), y actualmente se está convirtiendo en una alternativa factible al fenotipado tradicional, debido a su capacidad de realizar predicciones rápidas y precisas de resistencia. (Jeukens et al., 2019; Pataki et al., 2019;

Su et al., 2019). Además, debido a su capacidad de proporcionar información acerca de los mecanismos y patrones de transmisión de las RAM, la secuenciación ha tomado importancia en la salud pública y la práctica clínica (Macesic et al., 2017)

Últimamente uno de los enfoques para predecir resistencia a partir de los datos de secuenciación de genoma, consiste en aplicar modelos basados en aprendizaje de máquina, los cuales han ido tomando credibilidad debido a la capacidad de realizar predicciones precisas (Jeukens et al., 2019; Su et al., 2019). Estos modelos se basan en encontrar patrones en un conjunto de datos de entrenamiento en función de un resultado de interés, para posteriormente ser aplicados a nuevos datos y realizar predicciones (Macesic et al., 2017). Además, permite determinar mecanismos de resistencia no caracterizados o complejos, por su capacidad de encontrar patrones en los datos, los cuales podrían ser nuevas mutaciones, genes posiblemente asociado a resistencia (Hicks et al., 2019; Jeukens et al., 2019). Hasta la fecha, estos modelos se han centrado en predecir variables categóricas, es decir clasificar un aislamiento únicamente como sensible o resistente. Sin embargo, los puntos de corte utilizados para determinar fenotipos sensibles y resistentes pueden cambiar, por lo que estas categorías binarias no permiten ver los posibles cambios en la susceptibilidad con el transcurrir del tiempo (Pataki et al., 2019). Para ello la determinación de la concentración mínima inhibitoria (*MIC* por sus siglas en inglés) resulta conveniente, ya que permite un mayor grado de individualización de la terapia antimicrobiana., a diferencia de cuando utilizamos un formato cualitativo (sensible, intermedio, resistente) que se basa en puntos de corte; los mismos que cambian constantemente debido al incremento en la resistencia bacteriana, y son diferentes dependiendo de la agencia reguladora que lo determine, creando confusión a los microbiólogos y fabricantes de medicamentos (Labreche et al., 2015). Además mejorara la toma de decisiones en cuanto al tratamiento, puesto que se ha evidenciado que aislamientos determinados como susceptibles de acuerdo a los puntos de corte estándar pero con MICs elevadas están asociados a mayor riesgos de fracaso terapéutico (Hicks et al., 2019; Pataki et al., 2019). Los modelos que realicen predicciones de MICs resultan más convenientes para enfrentar esta problemática (Nguyen, Brettin, et al., 2018; Nguyen, Long, et al., 2018; Pataki et al., 2019).

Por otro lado, gracias al creciente conocimiento acerca de mecanismos de resistencia asociados a *Acinetobacter baumannii*, además de su tendencia de desarrollar multiresistencia e incluso panresistencia (Jeukens et al., 2019); nos brinda una alternativa para desarrollar modelos de predicción de resistencia. *A. baumannii* es una bacteria gram – negativa, muy importante en las infecciones asociadas a la atención en salud (IAAS) por su capacidad de resistencia, no solo a los antimicrobianos, sino también a los desinfectantes y la desecación (Garnacho-Montero & Timsit, 2019; Isler et al., 2019). Se caracteriza por su facilidad de adquirir nuevos determinantes de resistencia a través de elemento genéticos móviles (Garnacho-Montero & Timsit, 2019)

Por lo que el presente estudio tiene los siguientes objetivos:

- Caracterizar fenotípicamente los perfiles de resistencia de *Acinetobacter baumannii*.
- Identificar las características del perfil genómico de resistencia de *Acinetobacter baumannii*.
- Construir un modelo de aprendizaje de máquina que combine el perfil genómico y fenotípico para predicción de resistencia en *Acinetobacter baumannii*

Se analizaron un total de 343 aislamientos de *A. baumannii*, de los cuales 76 fueron provenientes del programa de vigilancia de Infecciones Asociadas a la Atención en salud realizado por el Instituto Nacional de Salud de Colombia en el periodo 2012-2015. Los datos de secuenciación de los 76 aislamientos fueron obtenidos de un trabajo previo realizado por el grupo de investigación de Bioinformática y Epidemiología molecular de la Universidad Nacional de Colombia (Prada Cardozo, 2018) y los valores de la concentración mínima inhibitoria fueron determinados en la presente investigación mediante microdilución en caldo. Los 267 genomas restantes fueron obtenidos de la base de datos *Genome del NCBI* y sus perfiles de susceptibilidad de la base de datos *Biosample del NCBI*

1. Capítulo 1: Generalidades

1.1 Enfermedades Infecciosas

Las enfermedades infecciosas siempre se han considerado como una problemática a nivel mundial, y en la actualidad parece ser que van ganando la batalla debido a la falta de innovación de nuevos antibióticos y a la capacidad evolutiva que poseen los microorganismos al desarrollar mecanismos de resistencia cuando se enfrentan a estos fármacos. Se encuentran relacionadas con altas tasas de morbilidad y mortalidad, además de elevados costos que se encuentran asociados a la atención en salud, ya que las infecciones generalmente requieren estadías prolongadas en el hospital, visitas recurrentes al personal médico de seguimiento y uso de tratamientos que pueden ser más costoso y posiblemente más tóxicos (Centers for Disease Control and Prevention (U.S.), 2019; Ríos et al., 2011; Yanpeng Zhang et al., 2018).

Hoy en día las enfermedades infecciosas son responsables de alrededor de 15 millones de los 57 millones de muertes anuales en todo el mundo, cifra alarmante que se debe al incremento de patógenos resistentes a los antibióticos (Morens et al., 2004). Las infecciones asociadas a la atención en salud (IAAS) constituyen uno de los mayores problemas para la seguridad del paciente, causando aumento de la morbilidad, mortalidad y la carga económica, debido a la aparición de microorganismos multi-resistentes, lo que disminuye las posibilidades de tratamiento, complicando aún más la situación. En este sentido la Organización Mundial de la Salud (OMS) publicó que la prevalencia para esta patología varía del 3 % al 20,7% (Samuel et al., 2010), y los centros para la prevención y control de enfermedades (CDC) informaron 2,8 millones de infecciones en Estados Unidos, y 35000 muertes causadas por bacterias resistentes cada año , con costos asociados a estas infecciones de 23 billones de dólares en Estados Unidos (Centers for Disease Control and Prevention (U.S.), 2019; Goldstein et al., 2018). Por otra parte, datos de la OMS muestran que 1 de cada 20 pacientes hospitalizados son afectados por IAAS; siendo esta cifra de 4,1 millones de pacientes al año de los cuales 37.000 mueren (Organización

Mundial de la Salud, 2018b). En Colombia, el Instituto Nacional de Salud reportó 6474 infecciones asociadas a dispositivos en el área de cuidados intensivos entre el 2015-2016, siendo las de mayor frecuencia las infecciones asociadas a catéter y ventilación mecánica (Corredor et al., 2018).

Debido a su relación con microorganismos multi-resistentes y sumado la falta de innovación de nuevos antibióticos, estas infecciones se tornan muy complejas de tratar; ya que los tratamientos se vuelven ineficaces y la transmisión de las infecciones persiste (Zheng et al., 2019). En los últimos 20 años, han sido aprobados únicamente 2 nuevos grupos de antibióticos, lipopéptidos y oxazolidinonas, cuyo espectro cubre únicamente a gram-positivas. En cuanto a gram-negativas, las quinolonas en 1962 fueron los últimos antibióticos activos contra este grupo descubiertos. En la actualidad, solamente 5 nuevos antibióticos activos contra gram-negativas se encuentran en fase 3 de prueba. El desarrollo de nuevos antibióticos es indispensable para combatir esta amenaza mundial, sin embargo la investigación en relación con este tema es bastante limitada por los altos costos que implica y la baja recuperación de la inversión (Casellas, 2011.; Organización Mundial de la Salud, 2016; Tacconelli et al., 2018). Por tales motivos la OMS publicó las medidas necesarias para un óptimo control de estas infecciones, dentro de las que se encuentra la reducción del uso indiscriminado y sin fórmula médica de los antimicrobianos, como una de las medidas planteadas para hacer frente este problema. (Organización Mundial de la Salud, 2011)

En el mismo sentido en el 2016 la Sociedad de Enfermedades Infecciosas de América (IDSA), la Sociedad para la Epidemiología de la Atención Médica de América (SHEA) y la Sociedad de Enfermedades Infecciosas Pediátricas (PIDS) implementaron un programa de administración de antibióticos (ASP) con el objetivo de que exista un uso racional de los antimicrobianos por parte del personal de salud, y así disminuir las tasas de mortalidad asociadas a este tipo de patología (Barlam et al., 2016)

En Colombia el Ministerio de Salud y Protección Social (MSPS) estableció un plan que tiene como objetivo reducir las infecciones asociadas a la atención en salud (IAAS) y contener la resistencia a los antimicrobianos, denominado “Plan Decenal de Salud Pública 2012-2021”, el cual tiene como estrategia la participación activa del programa de prevención y control de las IAAS, la Resistencia Antimicrobiana y el Uso prudente de antibióticos, y su visión es la implementación de estos programas de control en al menos

el 90% de las instituciones de salud de alta y mediana complejidad del país para el año 2021(Corredor et al., 2018).

1.2 Resistencia Bacteriana

La resistencia bacteriana ha sido una problemática que en los últimos años ha ido captando la atención de los investigadores, debido a las consecuencias que conlleva, particularmente al peligro que pone los avances de la medicina moderna tales como trasplantes de órgano, terapia cancerígena, entre otros. Estos procedimientos tienen un importante riesgo de infección, por eso su administración se ve condicionada por la disponibilidad de antibióticos eficaces (Centers for Disease Control and Prevention (U.S.), 2019). La resistencia bacteriana constituye un problema de salud pública que se encuentra condicionado por el abuso en el uso de antibióticos en situaciones donde no aportan ningún beneficio, y particularmente, por la utilización de terapias empíricas.

Con el paso del tiempo, esta situación se ha vuelto un verdadero desafío, no solo en el ámbito de la salud, sino también en el de la seguridad alimentaria. Debido a la gran amenaza que representa la resistencia bacteriana, y en busca de una respuesta de manera global, los líderes mundiales en el 2016 se enfocaron en buscar la manera de detener la diseminación de las infecciones asociadas a bacterias resistentes a antibióticos. El tema fue tratado en la Asamblea General de las Naciones Unidas, donde promover el uso responsable y prudente de los antimicrobianos y fomentar la investigación para el desarrollo de nuevas tecnologías para el diagnósticos y nuevos medicamentos, fueron las medidas que los diversos países se comprometieron a adoptar con el fin de detener esta problemática. (Organización Mundial de la Salud, 2016)

Desde el punto de vista clínico, es indispensable encontrar estas alternativas de diagnóstico que nos permitan obtener resultados tempranos que ayuden al médico en la toma de decisiones precisas en cuanto al tratamiento. Ya que el uso de terapias empíricas se debe a que el resultado de los antibiogramas se tarda entre 24-48 horas, y se ha demostrado que más del 25% de las infecciones se han tratado con un antibiótico incorrecto, ocasionando principalmente en las áreas de hospitalización aumento en la tasa de mortalidad (Martínez-Agüero et al., 2019).

Las bacterias se encuentran constantemente desarrollando nuevas estrategias de defensa que les permite sobrevivir a la acción de los antibióticos, denominados mecanismos de resistencia. Se encuentran codificados en el DNA, y es de vital importancia entender los mecanismos moleculares ya que permiten comprender este fenómeno, y encontrar estrategias para bloquear o disminuir la resistencia (Centers for Disease Control and Prevention (U.S.), 2019; Wilson et al., 2020). La resistencia se origina principalmente por dos procesos: transferencia horizontal de genes y mutaciones espontaneas (Biswas, Raju et al., 2019).

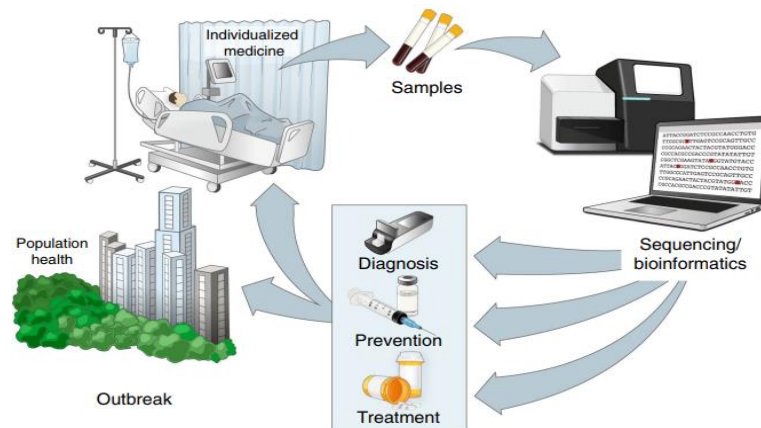
Es importante mencionar que existen de tres tipos de resistencia: la resistencia intrínseca, la resistencia inducida y la adquirida. Donde la intrínseca hace referencia a la resistencia que comparten un grupo de bacterias de la misma especie, independientemente de una previa exposición a un antimicrobiano; y en la inducida las bacterias contienen los genes de resistencia, pero se expresan únicamente tras una exposición al antibiótico (Sultan et al., 2018).

Por otra parte, la resistencia adquirida está relacionada con la capacidad que tiene un microorganismo de adquirir material genético foráneo que le permita desarrollar capacidad de resistencia frente a un cierto antibiótico. Existen varios mecanismos para adquirir esta resistencia, pero el principal es mediante plásmidos. Las mutaciones también se encuentran relacionadas con la resistencia adquirida, las secuencias de inserción y los integrones tienen la capacidad de desplazar el material genético y dar lugar a mutaciones (Sultan et al., 2018; Wilson et al., 2020). Los genes involucrados en desarrollar resistencia tras una mutación son los que codifican para un blanco farmacológico, para transporte de antibióticos y para enzimas modificadoras de antimicrobianos. Se ha agrupado a los mecanismos de resistencia antimicrobiana en 4 grupos: los que limitan la absorción del antibiótico, inactivan el blanco farmacológico, inactivan el fármaco y las bombas de eflujo. Estos mecanismos difieren entre bacterias gram positivas y gram negativas debido a las diferencias que existen entre sus estructuras (Centers for Disease Control and Prevention (U.S.), 2019; Sultan et al., 2018).

1.3 Secuenciación de genoma completo como tecnología de diagnóstico microbiológico

Desde la década pasada, la urgencia por encontrar la manera de detener la creciente aparición de enfermedades infecciosas asociadas a patógenos multi-resistentes, se incrementó. Con los avances en genómica y computación es posible innovar alternativas para el control de las mismas, y de la medicina en general. El desarrollo de la genómica de alto rendimiento y en tiempo real nos permite incluir a la secuenciación como una potente herramienta para tratar diversas enfermedades, incluyendo las infecciosas; ya que la velocidad y la facilidad con la que se realiza este procedimiento van mejorando continuamente. La implementación de estas técnicas tiene como objetivo tener una respuesta con precisión a las enfermedades infecciosas (Figura 1-1) (Ladner, 2019)

Figura 1-1: La secuenciación de patógenos durante enfermedades infecciosas puede informar intervenciones precisas (Ladner, 2019)



Adicionalmente, estos avances han originado una notable disminución de los costos de la secuenciación. Esta combinación, bajo costo y rápida velocidad de secuenciación de genoma, hace posible el uso de esta tecnología en la práctica de microbiología de diagnóstico y no solo en la investigación microbiana como sucede actualmente (Köser et al., 2012).

Dada la urgencia de reducir el uso erróneo de antibióticos, el desarrollo de pruebas rápidas de susceptibilidad se ha tornado necesario. Por tal motivo, los laboratorios de microbiología han ido implementando nuevas técnicas de diagnóstico basadas en genotipificación (C.

Zhang et al., 2019). Sin embargo, estas técnicas han estado limitadas por el lento ritmo de la secuenciación o enfocadas únicamente a la determinación de la presencia/ ausencia de genes de resistencia, lo que puede llevar a una correlación inadecuada con sus fenotipos (Didelot et al., 2012; Yamamoto et al., 2015). La secuenciación de genoma completo es actualmente la alternativa más prometedora debido a que tiene el potencial de caracterizar variantes genéticas menores en el genoma bacteriano presente en un paciente en particular, siendo de mucha importancia para diseñar un tratamiento acertado y personalizado, ya que nos permite entender los mecanismos moleculares que subyacen a la patogénesis de la infección incluyendo la resistencia, y es de utilidad para la identificación y caracterización de patógenos sin la necesidad de un conocimiento a priori del agente causal (Bradley et al., 2015; Ladner, 2019).

The National Center for Biotechnology Information (NCBI por sus siglas en inglés) es una importante fuente de información que ayuda a la comprensión de los procesos moleculares y genéticos fundamentales que controlan la salud y la enfermedad (Jenuth, 1999). Este recurso actualmente es de vital importancia porque pone a disponibilidad datos genómicos a gran escala de una variedad de bacterias de importancia clínica, además de disponer de datos fenotípicos. Esta herramienta permite recolectar datos de manera rápida, fácil y económica, pues su acceso es libre (Kim et al., 2020; Mailman et al., 2007). Esto hace posible realizar asociaciones genotipo-fenotipo, que brindan la oportunidad de desarrollar algoritmos que utilicen esta información para el desarrollo de modelos de predicción de resistencia que resultarían de gran utilidad a la hora de tratar infecciones (Kim et al., 2020).

Es importante mencionar que este continuo desarrollo de diagnóstico de precisión tiene como limitante que la secuenciación de alto rendimiento y la bioinformática se manejan únicamente en los laboratorios de investigación académicos; por lo que es indispensable el trabajo en colaboración de los investigadores con las agencias de salud para hacer posible la implementación de esta herramienta de diagnóstico.

1.4 Inteligencia artificial: predicción de resistencia

Como hemos visto, los métodos genómicos son fundamentales para el desarrollo de estrategias para la vigilancia y control de la resistencia antimicrobiana (Maguire et al., 2019). Actualmente, gracias al amplio conocimiento que se tiene acerca de mecanismos

de resistencia caracterizados a nivel genómico de una bacteria, sumado a la gran cantidad de herramientas bioinformáticas disponibles, existe la posibilidad de crear enfoques basados en inteligencia artificial para descifrar y combatir la resistencia antimicrobiana (Jeukens et al., 2019). Modelos basados en aprendizaje de máquina se han utilizado intensamente para investigar infecciones bacterianas, realizar predicciones y tomar decisiones a partir de conjuntos de datos biológicos (Macesic et al., 2017; C. Zhang et al., 2019). Generalmente se han desarrollado para mejorar el tratamiento de infecciones causadas por bacterias resistentes a múltiples fármacos (*MDR*, por sus siglas en inglés) (Macesic et al., 2017), para lograrlo se tiene como objetivo encontrar valores predictivos de la concentración mínima inhibitoria que conduzcan a una intervención temprana y una terapéutica eficaz (Jeukens et al., 2019; Köser et al., 2012).

El aprendizaje de máquina puede ser implementado de dos formas: supervisado o no supervisado. En el aprendizaje supervisado el conjunto de entrenamiento incluye el valor de salida de interés, es utilizado para construir modelos de predicción. Estos modelos tienen disponible un resultado de salida (por ejemplo, el fenotipo de una bacteria: resistente o sensible) y lo que hace es intentar aprender de los datos disponibles (mutaciones genómicas, genes de resistencia, etc.) para predecir dicho resultado. Posteriormente el modelo desarrollado es aplicado a datos nuevos (que no se han usado en el entrenamiento) para predecir el resultado de interés. Caso contrario, en el aprendizaje no supervisado no se dispone de una salida de interés, solamente un conjunto de datos con diferentes variables, y el objetivo de este método es encontrar la organización y estructura de los datos (Macesic et al., 2017).

Se han descrito como una buena alternativa de predicción de resistencia, modelos basados en aprendizaje de máquina supervisados utilizando genes asociados a la resistencia, datos del genoma completo o combinaciones de ambos. Estos modelos nos permiten predecir propiedades/comportamientos de los objetos de estudio a partir de sus características después de haber sido sometidos a un entrenamiento (C. Zhang et al., 2019). El uso de estos modelos se considera de mejor viabilidad y utilidad por: 1) su desempeño de predicción eficiente para cualquier bacteria a analizar, teniendo como único requisito disponer de suficientes datos para el proceso de entrenamiento del modelo; 2) la fácil actualización del modelo con la identificación de nuevos genes de resistencia, 3) la

probabilidad de descubrir nuevos elementos genómicos que estén asociados a la resistencia, ya que permite utilizar el genoma completo como dato de entrada, logrando detectar posibles mutaciones en regiones reguladoras de genes de resistencia que provocan cambios en la expresión del fenotipo (Gordon et al., 2014; Nguyen, Brettin, et al., 2018; Nguyen, Long, et al., 2018)

Las técnicas basadas en aprendizaje de máquina definitivamente podrían utilizarse para crear un sistema de apoyo en las decisiones clínicas, que ayuden a los médicos en la identificación de bacterias resistentes (Martínez-Agüero et al., 2019), en el control de brotes y en una prescripción de antibióticos más efectiva (Giacobbe et al., 2020; Qu et al., 2019). Gracias al continuo avance en esta tecnología su posible uso se vuelve cada vez más accesible, revolucionando la clásica microbiología clínica como difusión en disco y determinación de la concentración mínima inhibitoria (Bradley et al., 2015; Jeukens et al., 2019). Además su implementación promete disminuir notablemente los costos asociados a la atención en salud y las tasas de morbilidad (Ngiam & Khor, 2019).

1.5 *Acinetobacter baumannii* como modelo para predicción de resistencia antimicrobiana

Acinetobacter baumannii (*A. baumannii*) es un coccobacilo, Gram-negativo, aeróbico, no fermentador, inmóvil, no formador de esporas, catalasa positiva y oxidasa negativa. Es el patógeno humano más común del género *Acinetobacter* y se le ha incluido dentro del grupo ESKAPE por ser bacteria multi resistente potencialmente capaz de ocasionar infecciones en hospitales (Hamidian & Nigro, 2019; Tacconelli et al., 2018). Se ha catalogado en los últimos años como un patógeno oportunista de gran importancia clínica debido al papel que desempeña en las infecciones asociadas a la atención en salud (IAAS), habilidad que se le atribuye a la facilidad con la que se adapta y persiste en diferentes ambientes y a la amplia resistencia intrínseca que posee en virtud de los elementos genéticos presentes en su genoma. En los entornos hospitalarios, *A. baumannii* presenta un índice de mortalidad del 26% que puede llegar hasta el 43% en el servicio de cuidados intensivos (UCI) (Asif et al., 2018; Mobasser et al., 2018).

La Red Nacional de Seguridad Sanitaria reportó en el 2009 que *A. baumannii* es el responsable de aproximadamente el 1.8% de infecciones asociadas a la atención en salud,

y se estima que a nivel mundial un millón de infecciones por año son causadas por este microorganismo. Adicionalmente, los Centros para el Control y Prevención de Enfermedades (CDC) en 2013 en un informe donde trataban la problemática asociada a patógenos resistentes, informaron que 7300 infecciones ocasionadas por *A. baumannii* multi-resistente (MDR) se observan por año en Estados Unidos, de las cuales 500 terminan en muertes (Parquet et al., 2019). En el 2018, la OMS lo catalogo *A. baumannii* resistente a carbapenémicos como principal amenaza a la medicina moderna, por su capacidad de resistencia frente a múltiples antibióticos (Hamidian & Nigro, 2019; Organización Mundial de la Salud, 2018a).

A. baumannii se ha relacionado con procedimientos médicos invasivos, mecanismos de ventilación, inmunosupresión, quemaduras, heridas y traumas, desencadenando infecciones de importancia clínica como: neumonía asociada a ventilación, infecciones del tracto urinario y septicemia, entre otras. Las neumonías asociadas a *A. baumannii* representan el 15 % de todas las infecciones que ocurren a nivel hospitalario. Además, la hospitalización prolongada es un factor de predisposición para que se desencadenen algunas de estas infecciones asociadas a *A. baumannii* (Asif et al., 2018; Yanpeng Zhang et al., 2018).

Actualmente, existe una alta tasa de incidencia de *A. baumannii* resistente a carbapenémicos, por lo que las opciones terapéuticas se tornan limitadas ya que estos antibióticos son los más activos y resistentes frente a las β -lactamasas, incluidas las de espectro extendido. Las enzimas tipo OXA (oxacilinasas) representan uno de los principales mecanismos por el cual estas bacterias desarrollan resistencia frente a carbapenémicos, y por su fácil diseminación resultan ser una amenaza. Existe cuatro subgrupos de oxacilinasas; OXA-51 ubicado en el cromosoma; y OXA-23, OXA-40 y OXA-53 adquiridos. (Tabla 1-1). En Colombia la amplia diseminación de los genes *bla*_{OXA-51} y *bla*_{OXA-23} está relacionada con la creciente resistencia de *A. baumannii* a carbapenémicos (Correa et al., 2017; Gómez et al., 2017; Andrés Opazo et al., 2012).

Tabla 1-1: Carbapenemasas tipo OXA asociadas a *A. baumannii* en Suramérica (Andrés Opazo et al., 2012)

OTC subgroup	Countries	Genetic location
OXA-51	Argentina, Bolivia, Brazil, Chile*, Colombia, Venezuela	Chromosome
OXA-23	Argentina, Colombia, Brazil	Chromosome, Plasmid
OXA-40	Brazil, Chile	Plasmid
OXA-58	Argentina, Bolivia, Chile*, Colombia, Bolivia, Venezuela	Plasmid
OXA-143	Brazil	Plasmid

1.5.1 Mecanismos de resistencia en *A. baumannii*

La capacidad de resistencia de *A. baumannii* está dado por múltiples mecanismos que se ponen en marcha cuando detectan la presencia de un antibiótico, entre los que se encuentran: las enzimas de degradación, alteraciones genéticas (mutaciones, incorporación o pérdida de un gen, baja expresión o sobreexpresión genética) y las bombas de eflujo, cuyos genes han sido encontrados tanto en cromosomas como en plásmidos (Asif et al., 2018; Parquet et al., 2019).

La resistencia a antibióticos carbapenémicos está mediada por β -lactamasas, alteración en las proteínas de unión a la penicilina, disminución de la permeabilidad por cambios en las porinas de la membrana externa y por bombas de eflujo. Dentro de las β -lactamasas, la más común asociada a *A. baumannii* es la AmpC cefalosporinasa, β -lactamasa del grupo C, que le confiere resistencia a la penicilina y a las cefalosporinas de reducido y amplio espectro. Las metalo- β -lactamasas cumplen un papel importante en la degradación de carbapenémicos en *A. baumannii*, se le consideran como una amenaza ya que están presentes en elementos genéticos móviles, lo que les permite transferirse fácilmente entre bacterias (Eliopoulos et al., 2008). OXA-23, OXA-24, OXA-58, OXA-143 y OXA-235 están codificadas en plásmidos (Hamidian & Nigro, 2019; Múnera et al., 2014). Estas enzimas muestran una actividad débil contra carbapenémicos, sin embargo la presencia de una secuencia de inserción, *ISAbal1*, impulsan la sobre expresión originando reducción de la susceptibilidad (Hamidian & Nigro, 2019; Potron et al., 2015). Otro mecanismo que confiere resistencia a los carbapenémicos en *A. baumannii* son la reducción de las proteínas o

porinas de membrana externa (OMPs) y la alteración de las proteínas de unión a la penicilina (PBPs), se ha comprobado que una disminución en la expresión de estas proteínas disminuye la sensibilidad de los antibióticos (Eliopoulos et al., 2008)

En cuanto a la resistencia de aminoglicósidos, el mecanismo se basa en una alteración en el grupo amino y carboxilo por enzimas modificadoras de aminoglicósidos, como ANT(3'')-Ia, AAC(6')-Ib, APH(3')-1a, AAC(3)-Ia, APH(3')-VI. Alteraciones en las proteínas de unión ribosomal y disminución de la entrada del antibiótico por bombas de eflujo son mecanismos que también han sido reportados en *A. baumannii* que confieren la resistencia a aminoglicósidos como ArmA y RmtB (Eliopoulos et al., 2008; Múnera et al., 2014)(Potron et al., 2015)

A. baumannii ha reportado mutaciones en los genes *gyrA* y *parC*, que están involucrados en la resistencia a quinolonas, ya que originan un cambio fenotípico en la DNA girasa y topoisomerasa IV, lo que lleva a una disminución de la afinidad de las quinolonas por estas enzimas. Los genes *qnrA*, *qnrB*, y *qnrS* están codificados en plásmidos y participan activamente en la resistencia a quinolonas por parte de *A. baumannii* (Eliopoulos et al., 2008; Múnera et al., 2014).

Finalmente el mecanismo de resistencia a la colistina por parte *A. baumannii* se encuentra codificado en el DNA cromosomal. Se ha comprobado la existencia de 2 mecanismos que le confieren esta resistencia. El primero se basa en mutaciones en los genes que codifican para el lípido A (*lpxA*, *lpxC*, y *lpxD*) que da lugar a la pérdida del lipopolisacárido, blanco inicial de la colistina, y el segundo que consiste en un reordenamiento de la membrana externa resultado de mutaciones en los genes *pmrA* y *pmrB* (Múnera et al., 2014).

2. Capítulo 2: Caracterización del perfil fenotípico de resistencia de *Acinetobacter baumannii*.

2.1 Introducción

Las pruebas de susceptibilidad juegan un papel importante en el control de bacterias resistentes, ya que no solo nos permiten predecir el tratamiento sino también detectar la aparición de resistencia (Zankari et al., 2013). Hoy en día la resistencia antimicrobiana se evalúa mediante el uso de cultivos bacterianos y pruebas fenotípicas (Deelder et al., 2019), lo que nos permite medir la posible respuesta bacteriana frente a los antibióticos expresada en términos de susceptibilidad y resistencia. (Kester & Fortune, 2014)

Las principales pruebas fenotípicas utilizadas en la actualidad son difusión en disco y determinación de la concentración mínima inhibitoria (Hicks et al., 2019). Sin embargo, estas pruebas fenotípicas convencionales tienen una gran desventaja porque se requiere realizar la inoculación de la muestra en un cultivo primario hasta que exista un crecimiento suficiente que permita hacer la clasificación taxonómica y las pruebas de susceptibilidad, lo que en algunas ocasiones requiere de hasta 15 días, ya que a veces se necesita de subcultivos para purificar los aislamientos o mucho más tiempo en el caso de bacterias de crecimiento lento (Gordon et al., 2014; Nguyen, Brettin, et al., 2018; Zankari et al., 2013). Además, los resultados obtenidos en diferentes laboratorios no son comparables incrementando el error en estas metodologías. (Zankari et al., 2013)

La determinación de la concentración mínima inhibitoria (MIC por sus siglas en inglés) se define como la concentración mínima necesaria de antibiótico para inhibir el crecimiento de una población de bacterias en un periodo determinado de tiempo (Kester & Fortune, 2014). Se considera como la prueba de oro a la hora de determinar la resistencia, ya que

nos permite evaluar el grado de resistencia, lo cual se ve reflejado en aumento o disminución del MIC (Kester & Fortune, 2014; Wiegand et al., 2008).

Dada la necesidad de disponer datos fenotípicos de resistencia precisos y reproducibles de *A. baumannii*, que no generen ruido a la hora de desarrollar los modelos; además de disponer de medidas cuantitativas que reflejen el grado de resistencia, en este estudio se propuso el siguiente objetivo: Caracterizar fenotípicamente los perfiles de resistencia de *Acinetobacter baumannii*

2.2 Metodología

2.2.1 Aislamientos bacterianos

El presente estudio cuenta con 76 aislamientos de *A. baumannii* obtenidos durante el periodo 2012 – 2015, los cuales forman parte del programa de vigilancia de IASS del Nacional de Salud de Colombia. Los aislamientos se criopreservaron a – 80°C en medio *Skim Milk* y se depositaron en el cepario del Instituto de Biotecnología de la Universidad Nacional de Colombia, de donde se obtuvieron para ser utilizados en el presente estudio. Los aislamientos fueron previamente identificados como *A. baumannii* mediante amplificación y secuenciación del gen *rpoβ* (Prada Cardozo, 2018)

2.2.2 Selección de antibióticos

Se utilizó un panel amplio de antibióticos para la presente investigación, los cuales fueron elegidos a partir de varios criterios, los cuales se numeran a continuación

- Incluir los habitualmente usados en infecciones causadas por *A. baumannii* y disponibles en Colombia (Instituto Nacional de Salud, 2019)
- Tener por lo menos un representante de las diferentes familias de antibióticos disponibles actualmente
 - Costo
 - Incluir antibióticos que tengan mayor número de mecanismos de resistencia asociados

- Incluir aquellos que sean recomendados por el CLSI (Clinical and Laboratory Standards Institute & Weinstein, 2018)

Tabla 2-1: Antibióticos seleccionados

Antibiótico	Abreviatura	Familia de antibiótico
Amikacina	AMK	Aminoglicósidos
Ampicilina	AMP	Penicilinas
Cefazolina	CFZ	Cefalosporina de primera generación
Cefepime	FEP	Cefalosporina de cuarta generación
Ceftazidima	CAZ	Cefalosporina de tercera generación
Cefuroxima	CXM	Cefalosporina de segunda generación
Ciprofloxacina	CIP	Quinolona
Claritromicina	CLR	Macrólidos
Clindamicina	CLI	Lincosamida
Colistina	COL	Polimixina
Fosfomicina	FOS	Fosfonatos
Imipenem	IMP	Carbapenémicos
Linezolid		Oxazolidinona
Meropenem	MER	Carbapenémicos
Nitrofurantoína	NIT	Nitrofuranos
Oxacilina	OXA	Penicilinas
Tetraciclina	TCY	Tetraciclinas
Tianfenicol		Anfenicoles
Tigeciclina		Glicilclinas
Trimetoprim-Sulfametoxazol	SXT	Inhibidores de la vía del folato

Se seleccionaron un total de 21 antibióticos, los cuales se detallan en la tabla 2-1. Se logró contar con un representante de todas las familias de antibióticos, exceptuando a los monobactámicos y las rifamicinas

2.2.3 Preparación de soluciones madre de antibióticos

Se prepararon soluciones madre de los antibióticos seleccionados, cuyas diluciones se realizaron a una concentración dos veces mayor con respecto a la concentración final deseada y teniendo en consideración las instrucciones dadas por el fabricante del antibiótico. Los solventes y diluyentes utilizados para la preparación fueron los establecidos por el CLSI (Weinstein & Clinical and Laboratory Standards Institute, 2019). La concentración final se estableció con el fin de abarcar un total de 10 diluciones seriadas dobles de los antibióticos, rango en el cual se incluya el punto de corte (Tabla 2-2)

Los puntos de corte considerados fueron los establecidos por el CLSI (Weinstein & Clinical and Laboratory Standards Institute, 2019) (véase anexo A-1), excepto para la tigeciclina; cuyo punto de corte fue el establecido para *A. baumannii* en un consenso latinoamericano (Jiménez Pearson et al., 2019) . Para el trimetoprim-sulfametoxazol se tomó en cuenta solo el punto de corte del trimetoprim, dado que los estándares del CLSI y EUCAST vienen en una proporción 1:19, y nosotros disponíamos de una combinación en proporción 1:5.

Tabla 2-2: Rango de concentraciones de soluciones madre de antibióticos

Antibiótico	Concentración (µg/ml)	Antibiótico	Concentración (µg/ml)
Amikacina	1 – 512	Imipenem	0,250 – 128
Ampicilina	0,5 – 256	Linezolid	0,250 – 128
Cefazolina	0,125 - 64	Meropenem	0,250 – 128
Cefepime	0,5 – 256	Nitrofurantoína	2 - 1024
Ceftazidima	0,5 – 256	Oxacilina	0,125 - 64
Cefuroxima	0,5 – 256	Tetraciclina	0,250 – 128
Ciprofloxacina	0,250 – 128	Tianfenicol	0,5 – 256
Claritromicina	0,125 - 64	Tigeciclina	0,0625 - 32
Clindamicina	0,0625 - 32	Trimetoprim-Sulfametoxazol	0,250/1,25 - 128/640
Colistina	0,125 - 64	Vancomicina	0,250 – 128
Fosfomicina	4 - 2048		

2.2.4 Preparación de la suspensión bacteriana

Como primer paso se realizó la estandarización del inóculo, con el fin de garantizar un contenido de $\sim 1-2 \times 10^8$ UFC/ml en una absorbancia entre 0,130 – 0,250 medida por espectrofotometría a 600nm de longitud de onda.

El método directo recomendado por el CLSI fue el seleccionado para preparar el inóculo, para lo cual se tomaron 3-5 colonias frescas en un volumen de 3 ml de caldo Mueller – Hinton. Cada inóculo fue medido por espectrofotometría a 600nm de longitud de onda (véase anexo A-2). Si la absorbancia era superior a 0,250 se agregaba más caldo con el fin de disminuir la densidad óptica (DO), por el contrario, si la DO era muy baja se agregaba una nueva colonia hasta alcanzar una absorbancia mayor a 0,130. (Clinical and Laboratory Standards Institute & Weinstein, 2018)

2.2.5 Determinación de la concentración mínima inhibitoria

Se realizó la determinación de la concentración mínima inhibitoria (MIC) por el método de microdilución en caldo siguiendo los estándares establecidos por el CLSI (véase anexo B-1). Los ensayos se realizaron por triplicado tanto para cada uno de los 76 aislamientos, como para las cepas de referencia: *Escherichia coli* ATCC 25922 y *Staphylococcus aureus* ATCC 29213 (Clinical and Laboratory Standards Institute & Weinstein, 2018).

2.2.6 Recolección de datos públicos

Dado que el desempeño de los algoritmos de aprendizaje de máquina (*machine learning*) es más preciso y generalizado mientras mayor sea el conjunto de datos para entrenar (Babcock University et al., 2017; Giacobbe et al., 2020), además que es de crucial importancia disponer de conjunto de entrenamiento representativo de los casos (sensible / intermedio/ resistente) que queremos generalizar (Géron, 2019), se realizó una búsqueda en las bases de datos *Biosample* (BioSample – NCBI, 2020) y *Genome* del NCBI (Genome– NCBI, 2020), que correspondían a *Acinetobacter baumannii*, las cuales contenían la información del perfil fenotípico y cuyo genoma reportado estaba completo.

2.3 Resultados y discusión

2.3.1 Obtención de la suspensión bacteriana

Se realizó la estandarización de la suspensión bacteriana, en donde se obtuvo el número de células entre 1×10^8 y 2×10^8 UFC, como está establecido en CLSI (tabla 2-3)

Tabla 2-3: Recuento Celular

Aislamiento	Absorbancia	Recuento (UFC/ml)
GMR-RA687	0,173	1×10^8
RB147	0,211	$1,15 \times 10^8$
RB672	0,202	$1,75 \times 10^8$
RB1552	0,151	$9,82 \times 10^7$
RB640	0,130	$9,33 \times 10^7$
RB903	0,139	$1,03 \times 10^7$
RB979	0,217	$2,06 \times 10^8$
GMR-RA20	0,217	$1,9 \times 10^8$

Se ha informado que inóculos muy altos pueden conducir a un falso aumento de los valores de MIC, por otro lado un inóculo más pequeño da lugar a MICs más bajos; originando falsos resistentes y falsos susceptibles respectivamente, y esto podría afectar el desarrollo del modelo al emplear valores de MIC errados (Wiegand et al., 2008). Los inóculos obtenidos estuvieron dentro de los límites permitidos (tabla 2-3), lo que asegura la obtención de MICs confiables, y en consecuencia datos limpios para el entrenamiento de los modelos posteriormente.

2.3.2 Determinación de la concentración mínima inhibitoria

Se determinó la concentración mínima inhibitoria de 21 antibióticos para los 76 aislamientos de *A. baumannii* (Anexo A-3), y se los clasificó como sensibles, intermedios y resistentes según los puntos de corte establecidos en el CLSI (Weinstein & Clinical and Laboratory Standards Institute, 2019). La tabla 2-4 resume la susceptibilidad de los 76 aislamientos frente a los antibióticos probados menos trimetoprim-sulfametoxazol.

Tabla 2-4: Susceptibilidad de los 76 aislamientos colombianos frente a los antibióticos probados menos trimetoprim-sulfametoxazol

Antibiótico	Sensibles		Intermedios		Resistentes	
	N° aislamientos	%	N° aislamientos	%	N° aislamientos	%
Amikacina	15	20	27	35,53	34	44,74
Ampicilina	0	0	0	0,00	76	100,00
Cefazolina	0	0	0	0,00	76	100,00
Cefepime	5	7	3	3,95	68	89,47
Ceftazidima	13	17	28	36,84	35	46,05
Cefuroxima	0	0	0	0,00	76	100,00
Ciprofloxacina	2	3	1	1,32	73	96,05
Claritromicina	0	0	0	0,00	76	100,00
Clindamicina	0	0	0	0,00	76	100,00
Colistina	52	68	0	0,00	24	31,58
Fosfomicina	0	0	17	22,37	59	77,63
Imipenem	3	4	1	1,32	72	94,74
Linezolid	0	0	0	0,00	76	100,00
Meropenem	3	4	1	1,32	72	94,74
Nitrofurantoína	0	0	0	0,00	76	100,00
Oxacilina	0	0	0	0,00	76	100,00
Tetraciclina	2	3	19	25,00	55	72,37
Tianfenicol	0	0	0	0,00	76	100,00
Tigeciclina	76	100	0	0,00	0	0,00
Vancomicina	0	0	0	0,00	76	100,00

Los valores de la concentración mínima inhibitoria para el trimetoprim-sulfametoxazol se muestran en la tabla 2-5, dado que por lo mencionado anteriormente no se le puede clasificar como sensible, intermedio y resistente, utilizando los puntos de corte estándar del CLSI. Sin embargo, al considerar únicamente la concentración del trimetoprim se pudo observar que solo un aislamiento (RB69) se podría categorizar como sensible (MIC < = 2ug/ml).

Tabla 2-5: Resultados de la concentración mínima inhibitoria para el trimetoprim-sulfametoxazol

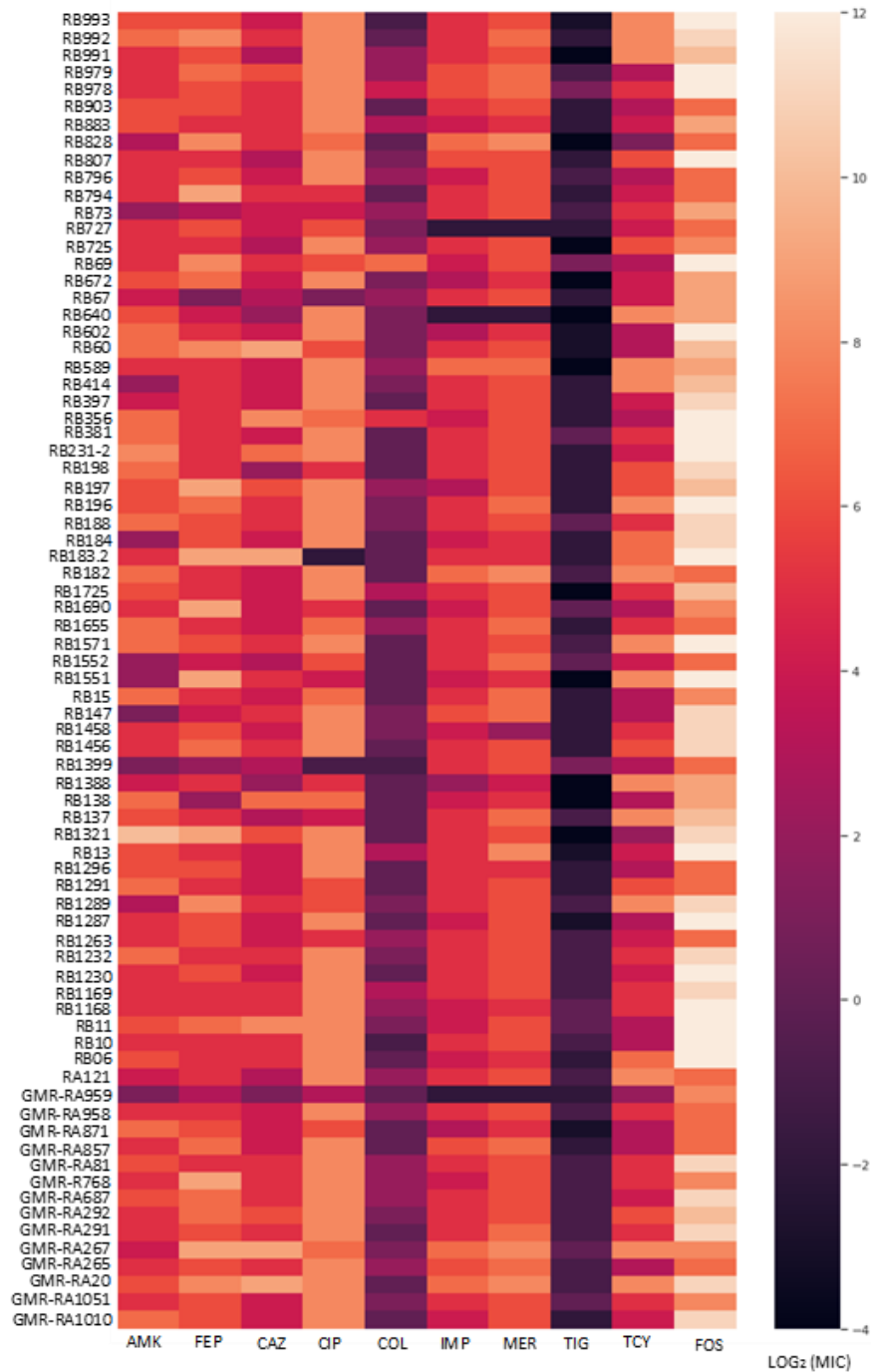
Concentración mínima inhibitoria (µg/ml)	Numero de aislamientos
0,250/1,25	0
0,500/2,5	0
1/5	0
2/10	1
4/20	3
8/40	1
16/80	6
32/160	18
64/320	12
128/640	16
>128/640	19

El presente estudio tiene como primer objetivo evaluar la eficacia de la mayor cantidad de antibióticos disponibles hoy en día para el tratamiento de enfermedades infecciosas, y por otro lado incluir la mayor cantidad de antibióticos que nos permitan evaluar los diferentes mecanismos de resistencia; por lo que se emplearon 21 antibióticos, teniendo un representante de cada una de las familias de antibióticos, excepto para el grupo de las penicilinas y carbapenémicos en donde se valoraron la ampicilina y oxacilina, imipenem y meropenem; respectivamente. Como muestra la tabla 2-4, el 100% de los aislamientos son resistentes a ampicilina, cefazolina, cefuroxima, claritromicina, clindamicina, linezolid, nitrofurantoína, oxacilina, tianfenicol y vancomicina. Comportamiento esperado en el caso de los antibióticos claritromicina, clindamicina, linezolid, nitrofurantoína, tianfenicol y

vancomicina, frente a los cuales *A. baumannii* es naturalmente resistente. En cuanto a los 4 restantes (ampicilina, oxacilina, cefazolina, cefuroxima) desde décadas pasadas se conocía de su baja efectividad, la cual es mucho más reducida en los 76 aislamientos colombianos estudiados, ya que los altos valores de MICs encontrados (anexo A-3) permiten clasificarlos como resistentes, lo cual puede deberse a la facilidad con que este microorganismo modifica sus mecanismos de resistencia por mutaciones o por transferencia horizontal de genes (Cai et al., 2019; Sader et al., 1998). Como muestra la tabla 2-4 algunos aislamientos mostraron susceptibilidad variable: 20, 7, 17, 3, 68, 4, 4 y 3% para amikacina, cefepime, ceftazidima, ciprofloxacina, colistina, imipenem, meropenem y tetraciclina respectivamente. Estos datos evidencian lo desafiante que se ha vuelto el manejo de infecciones causadas por *A. baumannii*, ya que la continua acumulación de mecanismos de resistencia por parte de este microorganismo hace que disminuya el número de antibióticos disponibles para tratar estas infecciones causadas por el microorganismo en la práctica clínica (Lee et al., 2017; A. Opazo et al., 2012)

Para la amikacina, cefepime, ceftazidima, ciprofloxacina, colistina, imipenem, meropenem, tigeciclina, tetraciclina y fosfomicina, se pudieron observar diversos grados de resistencia (figura 2-1), esto se ve reflejado por la variabilidad que se encontró en los valores de la MIC, lo cual permite el desarrollo de modelos de predicción de estos antibióticos. El éxito del aprendizaje de máquina depende en gran manera de los datos utilizados para el entrenamiento, y cuando se habla del valor de salida (*output*) es recomendable que estos valores sean variables, y que se encuentren representados por varias instancias (genomas) (Géron, 2019; Pedregosa et al., 2011). En la figura 2-1 podemos observar la variabilidad de los perfiles para 10 antibióticos. Los perfiles correspondientes a ciprofloxacina y la fosfomicina fueron los menos variables, lo cual podría afectar el desempeño del modelo, además de que nos da una idea de que estos antibióticos resultarían poco eficaces a la hora de tratar infecciones por esta bacteria. En cuanto a la colistina y la tigeciclina podemos ver que la mayoría de los aislamientos son sensibles, no obstante se pudo observar varios valores de MIC, los cuales son los suficientemente variables permitiéndonos desarrollar modelos.

Figura 2-1: Perfiles de resistencia de *A. baumannii*



- Mapas de calor de los valores de MIC transformados a logaritmo base dos. Las celdas están codificadas con gradiente de color de negro a blanco, con negro para valores más bajos y blanco para valores más altos.
- AMK: amikacina, FEP: cefepime, CAZ: ceftazidima, CIP: ciprofloxacina, IMP: imipenem, MER: meropenem, TIG: tigeciclina, TCY: tetraciclina, FOS: fosfomicina

La tigeciclina fue el único antibiótico que mostro una tasa de sensibilidad del 100% (76 aislamientos sensibles), seguido de la colistina con 53 aislamientos sensibles, lo que hace a estos antibióticos eficaces para combatir infecciones de *A. baumannii* a menos que se encuentren contraindicados o existan otras alternativas de antibióticos menos peligrosos (Isler et al., 2019; Rolain et al., 2016). Sin embargo, la tasa de resistencia para la colistina se ha visto incrementada (32%) en comparación con otro estudio realizado en el 2014 que reportó una resistencia máxima del 18% (Gao et al., 2017), lo cual podría deberse a la presión selectiva que se ha ejercido al implementar el uso de este antibiótico en la clínica (Munck et al., 2014). Para los demás antibióticos se observa (figura 2-1) una buena distribución de los valores de MIC, lo cual hace posible que los algoritmos puedan aprender adecuadamente de los datos, ya que durante el entrenamiento un valor es observado varias veces, logrando captar patrones y aprender de los datos, que es lo que permite el aprendizaje de máquina. (Géron, 2019; Nguyen, Long, et al., 2018).

En cuanto a la efectividad de los carbapenémicos frente a *A. baumannii*, este en los reportes más recientes ha evidenciado un constante incremento en su resistencia, (Cerezales et al., 2019; Alzuhairi et al., 2020), incluso frente al imipenem y meropenem, que han sido ampliamente utilizados para tratar estas infecciones por su efectividad (Lee et al., 2017). En este estudio únicamente se observaron tres aislamientos sensibles, y los demás aislamientos presentaron valores de MIC frente a imipenem y meropenem por encima de 8µg/ml que los clasifican como resistentes (Anexo A-3), lo que es un escenario bastante complicado ya que los microorganismos resistentes a estos antibióticos generalmente suelen ser resistentes a los demás antibióticos comúnmente utilizados, lo cual causa dificultad para controlar las infecciones causadas por estos microorganismos (Gao et al., 2017; Lee et al., 2017).

Varios estudios han reportado una alta tasa de resistencia a la ceftazidima y al cefepime, que concuerdan con los resultados obtenidos en el presente trabajo, cuyas tasas de resistencia fueron del 46% (35 aislamientos) y 89% (68 aislamientos), respectivamente. Dentro de las cefalosporinas, aparentemente una mejor opción sería el uso de una cefalosporina de tercera generación como ceftazidima, ya que presenta mayor eficacia según los resultados del presente estudio (tabla 2-4), lo cual concuerda con otra investigación realizada en China. (Gao et al., 2017). La amikacina presenta una tasa de

resistencia del 44% (34 aislamientos), pero el 36% de los aislamientos resultaron en la categoría de intermedio, lo que permite pensar que se puede manejar su eficacia con un adecuado ajuste de dosis siempre y cuando se evite alcanzar la dosis tóxica del antibiótico (Weinstein & Clinical and Laboratory Standards Institute, 2019).

2.3.3 Recolección de datos públicos

Cuando se desarrollan modelos utilizando métodos de aprendizaje de máquina, el número de instancias utilizadas para su desarrollo es crucial para obtener predicciones precisas (Nguyen, Brettin, et al., 2018; Pataki et al., 2019). Aunque los 76 aislamientos presentaron variabilidad en sus perfiles, estos resultan insuficientes para conseguir modelos con un rendimiento acertado (Her & Wu, 2018). Por lo anterior se recurrió a las bases de datos *NCBI* para la obtención de datos, y así aumentar conjunto de datos, logrando un conjunto de entrenamiento más robusto y balanceado.

Se seleccionaron 267 genomas que correspondían a *A. baumannii* (tabla 2-6) (Anexo A-4), a los cuales se adicionaron a los 76 aislamientos colombianos previamente mencionados, obteniendo así un conjunto de 343 secuencias genómicas completas.

Tabla 2-6 : Información de los genomas de referencia de *A. baumannii*

	NUMERO DE SECUENCIAS	PROYECTO	AÑO DE RECOLECCION	ANTIBIOTICOS
267	1	PRJNA234144	2013	9
	10	PRJNA472798		21
	100	PRJNA545079	2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015	14
	1	PRJNA317141		20
	9	PRJNA384065	2007, 2008, 2009	14
	146	PRJNA384060		14

- La información presentada es obtenida del *NCBI*, se muestra el número de secuencias obtenidas por proyecto, además del año de recolección y el número de antibióticos que fueron evaluados.

La tabla 2-7 resume los perfiles de susceptibilidad correspondientes a los 267 genomas de referencia obtenidos de la base de datos *Biosample del NCBI*, cuyos valores de MIC se encuentran en el Anexo A-4.

Tabla 2-7: Perfiles de susceptibilidad correspondientes a los 267 genomas de referencia obtenidos de la base de datos *Biosample del NCBI*

Antibióticos	Sensible	Intermedio	Resistente	Total
Amikacina	129	100	37	266
Aztreonam	0	0	10	10
Cefepime	105	105	57	267
Cefotaxime	0	0	11	11
Ceftazidima	99	106	61	266
Ceftriaxona	28	20	63	111
Ciprofloxacino	83	128	56	267
Colistina	252	0	15	267
Doripenem	155	0	1	156
Fosfomicina	0	0	10	10
Gentamicina	75	129	63	267
Imipenem	122	2	142	266
Levofloxacino	48	11	53	112
Meropenem	117	4	146	267
Minociclina	7	3	0	10
Polimixina B	145	0	11	156
Tetraciclina	55	6	40	101
Tigeciclina	141	23	3	167
Tobramicina	49	23	40	112

Para la selección de las secuencias se tuvo en consideración primero que las secuencias correspondan a *A. baumannii*, esperando que compartan mecanismos de resistencia con

nuestro conjunto de datos (76 aislamientos), lo que facilita el éxito de los modelos, ya que el aprendizaje depende de la cantidad de veces que una característica (gen) pueda verse durante el entrenamiento (Géron, 2019). Y por otra parte, se conoce que es indispensable el uso de datos balanceados para incrementar la posibilidad de obtener un buen modelo (Géron, 2019; Nguyen, Long, et al., 2018; Raschka, 2015); razón por la cual también se intentó recolectar genomas que nos permitiera obtener un conjunto de datos balanceado y variable, haciendo énfasis en los genomas susceptibles ya que los 76 aislamientos presentaban altas tasas de resistencia, como se observó previamente. Fueron seleccionados aislamientos de *A. baumannii*, cuyo genoma reportado estaba completo y cuyo perfil fenotípico de resistencia se encontraba disponible en la base de datos *Biosample* del *NCBI*. Los perfiles de susceptibilidad de los aislamientos elegidos diferían entre sí, es decir los antibióticos reportados no eran los mismos para todos, razón por la cual se desarrollaron modelos de predicción individuales por antibiótico.

3. Capítulo 3: Identificación de las características genómicas asociadas a resistencia utilizadas como entrada para el desarrollo de modelos de predicción

3.1 Introducción

La resistencia bacteriana se encuentra mediada por adquisición de mutaciones cromosomales y transferencia horizontal de genes (Aytan-Aktug et al., 2020; Munck et al., 2014), por tales motivos el análisis del perfil genómico, específicamente el asociado a resistencia, resulta útil para la determinación de resistencia antimicrobiana (Aytan-Aktug et al., 2020). Además de mejorar la velocidad en la detección de resistencia, nos permite obtener información importante que puede contribuir a la vigilancia (Bradley et al., 2015).

Enfoques basados en aprendizaje de máquina se están aplicando actualmente para realizar modelos de predicción de resistencia, cuya precisión está condicionada en gran medida por la calidad de los datos que se emplean como entrada (Aytan-Aktug et al., 2020). Usar los datos genotípicos como datos de entrada para el entrenamiento de un modelo de predicción representa un desafío computacional grande en términos de velocidad de procesamiento lenta y requerimientos de memoria (Bottou, 2010), si no se realiza un procesamiento previo. Este inconveniente ha sido tratado de dos formas: utilizando las anotaciones del genoma (evitando el uso de secuencias) y *k-mers* (Macesic et al., 2017). Hasta el momento, la mayoría de los estudios han estado centrados en el uso de genes y SNPs caracterizados y asociados a los diferentes mecanismos de resistencia, los cuales han dado lugar a un buen desempeño, logrando una exactitud de los modelos hasta un 93% (Eyre et al., 2017). Sin embargo esto resulta útil únicamente para aquellos antibióticos cuyos mecanismos de resistencia estén determinados. Para combatir esta problemática, el uso de variantes raras (no conocidas) asociadas a genes de resistencia se han empleado originando modelos con una exactitud mayor al 90% para predicción de

resistencia en *Mycobacterium tuberculosis*; también se han combinado estos dos parámetros, variantes raras y causales, y la exactitud incremento en un 4,2%. (M. L. Chen et al., 2019). Igualmente, los estudios destacan que los métodos de aprendizaje de máquina permiten identificar todas las mutaciones que podrían estar involucradas en la resistencia (Aytan-Aktug et al., 2020; M. L. Chen et al., 2019).

Por otro lado, el uso de *k-mers* se ha vuelto una buena alternativa ya que permiten emplear todo el genoma, incluyendo en el modelo regiones reguladoras, operones y genes que no se encuentren aun identificados, logrando mejores desempeños ya que el modelo podría captar patrones tales como: co-dependencia de genes, posibles mutaciones en regiones reguladoras de expresión de genes e incluso nuevos genes o SNPs implicados en la resistencia (Nguyen, Long, et al., 2018); además de que su uso ha sido eficaz tanto en modelos de clasificación como de regresión con promedios de exactitud alrededor del 97% (Nguyen, Brettin, et al., 2018; Nguyen, Long, et al., 2018). Su principal desventaja es el ruido que genera debido a la redundancia, es decir que muchos *k-mers* siempre están presentes o ausentes simultáneamente, además de que interpretarlos se vuelve complejo (Drouin et al., 2019).

Otros predictores que han sido utilizados son: la estructura poblacional, genes accesorios, indel (inserciones/deleciones), SNPs en todo el genoma e incluso el año del aislamiento, de los cuales el mejor desempeño entre estos se obtuvo con los genes accesorios, con una exactitud del 93% y un *recall* de 83% (Moradigaravand et al., 2018). Al agregar la estructura poblacional y el año de aislamiento como dato de entrada no se observó una mejora significativa en el modelo, aunque el año de aislamiento ha resultado útil para confirmar que tan sostenibles en el tiempo son las variaciones genéticas asociadas a resistencia. Por último, el empleo de *indels* ha dado lugar a desempeños muy pobres en los modelos, debido a su poca relación con la resistencia antimicrobia. (Moradigaravand et al., 2018).

Visto que se ha comprobado que la información del resistoma es suficiente para garantizar un desempeño eficiente de los modelos de predicción, el presente capítulo tiene como objetivo identificar las características genómicas asociadas a resistencia y determinar sus variaciones en los 343 genomas de *A. baumannii*, para lo cual se utilizaron la secuencias de las proteínas traducidas a partir de la secuencia genómica como datos de entrada para el desarrollo de los modelos de predicción de resistencia.

Para hacer la identificación de los elementos genómicos codificados en cada genoma se recurrió a la base de datos CARD DB (*The Comprehensive Antibiotic Resistance Database*), la cual fue seleccionada ya que es curada de forma manual lo que asegura el uso de datos de referencia de alta calidad que incluyen secuencias de elementos genómicos asociados a AMR, incluida la resistencia intrínseca (C. Chen et al., 2019). Para formar parte de la base de datos, las secuencias deben cumplir con 3 requisitos: estar descrita en una publicación científica, la secuencia de DNA debe estar disponible en el *GenBank* e incluir evidencia experimental de concentraciones mínimas inhibitorias elevadas (Alcock et al., 2020). Esta base de datos cuenta con la herramienta *Resistance Gene Identifier (RGI)* para la predicción del resistoma a partir de la secuencia del genoma. Adicionalmente, CARD DB se actualiza mensualmente lo que asegura una anotación actualizada (Jeukens et al., 2019; Jia et al., 2017).

3.2 Metodología

3.2.1 Predicción de resistoma

Para la predicción del resistoma de los 343 genomas de *A. baumannii* seleccionadas para el presente estudio se utilizó el software *Resistance Gene Identifier (RGI)* 5.1.0 de la base de datos CARD DB (*The Comprehensive Antibiotic Resistance Database*, 2020)

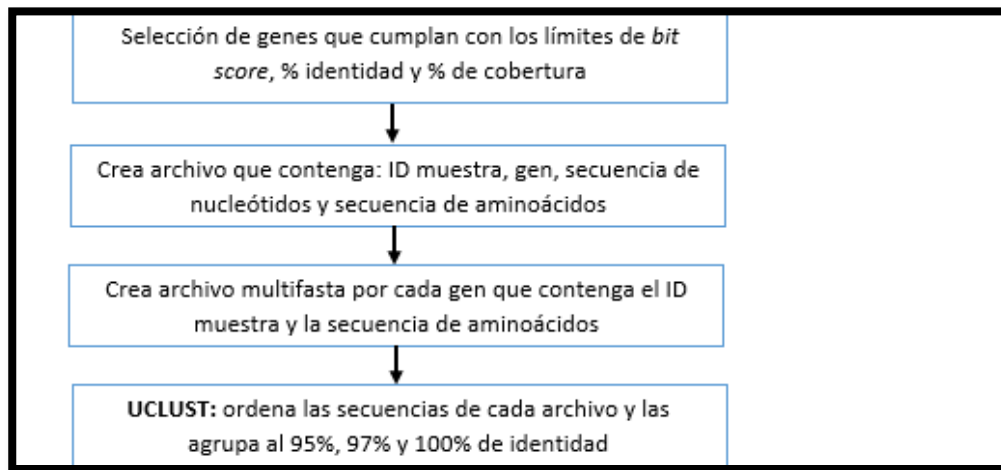
RGI se ejecutó en un servidor Dell con sistema operativo Linux. Los elementos genómicos asociados a la resistencia a antibióticos son identificados y clasificados por el RGI como *perfect* que corresponden a las secuencias de proteínas determinadas con el 100% de identidad en relación a una secuencia de referencia de la base de datos CARD, *strict* que corresponde a aquellas secuencias que presentan variaciones al compararlas con la secuencia de referencia de CARD pero se encuentra dentro de los límites del *bit score*, y *loose* funciona fuera de los límites del modelo con el fin de permitirnos detectar homólogos más distantes de genes de resistencia (Alcock et al., 2020), y son almacenados en dos archivos de salida: un archivo en formato .json y otro en formato texto delimitado por tabulaciones. De este último se obtuvo la secuencia de aminoácidos de cada uno de los elementos de resistencia identificados.

3.2.2 Determinación de las variaciones aminoacídicas de los elementos asociados a resistencia

A partir del resistoma de todos los aislamientos, se identificaron las variaciones a nivel de la secuencia de proteína de cada uno de los elementos genómicos asociados a resistencia identificados previamente.

Para esto el Grupo de Bioinformática del Instituto de Biotecnología de la Universidad Nacional de Colombia diseñó un flujo de trabajo (Figura 3-1), el cual se implementó a través de un Script escrito en Python 3.7 (Python Software Foundation, 2021) llamado *tablaAlelos.py*, que fue ejecutado en un servidor Dell con sistema operativo OpenSuse 15.2 (Sanchez, Marjanovic, 2021).

Figura 3-1: Flujo de trabajo del programa *tablaAlelos.py* para la determinación de las variaciones a nivel de secuencia de proteína de cada uno de los elementos asociados a resistencia encontrados en los genomas de los 343 aislamientos de *A. baumannii* analizados.



Tras realizar la anotación de genes asociados a resistencia a través del software *RGI* previamente descrito, se utilizó el archivo de salida en formato de texto plano que corresponde a la información de los *hits* identificados, para cada aislamiento en estudio, que incluye nombre del gen, secuencia de proteína, secuencia de DNA, porcentaje de cobertura, porcentaje de identidad, entre otros (Alcock et al., 2020)

Inicialmente, *tablaAlelosAB.py* leyó cada una de las tablas de resultados generada por *RGI*, fila por fila, seleccionando aquellos genes que tuvieran un porcentaje de identidad y

de cobertura mayores al 50% y 80% prospectivamente con respecto los genes de referencias en CARD DB (figura3-1).

Con estos datos *tablaAlelos.py* creó un archivo *tablageneral.txt* donde se almacenó una matriz contiene los identificadores de los aislamientos, nombre del gen (únicamente los que superaron los límites de *bit score*, porcentaje de identidad y cobertura previamente establecidos), el *bit score*, la secuencia de nucleótidos, la secuencia de aminoácidos, el porcentaje de identidad y el porcentaje de cobertura. Después de esto, partir del archivo *tablageneral.txt* el programa *tablaAlelos.py* generó un archivo multifasta con todas las secuencias de aminoácidos identificadas para un mismo gen ordenadas de mayor a menor tamaño incluyendo la identificación de cada aislamiento. En seguida mediante el programa UCLUST (Edgar, 2010) las secuencias de cada archivo multifasta fueron agrupadas al 100%, 97% y al 95% y los resultados de cada agrupamiento de cada gen fueron almacenados en la carpeta *gpalel*. De esta manera las secuencias con porcentajes de identidad del 100% quedan agrupadas bajo un mismo alelo; lo mismo sucede con el 97% y 95%. Estos resultados fueron leídos entonces por el programa para generar tres archivos de salida en formato de texto, cada uno correspondiente a las 3 agrupaciones mencionadas, cada archivo contenía el ID del aislamiento y los nombres de los genes encontrados en los 343 genomas, las columnas de los genes fueron llenas con su alelo correspondiente dependiendo de la agrupación utilizada, y con -1 en los casos donde un dicho gen estaba ausente. Estas matrices fueron utilizadas para el entrenamiento de los modelos (tabla 3-1).

Tabla 3-1: Ejemplo de la matriz de las variaciones de elementos asociados a resistencia generada utilizando el programa *tablaAlelos.py*

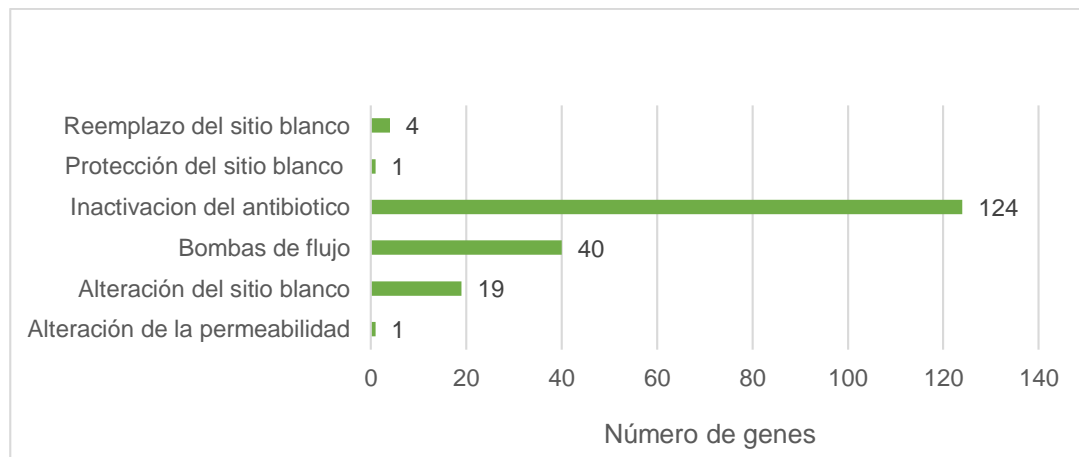
ID_aislamiento	adeF	OXA-23	adeR
INSRB-06	1	2	0
INSRB-10	0	1	1

3.3 Resultados

3.3.1 Predicción de resistoma

En los 343 aislamientos de *A. baumannii* seleccionados, los genes que podrían mediar resistencia fueron identificados utilizando la herramienta *RGI*, los genes seleccionados se hicieron estableciendo un umbral del 50% para el porcentaje de identidad y 80% para el porcentaje de cobertura porque la literatura afirma que se necesita como mínimo este valor para inferir con precisión la función de una proteína, además con el fin de aprovechar la capacidad de reconocer nuevas características en los datos por las técnicas basadas en aprendizaje de máquina; la elección del límite más bajo consideramos conveniente. (Giacobbe et al., 2020). Por otro lado se estableció también un punto de corte para la cobertura que fue del 80%. (Kankainen et al., 2012; Rost, 1999). Se identificaron 192 genes mediadores de resistencia, de estos 19 presentaron dos copias, y uno (*sul2*) tres copias. Se encuentran involucrados en 6 mecanismos de resistencia y se distribuyen así: 1 gen que participa en la alteración de la permeabilidad, 19 en alteración del sitio blanco (tabla 3-3), 40 bombas de eflujo (tabla 3-5), 124 en inactivación del antibiótico (tabla 3-2), 1 en protección del sitio blanco y 4 en reemplazo del sitio blanco (tabla 3-4) (Figura 3-2).

Figura 3-2: Mecanismos de resistencia de los 343 genomas de *A. baumannii*

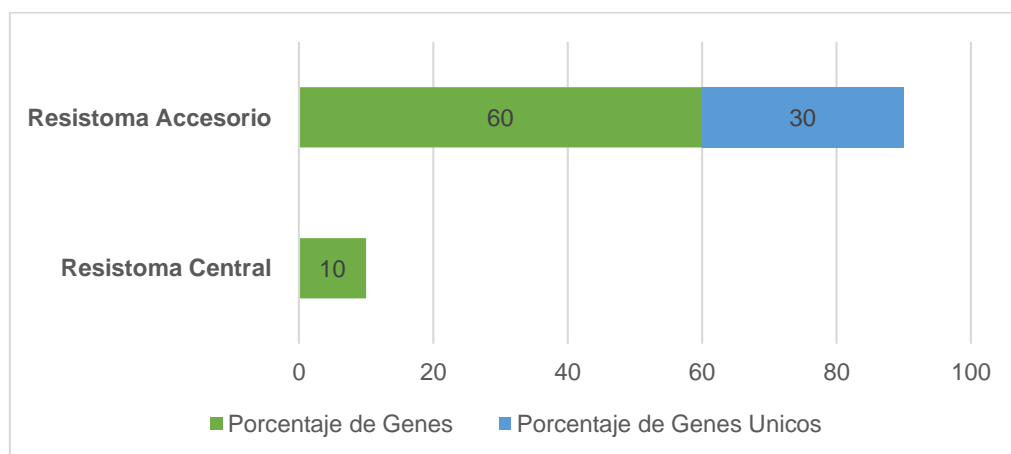


Estos resultados muestran la habilidad de *A. baumannii* por adquirir resistencia, lo que se ve representado por el elevado número de genes asociados a resistencia identificados, el mecanismo más importante descrito en *A. baumannii* son las β -lactamasas, cual concuerda con lo encontrado en este estudio (Lee et al., 2017). La acumulación de genes de resistencia está relacionada con cambios en la susceptibilidad, por lo que esto permite

desarrollar modelos de predicción, los cuales aprenden del comportamiento de los datos. El elevado número de genes de resistencia encontrados hacen posible el uso de estos como entrada de los modelos, no obstante, muchos de estos genes fueron encontrados en un solo genoma, siendo poco informativos para el algoritmo, por lo que uso sería restringido.

Con el fin de evaluar la variabilidad genómica y obtener una perspectiva más amplia de la resistencia a antibióticos, se ha definido como “resistoma central” a todo el conjunto de genes de resistencia que se encontraban presente en al menos el 95% de los aislamientos, como “resistoma accesorio” a los genes restantes, y de estos últimos llamamos “genes únicos” aquellos genes que estaban presente únicamente en un genoma (Harrison et al., 2020). Como muestra la figura 3-3 únicamente el 10% de genes (n=18) forman parte del resistoma central, y el 90% (n=174) se considera como resistoma accesorio. Se puede evidenciar que la mayor parte de los genes de resistencia forman parte del genoma accesorio, lo que puede indicar que fueron adquiridos principalmente por transferencia horizontal o que variaron significativamente sus secuencias por mutaciones acumuladas, fenómenos producidos como consecuencia de la adaptación de la bacteria al medio en que crece (Álvarez et al., 2020).

Figura 3-3: Resistoma de los 343 genomas de *A. baumannii* utilizados para el entrenamiento de los modelos de aprendizaje de máquina.



- La figura ilustra la extensa variabilidad genómica de *A. baumannii* con relación a los genes de resistencia, en donde los 343 genomas comparten únicamente el 10% de genes. Por otro lado el 30% de genes estaban presente en un único genoma.

La variabilidad genómica de *A. baumannii*, hace posible su uso para desarrollar modelos de predicción, pues es indispensable que existan variaciones que permitan diferenciar patrones que den lugar a un resultado en particular (Géron, 2019; Raschka, 2015), lo que

unido a que sus mecanismos de resistencia han sido muy estudiados, hace que el uso dicha variabilidad genómica sea muy útil para lograr predicciones precisas (Deelder et al., 2019). Además, dado el pequeño genoma central que *A. baumannii* presenta, estudios anteriores mencionan que la mayor parte del resistoma debería ser parte del genoma accesorio, lo cual concuerda con lo encontrado en el presente estudio (M. Corredor & Muñoz-Gómez, 2020). También, encontramos que del resistoma accesorio el 30% de genes eran genes únicos, cuyo mecanismo de resistencia más predominante fue inactivación del antibiótico (67%) (Figura 3-2). En el resistoma central, el 85% (n=16) de genes encontrados correspondían a bombas de eflujo (tabla 3-5), lo que concuerda con reportes previos donde han sido caracterizadas en *A. baumannii* asignándoles un rol importante en la resistencia antimicrobiana de esta bacteria e incluso se a mencionado que todas las bombas de eflujo conocidas de las bacterias gram negativas se han descrito en *A. baumannii*.(Abdi et al., 2020; Lin et al., 2017). Sin embargo, la sola presencia de estos genes no es causante de fenotipos resistentes, sino más bien su sobre-expresión es la que produce las altas tasas de resistencia, generalmente debido a mutaciones en sus promotores (Lin et al., 2017); por tales motivos resulta un tanto limitante el uso de este grupo de genes como predictores de resistencia, dado que para el presente estudio no se utilizara como entrada del modelo regiones reguladoras, sino únicamente genes de resistencia (Jeukens et al., 2019).

En cuanto a los genes que forman parte del resistoma accesorio, la mayoría de ellos se encontraban en un grupo limitado de genomas; lo que pone en evidencia la gran variabilidad de las cepas de *A. baumannii* con respecto a los genes de resistencia. Como se observa en las tablas 3-2, 3-3, 3-4 y 3-5, se encontraron genes de resistencia para las principales familias de antibióticos usados generalmente para tratar infecciones ocasionadas por esta bacteria (aminoglicósidos, carbapenémicos, cefalosporinas, fluoroquinolonas, glicopéptidos, tetraciclina, tigeciclina) (Lee et al., 2017). La inactivación del antibiótico fue el mecanismo que mayor cantidad de genes asociados encontramos (tabla 3-2), no obstante su distribución es variable entre los genomas de *A. baumannii*, lo que podría deberse a la capacidad natural que este microorganismo posee para incorporar DNA exógeno, de tal manera que los genes incorporados son dependientes del medio en donde se lleve a cabo este fenómeno, lo que explicaría las distribución tan variable de los genes en estos genomas (Lee et al., 2017). Las oxacilinasas son las principales β -lactamasas que se encuentran en *A. baumannii*, algunas de ellas están presentes de

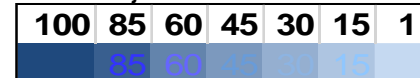
manera intrínseca como el subgrupo OXA-51-like, que confieren actividad leve de β -lactamasas (Mentasti et al., 2020). Sin embargo, se ha reportado que un aumento en la actividad oxacilinasas por la presencia de elementos de inserción tipo *IsAba1* ha dado lugar a cepas resistentes a carbapenémicos (Yakkala et al., 2019). En nuestro conjunto de datos hemos encontrado 55 enzimas de tipo OXA de las cuales 46 correspondieron a OXA-51-like, las cuales se encuentran codificadas en el cromosoma y son intrínsecas en *A. baumannii* (Chakravarty, 2020). La variante más predominante fue OXA-66, la cual no tiene relevancia clínica incluso con la presencia de secuencias de inserción, pues no se ha evidenciado una correlación entre los fenotipos resistentes y esta variante (Nigro & Hall, 2018), sin embargo la mayoría de los aislamientos que presentaban esta variante tenían un fenotipo resistente a carbapenémicos, lo cual puede deberse a otro tipo de β -lactamasas que se encuentran ampliamente distribuidas. El gen *bla-OXA-23* es la principal oxacilinasas adquirida mediante plásmidos (Mentasti et al., 2020), la cual representa el mecanismo más relevante en conferir resistencia a carbapenémicos en *A. baumannii*. En el presente estudio encontramos que el 81% (n=177) de los aislamientos resistentes tenían alguna oxacilinasas del subgrupo OXA-23-like, y el resto que no la contenía, albergaban otro tipo de oxacilinasas, y β -lactamasas como ADC, NDM, KPC, IMP, TEM, VEB y CTX-M, siendo la TEM la segunda más frecuente (49%) después de la OXA-23. La variante VEB-9 es reportada por primera vez en *A. baumannii* en el presente estudio, la cual se encontró en el aislamiento GMR-RA267, que forma parte de los 76 aislamientos colombianos recolectados por el Instituto Nacional de Salud de Colombia. Este gen fue previamente encontrado en *P. aeruginosa* y asociado con fenotipos resistentes a cefalosporinas de tercera y cuarta generación (Jahan et al., 2019), cuyas MICs son bastante elevadas, lo que concuerda con lo observado en el presente estudio donde las MICs de cefepime y ceftazidima resultaron >256 ug/ml en dicho aislamiento. Se identificó con un porcentaje de identidad y cobertura del 100%, lo que corrobora la presencia de esta variante en dicho aislamiento, y puede soportar los reportes previos de que la diseminación de resistencia puede ocurrir entre bacterias de diferente género, como sucede entre *A. baumannii* y *P. aeruginosa* (Sid Ahmed et al., 2020). Por otro lado, únicamente en el 6% (n=14) de los aislamientos se encontraron solamente mecanismos de resistencia intrínsecos (cromosomales). Asociados a estos se identificaron genes de tipo OXA-51-like y ADC, los cuales posiblemente están asociados a secuencias de inserción que ocasionan sobre-expresión de estos genes, dando como resultado incremento en las MICs de antibióticos betalactámicos, incluidos los carbapenémicos. (Nigro & Hall, 2018)

Tabla 3-2: Resistoma accesorio correspondiente a Inactivación del antibiótico de los 343 aislamientos de *A. baumannii*

Mecanismo de Resistencia	Gen	Porcentaje	Localización	Aminoglicósido	Anfenicol	Carbapenémico	Cefalosporina	Diaminopirimidino	Fluoroquinolona	Fosfomicina	Glicilciclina	Penicilinas	Péptidos	Rifamicinas	Sulfonamidas	Tetraciclinas	Otros
Inactivación del Antibiótico	APH(3'')-Ib	100	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	APH(6)-Id	100	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	OXA-23	100	Rosado	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	TEM-1	100	Rosado	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	mphE	100	Rosado	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	APH(3')-VIa	100	Amarillo	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	AAC(3)-IIe	100	Amarillo	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	OXA-66	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	APH(3')-Ia	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	aadA	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	ADC-5	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	ADC-30	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	AAC(6')-Ia	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	OXA-65	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	AAC(6')-Ib7	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	ANT(2'')-Ia	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	ADC-25	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	catB8	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	OXA-64	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	AAC(3)-Ia	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	ADC-79	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	ampC beta-lactamase	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	ADC-2	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	OXA-82	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	ADC-76	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	ADC-58	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	OXA-69	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
	ADC-3	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco
OXA-68	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	
OXA-100	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	
OXA-72	100	Rosado	Verde	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	Blanco	

El porcentaje se encuentra representado en escala de color. Localización plasmídica: verde, localización cromosomal: rosado, reportado en plásmido y cromosoma: amarillo. El negro representa que dicho gen confiere resistencia a un cierto antibiótico. Gen que presenta dos copias: letra en rojo

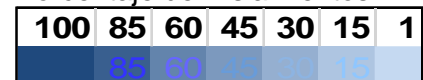
Porcentaje de Aislamientos



Mecanismo de Resistencia	Gen	Porcentaje	Localización	Aminoglicósidos	Anfenicol	Carbapenémicos	Cefalosporinas	Diaminopirimidina	Fluoroquinolonas	Fosfomicina	Gliciliciclina	Penicilinas	Péptidos	Rifamicinas	Sulfonamidas	Tetraciclinas	Otros
Inactivación del Antibiótico	OXA-90	100	Verde			■	■					■					
	AAC(6)-laf	100	Verde	■													
	aadA2	100	Rosado	■													
	BRP(MBL)	100	Rosado										■				
	NDM-1	100	Rosado				■	■									
	OXA-51	100	Verde				■	■					■				
	OXA-71	100	Verde				■	■					■				
	OXA-24	100	Rosado				■	■					■				
	ADC-77	100	Verde				■	■					■				
	OXA-94	100	Verde				■	■					■				
	catI	100	Amarillo	■	■		■	■					■				
	OXA-208	100	Verde				■	■					■				
	CARB-16	100	Rosado				■	■									■
	ADC-6	100	Verde				■	■									■
	SAT-2	100	Verde														■
	arr-2	100	Rosado												■		
	OXA-259	100	Verde				■	■					■				
	catB3	100	Rosado	■	■		■	■					■				
	OXA-121	100	Verde				■	■					■				
	FosA6	100	Rosado				■	■			■		■				
	OXA-314	100	Verde				■	■					■				
	OXA-98	100	Verde				■	■					■				
	ADC-74	100	Verde				■	■					■				
	APH(3')-VI	100	Rosado	■	■		■	■									
	OXA-120	100	Verde				■	■					■				
	PER-7	100	Rosado				■	■					■				
	ADC-56	100	Verde				■	■					■				
	OXA-407	100	Verde				■	■					■				
	PER-1	100	Rosado				■	■					■				
	Tet(X3)	100	Rosado									■					■
VIM-4	100	Rosado				■	■					■					
OXA-146	100	Rosado				■	■					■					
ADC-81	100	Verde				■	■					■					
OXA-235	100	Rosado				■	■					■					
OXA-223	100	Verde				■	■					■					

El porcentaje se encuentra representado en escala de color. Localización plasmídica: verde, localización cromosomal: rosado, reportado en plásmido y cromosoma: amarillo. El negro representa que dicho gen confiere resistencia a un cierto antibiótico

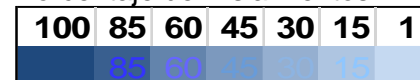
Porcentaje de Aislamientos



Mecanismo de Resistencia	Gen	Porcentaje	Localización	Aminoglicósido	Anfenicol	Carbapenémico	Cefalosporinas	Diaminopirimidin	Fluoroquinolonas	Fosfomicina	Glicilicilina	Penicilinas	Péptidos	Rifamicinas	Sulfonamidas	Tetraciclinas	Otros		
Inactivación del Antibiótico	ADC-82	100	Verde																
	OXA-338	100	Verde																
	OXA-58	100	Rosado																
	ADC-11	100	Verde																
	CTX-M-15	100	Rosado																
	CTX-M-115	100	Rosado																
	OXA-106	100	Verde																
	VEB-9	100	Morado																
	KPC-2	100	Rosado																
	OXA-113	100	Verde																
	OXA-317	100	Verde																
	OXA-371	100	Verde																
	ADC-7	100	Verde																
	OXA-78	100	Verde																
	OXA-217	100	Verde																
	OXA-172	100	Verde																
	ADC-43	100	Verde																
	OXA-263	100	Verde																
	CARB-3	100	Rosado																
AAC(3)-IId	100	Rosado																	

- El porcentaje se encuentra representado en escala de color. Localización plasmídica: verde, localización cromosomal: rosado, no reportado en *A. baumannii*: morado. El negro representa que dicho gen confiere resistencia a un cierto antibiótico

Porcentaje de Aislamientos



Los aminoglicósidos son una clase de antibióticos que han sido ampliamente utilizados para tratar infecciones ocasionadas por esta bacteria, no obstante su eficacia se ha visto reducida por el surgimiento y diseminación de la resistencia. Las enzimas modificadoras de aminoglicósidos (AMEs por sus siglas en inglés) son el principal factor responsable del incremento y diseminación de la resistencia a estos antibióticos (Hasani et al., 2016; Sheikhalizadeh et al., 2017), y se clasifican como acetiltransferasas (AACs), nucleotidiltransferasas (ANTs) y fosfotransferasas (APHs) (Yakkala et al., 2019). Entre los genes que codifican AMEs encontrados en esta investigación, *aph(3'')-Ib* y *aph(6)-Id* fueron los genes más predominantes con una tasa de distribución de 60% (n=203) y 57% (n=198); respectivamente (tabla 3-2). Cuando comparamos la ocurrencia de estos genes con los diversos fenotipos en el presente estudio, la presencia de *aph(3'')-Ib* y *aph(6)-Id* fue

significativamente alta en los aislamientos resistentes a la amikacina (58 de 71) y (57 de 71), lo cual concuerda con estudios previos que mencionan la amplia distribución de las fosfotransferasas en *A. baumannii* y su relación con tasas elevadas de resistencia a aminoglicósidos, en particular a la amikacina (Moniri et al., 2010; Sheikhalizadeh et al., 2017). También se pudo observar que existe una co-ocurrencia de estos genes asociados a fenotipos resistentes e intermedios, sin embargo no se pudo determinar si existe alguna importancia clínica ya que las MICs no presentaron ninguna diferencia significativa al compararlas con aquellas que no albergaban estos genes. Las posibles relaciones encontradas serán presentadas más adelante a partir de los resultados obtenidos por los algoritmos de aprendizaje de máquina que se aplicaron a los datos, lo cual es una de las ventajas que tiene el uso del aprendizaje de máquina (Deelder et al., 2019). En cuanto a la AAC(6')-Ia, si bien estaba presente únicamente en el 16% (n=58) de los aislamientos (tabla 3-2), fue la tercera AMEs más frecuente en los aislamientos resistentes (34 de 71). Estudios anteriores, al igual que nosotros han detectado la alta prevalencia de este gen asociado a fenotipos resistentes a la amikacina (Troyano-Hernández et al., 2018).

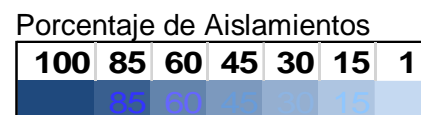
Otro mecanismo de resistencia evaluado fue alteración del sitio blanco para el cual únicamente se detectaron 9 genes, de los cuales 5 están involucrados en resistencia a fluoroquinolonas (tabla 3-3). Se pudo observar que todos los genes asociados a este mecanismo de resistencia se encuentran codificados cromosómicamente, habiendo sido reportado que las variaciones en la resistencia debida a ellos viene dada por mutaciones espontáneas que ocurren en respuesta al estrés antibiótico (Geisinger et al., 2019; Park et al., 2011). En este sentido, los niveles altos de resistencia a fluoroquinolonas en *A. baumannii* son debidos a mutaciones en los genes *gyrA*, *gyrB*, *parC* y *parE* (Park et al., 2011). En este trabajo se encontró que las mutaciones en *gyrB* fueron las más frecuentes, presente en el 93% (n=319) de los aislamientos, aunque no se encontró correlación entre los fenotipos resistentes a la ciprofloxacina y dichas mutaciones, ya que estuvieron presentes en aislamientos tanto sensibles como resistentes. Investigaciones previas han demostrado que la mayoría de las mutaciones reportadas en *gyrB* están asociadas a susceptibilidad a fluoroquinolonas, excepto a ácido nalidíxico (Fábrega et al., 2009), lo que explicaría la presencia de estas mutaciones incluso en aislamientos sensibles a ciprofloxacina. Por otro lado, se encontró que todos los aislamientos resistentes a ciprofloxacina albergaban mutaciones en *gyrA*, a diferencia de los aislamientos sensibles en donde no se encontró ninguna mutación importante, lo que concuerda con lo reportado

previamente por Warner (Warner et al., 2016), que sugiere que existe una asociación entre dichas mutaciones y la resistencia a ciprofloxacina. En cuanto a *parC* y *parE*, encontramos mutaciones únicamente en el 1%(n=4) y 0,5%(n=2) de los aislamientos, respectivamente. Además, es importante mencionar que dichas mutaciones estaban presentes solo en aislamientos resistentes. A pesar de que este resultado concuerda con lo reportado en la literatura, que reporta la correlación de las mutaciones en *parC* y la resistencia a fluoroquinolonas, (Geisinger et al., 2019; Honarmand Jahromy et al., 2020; Park et al., 2011), los resultados del presente estudio difieren en cuanto a la frecuencia, ya que ellos reportan una alta frecuencia de mutaciones en *parC* y en presente estudio se hallaron únicamente en 4 de los 129 aislamientos resistentes.

Tabla 3-3: Resistoma accesorio correspondiente a Alteración de la permeabilidad y del sitio blanco de los 343 aislamientos de *A. baumannii*

Mecanismo de Resistencia	Gen	Porcentaje	Localización	Aminoglicósido	Anfenicol	Carbapenémico	Cefalosporinas	Diaminopirimidinas	Fluoroquinolonas	Fosfomicina	Glicilciclina	Penicilinas	Péptidos	Rifamicinas	Sulfonamidas	Tetraciclinas	Otros
Alteración de la permeabilidad	tmrB																
Alteración del sitio blanco	gyrB																
	EF-Tu																
	gyrA																
	armA																
	ugd																
	Salmonella serovars gyrB																
	rpsL																
	parC																
	parE																

- El porcentaje se encuentra representado en escala de color. Localización plasmídica: verde, localización cromosomal: rosado, no reportado en *A. baumannii*: morado. El negro representa que dicho gen confiere resistencia a un cierto antibiótico. Gen que presenta dos copias: letra en rojo



Al momento de aplicar algoritmos basados en aprendizaje de máquina, disponer de variables informativas, es decir que su comportamiento sea causal de cambios en la variable dependiente, es crucial. (Géron, 2019). Se comprobó que las mutaciones en *gyrA* están fuertemente relacionadas con los fenotipos resistentes a ciprofloxacina en los aislamientos utilizados en esta investigación. Por lo que, podemos concluir que las mutaciones asociadas *gyrA* podrían ser el principal mecanismo de resistencia a ciprofloxacina, lo cual podría convertirlo en un excelente predictor a la hora de desarrollar modelos de predicción para este antibiótico.

Protección y reemplazo del sitio blanco fueron los mecanismos con menos cantidad de genes asociados, solo se encontraron: *msrE*, *sul1*, *sul2*, *dfrA1* y *dfrA19*, con una distribución variable como muestra la tabla 3-4. La mayoría de estos genes codifican mecanismos de resistencia a las sulfonamidas, a excepción de *msrE* cuyo mecanismo está relacionado con resistencia a las tetraciclinas. Un total de 73 aislamientos de *A. baumannii* fueron resistentes a trimetoprim-sulfametoxazol, donde los genes *sul1* y *sul2* se detectaron en el 99% (n= 72) de estos aislamientos, 68% (n=50) con *sul2*, 47% (n=35) con *sul1*, y únicamente el 13% (n=10) albergaban ambos. *sul1* y *sul2* estuvieron ausentes en todos los aislamientos sensibles. Como ya ha se reportado previamente, observamos el importante papel que juegan estos genes en resistencia a sulfonamidas (Hamidian et al., 2016; Khurshid et al., 2019), sin embargo nuestros hallazgos no concuerdan con otros resultados en donde mencionan que *sul1* es el gen más frecuente en *A. baumannii* (Khurshid et al., 2019). Tampoco se logró determinar si existen diferencias en las MICs dependiente de la *sul* que contienen, dado que los perfiles de resistencia fueron recolectados del *NCBI* en donde los valores de MIC para estos antibióticos no mostraron variabilidad.

Las bombas de eflujo, como previamente lo hemos mencionado juegan un papel importante en la resistencia de *A. baumannii*, y se encuentran ampliamente distribuidas en esta bacteria (Yakkala et al., 2019); lo que concuerda con nuestros resultados (tabla 3-5) donde se observa que los genes asociados a este mecanismo son los más frecuentes, principalmente relacionados con la resistencia a tetraciclinas. Se ha descrito que la resistencia a tetraciclinas viene dada principalmente por la adquisición de bombas de eflujo pertenecientes a la familia MFS como *tetA* y *tetB*, y a la familia (RND) tales como *adeABC*, *adeIJK*, *adeFGH*, *adeM*, *adeDE*. Cuando estos dos sistemas se hallan en asociación, se

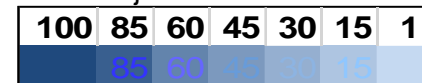
ha evidenciado que existe incremento en las MICs de las tetraciclinas (Khurshid et al., 2019; Ranjbar & Farahani, 2019), como se confirmó en este estudio, ya que *tetA* y *tetB* estuvieron presentes únicamente en aislamientos resistentes

Tabla 3-4: Resistoma accesorio correspondiente a protección y reemplazo del sitio blanco de los 343 aislamientos de *A. baumannii*

Mecanismo de Resistencia	Gen	Porcentaje	Localización	Aminoglicósido	Anfenicol	Carbapenémico	Cefalosporinas	Diaminopirimidin	Fluoroquinolonas	Fosfomicina	Glicilicilina	Penicilinas	Péptidos	Rifamicinas	Sulfonamidas	Tetraciclinas	Otros
Protección del sitio blanco	msrE	85	60		100											100	
Reemplazo del sitio blanco	<i>sul2</i>	85	60												100		
	<i>sul1</i>	85	45												100		
	dfrA1	85	60														
	dfrA19	85	60														

- El porcentaje se encuentra representado en escala de color. Localización plasmídica: verde, localización cromosomal: rosado, reportado en plásmido y cromosoma: amarillo. El negro representa que dicho gen confiere resistencia a un cierto antibiótico. Gen que presenta dos copias: letra en rojo

Porcentaje de Aislamientos

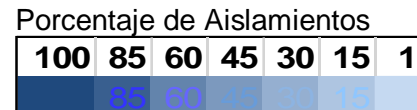


Por otra parte, el gen *tet(X3)* que se caracteriza por conferir altos niveles de resistencia a la tigeciclina (Cui et al., 2020), fue identificado con el 100% de identidad en 3 aislamientos, de los cuales 2 fueron recolectados por el Instituto Nacional de Salud de Colombia y si bien, los aislamientos que albergaban este gen eran susceptibles a tigeciclina, sus MICs estuvieron entre las más altas reportadas (1 µg/ml). Este gen se encuentra asociado a plásmidos y se ha aislado de muestras recolectadas de animales, humanos y del ambiente (Cao et al., 2020); en Colombia se ha encontrado esta variante únicamente en *A. nosocomialis* y *A. pittii* (He et al., 2019), y hasta lo actualmente revisado, el presente estudio es el primero en reportar la presencia de *tet(X3)* en aislamientos colombianos de *A. baumannii*. Este hallazgo se puede considerar como una alerta ya que el gen *tet(X3)* presente en plásmidos se ha convertido en la forma móvil más exitosa de diseminación de la resistencia a la tigeciclina (He et al., 2019), lo que podría restringir su uso.

Tabla 3-5: Resistoma accesorio correspondiente a bombas de eflujo de los 343 aislamientos de *A. baumannii*

Mecanismo de Resistencia	Gen	Porcentaje	Localización	Aminoglicósido	Anfenicol	Carbapenémico	Cefalosporinas	Diaminopirimidin	Fluoroquinolonas	Fosfomicina	Gliciliciclina	Penicilinas	Péptidos	Rifamicinas	Sulfonamidas	Tetraciclinas	Otros
Bombas de eflujo	adeR	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	adeB	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	adeS	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	adeA	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	AbaQ	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	soxR	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	AbaF	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	adeL	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	adeN	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	adeC	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	tetR	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	tet(B)	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	qacH	100	Rosado	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	cmlB1	100	Rosado	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	golS	100	Púrpura	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	mdsA	100	Púrpura	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	tet(D)	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	OprM	100	Púrpura	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	floR	100	Rosado	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	cmlA5	100	Rosado	■	■	■	■	■	■	■	■	■	■	■	■	■	■
	tet(A)	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■
tet(39)	100	Rosado	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
Escherichia coli emrE	100	Verde	■	■	■	■	■	■	■	■	■	■	■	■	■	■	

■ El porcentaje se encuentra representado en escala de color. Localización plasmídica: verde, localización cromosomal: rosado, reportado en plásmido y cromosoma: amarillo. El negro representa que dicho gen confiere resistencia a un cierto antibiótico. Gen que presenta dos copias: letra en rojo



Lo que encontramos en el presente estudio concuerda con otros hallazgos, que indican que *A. baumannii* tiene una amplia variedad de genes contribuyendo a la resistencia contra diversos grupos de antibióticos (Salih & Shafeek, 2019), y esto se debe particularmente a la capacidad por adquirir este tipo de determinantes (Honarmand Jahromy et al., 2020). Por lo que, se ha convertido en un desafío tratar de erradicar esta bacteria, dado el número

limitado de opciones terapéuticas efectivas que quedan disponibles. Además, con los resultados obtenidos podemos ver que existe una relación entre el genoma y el fenotipo, lo que nos permite avanzar al segundo paso que es la predicción de perfiles fenotípicos a partir de los datos genotípicos asociados a resistencia.

3.3.2 Determinación de las variaciones de los elementos asociados a resistencia

Con el fin de obtener datos más informativos y por todo lo mencionado anteriormente, se realizó la determinación de variaciones de los elementos asociados a resistencia identificados previamente. Para lo cual se empleó la secuencia de aminoácidos debido a que se ha descrito que el uso de la secuencia de nucleótidos podría disminuir la precisión de los modelos (Aytan-Aktug et al., 2020), además se ha demostrado que la concordancia con el perfil fenotípico está más relacionada con la información dada por las proteínas que el DNA (C. Chen et al., 2019).

Estudios anteriores han descrito que tener en consideración las variaciones de la secuencia de los genes asociados en resistencia mejoran sustancialmente los desempeños de los modelos de predicción (M. L. Chen et al., 2019; Deelder et al., 2019; Eyre et al., 2017), no obstante también se conoce que existe variaciones que no son causales de resistencia, lo que podría introducir ruido a la hora de entrenar estos modelos. Por tales motivos, se realizaron los 3 tipos de agrupaciones, las cuales ayudaran a responder esta interrogante, además que se espera que dependiendo del antibiótico que se esté modelando, resultara mejor una u otra agrupación.

Se identificaron y enumeraron todas las variantes de secuencia de aminoácidos (agrupación al 100%), observadas en cada gen encontrado para los 343 aislamientos de *A. baumannii* (tabla 3-6). El gen *adeS* es el que presentó mayor cantidad de alelos (n= 91), seguido de *adeB* (n=63), *Acinetobacter baumannii AmvA* (n=57), *MexK* (n=56), *adeH* (n=53) y *adeF* (n=47). Sin embargo, la mayoría de los genes (n=162) correspondían a β -lactamasas, y presentaron un solo alelo lo que sugiere que estos genes se encuentran bastante conservados en *A. baumannii*.

Tabla 3-6: Genes y alelos en el resistoma de los 343 aislamientos de *Acinetobacter baumannii* agrupados al 100%, 97% y 95% de identidad.

Agrupamiento	100%	97%	95%
Genes	192	192	192
Media de alelos	7	1	1
Máximo de alelos	91	10	6
Mínimo de alelos	1	1	1
Copias de genes	19	19	19

Como se había mencionado anteriormente también se realizaron dos agrupamientos más, una al 97% y otra al 95%; la tabla 3-6 muestra los resultados obtenidos. Se observó una reducción esperada en el número medio de alelos encontrados por gen, siendo este valor de 1 para los dos agrupamientos. El gen *adeS* mostro una reducción significativa al agruparle al 97% y 95%, con un número de alelos del 10 y 6 respectivamente, a diferencia del 100% en donde se encontraron 91 alelos.

La determinación de los genes de resistencia es un paso importante para la presente investigación, puesto que los mismos serán tomados como características de entrada del modelo (*input*). Por tales motivos es importante realizar este paso con la mayor precisión posible, ya que de esto depende la cantidad de ruido que se ingrese a los modelos y por ende al desempeño de los mismos (Géron, 2019). La base de datos CARD es una base curada, la cual se actualiza mensualmente, cuyas secuencias de referencia son de alta calidad y, con evidencia experimental de elevación de MICs (Alcock et al., 2020), por esta razón fue seleccionada para la identificación de los genes de resistencia.

4. Capítulo 4: Desarrollo del modelo de predicción del perfil de resistencia a partir de las secuencias del genoma de *Acinetobacter baumannii*

4.1 Introducción

Actualmente, la investigación biológica ha tenido un drástico cambio por el surgimiento de tecnologías genómicas de alto rendimiento; en consecuencia se ha evidenciado una creciente cantidad de datos los cuales por si solos resultan inútiles (X. Chen & Ishwaran, 2012; Obermeyer & Emanuel, 2016). Por tal motivo, la inteligencia artificial ha tomado un papel indispensable para el análisis de estos datos, los cuales podrían ayudar a encontrar respuestas a problemas relacionados con diversas ciencias como: astronomía, biomedicina, genómica, etc (Obermeyer & Emanuel, 2016).

Debido a lo desafiante que se ha vuelto el análisis de datos genómicos a gran escala y a la cantidad de fenómenos biológicos que aún resultan incomprendidos se han desarrollado métodos computacionales como aprendizaje de máquina que facilitan la obtención de datos más relevantes y significativos (Drouin et al., 2016). El aprendizaje de máquina permite analizar de manera más efectiva las variables genéticas y sus interacciones las cuales son más predictivas de rasgos fenotípicos complejos (Okser et al., 2014). Además, presenta múltiples ventajas frente a otros modelos que se han venido realizando como algoritmos basado en reglas, cuya limitante es que no puede aplicarse para detectar mecanismos moleculares que no hayan sido caracterizados previamente (Moradigaravand et al., 2018). El aprendizaje de máquina tiene como objetivo construir un modelo genotipo-fenotipo que permita reconocer patrones genéticos a partir de un conjunto de entrenamiento, logrando predecir de manera precisa nuevos casos con antecedentes genéticos similares (Drouin et al., 2016). Tiene varias aplicaciones a datos genómicos

como: predicción, selección de variables, análisis de vías, asociación genética y detección de epistasis (X. Chen & Ishwaran, 2012; Okser et al., 2014).

Se ha observado que el aprendizaje de máquina ayuda claramente en tres áreas de la medicina como: pronóstico, interpretación de imágenes y diagnóstico, por lo que su uso se ha vuelto indispensable para resolver problemas complejos en esta ciencia. Además, la creciente disponibilidad de genomas y herramientas bioinformáticas, en conjunto con la disminución de costos de secuenciación, permiten que el desarrollo e implementación de esta herramienta sea cada vez más accesible (Jeukens et al., 2019; Obermeyer & Emanuel, 2016).

Debido a las grandes ventajas que estos modelos proporcionan como método de diagnóstico, y a la urgencia de encontrar métodos alternativos a las pruebas convencionales de laboratorio para combatir la resistencia, varias investigaciones se han centrado en construir modelos de aprendizaje de máquina que permitan la predicción de perfiles de resistencia a partir de datos de secuenciación de genoma completo (Hicks et al., 2019). Estos estudios han estado enfocados en una variedad de bacterias especialmente aquellas que causan multi-resistencia adquirida en los hospitales, tales como *N. gonorrhoeae*, *Staphylococcus aureus*, *Enterobacteriaceae*, *Mycobacterium tuberculosis*, *Pseudomonas aeruginosa*, *Streptococcus pneumoniae*, *Klebsiella pneumoniae* entre otras (M. L. Chen et al., 2019; Davis et al., 2016; Eyre et al., 2017; Jeukens et al., 2019; Y. Li et al., 2017; Macesic et al., 2017b; Nguyen, Brettin, et al., 2018).

La construcción de estos modelos tiene tres etapas: entrenamiento, validación y prueba, y varios parámetros (características de entrada, forma de validación, número de muestras de entrenamiento, etc.) a considerar ya que el desempeño del modelo es dependiente de estos (Hicks et al., 2019; Nguyen, Long, et al., 2018).

Existen dos métodos principales de aprendizaje de máquina: aprendizaje supervisado y no supervisado (Xu & Jackson, 2019). El aprendizaje supervisado consiste en aprender relaciones entre un conjunto de datos empleados como ejemplos; incluyendo las soluciones deseadas, llamadas etiquetas (*labels*) (Géron, 2019). Por el contrario, el no supervisado infiere patrones a partir de datos sin etiquetas conocidas (Schridder & Kern, 2018; Xu & Jackson, 2019). Existen diversos métodos de aprendizaje de máquina supervisados que han resultado muy útiles en genómica, particularmente en predicción de resistencia, como por ejemplo: árboles de decisión, *support vector machine* (SVM), redes

neuronales, aprendizaje basado en reglas, regresiones, entre otros (Maglogiannis, 2007; Xu & Jackson, 2019).

El uso de algoritmos basados en reglas ha demostrado tener un eficiente desempeño en la predicción de fenotipos resistentes en bacterias a partir de genotipo, no solamente por la exactitud con la que realizan la predicción, sino también por ser altamente interpretables, ya que estos algoritmos se basan en un proceso de decisiones que son validados por expertos en el tema con el que se trabaja (Drouin et al., 2019). Sin embargo, su efectividad está limitada a que tan amplio es el conocimiento de los elementos genómicos mediadores de resistencia, es decir si no se encuentra bien caracterizado a nivel molecular los mecanismos de resistencia, entonces el desempeño del modelo será deficiente (Moradigaravand et al., 2018). Por otro lado su actualización resulta compleja, ya que con el creciente descubrimiento de nuevos genes de resistencia el modelo requiere incorporar nuevas reglas. Estos modelos no permiten predecir medidas cuantitativas como concentración mínima inhibitoria, las cuales nos permiten tomar decisiones más acertadas en el tratamiento (Pesesky et al., 2016).

En contraste, existen otros algoritmos que nos permiten superar algunas de las limitantes presentadas anteriormente por los métodos basados en reglas, en particular que identifican patrones del conjunto de datos de entrenamiento independientemente de que estos hayan sido o no caracterizados previamente (Pesesky et al., 2016). Hasta la fecha, varios algoritmos de este grupo se han empleado en el desarrollo de modelos predictivos de resistencia en bacterias, tales como: regresión logística, regresión lineal, redes neuronales, *support vector machine* y árboles de decisión, que incluyen *random forest*, *Set Covering Machines (SCM)* y *extreme gradient boosting (XGBoost)* (Hicks et al., 2019; Jeukens et al., 2019; Moradigaravand et al., 2018; Nguyen, Long, et al., 2018; Pesesky et al., 2016).

Las redes neuronales son modelos de aprendizaje de máquina profundo que permiten explicar relaciones no lineales y complejas, es decir permiten modelar la epistasis; sin embargo su uso no ha mostrado mejoras significativas en el desempeño frente a los demás modelos; adicionalmente interpretarlos resulta difícil lo que limita aún más su aplicabilidad en predicción de resistencia (M. L. Chen et al., 2019; Macesic et al., 2017b; Moradigaravand et al., 2018). Por otro lado, tenemos los modelos basados en arboles de decisión los cuales son muy empleados ya que permiten manejar fácilmente muchos

predictores e interacciones de alto orden, dejando controlar la complejidad de los modelos (Brennan et al., 2019). Tanto redes neuronales como arboles de decisión, son técnicas ampliamente dependientes de la información dada por clasificadores pre-determinados. En redes neuronales la clasificación es usada para determinar el error de la red y luego ajustar la red para minimizar este error. Por el contrario, en árboles de decisión la clasificación es usada para determinar que atributos proveen la mejor información que pueda ser empleada para resolver los problemas de clasificación (Ayodele, 2010).

Random forest es un método de clasificación basado en conjunto de cientos a miles de árboles los cuales son construidos utilizando particiones recursivas (Breiman, 2001; X. Chen & Ishwaran, 2012). Estos árboles explican la variación de una única variable de respuesta por una o más variables explicativas o predictoras; tiene la capacidad de explicar la correlación así como las interacciones entre las variables predictoras. Dependiendo del tipo de variable de respuesta los árboles pueden ser de clasificación o regresión, cuando esta sea categórica o numérica, respectivamente (De'ath & Fabricius, 2000). Este modelo resulta muy útil en genómica debido a la facilidad que proporciona para interpretar respuestas y a la flexibilidad para emplear variables de respuesta tanto numéricas como categóricas (X. Chen & Ishwaran, 2012; De'ath & Fabricius, 2000; Moradigaravand et al., 2018). En cuanto a predicción de resistencia, *random forest* se ha empleado para construir modelos de clasificación binaria, multiclase y de regresión (Hicks et al., 2019). Se ha afirmado que el uso de estos clasificadores multiclase y de regresión para predicción de la MIC en bacterias resulta muy útil porque permite obtener datos más relevantes que son necesarios a la hora de diseñar un tratamiento (Eyre et al., 2017; Hicks et al., 2019). Sin embargo, las tasas de coincidencia entre las MICs obtenidas en el laboratorio y las predichas son muy bajas, lo que podría deberse que hay pocos reportes y de variada calidad de las MICs de los microorganismos resistentes que han sido secuenciados hasta ahora; por lo que es necesario la determinación de la MIC con alta precisión para asegurar un modelo de alta calidad que permita relacionar este fenotipo con las características genómicas de cada microorganismo (Hicks et al., 2019; Nguyen, Long, et al., 2018). Hasta lo actualmente revisado, la mayoría de los estudios han empleado datos de fenotipo binario, sensible o resistente, no obstante se ha demostrado que esta categorización podría sesgar la realidad, ya que se ha evidenciado que aislamientos clasificados como sensibles pero con MICs altas, están asociados con riesgo de fracaso terapéutico, en comparación con MICs bajas. Además, los puntos de corte son establecidos por diferentes organizaciones, por lo que se generan diferencias en los puntos de corte que influyen en

la clasificación de la sensibilidad frente a un antibiótico afectando el rendimiento de los modelos predictivos, ya que estos datos son la base de su entrenamiento (Hicks et al., 2019; Y. Li et al., 2017; Moradigaravand et al., 2018).

Uno de los desafíos en el desarrollo de estos modelos de predicción es el sobreajuste, es decir el modelo se ajusta perfectamente a los datos de entrenamiento, pero su desempeño es pobre cuando se ingresan nuevos datos. Este hecho sucede cuando el tamaño de la muestra (número de genomas) es menor que el número de características (Drouin et al., 2016, 2019). El algoritmo *set covering machine* (SCM por sus siglas en inglés) se ha considerado prometedor a la hora de superar esta limitante y consiste en reducir lo más posible el número de características, seleccionando un número representativo del total de características disponibles en el modelo, a través de combinaciones lógicas que pueden ser una disyunción o conjunción de reglas producidas por el algoritmo. (Drouin et al., 2016; Marchand et al., 2003). En predicciones de fenotipo a partir de genomas, un modelo de conjunción asigna un genoma a una clase positiva siempre y cuando todas las reglas den como resultado verdadero; a diferencia del modelo de disyunción en donde para que un genoma sea asignado a una clase, al menos una regla debe dar verdadero (Drouin et al., 2016). Las limitantes a este modelo están asociadas a que maneja únicamente atributos booleanos (Marchand et al., 2003). A pesar de las desventajas que presenta el modelo, su desempeño para predecir resistencia en diferentes especies bacterianas ha mostrado ser equivalente o mejor que otros algoritmos como regresión logística, con exactitudes en un intervalo de 84% hasta 100%; además, sus resultados mostraron ser altamente interpretables y permitieron detectar nuevos mecanismos de resistencia que no se encontraban reportados (Drouin et al., 2019).

Por otro lado, la regresión logística ha sido actualmente muy empleada, y muchas investigaciones afirman que este algoritmo tiene el mejor desempeño, alcanzando valores del área bajo la curva (AUC) de 0,949, sensibilidad de 90.1% y especificidad de 90.5%, superando en muchos de los casos a *random forest* y *support vector machine* (Aun et al., 2018; M. L. Chen et al., 2019). Son clasificadores lineales que estiman la probabilidad de que un dato pertenezca a una clase (Drouin et al., 2019). Se han generado modelos de predicción de resistencia con este algoritmo que han mostrado varias ventajas, entre ellas, que admiten tanto fenotipos discretos como continuos y también, que han logrado discriminar las mutaciones causantes de resistencia de las mutaciones derivadas de la naturaleza clonal de las poblaciones bacterianas (Aun et al., 2018).

De todos los modelos desarrollados para predecir resistencia bacteriana, muy pocos se han enfocado en determinar fenotipos continuos, es decir predecir concentración mínima inhibitoria (Eyre et al., 2017; Y. Li et al., 2017; Nguyen, Brettin, et al., 2018; Nguyen, Long, et al., 2018). Los modelos que han predicho valores de concentración mínima inhibitoria han empleado dos algoritmos, regresión lineal y *gradient boosting*, con los cuales se han logrado desempeños muy efectivos, cuyos errores muy importantes (VME) y errores mayores (ME) fueron de 0,1% y < 3%, respectivamente; cumpliendo con lo establecido por la FDA para las herramientas de diagnóstico, incluso logrando predecir los valores del MIC en el 93% de las muestras, lo cual es un resultado muy alentador (Aun et al., 2018; Eyre et al., 2017; Hicks et al., 2019; Nguyen, Brettin, et al., 2018; Nguyen, Long, et al., 2018). Los ME ocurren cuando el modelo predice un aislamiento verdaderamente sensible como resistente, llevando a la restricción innecesaria de opciones terapéuticas. Los VME se ocasionan cuando el modelo predice un aislamiento verdaderamente resistente como susceptible, originando el fracaso terapéutico (Hombach et al., 2013).

Como se ha visto hasta el momento, la implementación de estos modelos ha resultado óptima ya que la exactitud con la que han realizado las predicciones ha alcanzado hasta el 100% y no ha bajado del 80% (Aun et al., 2018; M. L. Chen et al., 2019; Eyre et al., 2017; Nguyen, Long, et al., 2018; Pesesky et al., 2016). No obstante, algunos estudios han valorado la calidad de los modelos para predicción de un solo antibiótico o para varios antibióticos. Este último se le conoce como algoritmo multitarea, y se ha evidenciado que la exactitud mejora en 1-2% en estos modelos al compararlo con aquellos que predicen para solo un antibiótico, sin embargo para realizar un modelo multitarea es indispensable contar con miles de instancias, aumentando no solo la complejidad del modelo sino el gasto computacional (M. L. Chen et al., 2019; Nguyen, Long, et al., 2018).

Con todo lo mencionado, no es posible catalogar a un algoritmo como eficiente o deficiente, ya que su desempeño es dependiente de varios factores como: antibiótico, especie, métricas para valorar el desempeño, tipo de variable, muestreo, e incluso el tipo de predicción que realiza (Canovas-Segura et al., 2019; Hicks et al., 2019). El tamaño de los datos es crucial a la hora de desarrollar un modelo de predicción, se han informado buenos desempeños con tamaño de muestra (número genomas) mayores a 100. Sin embargo, lo que se debe considerar es tener balance en la muestra, lo que significa igual cantidad de fenotipos resistentes y sensibles, para garantizar precisión en los modelos (Davis et al., 2016; Hicks et al., 2019; Moradigaravand et al., 2018). También, se ha demostrado que la variabilidad del desempeño está determinada por las diferentes especies (Hicks et al.,

2019), y hasta la fecha se han reportado tres modelos de predicción de resistencia para *A. baumannii* frente a ciprofloxacina y a carbapenémicos, mediante *set covering machine*, *random forest* y *adaboost*, siendo estos de clasificación binaria (resistente o sensible), con una precisión de >90% (Davis et al., 2016; Giacobbe et al., 2020; Hicks et al., 2019). Puesto que hasta lo actualmente revisado únicamente se han realizado modelos de predicción abordándolos como tarea de clasificación, y dirigidos a dos antibióticos, el presente trabajo tiene como objetivo desarrollar modelos de predicción de resistencia para *A. baumannii* empleando varios algoritmos que realicen tareas de regresión, para determinar la concentración mínima inhibitoria (MIC) de 10 antibióticos, por lo que se desarrolló un modelo de predicción por cada antibiótico seleccionado, debido a que no se dispone de una cantidad de genomas suficientes con datos de MICs asociados, que permita realizar un solo modelo que realice la predicción de los 10 antibióticos al tiempo con una buena precisión.

4.2 Metodología

4.2.1 Pre-procesamiento de datos

Los datos de entrada (*features*) para cada modelo fueron los genes de resistencia con sus respectivas variaciones. Se realizaron matrices donde las filas representan los aislamientos y las columnas contienen las variaciones de los genes identificados previamente, considerándose cada gen como una característica del modelo (Tabla 4-1). Estas tablas fueron obtenidas mediante el script *tablaAlelos .py*.

Se realizaron 3 representaciones de los datos.

- La primera, en donde se marcaba con -1 si el gen no estaba presente, si estaba presente la matriz se llenaría con el número correspondiente de alelo detectado al realizar una agrupación al 100% de identidad
- La segunda, en donde se marcaba con -1 si el gen no estaba presente, y si estaba presente la matriz se llenaría con el número correspondiente de alelo detectado al realizar una agrupación al 97% de identidad
- La tercera, en donde se marcaba con -1 si el gen no estaba presente, si estaba presente la matriz se llenaría con el número correspondiente de alelo detectado al realizar una agrupación al 95% de identidad

Se desarrollaron varios modelos de predicción en donde se utilizaron cada una de estas tres representaciones por antibiótico previamente seleccionado.

Tabla 4-1: Estructura de los datos de entrada del modelo de aprendizaje de máquina

ID_aislamiento	adeF	OXA-23	adeR
INSRB-06	1	2	-1
INSRB-10	-1	1	1

Por otro lado, la salida (*output*) fueron valores que correspondían al valor de la MIC de cada antibiótico. Se ha determinado que la mayoría de los algoritmos de aprendizaje de máquina no funcionan bien con datos que tienen diferentes escalas, por tal motivo escalar los datos antes de entrenar los modelos es un paso importante (Géron, 2019). Por lo tanto se utilizó el log2 de las MICs obtenidas por microdilución en caldo y recolectadas del *NCBI*, para todas las tareas de aprendizaje de máquina. Es común que las MICs son expresadas utilizando los símbolos $>$, $<$, \geq y \leq , símbolos que se emplean para reportar aquellas MICs que se encuentran por encima o debajo de la dilución más alta o baja probada, respectivamente. Para eliminar estos símbolos de los datos como se ha establecido en trabajos previos (Eyre et al., 2017; Nguyen, Brettin, et al., 2018), si la MIC era $> x$, el valor de la MIC se cambió a $2x$, si era $< x$, la MIC se cambió a $\frac{x}{2}$, y si la MIC era $\leq x$ o $\geq x$, se eliminó el símbolo y el valor de la MIC permaneció sin cambios. Una vez obtenidos los valores predichos en escala log2, y se calculó la precisión de la predicción dentro de +/- una dilución doble. Para el entrenamiento de los modelos de predicción de la MIC de cada antibiótico, se utilizaron los genomas de *A. baumannii* colombianos más los que fueron descargados del *NCBI* que tenían asociado el dato de la MIC del respectivo antibiótico, por lo que el número de genomas empleados en el desarrollo de cada modelo es diferente.

Dado que las características de entrada se encontraban codificadas con números, los cuales indican la ausencia del gen o el alelo correspondiente, y que los modelos pueden considerarlas como una variable ordinal, es decir darle más peso al mayor número, lo cual sería un error (Raschka, 2015), se procedió a crear características *dummys* usando el método *get_dummies* implementado en la biblioteca pandas 1.0.5, el cual consiste en crear una variable binaria por cada categoría de la variable a convertir (Géron, 2019; Raschka,

2015). Por lo tanto, después de este paso la matriz contiene más características, y se encuentra codificada únicamente con valores de 1 y 0.

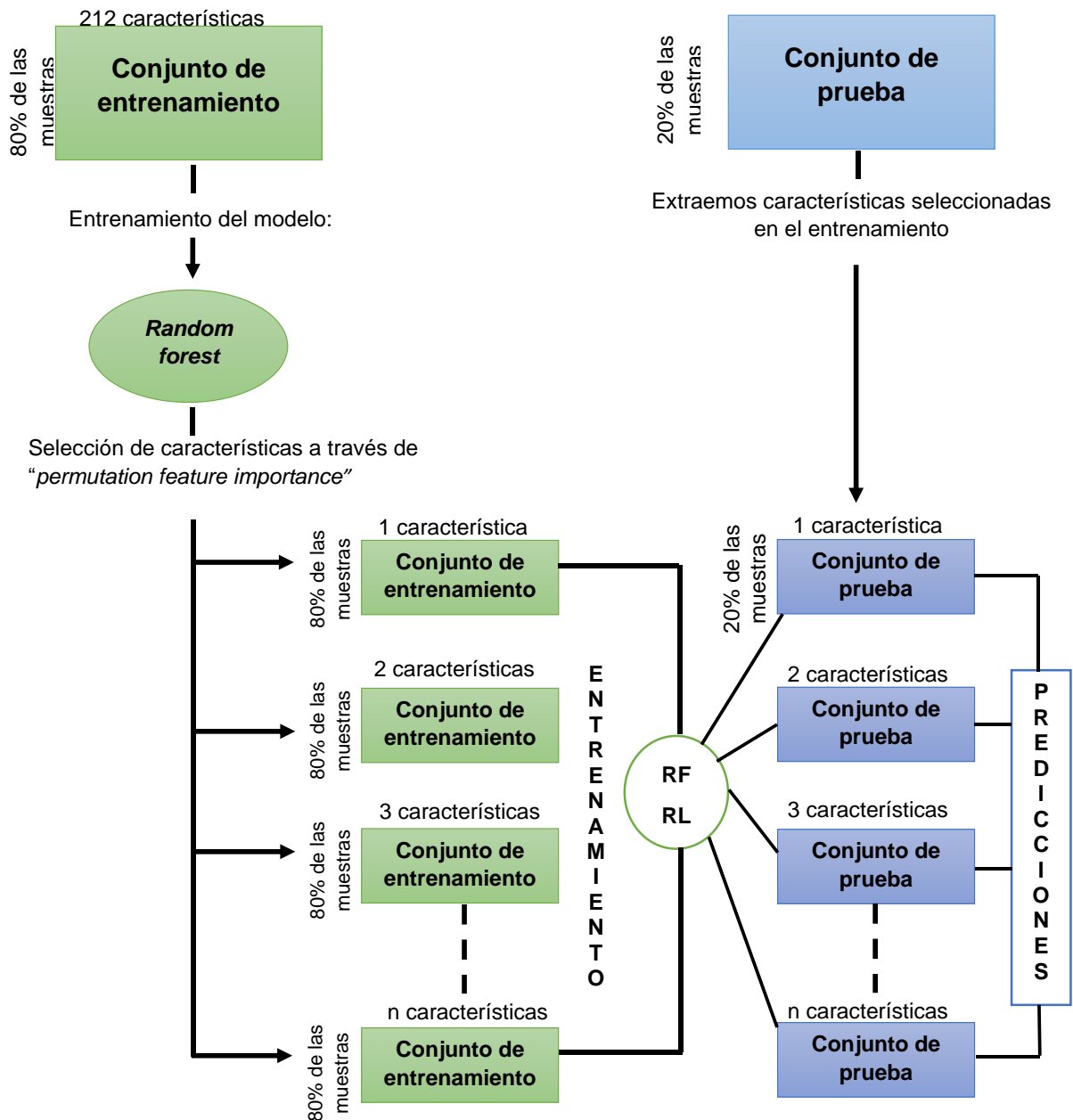
4.2.2 Características importantes (*feature importance*)

Con el fin de identificar los genes más importantes para predecir una MIC, fueron desarrollados modelos independientes para cada antibiótico (como se describió previamente). La identificación de las características predictoras se realizó utilizando *scikit-learn* v0.23.2 mediante dos aproximaciones. Se empleó una regularización penalizada con Lasso, (Géron, 2019; Nguyen, Brettin, et al., 2018), y por otro lado, se ordenaron en forma descendente las características según su importancia usando “*permutation feature importance*”, utilizando como estimador *random forest* y los parámetros por defecto. Esta función cambia de forma aleatoria los valores de las características, y evalúa el modelo; la disminución en el puntaje es indicativo de la importancia que representa dicha característica para el modelo (Pedregosa et al., 2011).

4.2.3 Construcción de modelos de aprendizaje de máquina

Se realizó la predicción de MICs para 10 diferentes antibióticos (amikacina, ceftazidima, cefepime, ciprofloxacina, colistina, fosfomicina, imipenem, meropenem, tetraciclina y tigeciclina), se utilizaron los algoritmos de aprendizaje de máquina: regresión *lasso*, *random forest* y *gradient boosting* disponibles en la librería *scikit-learn* de Python (Pedregosa et al., 2011). Se crearon 9 modelos por antibiótico, por cada antibiótico se utilizaron las tres matrices de entrada agrupadas al 100%, 97% y 95% de identidad, y con cada uno los diferentes algoritmos mencionados previamente, se emplearon todos los genes identificados. Para la evaluación de nuestros modelos se dividió el conjunto de datos en 80% para el entrenamiento y 20% para prueba. Para que los resultados fueran comparables se estableció durante la división un “*random state*”, función que permite obtener siempre el mismo conjunto de prueba (Géron, 2019; Martínez-Agüero et al., 2019; Pedregosa et al., 2011). La codificación *dummys* (ver sección de pre-procesamiento de datos) se realizó para el entrenamiento de los modelos realizados con regresión *lasso*. Los modelos basados en *random forest* y *gradient boosting*, se entrenaron con las matrices sin preprocesamiento, es decir se empleó todas las características de entrada (genes), sin previa selección. El ajuste de los hiperparámetros se detalla más adelante.

Figura 4-1: Esquema del estudio empleando selección de características



- Inicialmente se entrenó el modelo con todas las características descritas y se ordenaron en forma descendente según su importancia, la cual fue obtenida mediante "permutation feature importance" utilizando random forest. Posteriormente se creó un conjunto de datos con una característica, se entrenó tanto con random forest como con regresión lasso, realizamos predicciones y se evaluó. Se realizó el mismo procedimiento, pero esta vez con dos características, y así sucesivamente hasta probar todas las características que tuvieron un puntaje de importancia diferente de cero.
- RF = random forest, RL= regresión lasso, n = número de características seleccionadas por "permutation feature importance".

Por último, se realizó otro tipo de aproximación con el fin de evaluar el impacto que ejerce cada una de las características en el rendimiento del modelo (figura 4-1). Para esto se entrenó, se ajustaron los hiperparámetros y se evaluó cada modelo conforme se fueron agregando de una a una las características, empezando por la más importante de acuerdo con los resultados obtenidos usando *permutation feature importance* utilizando *random forest* y dejando fuera todas aquellas que tenían puntaje 0 (Anexo A-5). El número de modelos obtenidos depende del número de características que previamente han sido seleccionadas. Esta metodología fue aplicada únicamente para regresión *lasso* y *random forest*.

Todos los modelos fueron entrenados utilizando la raíz del error cuadrático medio (*RMSE* por sus siglas en inglés). Las curvas de aprendizaje también fueron determinadas para evaluar el sobreajuste de los modelos durante el entrenamiento y ajuste de los hiperparámetros (Géron, 2019; Raschka, 2015).

4.2.4 Ajuste de hiperparámetros

Los hiperparámetros de *random forest*, regresión lineal y *gradient boosting* fueron seleccionados y ajustados mediante *grid search*. Para evaluar la sensibilidad, estabilidad y precisión del modelo se utilizó *10-fold cross validation* (Géron, 2019; Pedregosa et al., 2011). Para regresión lineal se ajustó el parámetro *alpha* el cual permite controlar el grado de regularización, mientras mayor sea su valor más fuerte la regularización, es decir los coeficientes de un mayor número de características, las cuales son menos importantes (Géron, 2019), tenderán a ir a cero y no se los consideraría para el enteramiento del modelo (Raschka, 2015). Los hiperparámetros ajustados de *random forest* fueron: *maximum tree depth*, *minimum samples Split*, *minimum samples leaf* y *maximum features*. *Maximum tree depth* limita la profundidad que el árbol de decisión tendrá conforme se crea el bosque (*forest*) (Géron, 2019; Nguyen, Brettin, et al., 2018; Raschka, 2015). *Minimum samples Split* es el mínimo número de muestras que un nodo debe tener antes de poder dividirlo. *Minimum samples leaf* es el mínimo número de muestras que un nodo hoja debe tener. Por último, *maximum features* es el máximo número de características evaluadas para dividir cada nodo. En cuanto a *gradient boosting* se ajustaron los mismos hiperparámetros que *random forest*, además de *learning rate*. En la tabla 4-2 se resumen los valores probados para cada hiperparámetro seleccionado.

Tabla 4-2: Hiperparámetros ajustados en *random forest* y *gradient boosting*

Hiperparámetro	<i>Random forest</i>	<i>Gradient boosting</i>
<i>Learning_rate</i>	-	0.001, 0.01, 0.05, 0.1, 0.2
<i>Max_depth</i>	5, 6, 7, 8, 10	5, 6, 7, 8, 10
<i>Min_samples_split</i>	2, 3, 5, 6	2, 3, 5, 6
<i>Min_samples_leaf</i>	1,2,3	1, 2, 3
<i>Max_features</i>	"sqrt", "auto"0.1, 0.4, 0.5, 0.8,	'auto', 'sqrt', 0.4, 0.5, 0.2, 0.30, 0.6

4.2.5 Evaluación del modelo

Con el fin de seleccionar el modelo con mayor potencial diagnóstico, fueron utilizadas varias métricas para evaluar los modelos obtenidos. Por un lado, calculamos el *very major error* (VME) o falsos negativos, *major error* (ME) o falsos positivos, y la precisión dentro de +/- 1 dilución doble (ACC-2) con un intervalo de confianza del 95% para cada modelo, según lo establecido por la FDA (*Food and Drug Administration*) (Health, 2020) para los estándares de dispositivos de medición de MIC automatizados y consistente con las prácticas de microbiología clínica establecidas (Hicks et al., 2019; Nguyen, Brettin, et al., 2018; Nguyen, Long, et al., 2018; Pataki et al., 2019).

En los problemas de regresión es conveniente realizar las valoraciones empleando funciones de costo; es decir medir que tan malo es el modelo (Géron, 2019). La tabla 4-3 define cada una de las métricas seleccionadas en el presente estudio, y la fórmula empleada para su cálculo.

Tabla 4-3: Definición de los parámetros empleados para evaluación del modelo

Parámetro	Definición	Cálculo
VME	Cuando la muestra es resistente, pero se predice como sensible	$\%FN = \frac{FN}{FN + VN} \times 100$
ME	Cuando la muestra es sensible, pero se predice como resistente	$\%FP = \frac{FP}{FP + VP} \times 100$
ACC-2	Fracción de las muestras cuyos valores de MIC fueron correctamente predichos dentro de una dilución doble	$ACC - 2 = \frac{\text{valor predicho} \pm 1}{\text{valor real}} * 100$
MAE	Promedio del error absoluto entre el valor real y el predicho	$MAE = \frac{1}{m} \sum_{i=1}^m h(x^{(i)}) - y^i $
MSE	Promedio del error cuadrático entre el valor real y el predicho	$MSE = \frac{1}{m} \sum_{i=1}^m (h(x^{(i)}) - y^i) ^2$
RMSE	Promedio de la raíz del error cuadrático entre el valor real y el predicho	$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (h(x^{(i)}) - y^i) ^2}$

- VME: *very major error*, ME: *major error*, ACC-2: precisión dentro de +/- 1 dilución doble, MAE: error absoluto medio, MSE: error cuadrático medio, RMSE: raíz del error cuadrático medio
- FN: falsos negativos, VN: verdaderos negativos, FP: falsos positivos, VP: verdaderos positivos
- Valor predicho +/- 1: valor predicho correctamente dentro de +/- 1 dilución doble con relación al valor real.
- m = número de instancias, i = instancia en el dataset, $h(x^{(i)})$ =valor predicho, y^i = valor real (Géron, 2019; Health, 2020)

4.3 Resultados y discusión

4.3.1 Pre- procesamiento de datos

Dado que el objetivo del presente estudio es el desarrollo de modelos de predicción, los datos fueron organizados de tal manera que los algoritmos seleccionados fueran capaces de interpretarlos. Por lo cual cada uno de los genomas fue codificado en función de la presencia/ausencia de los alelos en cada uno de los genes de resistencia detectados; estos alelos fueron codificados con números, siendo el gen *adeS* el que tuvo el mayor número de alelos (91), cuando cada grupo de secuencias de proteínas pertenecientes a un mismo gen, fue subdividido en subgrupos de secuencias con el 100% de identidad entre ellas, lo que significa que para dicho gen se encontraron 91 secuencias diferentes (tabla 3-6).

Posterior a esto se generaron diversas matrices, las cuales fueron resultado del script previamente descrito que fue ejecutado en un sistema operativo Linux, las matrices fueron generadas en formato .csv (valores separados por comas). Como se indicó previamente, los genes fueron agrupados en 3 diferentes porcentajes de identidad (100%, 97% y 95%) de la secuencia proteica. Además, al recolectar los datos fenotípicos desde el *NCBI* se observó que no siempre se reportan los mismos antibióticos por aislamiento, por lo que el número de aislamientos disponibles para generar la matriz para cada antibiótico fue diferente (tabla 4-4). Según esto, en total las matrices generadas fueron 30, 10 matrices correspondientes al número de antibióticos probados por las 3 agrupaciones utilizadas. Los antibióticos fueron seleccionados en función de la variabilidad de sus MICs, y de la disponibilidad de estos a la hora de recolectar los datos.

El objetivo de estos modelos es lograr generalizar bien, por lo tanto es importante que los datos de entrenamiento sean representativos de los casos que se quiere predecir (Géron, 2019). En este caso es importante conseguir datos balanceados es decir, contar con genomas que sean representativos de los tres fenotipos: sensible, intermedio y resistente. Por todo esto, la recolección de datos públicos fue indispensable para el presente estudio, pues inicialmente contábamos únicamente con 76 aislamientos, de los cuales la mayoría presentaba fenotipo resistente. Como muestra la tabla 4-4, de cierto modo se logró balancear los datos, excepto para la fosfomicina, la colistina y tigeciclina, lo cual afectará en el desempeño de dichos modelos, pues el hecho de que estos conjuntos se encuentren

sesgados hacia un fenotipo, dificulta al modelo encontrar esos patrones subyacentes que es lo que le permite realizar predicciones precisas (Géron, 2019)

Tabla 4-4: Resumen de las 343 muestras utilizadas para construir y evaluar modelos de predicción de resistencia y el número de genomas disponibles para cada antibiótico

Antibióticos	Sensible	Intermedio	Resistente	Total
Amikacina	144	127	71	342
Cefepime	110	108	125	343
Ceftazidima	111	134	97	342
Ciprofloxacina	85	129	129	343
Colistina	319	0	24	343
Fosfomicina	0	17	69	86
Imipenem	125	3	214	342
Meropenem	120	5	218	343
Tetraciclina	57	25	95	177
Tigeciclina	214	26	3	243

4.3.2 Características importantes (*feature importance*)

Hasta lo actualmente revisado, es el primer reporte en donde se utiliza *permutation feature importance* para obtener el puntaje de importancia de cada característica y reducir la dimensionalidad de los datos empleando *random forest*, el cual ordena en orden decreciente la lista de genes de resistencia conocidos en función de la importancia que el modelo determine. Este procedimiento se realizó por separado para cada uno de los 10 antibióticos seleccionados (amikacina, cefepime, ceftazidima, ciprofloxacina, colistina, fosfomicina, imipenem, meropenem, tetraciclina, tigeciclina), utilizando únicamente el conjunto de entrenamiento. Este ensayo se realizó con cada una de las matrices generadas (n=30). En la figura 4-2 se resumen las principales características seleccionadas por medio de esta metodología.

Figura 4-2: Principales genes seleccionados por la función “*permutation feature importance*”

AMIKACINA	2	3	1	5	6	4		21	10	7		
CEFEPIME	24	29	6	4	3	1		38	41	10	2	
CEFTAZIDIMA	15	4	14	3	16	1				2	9	36
CIPROFLOXACINO	9		13	23	4	2		40		3	1	38
COLISTINA	22	1	3	21	2	23			37	4		17
FOSFOMICINA	4	32	3	15	16					2	1	
IMPENEM	6	33	9	1	13	4				4	2	15
MEROPENEM	8		12	1	14	7				4	2	13
TETRACICLINA	9	12	16	17	4	7	10	2	3	5	1	
TIGECICLINA	9	11	1	18	36	3				2	21	33
	ACCs	ANTs	APHs	OXAs	β -lacta	<i>gyrA</i>	<i>tetA</i>	<i>tetB</i>	<i>tetR</i>	B.F	Ef-TU	<i>gyrB</i>

- Importancia de los genes para el desarrollo de resistencia en 10 antibióticos según *permutation feature importance*. Las características expuestas son genes individuales o grupos de genes. Se encuentran coloreados de acuerdo con el rango de importancia (**azul** = los 5 más importantes, **verde** = 6-20, **amarillo**= 21-40, **negro**= sin importancia).
- Se muestran solo las características más relevantes determinadas por “*permutation feature importance*”
- **ACCs**= acetiltransferasas, **ANTs**= nucleotidiltransferasas, **APHs**= fosfotransferasas, **OXAs**= oxacilinasas, **β -lacta**= β -lactamasas, **B-F**= bombas de eflujo, **Ef-TU**= *Escherichia coli*_EF-Tu_mutants_conferring_resistance_to_Pulvomycin

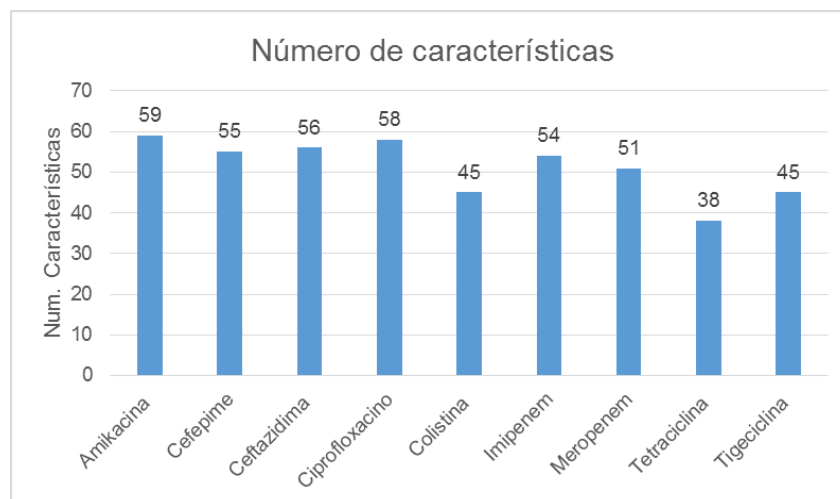
Se pudo observar que esta función es muy eficiente porque al determinar características importantes, para 7 de los 10 de antibióticos *permutation feature importance* pudo identificar dentro de las 5 características más predictivas, genes que se sabe están correlacionadas con la resistencia a cierto antibiótico. Por ejemplo, para amikacina las enzimas modificadoras de aminoglicósidos (ACCs, ANTs, APHs) fueron ubicadas en los tres primeros lugares (figura 4-2), mientras que para las cefalosporinas, carbapenémicos y las oxacilinasas, las β - lactamasas fueron las más predictivas (figura 4-2). El gen *gyrA* fue catalogado como la segunda característica más importante dada el papel que este cumple a la hora de desarrollar resistencia a ciprofloxacina.

Se ha evidenciado que existen deficiencias en la identificación de estas características cuando los datos de entrenamiento se encuentran imbalanceados, generalmente se detectan genes conocidos por conferir resistencia de manera más exitosa cuando el número de genomas resistentes y sensibles es equitativo (Hyun et al., 2020). Esto se vio reflejado en tigeciclina y colistina, en donde ninguno de los genes seleccionados por el

modelo como importantes está relacionados con la resistencia a estos antibióticos. La mayoría de los genomas que representaban estos antibióticos eran sensibles, y a pesar de los esfuerzos realizados por conseguir genomas resistentes, esto no fue posible (Van Camp et al., 2020), ya que al ser de último recurso, el uso de los mismos ha sido mejor controlado, lo que ha evitado el desarrollo de resistencia. Además, la elucidación de los mecanismos moleculares de resistencia a estos antibióticos aún está incompleta, lo que limita la identificación de los genes involucrados (Macesic et al., 2020). En cuanto a la fosfomicina, el número total de genomas empleado en el entrenamiento fue pequeño (n=86), lo que afectó no solo la identificación de características sino también el desempeño del modelo como veremos más adelante.

El número de características importantes seleccionado para cada antibiótico varió entre 38 y 59, lo que indicaría alrededor de 150 son poco informativas, lo que no quiere decir que estos genes estén contribuyendo o no a la resistencia, ya que es posible que por el bajo número de datos el modelo no los considere informativos durante el entrenamiento debido a que no capta un patrón que los relacione. La figura 4-3 muestra el número total de características obtenidas para cada uno de los diferentes modelos y en el anexo A-5 se detalla el nombre y la importancia de cada una de las características seleccionadas por antibiótico, siendo evidente que los genes que codifican para bombas de eflujo fueron seleccionados de manera recurrente dentro de las 10 primeras características predictivas para todos los antibióticos, lo que refleja el importante papel que juegan estos genes en la resistencia de *A. baumannii* como ya habíamos mencionado previamente.

Figura 4-3: Número de características seleccionadas por antibiótico



La selección de características mediante métodos “*wrapper*” como es *permutation feature importance*, ha mostrado ser una buena alternativa a la hora de trabajar con características genómicas (Hauskrecht et al., 2011), sin embargo se debe tener cuidado a la hora de interpretar estas características cuando se emplean modelos como *random forest*, dado que la elección de un gen como predictivo no necesariamente se debe a que está involucrado en la resistencia a dicho antibiótico, sino más bien puede deberse a la co-ocurrencia que tenga con un gen verdaderamente relevante involucrado en la resistencia (Nguyen, Brettin, et al., 2018; Nguyen, Long, et al., 2018; Van Camp et al., 2020). Por ejemplo esto explicaría que ciertas oxacilinasas (anexo A-5) estén catalogadas como predictivas para determinar la resistencia al aminoglicósido amikacina, ya que se ha comprobado una alta frecuencia de co-ocurrencia de este tipo de genes con genes de enzimas modificadoras de aminoglicósidos en *A. baumannii* (Nowak et al., 2014), los cuales generalmente son encontrados en islas de resistencia, plásmidos o integrones, que albergan simultáneamente estos dos tipos de genes de resistencia (Van Camp et al., 2020).

Como veremos más adelante (sección 4.3.4), la selección de características resulta ser un paso necesario en el aprendizaje de máquina, ya que nos permite obtener modelos con mejor desempeño. La identificación de los predictores más importantes resulta factible para los casos en que existen cientos a miles de características, como es nuestro caso (Pataki et al., 2019)

4.3.3 Construcción de modelos de aprendizaje de máquina

En el presente estudio entrenamos diversos modelos utilizando varios algoritmos como: regresión *lasso*, *random forest* y *gradient boosting*. Empleamos dos tipos de aproximaciones, la primera entrenamos los modelos utilizando todas las características de entrada, se crearon un total de 90 modelos (9 por antibiótico) (Anexo A-9, A-10 y A-11.), porque se aplicaron 3 algoritmos diferentes, cada uno fue entrenado con las 3 matrices (referentes al porcentaje de identidad aplicado 100%, 97%, y 95% para agrupar las secuencias de los genes), y esto se realizó por cada uno de los 10 antibióticos.

Como regresión *lasso*, *random forest* y *gradient boosting*, dependen de diferentes hiperparámetros, estos fueron seleccionados utilizando *gride search* (Pedregosa et al., 2011), los que aseguran el mejor rendimiento de acuerdo con la estrategia de validación

aplicada (*10-fold cross validation*) (anexos A-6, A-7 y A-8). Por otro lado, se realizó otra aproximación, en donde se obtuvieron diversos modelos por antibiótico conforme se aumentaba progresivamente las características, con el fin de determinar cuáles y cuantas son las características que nos permitan obtener el mejor modelo. En el anexo C-1 y C-2 se muestra las precisiones dentro de +/- una dilución doble de los diferentes modelos desarrollados con *random forest* y regresión *lasso* conforme se iba aumentando el número de características.

En la tabla 4-5 se presenta la precisión dentro de +/- 1 dilución doble del mejor modelo de cada antibiótico cuando utilizamos todas las características como entrada del modelo. La precisión +/- 1 dilución doble (ACC-2) media obtenida fue de 86%, 86% y 84% para regresión *lasso*, *random forest* y *gradient boosting*; respectivamente. Cuando hablamos de manera individual, los modelos de 6 de los 10 antibióticos obtuvieron una ACC-2 $\geq 0,90$ cuando se utilizó regresión *lasso*, mientras que el desempeño para cefepime y meropenem, cuando se empleó *random forest* y *gradient boosting*, disminuyó su precisión de 91% a 88% para meropenem, y de 92% a 87% para cefepime. Dos antibióticos: colistina y tigeciclina, tuvieron una ACC-2 88% con regresión *lasso*; 88% con *random forest*, y 80% con *gradient boosting*; no obstante, independiente del método ninguno de los desempeños estuvo por debajo del 80%, mientras que los desempeños más bajos fueron los de los modelos para la fosfomicina con un ACC-2 de 61% para los tres métodos.

Se sabe que la representación no uniforme del número de genomas con MIC para un determinado antibiótico influye en la precisión, de forma que cuando se tienen más genomas de aislamientos a los que se les ha determinado la MIC frente a un antibiótico los modelos tienden a tener precisiones más altas (Nguyen, Brettin, et al., 2018). Esto se puede corroborar en el presente trabajo ya que como en el caso de la fosfomicina y la tetraciclina, los ACC-2 fueron bajos, afectados por el número de datos disponibles: 86 y 177 respectivamente, siendo los conjuntos de datos más pequeños (tabla 4-4).

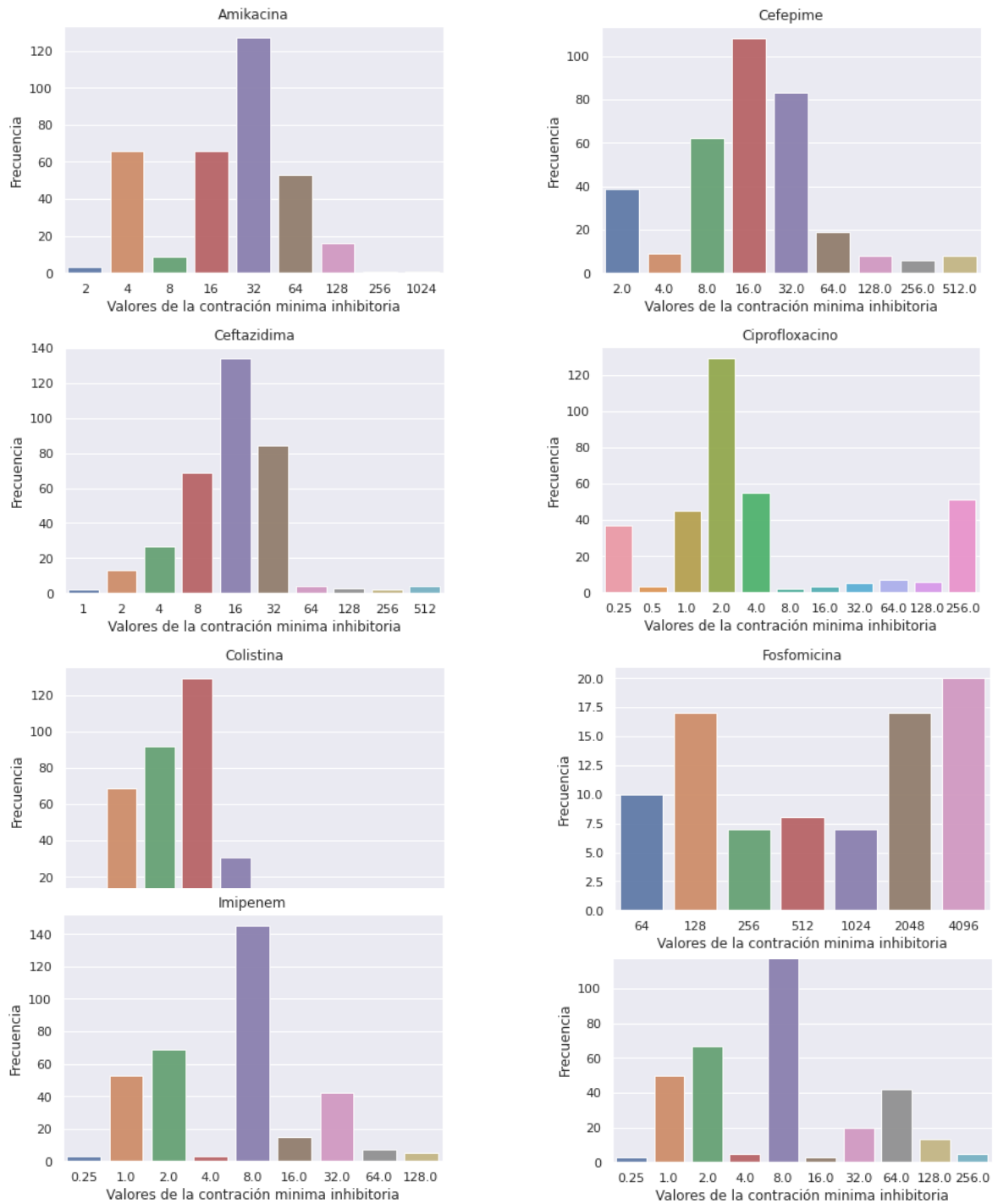
Tabla 4-5: Precisión de los modelos desarrollados con *random forest*, regresión *lasso* y *gradient boosting*

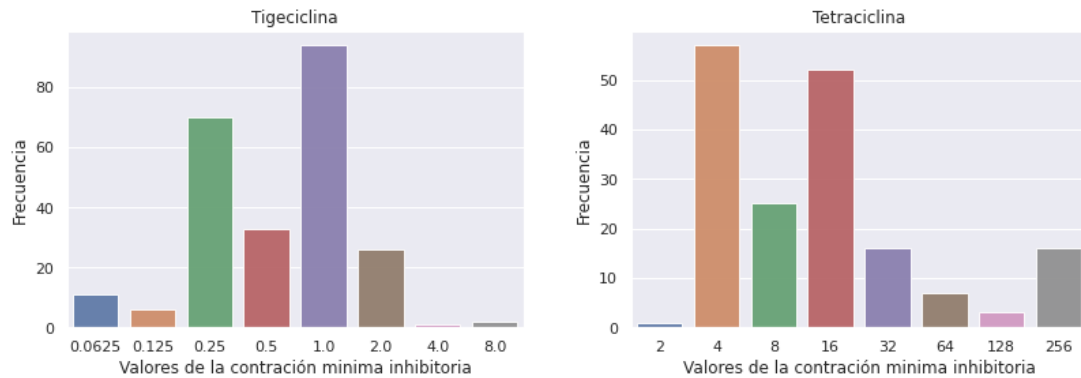
Modelo	RL		RF		GB	
	ACC-2	I.C 95%	ACC-2	I.C 95%	ACC-2	I.C 95%
Amikacina-97%	90	83- 97	91	84 -98	94	88 -100
Ceftazidima-95%	90	83 -97	90	83 -97	91	84 -98
Ciprofloxacina-95%	91	84 -98	94	88 -100	91	84 -98
Colistina-95%	88	80 -96	88	80 -96	80	71 -89
Cefepime-95%	92	86 -98	87	79 -95	80	71 -89
Fosfomicina-97%	61	50 -72	61	50 -72	61	50 -72
Imipenem-95%	94	88 -100	91	84 -98	87	79 -95
Meropenem-95%	91	84 -98	88	80 - 96	87	79 -95
Tetraciclina-100%	83	74 -92	83	74 -92	80	71 -89
Tigeciclina-95%	88	80 -96	88	80- 96	86	71 -89

- 95%, 97% y 100%: de agrupación
- **ACC-2**: precisión dentro de +/- dilución, **I.C 95%**: intervalo de confianza al 95%
- **RL**: regresión *lasso*, **RF**: *random forest*, **GB**: *gradient boosting*

Por otra parte, la precisión de los modelos se ve afectada por la forma como las MICs estén representadas, ya que, al momento de recolectar los valores de las MIC desde bases de datos con las del *NCBI*, estas se encontraban reportadas en algunos casos con símbolos como > o < entre otros, los cuales no permiten contar con el valor específico de la MIC, es decir que no se puede saber el valor de la MIC exacta por encima o debajo de la última dilución probada, afectando en general la precisión del modelo.

Figura 4-4: Distribución de los genomas por cada MIC y por antibiótico



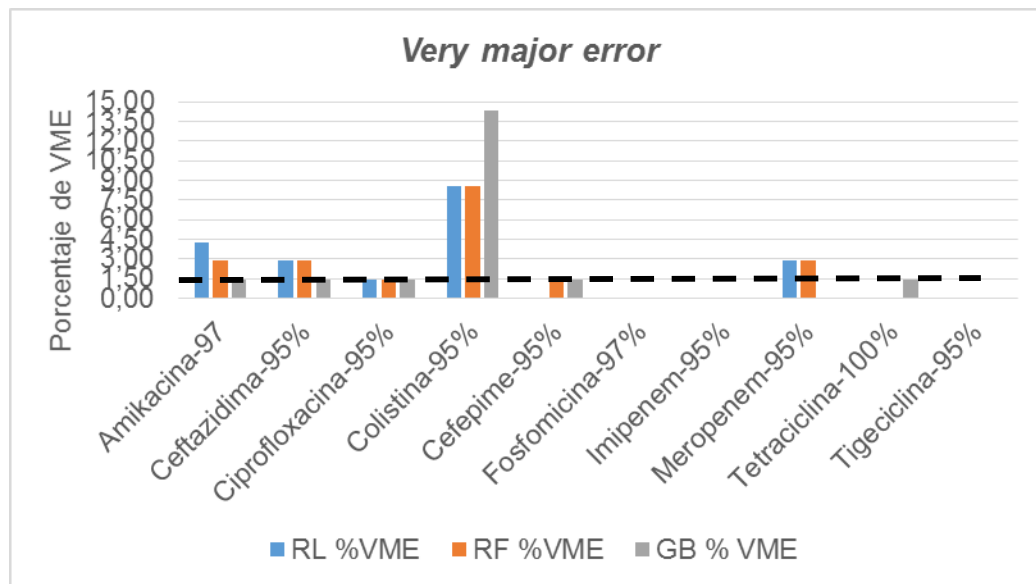


Como se puede observar (Figura 4-4), el número de genomas para cada valor de MIC se encontraban imbalanceados, y como ya se ha descrito, generalmente los modelos no lograron predecir correctamente los valores de MIC que estaban representados por pocos genomas.

El VME varió desde 0% hasta 8,57% para regresión *lasso* y *random forest*, y desde 0% hasta 14,29% para *gradient boosting*. El promedio de la tasa de VME para regresión *lasso*, *random forest* y *gradient boosting* fue 2%, el cual fue obtenido de los mejores modelos desarrollados (10 modelos por método, 1 por cada antibiótico, tabla 4-5). A excepción de la colistina, todos los antibióticos cumplieron con los lineamientos establecidos por el FDA para el VME ($\geq 1,5\%$) que permite aceptarlos como método de diagnóstico clínico (Health, 2020; Nguyen, Brettin, et al., 2018) (Figura 4-5). Los resultados presentados corresponden a los modelos que mejor desempeño obtuvieron, en los anexos A-9, A-10, A-1 se detalla el resultado para los 90 modelos desarrollados. El porcentaje de VME mas alto estuvo asociado a la colistina (figura 4-5), que fue de 8,57% con *random forest* y regresión *lasso*, y 14,29% con *gradient boosting*, lo cual posiblemente sea por la baja representación de genomas resistentes en el conjunto de datos (7%) (Figura 4-4). Por otra parte, la fosfomicina y la tetraciclina a pesar de tener las ACC-2 más bajas, 61% y 83% respectivamente, tienen un VME fue 0% (figura 4-5). Particularmente en el caso de la fosfomicina, esto podría haber ocurrido porque el conjunto de datos no contenía genomas sensibles, por lo que el modelo únicamente logro predecir correctamente la resistencia a este antibiótico. De igual manera sucede con la tetraciclina, ya que la mayor parte de los genomas son resistentes, y posiblemente a esto se deba su pequeña tasa de VME, y su ACC-2 baja podría deberse también a la posible falta de características genéticas discriminantes entre diluciones de MIC (Hicks et al., 2019), es decir que no hay diferencias en los genes de resistencia existentes en aislamientos que difieren en una sola dilución de

MIC. En general cuando se observan los resultados de los 90 modelos (30 por método), 15 modelos desarrollados con regresión *lasso* no cumplen con el límite permitido ($\geq 1,5\%$), sin embargo, en la mayoría de estos el error fue del 2,86% lo que representa en valores absolutos que solo 3 aislamientos fueron mal clasificados (anexo A-10). Con *random forest* 11 modelos obtuvieron tasas de errores superiores al 1,5% (anexo A-9), y con *gradient boosting* únicamente 5 modelos superaron este límite (anexo A-11), aunque los modelos desarrollados por este algoritmo resultaron con las precisiones (ACC-2) bajas con relación a los otros dos algoritmos empleados. Todos los modelos desarrollados para la colistina, independientemente de la agrupación o algoritmo utilizado, tuvieron un error por encima del límite permitido.

Figura 4-5: *Very major error*

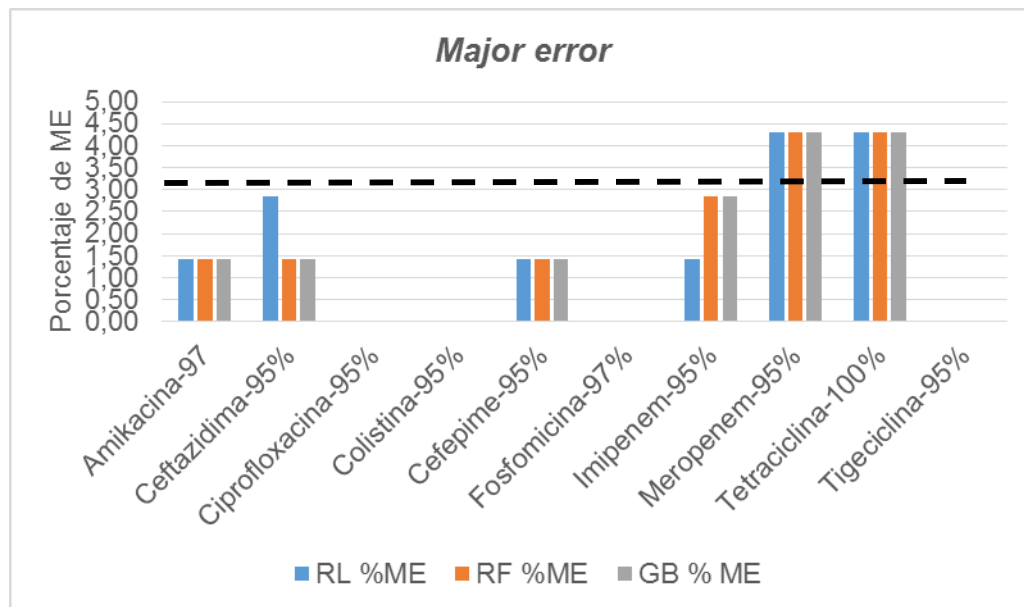


- 95%, 97% y 100%: % de agrupación
- **RL**: regresión *lasso*, **RF**: *random forest*, **GB**: *gradient boosting*
- - - - - Hace referencia al límite permitido por el FDA 1,5%

Con relación al ME el promedio para regresión *lasso*, *random forest* y *gradient boosting* fue de 2,15%, 1,86% y 1,57%, respectivamente. La tetraciclina fue el antibiótico que tuvo los valores más altos de ME (figura 4-6), 4,29% en los tres métodos evaluados, y a diferencia del VME la alta frecuencia de genomas resistentes (53,68%) en este caso son causantes de tasas elevadas de ME, este error hace referencia a la dificultad del modelo en clasificar un aislamiento como sensible, por lo que el modelo no pudo hacerlo correctamente ya que la cantidad limitante de genomas sensible dificulta el aprendizaje de esta clase. Según el FDA es necesario que el porcentaje de ME sea $\leq 3\%$ para poder

utilizarlo como método diagnóstico en la clínica (Health, 2020), y que es importante determinarlo puesto que este error tiende a restringir las opciones terapéuticas de una manera innecesaria, que privan al paciente de una terapia exitosa (Hombach et al., 2013). Al analizar los mejores modelos por antibiótico y por algoritmo, cuando empleamos regresión *lasso* 2 modelos correspondientes a ceftazidima y tetraciclina no cumplen con lo establecido por la FDA, debido a que las tasas de ME fueron de 4,29% para ambos casos. Datos similares se observaron cuando se utilizó *random forest*, con la diferencia que los modelos con error de 4,29% correspondieron al meropenem y tetraciclina. Finalmente, con *gradient boosting* únicamente el modelo desarrollado para tetraciclina fue el que no cumplió con el límite permitido por el FDA, cuyo error fue de 4,29%. De los 90 modelos desarrollados, 13 tienen errores superiores al 3%, 3 con regresión *lasso* (anexo A-10), 6 con *random forest* (anexo A-9), y 4 con *gradient boosting* (anexo A-11).

Figura 4-6: Major Error



- 95%, 97% y 100%: % de agrupación
- **RL**: regresión *lasso*, **RF**: *random forest*, **GB**: *gradient boosting*
- - - - - Hace referencia al límite permitido por el FDA 3%

En general, 7 modelos (amikacina, cefepime, ceftazidima, ciprofloxacina, fosfomicina, imipenem y tigeciclina) de los 10 desarrollados podrían ser aprobados para predecir la concentración mínima inhibitoria ya que cumplen requerimientos establecidos por el FDA (Health, 2020). De estos, la amikacina, cefepime, ceftazidima, ciprofloxacina e imipenem con ACC-2 > 90%, la mayoría de estos obtuvieron sus mejores desempeños con las

matrices agrupadas al 95%, lo que podría significar que muchas de las mutaciones en la secuencia de aminoácidos son silenciosas, es decir que no influyen en la resistencia (Aytan-Aktug et al., 2020; C. Chen et al., 2019). Por otra parte la fosfomicina y la tigeciclina con ACC-2 de 61% y 88%, respectivamente, tuvieron el mismo rendimiento independientemente del algoritmo utilizado para desarrollar el modelo. Los modelos que predicen la MIC de meropenem, tetraciclina y colistina, no cumplen con los límites de VME y ME, aunque como en el caso de la tetraciclina y el meropenem con un ME de 4,29%, este error se debe únicamente a un aislamiento mal categorizado. Estos errores son fácilmente solucionables hacia el futuro con la adición de nuevos datos que ayuden al balance adecuado de los mismos (Géron, 2019; Nguyen, Brettin, et al., 2018; Raschka, 2015). El modelo menos confiable es el de la colistina, debido a que no se pudo mejorar su rendimiento independientemente de las variaciones que se realizaron para optimizarlo, obteniéndose en el mejor de los modelos un VME de 8,57%, lo que implica que se requiere seguir ampliando el número de datos balanceados para obtener modelos que hagan una predicción confiable de la MIC para este antibiótico y reducir así la probabilidad de fracaso terapéutico (Hombach et al., 2013)

Por otra parte, se evaluó el efecto en los modelos de aprendizaje de máquina que tiene el utilizar las diferentes variantes o alelos que se producen al agrupar sus secuencias proteicas al 100%, 97% y 95% de identidad, observándose que aunque existen diferencias, están son mínimas en términos de precisión (~ 2 – 4%). Para los modelos realizados para la amikacina, se observa que el mejor rendimiento se obtuvo con la agrupación al 97%, teniendo ACC-2 de 90% y 94% con regresión *lasso* y *gradient boosting* respectivamente, mientras que con *random forest* el desempeño fue el mismo con las tres agrupaciones, 91%. En el caso de la fosfomicina la ACC-2 fue de 61% con las tres agrupaciones con los tres algoritmos, excepto con la agrupación del 95% cuando se empleó *gradient boosting* en donde el ACC-2 disminuyó al 55%. Los resultados muestran en general que al reducir los alelos, agrupando las variantes de la secuencia proteica de cada gen al 97% o 95%, los genes mantienen o mejoran su función y que estas agrupaciones permiten una discriminación suficiente por parte del modelo dando lugar a una predicción correcta. Por otra parte, la mayor diferencia se vio en la tetraciclina cuyas precisiones diferían con ~7% entre uno y otra (Anexo A- 9, A-10, A-11), obteniéndose el mejor modelo cuando se utiliza el conjunto de datos agrupados al 100% de identidad, posiblemente debido a que al

contrario a otros antibióticos las mutaciones en los genes asociados a esta resistencia influyen en su variación.

Un verdadero desafío al predecir resistencia a partir del genoma completo de una bacteria es la gran cantidad de genes (características) que contiene el conjunto de datos, lo que genera un problema que en aprendizaje de máquina se lo conoce como “dimensionalidad de los datos” (Géron, 2019; Macesic et al., 2020; Raschka, 2015). Una apropiada selección de características previa al entrenamiento es una solución para afrontar este problema. Considerando que la selección de características que tengan importancia predictiva mejoraría la precisión de los modelos, se utilizó “*permutation feature importance*” (Pedregosa et al., 2011). Esto se realizó únicamente con regresión *lasso* y *random forest*, debido a que a pesar de haber realizado diferentes ajustes de los hiperparámetros para *gradient boosting*, la mayoría de los modelos desarrollados por esta metodología tendían al sobreajuste. Esto puede ser debido a que el número de genomas que se empleó para el entrenamiento fue pequeño, y se ha visto que este algoritmo requiere más de 1000 genomas para hacer buenas predicciones (De Bruyne et al., 2020).

Se escogieron los mejores modelos en términos de ACC-2, VME y ME, en los cuales cada uno contiene diferente número de características puesto que se evaluó el efecto que causa cada una de ellas al ir las agregando de una en una al modelo (Anexo C-1 y C-2). En la tabla 4-6 se detallan los resultados obtenidos para los mejores modelos y el número de características de cada uno. El uso de los genes más importantes resultó en un incremento promedio de la ACC-2 de ~5% en el rendimiento de los modelos al compararlos con los modelos previamente descritos, los cuales fueron entrenados con 212 características simultáneamente. La diferencia más marcada se vio en el modelo desarrollado para predecir resistencia a fosfomicina para el cual el ACC-2 cambió de 61% a 78% empleando *random forest*, mostrando que los bajos ACC-2 debidos a un conjunto de datos pequeños se pueden mejorar con una adecuada selección de características como ha sido reportado previamente (Géron, 2019). En este caso de 212 genes inicialmente identificados y utilizados en la primera aproximación, se utilizaron únicamente 14 y 10 genes (características) para regresión *lasso* y *random forest*, y lográndose un aumento en la precisión del 7% y el 17% respectivamente. Posiblemente el menor desempeño con regresión *lasso* se deba a la dificultad que este algoritmo tiene para manejar patrones complejos (De Bruyne et al., 2020), como los asociados al conjuntos de datos relacionados con resistencia, pues un gen puede estar involucrado en la resistencia de varios

antibióticos, y a su vez varios genes asociados a la resistencia de un antibiótico (Alcock et al., 2020; Nguyen, Long, et al., 2018). Los modelos para la predicción de la MIC de tetraciclina, imipenem y meropenem, requirieron del menor número de características: 4, 8 y 8 respectivamente, cuando se empleó regresión *lasso*, lo que posiblemente se debe a que sus mecanismos de resistencia son menos complejos y posiblemente interviene un menor número de genes (Hicks et al., 2019; Macesic et al., 2020).

Tabla 4-6: Precisiones dentro de +/- una dilución doble y tasas de error de los modelos desarrollados con “*permutation feature importance*”

Modelo	RL						RF					
	# carac	%VME	%ME	ACC-2	I.C 95%		# caract	%VME	%ME	ACC-2	I.C 95%	
Amikacina-97%	67	2,86	1,43	94	99	89	40	0,00	1,43	96	91	100
Ceftazidima-95%	21	2,86	2,86	90	83	97	9	1,43	0,00	94	89	99
Ciprofloxacina-95%	15	1,43	0,00	96	91	100	46	1,43	0,00	96	91	100
Colistina-95%	16	8,57	0,00	88	80	96	25	8,57	0,00	90	83	97
Cefepime-95%	5	0,00	0,00	96	91	100	23	2,86	1,43	96	91	100
Fosfomicina-97%	14	0,00	0,00	67	56	78	10	0,00	0,00	78	68	88
Imipenem-95%	8	0,00	1,43	97	93	100	15	0,00	1,43	94	89	99
Meropenem-95%	8	2,86	2,86	91	84	98	17	1,43	2,86	90	83	97
Tetraciclina-100%	4	0,00	1,43	89	82	96	3	0,00	0,00	92	85	98
Tigeciclina -95%	17	0,00	0,00	88	80	95	4	0,00	0,00	88	80	95

- 95%, 97% y 100%: % de agrupación, **RL**: regresión *lasso*, **RF**: *random forest*, **ACC-2**: precisión dentro de +/- dilución, **I.C 95%**: intervalo de confianza al 95%, **VME**: *very major error* y **ME**: *major error*

En el otro extremo, el modelo para la predicción de la MIC de Amikacina fue el que requirió el mayor de características (67), posiblemente debido a que *A. baumannii* utiliza múltiples mecanismos resistencia frente a este antibiótico, y que los genes involucrados en dichos mecanismos se encontraban bien representados en el conjunto de datos de

entrenamiento, lo que hizo posible que el modelo logrará captar el efecto que ejerce cada uno de ellos en la resistencia, permitiendo incrementar el rendimiento del mismo (Géron, 2019; Hicks et al., 2019)

En general de acuerdo con los resultados obtenidos para los modelos utilizando *regresión lasso* y *random forest*, estos mejoraron después de la reducción de la dimensionalidad de los datos.

Considerando los mejores modelos por antibiótico al realizar una selección de características, el promedio del VME fue el mismo que cuando se utilizaron todos los genes como entrada del modelo para regresión *lasso*, 1,86%, sin embargo, en *random forest* disminuyó de 2% a 1,71%, cuando se realizó la selección de características. En cuanto al ME, éste disminuyó tanto para regresión *lasso* como para *random forest* de 2,15% a 1% y de 1,86% a 0,7%, respectivamente. Si bien es cierto las diferencias no son significativas, podemos observar que una adecuada selección de características previa al entrenamiento permite incrementar el rendimiento del modelo. El modelo para predecir la resistencia de *A. baumannii* a la colistina fue el que mayor VME presentó, 8,57%, seguido del cefepime con 2,86% (Tabla 4-6), se ha descrito que la causa principal para este tipo de errores es la falta de aislamientos susceptibles en el conjunto de datos, lo que conlleva a que el modelo no logre discriminar patrones que dan lugar a un aislamiento sensible o resistente (Martínez-Agüero et al., 2019; Nguyen, Brettin, et al., 2018; Nguyen, Long, et al., 2018). El resto de antibióticos estuvieron por debajo del límite permitido por el FDA para aprobar la implementación de un método diagnóstico (Health, 2020). Y en cuanto al ME todos los antibióticos obtuvieron errores por debajo de 3%. Estos resultados demuestran que es posible realizar con precisión predicciones de la concentración mínima inhibitoria a partir de las características genómicas relacionadas con los mecanismos de resistencia que utiliza *A. baumannii*

En términos generales *random forest* fue el algoritmo que permitió obtener los mejores modelos en combinación con la selección de características, modelos en los cuales 8 de los 10 obtenidos tuvieron una ACC-2 $\geq 0,90$ (tabla 4-6).

Hasta la fecha la secuenciación de genoma completo ha estado enfocada principalmente en bacterias resistentes, por lo que es difícil encontrar datos de MIC pequeñas asociadas a genomas de aislamientos sensibles a los diferentes antibióticos utilizados para tratar

infecciones causadas por de *A. baumannii* y otras bacterias. Adicional, la mayoría de genomas no están acompañados del perfil fenotípico de resistencia, y los que lo tienen, generalmente están restringidos a los antibióticos β -lactámicos, dejando fuera otros antibióticos poco comunes pero no menos importantes como por ejemplo fosfomicina y tetraciclina, lo cual afecta al desempeño de los modelos desarrollados para estos antibióticos. Todo esto trae como consecuencia que aunque haya muchos genomas disponibles, hay pocos con datos fenotípicos de resistencia. Sería óptimo recolectar un número balanceado de genomas que represente los diversos valores de MIC para mejorar la precisión de los modelos en general, y además disminuir las tasas de VME y ME. Los genomas empleados para el presente estudio fueron seleccionados de diferentes países, pudiéndose observar que se pueden obtener buenas precisiones independientemente de la localización geográfica de origen de los genomas. Adicionalmente, los resultados sugieren que se pueden lograr modelos precisos sin tener una gran cantidad de genomas, lo que permite optimizar el gasto computacional, disminuyendo notablemente el tiempo de entrenamiento.

La mayoría de los estudios reportados hasta el momento han estado enfocados en predecir las categorías resistente o sensible, mientras que el objetivo de este trabajo fue predecir la MIC, pues proporciona información más precisa del fenotipo. Se emplearon diversos algoritmos como: regresión *lasso*, *random forest* y *gradient boosting*. A pesar de que las diferencias en términos de ACC-2 no son significativas, se pudo observar que *random forest* es una buena alternativa a la hora de desarrollar modelos que realicen tareas de regresión, no obstante regresión *lasso* también resulta útil ya que es un algoritmo más sencillo, es decir al tener menos grados de libertad, el ajuste es bastante sencillo pues únicamente tiene un hiperparámetro que ajustar, y requiere menos tiempo de entrenamiento (De Bruyne et al., 2020; Géron, 2019; Raschka, 2015). Por otro parte, *gradient boosting* es un algoritmo prometedor para realizar tareas de regresión, no obstante su ajuste es catalogado como complejo, aunque ha resultado útil en conjuntos de datos extensos, pues una de sus ventajas es su capacidad por optimizar el tiempo de entrenamiento (De Bruyne et al., 2020; Géron, 2019; Raschka, 2015). Al aplicar este algoritmo en nuestro conjunto de datos, se obtuvieron buenas precisiones (anexo A-11) pero la mayoría de ellas tendían al sobreajuste-

En este trabajo se consideró que el uso de los genes era una buena opción para el desarrollo de estos modelos, es que al usar genes asociados a los mecanismos de resistencia los modelos que se obtienen son más interpretables, puesto que al observar los genes que más influyen en el modelo, se puede inferir el mecanismo o los mecanismos de resistencia más involucrados, ayudando al conocimiento básico que permitiría inferir posibles tratamientos a las infecciones causadas por bacterias como *A. baumannii*.

También se observó que la reducción de la dimensionalidad de los datos es un paso esencial para asegurar modelos de predicción de resistencia exitosos, y que “*permutation feature importance*” que emplea *random forest* como estimador hace esta etapa de una manera adecuada. Sin embargo, esta reducción de características no mejora significativamente los modelos en los casos donde los mecanismos de resistencia son complejos o desconocidos, o donde un mecanismo no se encuentre representando por un número considerable de genomas (Macesic et al., 2020), como en el caso de la colistina en el presente trabajo.

Es importante considerar las variaciones en la secuencia de los genes de resistencia utilizados como características del modelo, ya que estas tienen impacto en el desarrollo de modelos, especialmente para tigeciclina, en donde la mejor precisión se vio cuando se empleaba las agrupaciones al 100% de identidad. Resulta difícil realizar comparaciones de los rendimientos entre los antibióticos probados, pues el desempeño que cada uno de sus modelos es fuertemente dependiente de sus mecanismos de resistencia asociados, es decir si estos son complejos, diversos o que tan caracterizados se encuentren (Hicks et al., 2019; Van Camp et al., 2020).

La limitante de estudio en la mayoría de las investigaciones centradas en aprendizaje de máquina, es lo complejo que se vuelve la recolección de los datos, pues están limitadas a utilizar únicamente los datos que se encuentren disponibles en las bases de datos, y como se mencionó antes, generalmente solo se encuentran genomas de bacterias resistentes, lo que lleva contar con conjuntos de datos imbalanceados. Por otra parte, dado que los valores de las MICs fueron obtenidos en su gran mayoría asociados a los genomas de aislamientos *A. baumannii* descargados de la base de datos *Biosample* del NCBI y que estos provienen de diversos países, podría ser que se haya introducido un poco de ruido en los modelos, consecuencia de las posibles diferencias que existan en los protocolos aplicados para su obtención. Finalmente, a pesar de no utilizar secuencias de los genomas

completos de *A. baumannii* como datos de entrada para nuestro modelo, los resultados obtenidos aquí sugieren que los genes de resistencia proporcionan suficiente información para que el modelo pueda captar patrones y realizar predicciones acertadas.

Hasta el momento son pocos los modelos desarrollados para predecir MIC (Nguyen, Brettin, et al., 2018; Nguyen, Long, et al., 2018; Pataki et al., 2019), y los existentes generalmente utilizan miles de genomas. El presente trabajo propone una alternativa para desarrollar modelos de predicción de MIC precisos empleando conjuntos de datos más pequeños, permitiendo así no solo optimizar el tiempo de recolección de datos sino también disminuir los requerimientos computacionales para procesar dichos datos. Se observó que la secuenciación de genoma completo en combinación con el aprendizaje de máquina, pueden ser una alternativa prometedora que permita obtener estimaciones tempranas del antibiograma para *A. baumannii*, lo cual podría disminuir el tiempo necesario para iniciar una terapia antimicrobiana eficaz en pacientes infectados con este microorganismo.

El tratamiento óptimo para tratar infecciones por *A. baumannii* es en gran parte desconocido (Maasdorp et al., 2021), además debido a la alta tasa de cepas multiresistentes las opciones terapéuticas se tornan limitadas. Sin embargo, se ha mencionado que los principales antibióticos activos contra esta bacteria son: cefalosporinas de amplio espectro (ceftazidima y cefepime), carbapenémicos (imipenem y meropenem), fosfomicina y tigeciclina, y los aminoglicósidos (amikacina y gentamicina) en terapia combinada (Albornoz et al., 2018).

En el presente estudio se logró desarrollar modelos que realizan predicciones confiables para todos estos antibióticos, los cuales podrían ser aceptados por el FDA al cumplir con los lineamientos establecidos (tabla 4-7). Por otro lado, estudios han mostrado la existencia de *A. baumannii* sensible a ciprofloxacina (AL-Enawey et al., 2020; Barletta Farías et al., 2018), por lo que resulta de mucha utilidad realizar pruebas fenotípicas a este antibiótico ya que incrementa las opciones terapéuticas; el presente estudio pone a disponibilidad un modelo que realiza la predicción con ACC-2 del 96% (tabla 4-7).

Los modelos desarrollados en el presente trabajo que tuvieron el mejor desempeño se presentan a continuación:

Tabla 4-7: Modelos que cumplen con las limitaciones del FDA.

Antibiótico	Método	Agrupamiento	ACC-2	%VME	%ME
Amikacina	RF	97%	96%	0	1,43
Cefepime	RL	95%	96%	0	0
Ceftazidma	RF	95%	94%	1,43	0
Ciprofloxacina	RF	95%	96%	1,43	0
Fosfomicina	RF	97%	78%	0	0
Imipenem	RL	95%	97%	0	1,43
Meropenem	RF	95%	90%	1,43	2,86
Tetraciclina	RF	100%	92%	0	0
Tigeciclina	RF,RL	95%	88%	0	0

- **ACC-2:** precisión dentro de +/- 1 doble dilución, **%VME:** porcentaje de very major error, **%ME:** porcentaje de major error
- **RF:** random forest, **RL:** regresión lasso.
- Agrupamiento: corresponde al porcentaje de identidad con el que los genes se agruparon
- Tigeciclina independientemente del método empleado no tuvo diferencias

Estos pueden ser de gran utilidad para tratar las infecciones causadas por *A. baumannii*, ya que cubren la mayoría de antibióticos que usualmente se emplean para tratar estas infecciones, permitiendo obtener el valor de la MIC, lo cual resulta valioso, pues usualmente la selección del antibiótico en clínica se realiza en función de este valor, que permite evaluar que tan activo es cada uno de los antibióticos frente a un aislamiento específico de *A. baumannii*. (Albornoz et al., 2018)

4.3.4 Acceso a los datos

El código en lenguaje Python utilizando la librería scikit-learn para la generación de los modelos desarrollados para predecir la concentración mínima inhibitoria empleando los genes de resistencia encontrados en *A. baumannii*, se puede encontrar en el siguiente enlace: <https://github.com/kaquilarg/Predicci-n-del-perfil-de-resistencia-a-partir-de-las-secuencias-del-genoma-de-aislamientos-colombian>. En el mismo enlace se encuentra los modelos disponibles para hacer la predicción de la MIC a partir de un nuevo genoma,

adicionalmente encontraremos los scripts necesarios para generar las matrices de entrada, a partir del cual se realiza la predicción. El manual que detalla la manera de emplear los modelos generados en el presente estudio esta disponible en el link mencionado previamente.

5. Conclusiones y recomendaciones

5.1 Conclusiones

Se evaluó la efectividad de 21 antibióticos frente a 76 aislamientos de *A. baumannii* recolectados en Colombia encontrándose que ampicilina, cefazolina, cefuroxima, claritromicina, clindamicina, linezolid, nitrofurantoína, oxacilina, tianfenicol y vancomicina fueron ineficaces pues el 100% de los aislamientos resultaron resistentes.

Se evidenció una baja de tasa de susceptibilidad de los 76 aislamientos de *A. baumannii* recolectados en Colombia frente a amikacina (20%), cefepime (7%), ceftazidima (17%), ciprofloxacina (3%), imipenem (4%), meropenem (4%) y tetraciclina (3%). Únicamente, para la tigeciclina y colistina las tasas de susceptibilidad fueron altas, del 100% y 68%, respectivamente.

Se identificaron 192 genes involucrados en diferentes mecanismos de resistencia de *A. baumannii*, siendo el mecanismo más importante el de inactivación del antibiótico, con 124 genes asociados.

El gen VEB-9 cuyo mecanismo de resistencia corresponde a inactivación del antibiótico y reportado en cromosoma de *P. aeruginosa* fue encontrado en el aislamiento GMR-RA267 recolectado por el Instituto Nacional de Salud de Colombia, hasta lo actualmente revisado es reportado por primera vez en *A. baumannii* a nivel mundial. Igualmente, en los aislamientos colombianos de *A. baumannii* RB183-2 y GMR-RA267, fue encontrado por primera vez en Colombia el gen *tet(X3)*, bomba de eflujo, involucrado en la resistencia a tigeciclina.

Los genes de resistencia encontrados en los genomas de *A. baumannii* fueron útiles para para desarrollar modelos de predicción de resistencia precisos y tiempos de entrenamiento cortos.

La selección de genes predictivos utilizando *Permutation feature importance* incrementó el rendimiento de los modelos obtenidos en términos del ACC-2 en ~5% en comparación con los modelos realizados sin previa selección de características.

Los modelos utilizando *random forest* desarrollados para predecir la MIC de amikacina, ciprofloxacina, cefepime, ceftazidima, fosfomicina, imipenem, meropenem, tetraciclina y tigeciclina, pueden ser útiles como herramientas de diagnóstico de MIC de *A. baumannii* por que cumplen con las especificaciones de la FDA, VME $\leq 1,5\%$ y ME $\leq 3\%$.

A excepción de la tetraciclina que obtuvo los mejores resultados al emplear genes agrupados al 100% de identidad, los antibióticos restantes arrojaron mejores resultados con las agrupaciones de los genes al 95 y 97%.

Random forest resulto ser la mejor opción a la hora de desarrollar modelos de predicción de resistencia a partir de datos genómicos para *A. baumannii*, adicionalmente el uso de *gride search* permite seleccionar adecuadamente los hiperparámetros que eviten el sobreajuste. *Gradient boosting* resultó un algoritmo muy complejo para los 343 de genomas disponibles para el desarrollo del modelo, si bien sus ACC-2 fueron superiores al 90%, muchos de ellos se encontraban sobreajutados.

5.2 Recomendaciones

Dado que la principal limitante en este estudio fue conseguir genomas que representen valores pequeños de MIC, se recomienda que en posteriores estudios también se reporten genomas susceptibles, lo cual permitiría obtener datos balanceados.

Por otro lado, se recomienda realizar mayor seguimiento de resistencia de colistina y tigeciclina, dado que al ser de último recurso sería útil lograr predecir con precisión la resistencia a estos antibióticos, y así evitar el desarrollo de resistencia por el uso no adecuado de los mismos

También, se recomienda a los investigadores reportar los perfiles fenotípicos de resistencia, si es posible la concentración mínima inhibitoria, ya que cuando se realizó la búsqueda de los genomas muchos de ellos quedaron fuera del estudio debido a la falta de información fenotípica.

Desarrollar modelos de predicción de resistencia multiespecies basado en aprendizaje de máquina, dado que fuera útil contar con una herramienta de diagnóstico que logre predecir resistencia independientemente de la bacteria que estemos analizando

Y finalmente, se recomienda si es posible utilizar el aprendizaje de máquina para desarrollar modelos de predicción de resistencia a partir de muestras metagenómicas, lo que podría evitar la necesidad de subcultivos en caso de muestras contaminadas. Esto daría lugar a intervenciones tempranas con tratamientos acertados.

A. Anexo: Tablas

1. Puntos de corte considerados para la preparación de soluciones madre de antibióticos

Antibiótico	Especie de referencia	S	I	R	Referencia
Amikacina	<i>Acinetobacter spp.</i>	≤4	8	≥16	CLSI
Ampicilina	<i>Enterobacteriaceae</i>	≤8	16	≥32	CLSI
Cefazolina	<i>Enterobacteriaceae</i>	≤2	4	≥8	CLSI
Cefepime	<i>Acinetobacter spp.</i>	≤8	16	≥32	CLSI
Ceftazidima	<i>Acinetobacter spp.</i>	≤8	16	≥32	CLSI
Cefuroxima	<i>Enterobacteriaceae</i>	≤8	16	≥32	CLSI
Ciprofloxacina	<i>Acinetobacter spp.</i>	≤1	2	≥4	CLSI
Claritromicina	<i>Staphylococcus spp</i>	≤2	4	≥8	CLSI
Clindamicina	<i>Staphylococcus spp</i>	≤0,5	1-2	≥4	CLSI
Colistina	<i>Acinetobacter spp.</i>	≤2		≥4	CLSI
Fosfomicina	<i>Enterobacteriaceae</i>	≤64	128	≥256	CLSI
Imipenem	<i>Acinetobacter spp.</i>	≤2	4	≥8	CLSI
Linezolid	<i>Staphylococcus spp</i>	≤4		≥8	CLSI
Meropenem	<i>Acinetobacter spp.</i>	≤2	4	≥8	CLSI
Nitrofurantoína	<i>Enterobacteriaceae</i>	≤32	64	≥128	CLSI
Oxacilina	<i>Staphylococcus spp</i>	≤2		≥4	CLSI
Tetraciclina	<i>Acinetobacter spp.</i>	≤4	8	≥16	CLSI
Tianfenicol	<i>Enterobacteriaceae</i>	≤8	16	≥32	CLSI

Antibiótico	Especie de referencia	S	I	R	Referencia
Tigeciclina	<i>Acinetobacter spp.</i>	≤1	2	≥4	(Jiménez Pearson et al., 2019)
Trimetoprim-Sulfametoxazol	<i>Acinetobacter spp.</i>	≤2/38		≥4/76	CLSI
Vancomicina	<i>Staphylococcus spp</i>	≤2	4-8	≥16	CLSI

- Los valores correspondiente a los puntos de corte están en µg/ml
- S: sensible, I: intermedio, R: resistente

2. Absorbancias de los inóculos correspondientes a *A. baumannii*

Muestra	Absorbancia	Muestra	Absorbancia
RB993	0,135	RB15	0,246
RB992	0,185	RB147	0,218
RB991	0,199	RB1458	0,191
RB979	0,217	RB1456	0,125
RB903	0,137	RB1399	0,159
RB883	0,158	RB1388	0,17
RB828	0,265	RB138	0,189
RB807	0,214	RB137	0,27
RB796	0,12	RB1321	0,155
RB794	0,212	RB13	0,22
RB73	0,219	RB1296	0,24
RB727	0,154	RB1291	0,175
RB725	0,288	RB1289	0,221
RB69	0,125	RB1287	0,171
RB672	0,272	RB1263	0,166
RB67	0,15	RB1232	0,15
RB640	0,199	RB1230	0,131
RB602	0,175	RB1169	0,138
RB60	0,309	RB1168	0,24
RB589	0,208	RB11	0,18
RB414	0,184	RB10	0,133
RB397	0,199	RB06	0,212
RB381	0,196	GMR-RA959	0,208
RB356	0,13	GMR-RA958	0,17

RB231-2	0,154	GMR-RA871	0,213	
RB198	0,23	GMR-RA857	0,164	
RB197	0,199	GMR-RA81	0,156	
RB196		0,21	GMR-RA768	0,205
RB188	0,249	GMR-RA687	0,14	
RB184	0,154	GMR-RA292	0,148	
RB183,2	0,187	GMR-RA291	0,151	
RB182	0,136	GMR-RA267	0,156	
RB1725	0,131	GMR-RA265	0,194	
RB1690	0,168	GMR-RA20	0,217	
RB1655	0,188	GMR-RA121	0,249	
RB1571	0,154	GMR-RA1051	0,134	
RB1552	0,177	GMR-RA1010	0,223	
RB1551	0,163	RB978	0,147	

3. Valores de la concentración mínima inhibitoria de los 76 aislamientos de *A. baumannii* obtenido por microdilución en caldo

Aislamiento	Concentración mínima inhibitoria (ug/ml)																				
	AMK	AMP	CFZ	FEP	CAZ	CXM	CIP	CLR	CLI	COL	FOS	IMP	LIN	MER	NIT	OXA	TCY	Tianfenicol	Tigeciclina	STX	VAN
RB993	64	>256	>64	64	16	>256	>128	64	>32	0,5	>2048	32	>128	64	>1024	>64	>128	>256	0,5	128/640	>128
RB992	128	>256	>64	256	32	>256	>128	>64	>32	1	2048	32	>128	64	>1024	>64	16	>256	0,5	32/160	128
RB991	32	>256	>64	64	8	>256	>128	32	>32	4	1024	32	>128	128	>1024	>64	>128	>256	0,5	64/320	128
RB979	32	>256	>64	128	64	>256	>128	>64	>32	4	>2048	32	>128	64	>1024	>64	4	>256	0,0625	128/640	128
RB903	64	>256	>64	64	32	>256	>128	>64	>32	1	128	32	>128	64	>1024	>64	32	>256	1	128/640	128
RB883	64	>256	>64	32	32	>256	>128	>64	>32	8	512	64	>128	128	>1024	>64	8	>256	0,25	16/80	128
RB828	8	>256	>64	256	32	>256	128	64	>32	1	128	32	>128	64	>1024	>64	16	>256	0,25	>128/640	32
RB807	32	>256	>64	32	8	>256	>128	>64	>32	2	>2048	32	>128	64	>1024	>64	8	>256	0,125	>128/640	128
RB796	32	>256	>64	64	16	>256	>128	>64	>32	4	128	128	>128	>1024	>64	2	>256	0,0625	>128/640	>128	
RB794	32	>256	>64	>256	32	>256	32	64	>32	1	128	32	>128	64	>1024	>64	8	>256	0,25	16/80	128
RB73	4	>256	>64	8	16	>256	16	16	>32	4	512	8	>128	32	>1024	>64	8	>256	0,0625	64/320	64
RB727	32	>256	>64	64	16	>256	64	64	>32	2	128	32	>128	64	>1024	>64	>128	>256	0,25	16/80	128
RB725	32	64	>64	32	8	>256	>128	16	>32	4	256	0,3	>128	0,25	>1024	>64	16	>256	0,0625	128/640	>128
RB69	32	>256	>64	256	32	>256	64	32	>32	>64	>2048	64	>128	128	>1024	>64	8	>256	0,25	2/10	128
RB672	64	>256	>64	128	16	>256	>128	64	>32	2	512	16	>128	64	>1024	>64	8	>256	0,5	128/640	128
RB67	16	>256	>64	2	8	128	2	64	>32	4	512	32	>128	64	>1024	>64	64	>256	0,5	4/20	>128
RB640	64	>256	>64	16	4	64	>128	64	>32	2	512	64	>128	128	>1024	>64	>128	>256	2	64/320	128
RB602	128	>256	>64	32	16	>256	>128	64	>32	2	>2048	32	>128	64	>1024	>64	>128	>256	0,5	32/160	>128
RB60	128	>256	>64	256	>256	128	64	>64	>32	2	1024	16	>128	64	>1024	>64	16	>256	0,5	16/80	128
RB589	32	>256	>64	32	16	>256	>128	64	>32	4	512	16	>128	32	>1024	>64	>128	>256	0,0625	64/320	128
RB414	4	>256	>64	32	16	>256	>128	32	>32	2	1024	32	>128	64	>1024	>64	8	>256	2	32/160	64
RB397	16	>256	>64	32	16	>256	>128	64	>32	1	2048	32	>128	64	>1024	>64	32	>256	0,5	128/640	>128
RB381	128	>256	>64	32	256	>256	128	64	>32	32	>2048	8	>128	32	>1024	>64	8	>256	0,125	64/320	128
RB356	128	>256	>64	32	16	>256	>128	64	>32	1	>2048	32	>128	64	>1024	>64	32	>256	0,5	32/160	>128
RB231-2	256	>256	>64	32	128	>256	>128	64	>32	1	>2048	16	>128	64	>1024	>64	16	>256	0,25	32/160	>128

- **AMK:** amikacina, **AMP:** ampicilina, **CFZ:** cefazolina, **FEP:** cefepime, **CAZ:** ceftazidima, **CXM** cefuroxima, **CIP:** ciprofloxacina, **CLR:** claritromicina, **COL:** colistina, **FOS:** fosfomicina, **IMP:** imipenem, **LIN:** linezolid, **MER:** meropenem, **NIT:** nitrofurantoína, **OXA:** oxacilina, **TCY:** tetraciclina, **STX:** trimetoprim-sulfametoxazol, **VAN:** vancomicina

Aislamiento	Concentración mínima inhibitoria (ug/ml)																				
	AMK	AMP	CFZ	FEP	CAZ	CXM	CIP	CLR	CLI	COL	FOS	IMP	LIN	MER	NIT	OXA	TCY	Tianfenicol	Tigeciclina	STX	VAN
RB198	128	>256	>64	32	4	128	32	>64	>32	1	2048	32	>128	128	>1024	>64	>128	>256	0,5	>128/640	>128
RB197	64	>256	>64	>256	64	128	>128	64	>32	4	1024	32	>128	64	>1024	>64	32	>256	0,5	>128/640	>128
RB196	64	>256	>64	128	32	>256	>128	64	>32	2	>2048	32	>128	64	>1024	>64	32	>256	0,5	>128/640	128
RB188	128	>256	>64	64	32	>256	>128	>64	>32	2	2048	16	>128	64	>1024	>64	64	>256	2	32/160	128
RB184	4	>256	>64	64	16	>256	>128	64	>32	1	2048	32	>128	64	>1024	>64	64	>256	0,25	>128/640	64
RB183,2	32	>256	>64	>256	>256	>256	0,25	64	>32	1	>2048	32	>128	128	>1024	>64	>128	>256	0,25	>128/640	32
RB182	128	>256	>64	32	16	>256	>128	64	>32	1	128	128	>128	>128	>1024	>64	>128	>256	0,5	>128/640	128
RB1725	64	>256	>64	32	16	>256	>128	16	>32	8	1024	16	>128	32	>1024	>64	16	>256	0,25	>128/640	128
RB1690	32	>256	>64	>256	16	>256	32	64	>32	1	256	32	>128	128	>1024	>64	>128	>256	0,25	32/160	128
RB1655	128	>256	>64	32	16	>256	128	>64	>32	4	128	128	>128	128	>1024	>64	>128	>256	0,0625	32/160	128
RB1571	128	>256	>64	64	32	>256	>128	64	>32	1	>2048	32	>128	64	>1024	>64	16	>256	0,25	>128/640	128
RB1552	4	>256	>64	16	8	>256	64	>64	>32	1	128	32	>128	64	>1024	>64	32	>256	0,5	128/640	>128
RB1551	4	>256	>64	>256	32	>256	16	32	>32	1	>2048	32	>128	64	>1024	>64	16	>256	0,25	>128/640	64
RB15	128	>256	>64	32	16	>256	128	>64	>32	1	256	16	>128	32	>1024	>64	32	>256	1	>128/640	128
RB147	2	>256	>64	16	32	>256	>128	>64	>32	2	2048	32	>128	128	>1024	>64	8	>256	0,25	32/160	64
RB1458	32	>256	>64	64	16	>256	>128	64	>32	2	2048	32	>128	64	>1024	>64	16	>256	0,0625	32/160	>128
RB1456	32	>256	>64	128	32	>256	>128	64	>32	1	2048	32	>128	64	>1024	>64	32	>256	1	128/640	64
RB1399	2	>256	>64	4	8	>256	0,5	64	>32	0,5	128	16	>128	64	>1024	>64	8	>256	0,125	4/20	>128
RB1388	16	>256	>64	32	4	128	32	>64	>32	1	512	32	>128	32	>1024	>64	128	>256	0,25	64/320	>128
RB138	128	>256	>64	4	128	128	128	64	>32	1	512	128	>128	>128	1024	>64	8	>256	1	128/640	128
RB137	64	>256	>64	32	8	128	16	64	>32	1	1024	64	>128	128	>1024	>64	>128	>256	0,5	4/20	>128
RB1321	>512	>256	>64	>256	64	>256	>128	64	>32	1	2048	32	>128	64	>1024	>64	16	>256	0,5	32/160	128
RB13	64	>256	>64	32	16	>256	>128	64	>32	8	>2048	32	>128	64	>1024	>64	8	>256	0,5	64/320	32
RB1296	64	>256	>64	64	16	>256	>128	64	>32	1	128	32	>128	64	>1024	>64	>128	>256	0,0625	32/160	128
RB1291	128	>256	>64	32	16	>256	64	>64	>32	1	128	64	>128	128	>1024	>64	8	>256	0,5	32/160	128
RB1289	8	>256	>64	256	32	>256	64	>64	>32	2	2048	32	>128	>128	>1024	>64	16	>256	0,125	16/80	128
RB1287	32	>256	>64	64	16	>256	>128	64	>32	1	>2048	32	>128	32	>1024	>64	8	>256	0,25	128/640	128
RB1263	32	>256	>64	64	16	>256	32	32	>32	4	128	0,3	>128	0,25	>1024	>64	4	>256	0,25	128/640	64
RB1232	128	>256	>64	32	32	>256	>128	64	>32	2	2048	16	>128	32	>1024	>64	8	>256	0,0625	32/160	128

- **AMK:** amikacina, **AMP:** ampicilina, **CFZ:** cefazolina, **FEP:** cefepime, **CAZ:** ceftazidima, **CXM** cefuroxima, **CIP:** ciprofloxacina, **CLR:** claritromicina, **COL:** colistina, **FOS:** fosfomicina, **IMP:** imipenem, **LIN:** linezolid, **MER:** meropenem, **NIT:** nitrofurantoína, **OXA:** oxacilina, **TCY:** tetraciclina, **STX:** trimetoprim-sulfametoxazol, **VAN:** vancomicina

Aislamiento	Concentración mínima inhibitoria (ug/ml)																				
	AMK	AMP	CFZ	FEP	CAZ	CXM	CIP	CLR	CLI	COL	FOS	IMP	LIN	MER	NIT	OXA	TCY	Tianfenicol	Tigeciclina	STX	VAN
RB1230	32	>256	>64	64	16	>256	>128	64	>32	1	>2048	32	>128	64	>1024	>64	64	>256	0,25	>128/640	128
RB1169	32	>256	>64	32	32	>256	>128	>64	>32	8	2048	32	>128	64	>1024	>64	64	>256	0,25	128/640	>128
RB1168	32	>256	>64	32	32	>256	>128	64	>32	4	>2048	16	>128	4	>1024	>64	32	>256	0,25	>128/640	>128
RB11	64	>256	>64	128	256	>256	>128	32	>32	2	>2048	16	>128	32	>1024	>64	128	>256	0,25	32/160	128
RB10	32	>256	>64	32	32	>256	>128	64	>32	0,5	>2048	16	>128	64	>1024	>64	8	>256	1	32/160	64
RB06	64	>256	>64	32	32	>256	>128	64	>32	1	>2048	32	>128	64	>1024	>64	32	>256	0,5	128/640	128
GMR-RA959	2	>256	>64	8	2	>256	8	>64	>32	1	256	32	>128	64	>1024	>64	>128	>256	0,125	8/40	64
GMR-RA958	32	>256	>64	32	16	>256	>128	32	>32	4	128	16	>128	64	>1024	>64	8	>256	0,25	64/320	64
GMR-RA871	128	>256	>64	64	16	>256	64	64	>32	1	128	8	>128	32	>1024	>64	8	>256	0,125	32/160	128
GMR-RA857	32	>256	>64	128	16	>256	>128	>64	>32	1	128	64	>128	64	>1024	>64	64	>256	0,25	16/80	>128
GMR-RA81	64	>256	>64	32	32	>256	>128	64	>32	4	2048	16	>128	64	>1024	>64	8	>256	1	64/320	128
GMR-RA768	32	>256	>64	>256	32	>256	>128	>64	>32	4	256	32	>128	64	>1024	>64	32	>256	0,5	32/160	128
GMR-RA687	64	>256	>64	128	32	>256	>128	>64	>32	4	2048	32	>128	128	>1024	>64	16	>256	1	>128/640	32
GMR-RA292	32	>256	>64	128	64	>256	>128	64	>32	2	1024	32	>128	64	>1024	>64	32	>256	0,0625	128/640	128
GMR-RA291	32	>256	>64	64	32	>256	>128	64	>32	1	2048	16	>128	32	>1024	>64	128	>256	0,25	64/320	128
GMR-RA267	16	>256	>64	>256	>256	>256	128	64	>32	2	256	32	>128	64	>1024	>64	32	>256	1	>128/640	>128
GMR-RA265	32	>256	>64	64	32	>256	>128	64	>32	4	128	128	>128	>128	>1024	>64	>128	>256	0,5	>128/640	32
GMR-RA20	64	>256	>64	256	>256	>256	>128	32	>32	1	2048	8	>128	64	>1024	>64	64	>256	0,25	>128/640	64
GMR-RA121	16	>256	>64	32	8	64	>128	64	>32	4	128	32	>128	128	>1024	>64	32	>256	0,25	64/320	>128
GMR-RA1051	32	>256	>64	64	16	>256	>128	64	>32	2	256	32	>128	64	>1024	>64	16	>256	0,25	128/640	128
GMR-RA1010	128	>256	>64	64	16	>256	>128	64	>32	1	2048	0,3	>128	0,25	>1024	>64	32	>256	0,25	64/320	>128
RB978	32	>256	>64	64	32	>256	>128	>64	>32	16	>2048	4	>128	16	>1024	>64	>128	>256	0,0625	128/640	128

- **AMK:** amikacina, **AMP:** ampicilina, **CFZ:** cefazolina, **FEP:** cefepime, **CAZ:** ceftazidima, **CXM** cefuroxima, **CIP:** ciprofloxacina, **CLR:** claritromicina, **COL:** colistina, **FOS:** fosfomicina, **IMP:** imipenem, **LIN:** linezolid, **MER:** meropenem, **NIT:** nitrofurantoína, **OXA:** oxacilina, **TCY:** tetraciclina, **STX:** trimetoprim-sulfametoxazol, **VAN:** vancomicina

Aislamiento	Concentración mínima inhibitoria (ug/ml)																				
	AMK	AMP	CFZ	FEP	CAZ	CXM	CIP	CLR	CLI	COL	FOS	IMP	LIN	MER	NIT	OXA	TCY	Tianfenicol	Tigeciclina	STX	VAN
RB993	64	>256	>64	64	16	>256	>128	64	>32	0,5	>2048	32	>128	64	>1024	>64	>128	>256	0,5	128/640	>128
RB992	128	>256	>64	256	32	>256	>128	>64	>32	1	2048	32	>128	64	>1024	>64	16	>256	0,5	32/160	128
RB991	32	>256	>64	64	8	>256	>128	32	>32	4	1024	32	>128	128	>1024	>64	>128	>256	0,5	64/320	128
RB979	32	>256	>64	128	64	>256	>128	>64	>32	4	>2048	32	>128	64	>1024	>64	4	>256	0,0625	128/640	128
RB903	64	>256	>64	64	32	>256	>128	>64	>32	1	128	32	>128	64	>1024	>64	32	>256	1	128/640	128
RB883	64	>256	>64	32	32	>256	>128	>64	>32	8	512	64	>128	128	>1024	>64	8	>256	0,25	16/80	128
RB828	8	>256	>64	256	32	>256	128	64	>32	1	128	32	>128	64	>1024	>64	16	>256	0,25	>128/640	32
RB807	32	>256	>64	32	8	>256	>128	>64	>32	2	>2048	32	>128	64	>1024	>64	8	>256	0,125	>128/640	128
RB796	32	>256	>64	64	16	>256	>128	>64	>32	4	128	128	>128	>128	>1024	>64	2	>256	0,0625	>128/640	>128
RB794	32	>256	>64	>256	32	>256	32	64	>32	1	128	32	>128	64	>1024	>64	8	>256	0,25	16/80	128
RB73	4	>256	>64	8	16	>256	16	16	>32	4	512	8	>128	32	>1024	>64	8	>256	0,0625	64/320	64
RB727	32	>256	>64	64	16	>256	64	64	>32	2	128	32	>128	64	>1024	>64	>128	>256	0,25	16/80	128
RB725	32	64	>64	32	8	>256	>128	16	>32	4	256	0,3	>128	0,25	>1024	>64	16	>256	0,0625	128/640	>128
RB69	32	>256	>64	256	32	>256	64	32	>32	>64	>2048	64	>128	128	>1024	>64	8	>256	0,25	2/10	128
RB672	64	>256	>64	128	16	>256	>128	64	>32	2	512	16	>128	64	>1024	>64	8	>256	0,5	128/640	128
RB67	16	>256	>64	2	8	128	2	64	>32	4	512	32	>128	64	>1024	>64	64	>256	0,5	4/20	>128
RB640	64	>256	>64	16	4	64	>128	64	>32	2	512	64	>128	128	>1024	>64	>128	>256	2	64/320	128
RB602	128	>256	>64	32	16	>256	>128	64	>32	2	>2048	32	>128	64	>1024	>64	>128	>256	0,5	32/160	>128
RB60	128	>256	>64	256	>256	128	64	>64	>32	2	1024	16	>128	64	>1024	>64	16	>256	0,5	16/80	128
RB589	32	>256	>64	32	16	>256	>128	64	>32	4	512	16	>128	32	>1024	>64	>128	>256	0,0625	64/320	128
RB414	4	>256	>64	32	16	>256	>128	32	>32	2	1024	32	>128	64	>1024	>64	8	>256	2	32/160	64

5. Valores de la concentración minina inhibitoria correspondientes a las secuencias de referencia

CEPA	WGS	AMK	ATM	CAZ	CIP	COL	CRO	CTX	Doripenem	FEP	FOS	GEN	IMP	LVX	MER	MINO	TCY	Tigeciclina	TOB
BIDMC56	JMUH				2	2				32		8		4	32			1	8
29282	VYJC	32	32	32	2	4	64	64		32	64	8	8	4	32	4		2	8
34301	VYJB	32	32	32	2	2	64	64		32	64	8	8	4	16	4		2	8
20858	VYJF	32	32	32	2	2	64	64		32	64	8	8	4	32	8		2	8
16533	VYJG	32	32	32	2	2	64	64		32	64	8	8	4	32	8		2	8
6507	6507	32	32	32	2	2	64	64		32	64	8	8	4	32	8		2	8
5464	VYIZ	32	32	32	2	2	64	64		32	64	8	8	4	32	4		2	8

5457	5457	32	32	32	2	2	64	64		32	64	8	8	4	32	2		2	8
5326	VYJA	32	32	32	2	2	64	64		32	64	8	8	4	32	2		2	2
2992	VYJD	32	32	32	2	2	64	64		32	64	8	8	4	32	2		2	8
2375	VYJE	32	32	32	2	2	64	64		32	64	8	8	4	16	4		2	8
MSRN23390	VHGI	64		32	4	2	64			8		16	2	8	2		4		16
MRSN489669	VHEO	64		32	4	2	64			32		16	8	8	8		4		8
MRSN489678	VHEN	16		32	4	2	64			8		16	2	8	2		4		8
MSRN21681	VHGK	64		32	4	2	64			32		16	8	8	8		16		16
MRSN423159	VHES	32		32	4	2	64			8		16	8	8	8		4		16
MRSN32304	MRSN32304	16		8	1	2	8			8		4	2	2	2		4		4
MSRN29908	VHGD	64		8	4	2	32			32		16	8	8	8		16		16
MRSN24008	VHGH	16		32	4	2	64			32		16	2	8	2		16		4
MRSN17493	VHGN	64		32	4	256	64			32		16	8	8	8		16		16
MRSN16880	VHGO	32		8	4	2	8			8		16	2	8	2		16		8
MRSN15088	VHGT	16		8	1	2	8			8		4	2	2	2		4		4
MRSN15075	VHGU	16		32	1	2	32			8		8	2	2	2		4		4
MRSN15070	VHGV	16		8	1	2	8			8		4	2	2	2		4		4
MRSN11816	VHHE	16		8	1	2	8			8		8	2	2	2		4		4
MRSN11703	VHHH	64		32	4	2	64			32		16	2	8	4		16		16
MRSN11695	VHHI	64		32	4	2	64			32		16	8	8	8		8		16

- **AMK:** amikacina, **ATM:** aztreonam, **CAZ:** ceftazidima, **CIP:** ciprofloxacina, **COL:** colistina, **CRO:** ceftriaxona, **CTX:** cefotaxima, **FEP:** cefepime, **FOS:** fosfomicina, **GEN:** gentamicina, **IMP:** imipenem, **LVX:** levofloxacino **MER:** meropenem, **MINO:** minociclina, **TCY:** tetraciclina, **TOB:** tobramicina

CEPA	WGS	AMK	ATM	CAZ	CIP	COL	CRO	CTX	Doripenem	FEP	FOS	GEN	IMP	LVX	MER	MINO	TCY	Tigeciclina	TOB
MRSN11669	VHHJ	16		32	4	2	32			32		16	2	8	2		4		4
MRSN11663	VHHK	64		32	4	2	64			32		16	8	8	8		16		16
MRSN11224	VHHL	64		32	4	2	64			32		16	8	8	8		16		16
MRSN10372	VHHM	16		8	1	2	32			8		8	2	2	2		4		4
MRSN7735	VHDV	16		8	1	2	8			8		4	2	2	2		4		4
MRSN7725	VHDW	16		8	1	2	8			8		4	2	2	2		4		4
MRSN7690	VHDX	16		32	4	2	64			32		16	2	2	2		4		16
MRSN7576	VHDY	16		8	1	2	8			8		4	2	2	2		4		4
MRSN7521	VHDZ	64		32	4	2	64			8		16	8	8	8		16		16
MRSN7460	VHEA	16		8	1	2	32			8		8	2	2	2		16		4
MRSN7446	VHEB	16		8	1	2	8			8		4	2	2	2		4		4
MRSN7431	VHEC	16		16	4	2	64			32		16	2	8	2		16		4
MRSN7251	VHED	16		32	4	2	64			32		16	2	2	2		16		16
MRSN7213	VHEE	16		32	1	2	64			8		16	2	2	2		4		16
MRSN7153	VHEF	16		8	1	2	32			8		4	2	2	2		4		4
MRSN7137	VHEG	16		16	2	2	64			8		16	2	2	8		4		16
MRSN7124	VHEH	16		16	1	2	64			8		16	2	2	2		16		16
MRSN7113	VHEI	16		8	1	2	32			8		4	2	2	2		4		4
MRSN7067	VHEJ	16		32	4	2	64			8		16	2	8	2		16		16
MRSN6541	VHEK	16		8	4	2	32			32		4	8	8	8		4		4
MRSN4943	VHEM	32		32	4	2	64			32		16	2	8	2		16		8
MRSN4484	VHER	64		32	4	2	64			32		16	8	8	8		16		16
MRSN3360	VHEZ	16		8	4	2	8			32		16	8	8	8		8		8
MRSN2821	VHGE	16		8	1	2	32			8		4	2	2	2		8		4
MRSN1551	VHGQ	16		8	4	2	32			8		8	2	2	2		16		4
MRSN1311	VHHA	16		32	4	2	64			32		16	8	8	8		16		4
MRSN1196	VHHB	32		32	4	2	64			32		16	2	8	2		4		8
MRSN1187	VHHC	64		32	4	2	64			8		16	2	8	2		4		16
MRSN1183	VHHD	64		8	4	2	32			8		16	2	8	2		16		16
MRSN1174	VHHF	64		32	4	2	64			32		16	8	8	8		8		16

- **AMK:** amikacina, **ATM:** aztreonam, **CAZ:** ceftazidima, **CIP:** ciprofloxacina, **COL:** colistina, **CRO:** ceftriaxona, **CTX:** cefotaxima, **FEP:** cefepime, **FOS:** fosfomicina, **GEN:** gentamicina, **IMP:** imipenem, **LVX:** levofloxacin, **MER:** meropenem, **MINO:** minociclina, **TCY:** tetraciclina, **TOB:** tobramicina

CEPA	WGS	AMK	ATM	CAZ	CIP	COL	CRO	CTX	Doripenem	FEP	FOS	GEN	IMP	LVX	MER	MINO	TCY	Tigeciclina	TOB
MRSN1171	VHHG	16		32	4	2	64			32		16	2	8	2		16		8
MRSN960	VHDR	64		32	4	2	64			32		16	8	8	8		16		4
MRSN959	VHDS	64		32	4	2	64			32		16	8	8	8		4		16
MRSN918	VHDT	64		32	4	2	64			32		16	8	8	8		16		4
MRSN843	VHDU	16		8	4	2	32			8		16	2	8	8		16		8
MRSN334	VHFA	16		32	4	2	64			32		16	8	8	8		8		4
MRSN22112	VHGJ	32		32	1	2	64			32		16	8	2	8		16		8
MRSN19482	VHGM	64		32	4	2	64			32		16	8	8	8		4		16
MRSN15129	VHGR	16		8	4	2	32			8		16	2	8	2		16		8
MRSN15093	VHGS	16		8	1	2	8			8		4	2	2	2		4		4
MRSN351524	VHEW	16		8	1	2	8			8		16	2	2	2		4		4
MRSN351162	VHEX	16		8	1	2	8			8		8	2	2	2		4		4
MRSN337038	VHEY	64		32	4	2	64			32		16	8	8	8		4		8
MRSN32915	VHFB	64		32	4	2	64			32		16	8	8	8		16		16
MRSN32892	VHFC	64		8	1	2	64			8		16	2	2	2		4		16
MRSN32875	VHFD	16		8	1	2	8			8		4	2	2	2		4		4
MRSN32866	VHFE	64		32	4	2	64			32		16	2	8	2		16		4
MRSN32865	VHFF	64		32	4	2	64			32		16	2	8	2		4		16
MRSN32842	VHFG	16		8	1	2	8			8		8	2	2	2		4		4
MRSN32797	VHFH	16		8	1	2	32			8		4	2	2	2		4		4
MRSN32142	VHFJ	16		32	1	2	64			32		16	2	2	2		4		16
MRSN32108	VHFK	64		32	4	2	64			8		16	2	8	2		16		16
MRSN32104	VHFL	64		8	4	2	64			8		16	8	8	8		16		16
MRSN32076	VHFM	16		8	1	2	8			8		4	2	2	2		4		4
MRSN31947	VHFN	16		8	1	2	32			8		4	2	2	2		4		4
MRSN31942	VHFO	16		8	1	2	8			8		4	2	2	2		8		4
MRSN31937	VHFP	16		8	1	2	8			8		4	2	2	2		4		4
MRSN31915	VHFQ	16		8	1	2	8			8		4	2	2	2		4		4
MRSN31523	VHFR	16		32	1	2	64			32		16	2	2	2		4		16
MRSN31468	VHFS	16		8	1	2	32			8		4	2	2	2		4		4

- **AMK:** amikacina, **ATM:** aztreonam, **CAZ:** ceftazidima, **CIP:** ciprofloxacina, **COL:** colistina, **CRO:** ceftriaxona, **CTX:** cefotaxima, **FEP:** cefepime, **FOS:** fosfomicina, **GEN:** gentamicina, **IMP:** imipenem, **LVX:** levofloxacin, **MER:** meropenem, **MINO:** minociclina, **TCY:** tetraciclina, **TOB:** tobramicina

CEPA	WGS	AMK	ATM	CAZ	CIP	COL	CRO	CTX	Doripenem	FEP	FOS	GEN	IMP	LVX	MER	MINO	TCY	Tigeciclina	TOB
MRSN31461	VHFT	16		8	1	2	8			8		4	2	2	2		4		4
MRSN31196	VHFU	64		32	4	2	64			32		16	8	8	8		16		8
MRSN31159	VHFV	16		8	1	2	8			8		4	2	2	2		4		4
MRSN30945	VHFW	16		8	1	2	8			8		4	2	2	2		4		4
MRSN30912	VHFX	16		8	1	2	8			8		4	2	2	2		4		4
MRSN30909	VHFY	16		8	1	2	32			8		4	2	2	2		4		4
MRSN30896	VHFZ	16		8	1	2	32			8		16	2	2	2		16		16
MRSN30885	VHGA	16		8	1	2	8			8		4	2	2	2		4		4
MRSN30000	VHGB	64		32	4	2	64			32		16	2	8	2		16		16
MRSN29999	VHGC	16		8	1	2	8			8		4	2	2	2		4		4
MRSN25547	VHGF	64		32	4	2	64			32		16	8	8	8		16		16
MRSN24603	VHGG	16		8	1	2	8			8		4	2	2	2		4		4
MRSN15574	VHGP	16		32	4	2	64			32		16	8	8	8		16		16
MRSN5969	VHEL	16		8	1	2	8			8		4	2	2	2		4		4
MRSN3874	VHET	64		32	4	2	64			32		16	8	8	8		4		16
MRSN3692	VHEU	16		8	1	2	32			8		4	2	2	2		4		4
MRSN3658	VHEV	16		8	1	2	8			8		4	2	2	2		4		4
MRSN21660	VHGL	64		32	4	2	64			32		16	8	8	8		16		16
MRSN14237	VHGY	64		32	4	2	64			32		16	8	8	8		16		16
MRSN14193	VHGZ	64		32	4	2	64			32		16	8	8	8		16		8
MRSN480622	VHEP	64		32	4	2	64			32		16	8	8	8		16		16
MRSN480561	VHEQ	64		32	4	2	64			32		16	8	8	8		4		16
MRSN15049	VHGW	64		32	4	2	64			32		16	8	8	8		16		4
MRSN14427	VHGX	64		32	4	2	64			32		16	8	8	8		4		16
FDA-CDC-AR_0275	VYVC	64		128	8	0,50	32	64	8	32		16	64	8	8	8	32	1	16
ARLG-1912	NEQM	32		16	2	0,25			2	16		8	8		8	4		2	
ARLG-1818	NEQE	32		16	2	0,50			2	16		8	8		8	8		1	
ARLG-1775	NEQF	8		1	2	1			2	2		8	8		4	2		1	

- **AMK:** amikacina, **ATM:** aztreonam, **CAZ:** ceftazidima, **CIP:** ciprofloxacina, **COL:** colistina, **CRO:** ceftriaxona, **CTX:** cefotaxima, **FEP:** cefepime, **FOS:** fosfomicina, **GEN:** gentamicina, **IMP:** imipenem, **LVX:** levofloxacin, **MER:** meropenem, **MINO:** minociclina, **TCY:** tetraciclina, **TOB:** tobramicina

CEPA	WGS	AMK	ATM	CAZ	CIP	COL	CRO	CTX	Doripenem	FEP	FOS	GEN	IMP	LVX	MER	MINO	TCY	Tigeciclina	TOB
ARLG-1788	NEQG	32		16	2	2			2	16		8	8		8	2		1	
ARLG-1777	NEQH	4		16	2	1			2	16		8	8		8	2		1	
ARLG-1854	NEQI	32		16	2	4			2	16		8	8		8	8		1	
ARLG-1915	NEQJ	8		16	2	1			2	16		8	8		8	8		1	
ARLG-1860	NEQK	32		8	2	1			2	16		8	8		8	2		2	
ARLG-1894	NEQL	4		16	2	1			2	16		8	8		8	16		1	
PR336	NGDZ	4		4	2	0,50			0,5	2		1	1		1	2		0,25	
ARLG-1862	NGGF	4		2	0,25	1,00			0,25	2		1	1		1	2		0,25	
PR391	NGEY	4		4	0,25	1			0,12	2		1	1		1	2		0,25	
PR356	NGCJ	4		8	0,25	1			0,5	2		2	1		2	2		0,25	
PR308	NGBY	4		2	0,25	0,50			0,12	2		1	1		1	2		0,25	
ARLG1314	NGHN	32		16	2	0,50			2	16		8	8		8	2		1	
ARLG1933	NGII	4		4	0,25	0,50			0,12	2		1	1		1	2		0,25	
ARLG1936	NGIK	4		4	0,25	0,25			0,25	2		1	1		1	2		0,25	
PR355	NGCI	32		16	2	0,50			2	16		8	1		2	2		1	
ARLG1935	NGIJ	4		4	0,25	0,25			0,12	2		1	1		1	2		0,25	
PR311	NGCB	4		4	2	0,50			0,5	2		1	1		1	2		0,25	
PR310	NGCA	4		8	0,25	1			0,25	2		1	1		1	2		0,25	
ARLG1326	NGHQ	4		4	0,25	0,25			0,12	2		1	1		1	2		0,25	
ARLG1902	NGIB	4		4	0,25	0,50			0,25	2		1	1		1	2		0,25	
ARLG1880	NGJD	4		4	0,25	0,50			0,12	2		1	1		1	2		0,25	
PR373	NGDA	4		4	0,25	1			0,12	2		1	1		1	2		0,25	
ARLG1941	NGIN	4		8	0,25	2			0,5	8		4	1		1	2		0,25	
ARLG1833	NGHF	4		16	2	0,50			2	16		8	8		8	2		0,5	
PR379	NGDF	4		2	2	1			1	4		8	1		1	2		1	
ARLG1830	NGHE	4		4	0,25	0,50			0,25	2		1	1		1	2		0,25	
PR306	NGBW	4		16	2	1			2	16		8	1		2	2		1	
PR337	NGEA	4		2	0,25	0,50			0,12	2		1	1		1	2		0,25	
ARLG1871	NGIY	32		8	2	1			2	16		1	8		8	2		0,25	

- **AMK:** amikacina, **ATM:** aztreonam, **CAZ:** ceftazidima, **CIP:** ciprofloxacina, **COL:** colistina, **CRO:** ceftriaxona, **CTX:** cefotaxima, **FEP:** cefepime, **FOS:** fosfomicina, **GEN:** gentamicina, **IMP:** imipenem, **LVX:** levofloxacino **MER:** meropenem, **MINO:** minociclina, **TCY:** tetraciclina, **TOB:** tobramicina

CEPA	WGS	AMK	ATM	CAZ	CIP	COL	CRO	CTX	Doripenem	FEP	FOS	GEN	IMP	LVX	MER	MINO	TCY	Tigeciclina	TOB
PR332	NGDV	4		16	2	4			1	16		8	1		1	2		2	
ARLG1872	NGIZ	4		8	2	0,50			2	16		1	8		8	4		0,5	
PR307	NGBX	4		4	0,25	0,50			0,25	2		1	1		1	2		0,25	
PR333	NGDW	8		4	0,25	1			0,12	2		1	1		1	2		0,25	
PR389	NGEW	4		4	1	1			1	16		1	1		1	2		0,5	
PR315	NGEM	4		4	0,25	0,25			0,5	2		1	1		1	2		0,25	
PR376	NGDD	32		16	2	1			2	16		8	8		8	8		1	
ARLG1787	NGGK	32		16	2	1			2	16		8	8		8	2		1	
ARLG1929	NGIG	4		4	0,25	0,25			0,25	4		1	1		1	2		0,25	
ARLG1940	NGIM	4		8	2	0,25			0,5	2		1	1		1	2		0,25	
PR388	NGEV	4		2	0,25	0,50			0,5	2		1	1		1	2		0,25	
PR340	NGEC	4		16	2	0,50			0,5	16		1	1		1	2		0,25	
ARLG1790	NGGM	32		16	2	1			2	16		8	8		8	4		1	
ARLG1810	NGGX	32		16	2	0,50			2	16		8	8		8	8		1	
PR361	NGCO	4		2	0,25	0,50			0,12	2		1	1		1	2		0,25	
ARLG1891	NGJI	32		16	2	1			2	16		8	8		8	2		1	
PR382	NGEP	32		16	2	1			2	16		8	8		8	8		1	
ARLG1786	NGGJ	32		16	2	0,50			2	16		8	8		8	8		1	
PR392	NGEZ	32		16	2	1			2	16		8	8		8	2		2	
PR325	NGDP	32		16	2	0,50			2	16		8	8		8	8		1	
ARLG1886	NGJH	4		16	2	1			2	16		8	8		8	16		2	
PR400	NGFG	32		16	2	1			2	16		8	8		8	8		0,5	
ARLG1793	NGGO	32		16	2	0,50			2	16		8	8		8	8		1	
ARLG1814	NGGY	4		8	0,25	0,50			0,25	4		2	1		1	2		0,25	
ARLG1275	NGHK	32		16	2	0,50			1	16		8	2		1	4		1	
PR364	NGCR	32		16	2	1			2	16		8	8		8	8		1	
ARLG1294	NGHL	4		16	2	0,50			2	16		8	2		1	4		0,5	
PR359	NGCM	32		16	2	1			2	16		8	8		8	8		2	
ARLG1901	NGIA	32		16	2	0,50			2	16		8	8		8	8		1	
ARLG1914	NGIF	4		16	2	0,50			2	16		8	8		8	8		1	

- **AMK:** amikacina, **ATM:** aztreonam, **CAZ:** ceftazidima, **CIP:** ciprofloxacina, **COL:** colistina, **CRO:** ceftriaxona, **CTX:** cefotaxima, **FEP:** cefepime, **FOS:** fosfomicina, **GEN:** gentamicina, **IMP:** imipenem, **LVX:** levoflaxacino **MER:** meropenem, **MINO:** minociclina, **TCY:** tetraciclina, **TOB:** tobramicina

CEPA	WGS	AMK	ATM	CAZ	CIP	COL	CRO	CTX	Doripenem	FEP	FOS	GEN	IMP	LVX	MER	MINO	TCY	Tigeciclina	TOB
PR387	NGEU	32		16	2	1			2	16		8	8		8	8		1	
ARLG1836	NGHH	32		16	2	0,50			2	16		8	8		8	8		1	
ARLG1842	NGFS	32		16	2	1			2	16		8	8		8	8		1	
ARLG1800	NGGR	4		8	0,25	0,25			0,25	2		1	1		1	2		0,25	
PR345	NGEG	32		16	2	0,50			2	16		8	8		8	8		1	
ARLG1906	NGID	32		16	2	1			2	16		8	8		8	2		1	
ARLG1898	NGHX	32		16	2	1			2	2		8	8		8	2		1	
ARLG1893	NGHV	32		16	2	1			2	16		8	8		8	2		1	
ARLG1847	NGFW	4		16	2	0,50			2	16		8	8		8	8		1	
ARLG1846	NGFV	4		2	2	4			0,12	2		1	1		1	2		0,25	
ARLG1899	NGHY	32		16	2	0,50			2	16		8	8		8	8		1	
ARLG1792	NGGN	8		16	2	0,50			1	16		8	1		1	2		0,5	
ARLG1802	NGGS	32		16	2	1			2	16		8	8		8	8		2	
ARLG1855	NGGA	32		16	2	0,50			2	16		8	8		8	8		1	
ARLG1824	NGHC	32		16	2	0,50			2	16		8	8		8	4		2	
PR309	NGBZ	32		16	2	0,50			2	16		8	8		8	8		1	
PR360	NGCN	4		8	0,5	0,50			0,25	4		1	1		1	2		0,25	
PR341	NGED	32		16	2	1			2	16		8	8		8	8		1	
PR321	NGDL	32		16	2	1			2	16		8	8		8	2		0,5	
ARLG1827	NGHD	32		16	2	0,50			2	16		8	8		8	8		1	
PR343	NGEE	4		2	0,25	0,50			0,25	2		1	1		1	2		0,25	
PR385	NGES	8		4	2	1			0,25	8		8	1		1	2		1	
PR351	NGCE	32		16	2	0,50			2	16		8	8		8	8		1	
PR380	NGDG	32		16	2	1			2	16		8	8		8	8		1	
ARLG1306	NGHM	4		8	0,25	0,25			0,5	4		8	1		1	2		1	
PR312	NGCC	32		16	2	0,50			2	16		8	8		8	8		1	
PR399	NGFF	32		16	2	1			2	16		8	8		8	8		1	
PR377	NGDE	32		16	2	4			2	16		8	8		8	8		0,5	
ARLG1806	NGGV	32		16	2	0,50			2	16		8	8		8	8		1	
PR394	NGFB	32		16	2	4			2	16		8	8		8	16		1	

- **AMK:** amikacina, **ATM:** aztreonam, **CAZ:** ceftazidima, **CIP:** ciprofloxacina, **COL:** colistina, **CRO:** ceftriaxona, **CTX:** cefotaxima, **FEP:** cefepime, **FOS:** fosfomicina, **GEN:** gentamicina, **IMP:** imipenem, **LVX:** levoflaxacino **MER:** meropenem, **MINO:** minociclina, **TCY:** tetraciclina, **TOB:** tobramicina

CEPA	WGS	AMK	ATM	CAZ	CIP	COL	CRO	CTX	Doripenem	FEP	FOS	GEN	IMP	LVX	MER	MINO	TCY	Tigeciclina	TOB
ARLG1851	NGFY	32		16	2	4			2	16		8	8		8	8		1	
PR353	NGCG	32		16	2	1			0,5	16		8	1		1	8		1	
ARLG1819	NGHA	32		16	2	0,50			2	16		8	8		8	8		1	
ARLG1838	NGHI	32		16	2	0,50			2	16		8	8		8	8		1	
PR358	NGCL	32		16	2	1			2	16		8	8		8	8		1	
ARLG1821	NGHB	32		16	2	1			2	16		8	8		8	16		1	
PR383	NGEQ	32		16	2	4			2	16		8	8		8	8		1	
ARLG1317	NGHO	4		8	2	0,50			2	16		8	8		8	2		2	
PR334	NGDX	32		16	2	4			2	16		8	8		8	8		1	
ARLG1882	NGJF	32		16	2	0,50			2	16		8	8		8	2		8	
ARLG1859	NGGE	4		16	2	1			2	16		8	8		8	16		1	
PR396	NGFD	32		16	2	1			2	16		8	8		8	2		1	
PR349	NGEK	32		16	2	0,50			2	16		8	8		8	8		1	
PR363	NGCQ	32		16	2	1			2	16		8	8		8	2		1	
PR335	NGDY	32		16	2	0,50			2	8		8	8		8	2		1	
ARLG1853	NGFZ	32		16	2	4			2	16		8	8		8	16		1	
PR381	NGDH	16		16	2	0,50			2	16		8	8		8	4		1	
ARLG1845	NGFU	32		16	2	0,50			2	16		8	8		8	16		1	
PR390	NGEX	32		16	2	1			2	16		8	8		8	16		1	
PR322	NGDM	4		16	2	4			2	16		8	1		4	8		1	
ARLG1857	NGGC	4		16	2	0,50			2	16		8	8		8	16		1	
ARLG1858	NGGD	16		16	2	0,50			2	16		8	8		8	2		0,25	
PR398	NGFE	32		16	2	1			2	16		8	8		8	4		4	
ARLG1323	NGHP	32		8	0,25	0,50			1	4		8	4		2	2		1	
PR304	NGBU	4		4	0,25	0,25			0,12	2		1	1		1	2		0,5	
PR357	NGCK	32		16	2	1			2	16		8	8		8	8		1	
ARLG1835	NGHG	32		16	2	1			2	16		8	8		8	16		1	
PR372	NGCZ	32		16	2	4			2	16		8	8		8	2		1	
PR347	NGEI	32		16	2	0,50			2	16		8	2		4	8		2	
PR328	NGDR	32		16	2	4			2	16		8	8		8	2		1	

- **AMK:** amikacina, **ATM:** aztreonam, **CAZ:** ceftazidima, **CIP:** ciprofloxacina, **COL:** colistina, **CRO:** ceftriaxona, **CTX:** cefotaxima, **FEP:** cefepime, **FOS:** fosfomicina, **GEN:** gentamicina, **IMP:** imipenem, **LVX:** levoflaxacino **MER:** meropenem, **MINO:** minociclina, **TCY:** tetraciclina, **TOB:** tobramicina

CEPA	WGS	AMK	ATM	CAZ	CIP	COL	CRO	CTX	Doripenem	FEP	FOS	GEN	IMP	LVX	MER	MINO	TCY	Tigeciclina	TOB
PR327	NGDQ	8		16	2	1			2	16		8	8		8	8		1	
ARLG1856	NGGB	4		16	2	0,50			2	16		8	8		8	8		1	
ARLG1896	NGHW	32		16	2	1			2	8		8	8		8	2		2	
PR371	NGCY	8		4	2	1			0,25	8		8	1		1	2		0,5	
ARLG1843	NGFT	32		16	2	0,50			2	16		8	8		8	8		1	
ARLG1841	NGFR	32		16	2	0,50			2	16		8	8		8	16		1	
PR395	NGFC	32		16	2	1			2	16		8	8		8	2		1	
PR352	NGCF	4		2	0,25	0,50			0,12	2		1	1		1	2		0,25	
PR317	NGEO	32		16	2	4			2	16		8	8		8	4		2	
ARLG1864	NGIV	32		16	2	0,50			2	16		8	8		8	2		1	
ARLG1885	NGJG	32		16	2	0,50			2	16		8	8		8	8		2	
ARLG1840	NGFQ	32		16	2	1			2	16		8	4		8	8		1	
ARLG1874	NGJB	4		4	0,25	0,25			0,25	2		1	1		1	2		0,25	
PR384	NGER	4		16	2	0,50			2	16		8	8		8	2		0,25	
PR319	NGDJ	32		16	2	0,50			2	16		8	8		8	4		1	
ARLG1327	NGHR	4		4	0,25	0,50			0,12	2		1	1		1	2		0,25	
ARLG1816	NGGZ	32		16	2	1			2	16		8	1		8	16		1	
PR305	NGBV	32		16	2	1			2	16		8	8		8	16		1	
ARLG1900	NGHZ	32		16	2	2			2	16		8	8		8	8		1	
ARLG1892	NGHU	4		16	2	1			2	16		8	8		8	8		1	
PR323	NGDN	32		16	2	1			2	16		8	8		8	2		1	
ARLG1907	NGIE	32		16	2	1			2	16		8	8		8	16		1	
ARLG1849	NGFX	32		16	2	1			2	16		8	8		8	8		0,5	
ARLG1881	NGJE	32		16	2	1			2	16		8	8		8	16		8	
ARLG1794	NGGP	4		4	0,25	0,50			0,25	2		1	1		1	2		0,25	
PR350	NGEL	4		2	0,25	0,50			0,25	2		1	1		1	2		0,25	
PR329	NGDS	4		4	0,5	0,25			0,5	2		1	1		1	2		0,5	
ARLG1904	NGIC	32		16	2	0,50			2	16		8	8		8	2		1	

- **AMK:** amikacina, **ATM:** aztreonam, **CAZ:** ceftazidima, **CIP:** ciprofloxacina, **COL:** colistina, **CRO:** ceftriaxona, **CTX:** cefotaxima, **FEP:** cefepime, **FOS:** fosfomicina, **GEN:** gentamicina, **IMP:** imipenem, **LVX:** levofloxacin, **MER:** meropenem, **MINO:** minociclina, **TCY:** tetraciclina, **TOB:** tobramicina

CEPA	WGS	AMK	ATM	CAZ	CIP	COL	CRO	CTX	Doripenem	FEP	FOS	GEN	IMP	LVX	MER	MINO	TCY	Tigeciclina	TOB
PR374	NGDB	4		1	0,25	1			0,12	2		1	1		1	2		0,25	
ARLG1873	NGJA	4		4	0,25	0,25			0,12	2		1	1		1	2		0,25	
ARLG1946	NGIP	4		2	0,25	0,50			0,12	2		1	1		1	2		0,25	
PR375	NGDC	4		2	0,25	1			0,12	2		1	1		1	2		0,25	
PR354	NGCH	4		8	2	0,25			2	4		8	8		8	2		0,25	

- **AMK:** amikacina, **ATM:** aztreonam, **CAZ:** ceftazidima, **CIP:** ciprofloxacina, **COL:** colistina, **CRO:** ceftriaxona, **CTX:** cefotaxima, **FEP:** cefepime, **FOS:** fosfomicina, **GEN:** gentamicina, **IMP:** imipenem, **LVX:** levoflaxacino **MER:** meropenem, **MINO:** minociclina, **TCY:** tetraciclina, **TOB:** tobramicina

6. Lista de las características seleccionadas por *permutation feature importance* y el valor de su importancia

AMIKACINA

Característica (gen)	Importancia	Característica (gen)	Importancia	Característica (gen)	Importancia	Característica (gen)	Importancia
APH(3')-Vla	1.29	APH(6)-ld	0.02	tet(B)	0.01	OXA-171	0.01
AAC(6')-lan	0.68	APH(3'')-lb	0.02	OXA-65	0.01	ADC-30	0.01
Escherichia_coli_gyrA	0.56	tetR	0.02	Salmonella_serovars_gyrB_	0.01	Ef-Tu_cp1	0.01
armA	0.17	cmlB1	0.02	ADC-2	0.01	ADC-76	0.01
OXA-23	0.14	mphE	0.02	CARB-16	0.01	OXA-106	0.01
AAC(3)-lle	0.1	msrE	0.02	ANT(2'')-la	0.01	golS	0.01
sul2	0.07	APH(3')-la	0.02	ADC-76	0.01	sul1	0.01
ugd	0.06	adeC	0.02	OXA-106	0.01	OXA-259	0.01
adeN	0.05	AAC(6')-lb7	0.02	golS	0.01	OXA-146	0.01
AbaF	0.05	AAC(6')-laf	0.02	ADC-25	0.01	OXA-100	0.01
AAC(3)-la	0.04	abeS	0.01	OXA-259	0.01	OXA-338	0.01
adeS	0.03	abeS_cp1	0.01	OXA-146	0.01	OXA-171	0.01
ADC-5	0.03	ANT(3'')-llc	0.01	ADC-30	0.01	OXA-69	0.01
TEM-1	0.03	adeB	0.01	ADC-25	0.01		
OXA-121	0.03	adeA	0.01	OXA-100	0.01		
OXA-64	0.02	adeR	0.01	OXA-338	0.01		
adeL	0.02	Mycobacterium_tuberculosis_rpoB	0.01	EF-Tu	0.01		
AbaQ	0.02	adeJ	0.01	OXA-69	0.01		
Clostridioides_difficile_gyrB	0.02	soxR	0.01	sul1	0.01		

CEFEPIME

Característica (gen)	Importancia	Característica (gen)	Importancia	Característica (gen)	Importancia
Escherichia_coli_gyrA	1.88	adeN	0.02	soxR	0.01
EF-Tu	0.38	abeS_cp1	0.02	ADC-76	0.01
CTX-M-115	0.2	AAC(3)-Ile	0.02	BRP(MBL)	0.01
OXA-146	0.2	AAC(6')-Ian	0.02	NDM-1	0.01
ADC-5	0.18	cmIB1	0.02	OXA-385	0.01
APH(3')-VIa	0.12	ADC-6	0.02	dfrA20	0.01
OXA-23	0.06	abeS	0.01	Tet(X3)	0.01
TEM-1	0.06	ANT(3'')-IIc	0.01	VEB-9	0.01
msrE	0.06	adeA	0.01	APH(3')-VI	0.01
adeC	0.05	adeA_cp1	0.01	ADC-58	0.01
ugd	0.04	adeR	0.01	OXA-51	0.01
mphE	0.04	adeS_cp1	0.01	AAC(3)-Ia	0.01
aadA	0.04	adeJ	0.01	ADC-79	0.01
OXA-64	0.03	sul2	0.01	ADC-77	0.01
adeS	0.03	tet(B)	0.01	OXA-223	0.01
AbaF	0.03	APH(6)-Id	0.01	OXA-338	0.01
OXA-65	0.03	APH(3'')-Ib	0.01	OXA-69	0.01
ADC-2	0.03	tetR	0.01	catI	0.01
EF-Tu_cp1	0.03	Mycobacterium_tuberculosis_rpoB	0.01	OXA-407	0.01
OXA-121	0.03	APH(3')-Ia	0.01		
OXA-94	0.03	ANT(2'')-Ia	0.01		

CEFTAZIDIMA

Característica (gen)	Importancia	Característica (gen)	Importancia	Característica (gen)	Importancia
Escherichia_coli_gyrA	1.16	AmvA	0.02	Clostridioides_difficile_gyrB	0.01
adeB	0.1	OXA-23	0.02	ugd	0.01
OXA-146	0.09	ADC-2	0.02	cmlB1	0.01
ANT(3'')-Ilc	0.07	msrE	0.02	Mycobacterium_tuberculosis_thyA	0.01
AbaQ	0.06	ampC_beta-lactamase	0.02	Mycobacterium_tuberculosis_katG	0.01
adeF	0.05	ANT(2'')-Ia	0.02	BRP(MBL)	0.01
adeS	0.05	ADC-77	0.02	adeC	0.01
MexK	0.05	OXA-65	0.01	Tet(X3)	0.01
EF-Tu	0.05	Mycobacterium_tuberculosis_rpoB	0.01	golS	0.01
macB	0.04	OXA-64	0.01	mdsC	0.01
adeA	0.04	rosA	0.01	AAC(3)-Ia	0.01
adeR	0.04	adeN	0.01	aadA2	0.01
APH(3'')-Ib	0.04	adeH	0.01	EF-Tu_cp1	0.01
AAC(6'')-Ian	0.04	adeG	0.01		
NDM-1	0.04	adeL	0.01		
adeJ	0.03	adel	0.01		
soxR	0.03	Staphylococcus_aureus_fusA	0.01		
mphE	0.03	TEM-1	0.01		
APH(3'')-VIa	0.03	AbaF	0.01		

CIPROFLOXACINO

Característica (gen)	Importancia	Característica (gen)	Importancia	Característica (gen)	Importancia
EF-Tu	9.9	AAC(6')-Ib7	0.03	Clostridioides_difficile_gyrB	0.01
Escherichia_coli_gyrA	1.72	OXA-121	0.03	soxR	0.01
msrE	0.32	adeA	0.02	ADC-76	0.01
CTX-M-115	0.22	APH(6)-Id	0.02	BRP(MBL)	0.01
mphE	0.14	OXA-23	0.02	OXA-385	0.01
OXA-146	0.13	AAC(3)-Ile	0.02	dfrA20	0.01
OXA-65	0.11	cmIB1	0.02	Tet(X3)	0.01
adeN	0.07	VEB-9	0.02	APH(3')-VI	0.01
AAC(6')-Ilan	0.06	OXA-64	0.01	APH(3')-VIa	0.01
ADC-2	0.06	adeF_cp1	0.01	OXA-259	0.01
adeJ	0.05	adeL	0.01	mdsA	0.01
TEM-1	0.05	abeS	0.01	mdsC	0.01
APH(3')-Ia	0.04	abeS_cp1	0.01	AAC(3)-Ia	0.01
goIS	0.04	AbaQ	0.01	OXA-69	0.01
ADC-5	0.03	adeB	0.01		
AbaF	0.03	adeB_cp1	0.01		
sul2	0.03	adeA_cp1	0.01		
ugd	0.03	adeR	0.01		
EF-Tu_cp1	0.03	adeS_cp1	0.01		
ampC_beta-lactamase	0.03	tet(B)	0.01		
adeC	0.03	APH(3'')-Ib	0.01		
VIM-4	0.03	tetR	0.01		

COLISTINA

Característica (gen)	Importancia	Característica (gen)	Importancia	Característica (gen)	Importancia
ANT(3'')-IIc	0.29	AAC(6')-Ian	0.02	tetR	0.01
ADC-73	0.22	Escherichia_coli_gyrA	0.02	mphE	0.01
APH(3'')-Ib	0.07	Mycobacterium_tuberculosis_thyA	0.02	ANT(2'')-Ia	0.01
adeN	0.05	ugd	0.02	ADC-58	0.01
adeS	0.05	APH(3')-Ia	0.02	ADC-30	0.01
MexK	0.05	OXA-121	0.02	AAC(3)-Ia	0.01
APH(3')-VIa	0.05	rosA	0.01	ADC-79	0.01
ADC-25	0.05	macB	0.01		
adeH	0.04	adeF	0.01		
cmIB1	0.04	adeG	0.01		
adeB	0.03	adeL	0.01		
adeA	0.03	ADC-5	0.01		
Mycobacterium_tuberculosis_r poB	0.03	AbaQ	0.01		
AmvA	0.02	Mycobacterium_tuberculosis_katG	0.01		
adeR	0.02	adeI	0.01		
Clostridioides_difficile_gyrB	0.02	adeJ	0.01		
soxR	0.02	AAC(3)-Ile	0.01		
AbaF	0.02	TEM-1	0.01		
sul2	0.02	OXA-24	0.01		
OXA-23	0.02	aadA2	0.01		

FOSFOMICINA

Característica (gen)	Importancia	Característica (gen)	Importancia
EF-Tu_cp1	2.02	Mycobacterium_tuberculosis_rpoB	0.02
macB	0.83	AbaF	0.02
APH(3'')-Ib	0.43	Mycobacterium_tuberculosis_katG	0.02
AAC(6')-Ian	0.43	ugd	0.02
adeB	0.3	OXA-64	0.01
cmIB1	0.3	rosA	0.01
adeI	0.15	abeM	0.01
AAC(3)-Ile	0.11	adeF	0.01
adeN	0.1	adeG	0.01
Staphylococcus_aureus_fusA	0.09	ANT(3'')-Iic	0.01
adeR	0.08	adeJ	0.01
adeS	0.07	EF-Tu	0.01
adeK	0.06	sul1	0.01
adeL	0.05	OXA-146	0.01
OXA-23	0.05		
TEM-1	0.05		
adeH	0.04		
OXA-65	0.04		
AmvA	0.03		
adeA	0.03		
ADC-5	0.03		
APH(6)-Id	0.03		

IMIPENEM

Característica (gen)	Importancia	Característica (gen)	Importancia	Característica (gen)	Importancia
OXA-23	2.16	Tet(X3)	0.02	KPC-2	0.01
EF-Tu	0.86	mphE	0.02	Mycobacterium_tuberculosis_thyA	0.01
OXA-65	0.25	abeM	0.01	ADC-30	0.01
Escherichia_coli_gyrA	0.21	AmvA	0.01	OXA-24	0.01
OXA-72	0.12	adeG	0.01	OXA-82	0.01
AAC(6)-Ian	0.09	ANT(3')-IIc	0.01	OXA-58	0.01
adeB	0.08	AbaQ	0.01	OXA-66	0.01
adeA	0.08	adeR	0.01	sul1	0.01
APH(3'')-Ib	0.07	Salmonella_serovars_gyrB	0.01	macB	0.01
Staphylococcus_aureus_fusA	0.06	sul2	0.01		
OXA-64	0.05	TEM-1	0.01		
adeS	0.04	cmIB1	0.01		
NDM-1	0.04	MexK	0.01		
adeL	0.03	Mycobacterium_tuberculosis_katG	0.01		
Clostridioides_difficile_gyrB	0.03	msrE	0.01		
AAC(3)-IIe	0.03	rosA	0.01		
OXA-259	0.03	CARB-16	0.01		
adeN	0.02	adeC	0.01		
adeH	0.02	APH(3')-VI	0.01		
ADC-5	0.02	golS	0.01		
adeI	0.02	qacH	0.01		
adeJ	0.02	APH(3')-VIa	0.01		

MEROPENEM

Característica (gen)	Importancia	Característica (gen)	Importancia	Característica (gen)	Importancia
EF-Tu	2,23	tet(B)	0,02	AbaF	0,01
OXA-23	1,79	ugd	0,02	qacH	0,01
OXA-65	0,31	adeC	0,02	VIM-4	0,01
mphE	0,21	OXA-146	0,02	AAC(6')-Ib7	0,01
EF-Tu_cp1	0,19	adeN	0,01	OXA-235	0,01
APH(3')-VI	0,16	AAC(3)-Ile	0,01	OXA-144	0,01
AAC(6')-Ian	0,15	TEM-1	0,01	OXA-24	0,01
adeS	0,14	tetR	0,01	OXA-121	0,01
Escherichia_coli_gyrA	0,14	Mycobacterium_tuberculosis_thyA	0,01		
OXA-72	0,14	CTX-M-115	0,01		
msrE	0,1	ANT(2'')-Ia	0,01		
OXA-259	0,08	OXA-255	0,01		
abeS	0,07	BRP(MBL)	0,01		
cmlB1	0,05	NDM-1	0,01		
ADC-2	0,04	OXA-385	0,01		
ADC-5	0,03	dfrA20	0,01		
Salmonella_serovars_gyrB	0,03	ampC_beta-lactamase	0,01		
OXA-64	0,02	Tet(X3)	0,01		
adeL	0,02	VEB-9	0,01		
Clostridioides_difficile_gyrB	0,02	golS	0,01		
sul2	0,02	sul1	0,01		

TETRACICLINA

Característica (gen)	Importancia	Característica (gen)	Importancia
EF-Tu	1.39	adeF	0.01
tet(B)	0.79	adeR	0.01
tetR	0.21	adeS	0.01
TEM-1	0.18	ADC-5	0.01
AbaQ	0.11	AbaF	0.01
macB	0.08	sul2	0.01
Escherichia_coli_gyrA	0.06	Mycobacterium_tuberculosis_katG	0.01
adeB	0.04	AAC(3)-Ile	0.01
AAC(6')-Iaa :	0.04	APH(6)-Id	0.01
AmvA	0.03	ugd	0.01
adeI	0.03	cmIB1	0.01
tet(D)	0.03	mphE	0.01
adeH	0.02	msrE	0.01
ANT(3'')-Iic	0.02	APH(3')-Ia	0.01
adeA	0.02	adeC	0.01
MexK	0.02	APH(3')-VI	0.01
Mycobacterium_tuberculosis_rpoB	0.02	mdsC	0.01
APH(3'')-Ib	0.02	tet(39)	0.01
OXA-23	0.02		
OXA-65	0.02		
OXA-64	0.01		
rosA	0.01		
abeM	0.01		

TIGECICLINA

Característica (gen)	Importancia	Característica (gen)	Importancia
APH(3'')-Ib	0.22	OXA-24	0.01
AmvA	0.18	tet(39)	0.01
Escherichia_coli_gyrA	0.14	rosA	0.01
adeS	0.11	adeL	0.01
adeN	0.1	adeA	0.01
sul1	0.1	adeJ	0.01
adeH	0.07	Clostridioides_difficile_gyrB	0.01
adeR	0.07	sul2	0.01
AAC(6')-Ian	0.07	APH(6)-Id	0.01
adeG	0.05	TEM-1	0.01
ANT(3'')-IIc	0.04	OXA-146	0.01
adeB	0.04	Mycobacterium_tuberculosis_katG	0.01
adeF	0.03	ugd	0.01
MexK	0.03	OXA-65	0.01
ANT(2'')-Ia	0.03	msrE	0.01
macB	0.02	armA	0.01
adeI	0.02	OXA-72	0.01
OXA-23	0.02	adeC	0.01
mphE	0.02	ADC-30	0.01
APH(3')-Ia	0.02	AAC(3)-Ia	0.01
EF-Tu_cp1	0.02	Mycobacterium_tuberculosis_thyA	0.01
APH(3')-VIa	0.02		
AAC(6')-Iaf	0.02		

7. Hiperparámetros seleccionados por *grid search* para el entrenamiento de modelos con *random forest*

Modelo	Parámetros			
	Profundidad	min_samples_split	max_features	min_samples_leaf
Amikacina-100%	6	6	0.4	3
Amikacina-97%	10	2	0.4	2
Amikacina-95%	7	6	0.8	2
Ceftazidima-100%	6	6	0.1	1
Ceftazidima-97%	5	2	0.1	1
Ceftazidima-95%	6	2	sqrt	1
Ciprofloxacina-100%	5	2	0.4	2
Ciprofloxacina-97%	10	2	0.4	1
Ciprofloxacina-95%	8	2	0.5	1
Colistina-100%	10	2	sqrt	1
Colistina-97%	7	2	10	2
Colistina-95%	6	2	10	2
Cefepime-100%	5	2	sqrt	1
Cefepime-97%	5	5	sqrt	1
Cefepime-95%	6	2	sqrt	1
Fosfomicina-100%	5	5	0,5	1
Fosfomicina-97%	5	6	0,4	1
Fosfomicina-95%	6	6	auto	1
Imipenem-100%	5	6	sqrt	2
Imipenem-97%	7	2	sqrt	3
Imipenem-95%	8	5	sqrt	2
Meropenem-100%	10	3	sqrt	3
Meropenem-97%	6	2	sqrt	2
Meropenem-95%	5	3	0.4	3

- 100, 97 Y 95: hace referencia al porcentaje de identidad con el que fueron agrupado los genes

Anexo A. Tablas

Modelo	Parámetros			
	Profundidad	min_samples_split	max_features	min_samples_leaf
Tigeciclina-100%	5	3	sqrt	3
Tigeciclina-97%	10	6	sqrt	2
Tigeciclina-95%	6	3	sqrt	3
Tetraciclina-100%	6	2	0.4	3
Tetraciclina-97%	7	3	0.4	3
Tetraciclina-95%	10	5	sqrt	1

- 100, 97 Y 95: hace referencia al porcentaje de identidad con el que fueron agrupado los genes

8. Hiperparámetros seleccionados por *gride search* para el entrenamiento de modelos con regresión *lasso*

Antibiótico	Agrupación 100%		Agrupación 97%		Agrupación 95%	
	Alfa	# características	Alfa	# características	Alfa	# características
Amikacina	0,01	61	0,01	42	0,03	20
Ceftazidima	0,01	57	0,01	40	0,01	40
Ciprofloxacina	0,03	26	0,005	80	0,005	68
Colistina	0,01	50	0,01	32	0,01	30
Cefepime	0,03	21	0,01	49	0,01	51
Fosfomicina	0,07	18	0,1	4	0,05	9
Imipenem	0,005	100	0,005	68	0,005	67
Meropenem	0,01	55	0,01	52	0,01	42
Tetraciclina	0,05	11	0,01	42	0,01	42
Tigeciclina	0,01	55	0,01	45	0,01	44

- 100, 97 Y 95: hace referencia al porcentaje de identidad con el que fueron agrupado los genes

9. Hiperparámetros seleccionados por *grid search* para el entrenamiento de modelos con *gradient boosting*

Modelo	Profundidad	min_sampl es_split	min_samples_le af	max_features	learning_rate
Amikacina-100%	10	6	3	sqrt	0.01
Amikacina-97%	8	2	2	0.6	0.01
Amikacina-95%	10	2	3	sqrt	0.01
Ceftazidima-100%	6	3	1	sqrt	0.01
Ceftazidima-97%	5	6	2	sqrt	0.01
Ceftazidima-95%	7	3	1	sqrt	0.01
Ciprofloxacina-100%	7	5	3	sqrt	0.05
Ciprofloxacina-97%	5	5	1	sqrt	0.05
Ciprofloxacina-95%	5	5	1	sqrt	0.05
Colistina-100%	10	6	1	auto	0.001
Colistina-97%	10	3	2	sqrt	0.001
Colistina-95%	10	2	2	sqrt	0.001
Cefepime-100%	5	6	1	auto	0.01
Cefepime-97%	8	2	1	sqrt	0.01
Cefepime-95%	7	2	1	sqrt	0.01
Fosfomicina-100%	6	6	2	auto	0.01
Fosfomicina-97%	8	6	2	sqrt	0.05
Fosfomicina-95%	6	3	2	sqrt	0.05
Imipenem-100%	5	5	3	sqrt	0.01
Imipenem-97%	10	3	2	sqrt	0.01
Imipenem-95%	7	5	3	sqrt	0.01
Meropenem-100%	8	3	3	sqrt	0.01
Meropenem-97%	5	2	3	sqrt	0.05
Meropenem-95%	8	6	3	0,2	0.01

- 100, 97 Y 95: hace referencia al porcentaje de identidad con el que fueron agrupado los genes

Anexo A. Tablas

Modelo	Profundidad	min_samples_split	min_samples_leaf	max_features	learning_rate
Tigeciclina-100%	8	5	3	sqrt	0.05
Tigeciclina-97%	5	3	2	sqrt	0,05
Tigeciclina-95%	5	3	2	sqrt	0,05
Tetraciclina-100%	10	2	3	sqrt	0.01
Tetraciclina-97%	8	2	3	0,5	0.01
Tetraciclina-95%	6	2	1	sqrt	0.01

- 100, 97 Y 95: hace referencia al porcentaje de identidad con el que fueron agrupado los genes

9. Resultados de modelos desarrollados con *random forest*

Modelo	MAE_test	MSE_test	RMSE_test	VME%	ME%	ACC-2	95% IC	
Amikacina-100%	0,72	0,83	0,91	2,86	1,43	91	84	98
Amikacina-97%	0,70	0,91	0,95	2,86	1,43	91	84	98
Amikacina-95%	0,73	0,94	0,96	2,86	1,43	91	84	98
Ceftazidima-100%	0,65	0,91	0,95	2,86	1,43	90	83	97
Ceftazidima-97%	0,70	1,06	1,03	2,86	1,43	90	83	97
Ceftazidima-95%	0,66	1,00	1,00	2,86	1,43	90	83	97
Ciprofloxacina-100%	0,69	1,1	1,06	1,43	0,00	90	83	97
Ciprofloxacina-97%	0,69	1,22	1,10	1,43	0,00	93	87	99
Ciprofloxacina-95%	0,64	1,23	1,10	1,43	0,00	94	88	100
Colistina-100%	0,90	1,82	1,35	8,57	0,00	88	80	96
Colistina-97%	0,97	1,88	1,37	8,57	0,00	88	84	92
Colistina-95%	0,97	1,89	1,38	8,57	0,00	88	84	92
Cefepime-100%	0,98	1,66	1,28	7,14	0,00	80	71	89
Cefepime-97%	0,76	1,26	1,12	1,43	1,43	83	74	92
Cefepime-95%	0,76	1,25	1,11	1,43	1,43	87	79	95
Fosfomicina-100%	1,26	2,27	1,50	0,00	0,00	61	50	72
Fosfomicina-97%	1,18	2,29	1,50	0,00	0,00	61	50	72
Fosfomicina-95%	1,28	2,32	1,52	0,00	0,00	61	50	72
Imipenem-100%	0,77	1,49	1,20	0,00	2,86	91	84	98
Imipenem-97%	0,76	1,50	1,22	0,00	2,86	88	80	96
Imipenem-95%	0,78	1,48	1,21	0,00	2,86	91	84	98

- 100, 97 Y 95: hace referencia al porcentaje de identidad con el que fueron agrupado los genes
- **MAE**: error absoluto medio, **MSE**: error cuadrático medio, **RMSE**: raíz del error cuadrático medio, **VME**: *very major error* expresado en porcentaje, **ME**: *major error* expresado en porcentaje, **ACC-2**: precisión dentro de +/- 1 dilución doble, **95% I.C** = intervalo de confianza al 95%

Anexo A. Tablas

Modelo	MAE_test	MSE_test	RMSE_test	VME%	ME%	ACC-2	95% IC	
Meropenem-100%	0,71	1,93	1,39	0,00	5,71	88	80	96
Meropenem-97%	0,74	1,88	1,37	1,43	4,29	87	79	95
Meropenem-95%	0,70	1,87	1,37	2,86	4,29	88	80	96
Tigeciclina-100%	0,77	1,18	1,08	0,00	0,00	86	78	94
Tigeciclina-97%	0,75	1,19	1,09	0,00	0,00	86	78	94
Tigeciclina-95%	0,80	1,30	1,14	0,00	0,00	86	78	94
Tetraciclina-100%	0,86	1,23	1,11	0,00	4,29	83	74	92
Tetraciclina-97%	0,86	1,18	1,08	0,00	4,29	77	67	87
Tetraciclina-95%	0,89	1,28	1,13	0,00	4,29	83	74	92

- 100, 97 Y 95: hace referencia al porcentaje de identidad con el que fueron agrupado los genes
- **MAE**: error absoluto medio, **MSE**: error cuadrático medio, **RMSE**: raíz del error cuadrático medio, **VME**: *very major error* expresado en porcentaje, **ME**: *major error* expresado en porcentaje, **ACC-2**: precisión dentro de +/- 1 dilución doble, **95% I.C** = intervalo de confianza al 95%

10. Resultados de modelos desarrollados con regresión *lasso*

Modelo	MAE	MSE	RMSE	VME%	ME%	ACC-2	ACC-2 95% IC	
Amikacina-100%	0,80	1,08	1,04	4,29	2,86	89	82	96
Amikacina-97%	0,81	1,1	1,05	4,29	1,43	90	83	97
Amikacina-95%	0,82	1,14	1,07	7,14	1,43	86	78	94
Ceftazidima-100%	0,65	1,12	1,06	2,86	4,29	90	83	97
Ceftazidima-97%	0,6	1,06	1,03	2,86	2,86	90	83	97
Ceftazidima-95%	0,59	1,06	1,03	2,86	2,86	90	83	97
Ciprofloxacina-100%	0,78	1,6	1,26	2,86	0,00	90	83	97
Ciprofloxacina-97%	0,65	1,2	1,10	1,43	0,00	93	87	99
Ciprofloxacina-95%	0,62	1,26	1,12	1,43	0,00	91	84	98
Colistina-100%	0,8	1,97	1,40	8,57	0,00	86	78	94
Colistina-97%	0,88	1,9	1,38	8,57	0,00	88	80	96
Colistina-95%	0,8	1,91	1,38	8,57	0,00	88	80	96
Cefepime-100%	0,95	1,88	1,37	4,29	0,00	81	72	90
Cefepime-97%	0,71	1,19	1,09	1,43	1,43	90	83	97
Cefepime-95%	0,72	1,16	1,08	0,00	1,43	92	86	98
Fosfomicina-100%	1,28	2,28	1,51	0,00	0,00	61	50	72
Fosfomicina-97%	1,38	2,83	1,68	0,00	0,00	61	50	72
Fosfomicina-95%	1,33	2,78	1,67	0,00	0,00	61	50	72
Imipenem-100%	0,72	1,5	1,22	1,43	2,86	90	83	97
Imipenem-97%	0,59	1,34	1,16	0,00	1,43	93	87	99
Imipenem-95%	0,55	1,2	1,10	0,00	1,43	94	88	100

- 100, 97 Y 95: hace referencia al porcentaje de identidad con el que fueron agrupado los genes
- **MAE**: error absoluto medio, **MSE**: error cuadrático medio, **RMSE**: raíz del error cuadrático medio, **VME**: *very major error* expresado en porcentaje, **ME**: *major error* expresado en porcentaje, **ACC-2**: precisión dentro de +/- 1 dilución doble, **95% I.C** = intervalo de confianza al 95%

Anexo A. Tablas

Modelo	MAE	MSE	RMSE	VME%	ME%	ACC-2	ACC-2 95% IC	
Meropenem-100%	0,72	1,8	1,34	1,43	2,86	87	79	95
Meropenem-97%	0,66	1,9	1,38	2,86	1,43	88	80	96
Meropenem-95%	0,62	1,78	1,33	2,86	4,29	91	84	98
Tigeciclina-100%	0,7	1,27	1,13	0,00	0,00	88	80	96
Tigeciclina-97%	0,67	1,29	1,14	0,00	0,00	88	80	96
Tigeciclina-95%	0,67	1,24	1,11	0,00	0,00	88	80	96
Tetraciclina-100%	0,83	1,17	1,08	0,00	4,29	83	74	92
Tetraciclina-97%	0,88	1,39	1,18	2,86	1,43	75	65	85
Tetraciclina-95%	0,83	1,33	1,15	2,86	1,43	75	65	85

- 100, 97 Y 95: hace referencia al porcentaje de identidad con el que fueron agrupado los genes
- **MAE**: error absoluto medio, **MSE**: error cuadrático medio, **RMSE**: raíz del error cuadrático medio, **VME**: *very major error* expresado en porcentaje, **ME**: *major error* expresado en porcentaje, **ACC-2**: precisión dentro de +/- 1 dilución doble, **95% I.C** = intervalo de confianza al 95%

11. Resultados de modelos desarrollados con *gradient boosting*

Modelo	MAE	MSE	RMSE	VME%	ME%	ACC-2	ACC-2 95% IC	
Amikacina-100%	0,8	0,99	0,99	10,00	8,57	86	78	94
Amikacina-97%	0,68	0,76	0,87	1,43	1,43	94	88	00
Amikacina-95%	0,83	1,02	1,01	4,29	0,00	80	71	89
Ceftazidima-100%	0,66	1,07	1,03	0,00	1,43	93	87	99
Ceftazidima-97%	0,67	1,02	1,01	1,43	1,43	91	84	98
Ceftazidima-95%	0,66	1,06	1,03	1,43	1,43	91	84	98
Ciprofloxacina-100%	0,7	1,18	1,08	1,43	1,43	88	80	96
Ciprofloxacina-97%	0,73	1,39	1,18	1,43	0,00	93	87	99
Ciprofloxacina-95%	0,71	1,24	1,11	1,43	0,00	91	84	98
Colistina-100%	1,17	2,8	1,67	14,29	0,00	80	71	89
Colistina-97%	1,18	2,92	1,67	14,29	0,00	80	71	89
Colistina-95%	1,18	2,83	1,68	14,29	0,00	80	71	89
Cefepime-100%	0,99	1,72	1,31	1,43	0,00	83	74	92
Cefepime-97%	0,72	1,3	1,14	1,43	0,00	83	74	92
Cefepime-95%	0,71	1,32	1,15	1,43	1,43	80	71	89
Fosfomicina-100%	1,34	2,4	1,55	0,00	0,00	61	50	72
Fosfomicina-97%	1,23	2,38	1,54	0,00	0,00	61	50	72
Fosfomicina-95%	1,19	2,35	1,53	0,00	0,00	55	43	67
Imipenem-100%	0,95	1,76	1,32	0,00	2,86	84	75	93
Imipenem-97%	0,88	1,64	1,28	0,00	2,86	90	83	97
Imipenem-95%	0,91	1,67	1,29	0,00	2,86	87	79	95

- 100, 97 Y 95: hace referencia al porcentaje de identidad con el que fueron agrupado los genes
- **MAE**: error absoluto medio, **MSE**: error cuadrático medio, **RMSE**: raíz del error cuadrático medio, **VME**: *very major error* expresado en porcentaje, **ME**: *major error* expresado en porcentaje, **ACC-2**: precisión dentro de +/- 1 dilución doble, **95% I.C** = intervalo de confianza al 95%

Anexo A. Tablas

Modelo	MAE	MSE	RMSE	VME%	ME%	ACC-2	ACC-2 95% IC	
Meropenem-100%	0,92	2,05	1,43	0,00	2,86	83	74	92
Meropenem-97%	0,7	1,96	1,4	1,43	2,86	90	83	97
Meropenem-95%	0,85	1,9	1,38	0,00	4,29	87	79	95
Tetraciclina-100%	0,92	1,61	1,27	1,43	4,29	80	71	89
Tetraciclina-97%	0,88	1,31	1,15	1,43	4,29	73	63	83
Tetraciclina-95%	0,86	1,31	1,14	0,00	1,43	77	67	87
Tigeciclina-100%	0,84	1,38	1,18	0,00	0,00	86	78	94
Tigeciclina-97%	0,85	1,42	1,19	0,00	0,00	83	74	92
Tigeciclina-95%	0,85	1,38	1,16	0,00	0,00	86	78	94

- 100, 97 Y 95: hace referencia al porcentaje de identidad con el que fueron agrupado los genes
- **MAE**: error absoluto medio, **MSE**: error cuadrático medio, **RMSE**: raíz del error cuadrático medio, **VME**: *very major error* expresado en porcentaje, **ME**: *major error* expresado en porcentaje, **ACC-2**: precisión dentro de +/- 1 dilución doble, **95% I.C** = intervalo de confianza al 95%

B. Anexo: Procedimientos

1. Determinación de la concentración mínima inhibitoria: micro dilución en caldo

- Realizar diluciones seriadas dobles de los antibióticos, ~ 10 diluciones.

Nota: tener en consideración las instrucciones dadas por el fabricante del antibiótico. Además, como la solución de antibiótico es inoculada después con una igual cantidad de suspensión bacteriana, las diluciones se realizaran a una concentración dos veces mayor con respecto a la concentración final deseada.

- Retire la placa de microtitulación del envase estéril. Si la placa no tiene tapa, cubrir con otra placa estéril o bien limpia con etanol al 70% para evitar contaminación

Nota: mantener siempre cerrada la tapa.

- Utilizar una fila para cada antibiótico con sus correspondientes 10 diluciones. Toda prueba debe tener un control de esterilidad y control de crecimiento. En el control de esterilidad agregar 200µl y en el de crecimiento 100µl. Figura 1
- Por cada aislamiento bacteriano, agregar 100µl de cada dilución de los diferentes antibióticos en sus correspondientes pocillos.
- Mezclar la suspensión bacteriana por vortex y realizar una dilución 1:100 (800 µl de suspensión bacteriana + 79.200 µl)

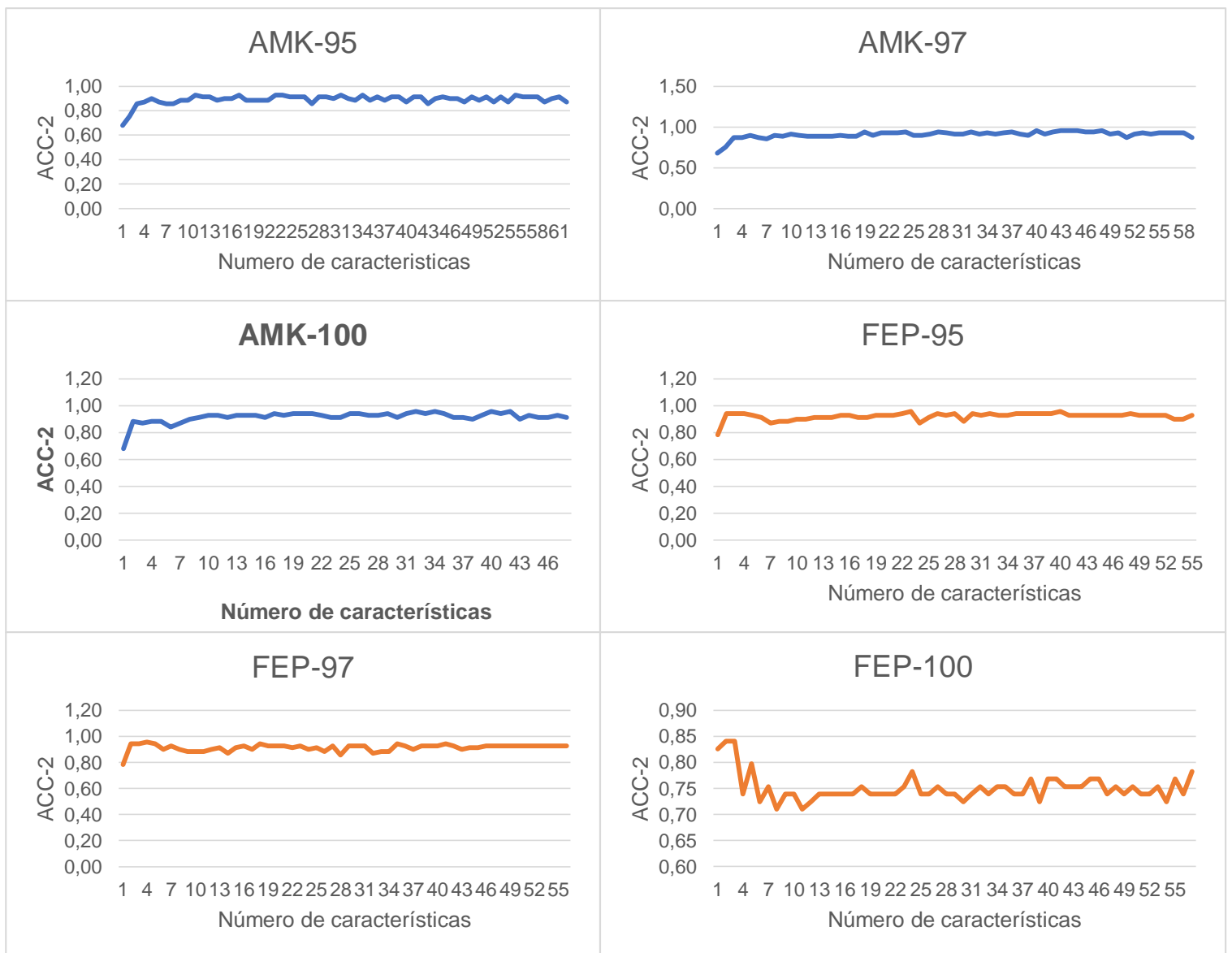
- Inocular en cada pocillo que contiene las diluciones de antibióticos y el control de crecimiento 100µl de suspensión bacteriana. Esto resulta en un inóculo final deseado de 5×10^5 UFC/ml
- Incubar la placa de microtitulación a 37°C por 16 horas

Nota: tener en cuenta que durante la incubación no se evapore el contenido de la placa de microtitulación. No apilar más de 4 placas.

- Revisar resultados al siguiente día.
- Limpiar la parte de debajo de la placa para eliminar condensación que pueda haber
- Para que la prueba sea válida se requiere que el control de crecimiento contenga suficiente crecimiento y que el control de esterilidad no se presente turbio. (Weinstein & Clinical and Laboratory Standards Institute, 2019; Wiegand et al., 2008)

C. Anexo: Figuras

1. Precisión dentro de +/- 1 dilución doble en modelos de *random forest*



CAZ-95



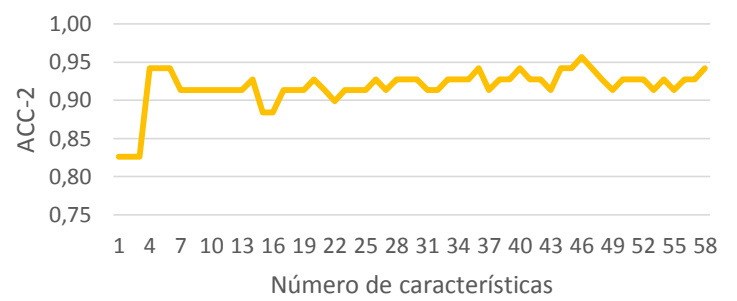
CAZ-97



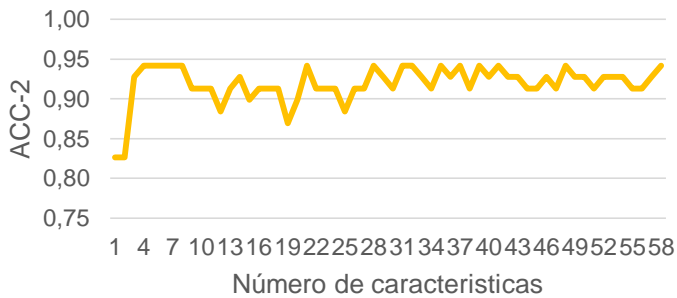
CAZ-100



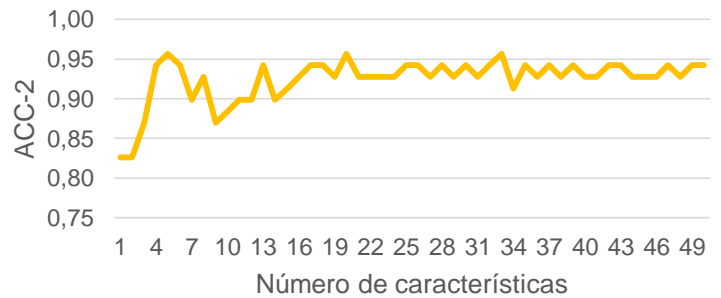
CIP-95



CIP-97



CIP-100

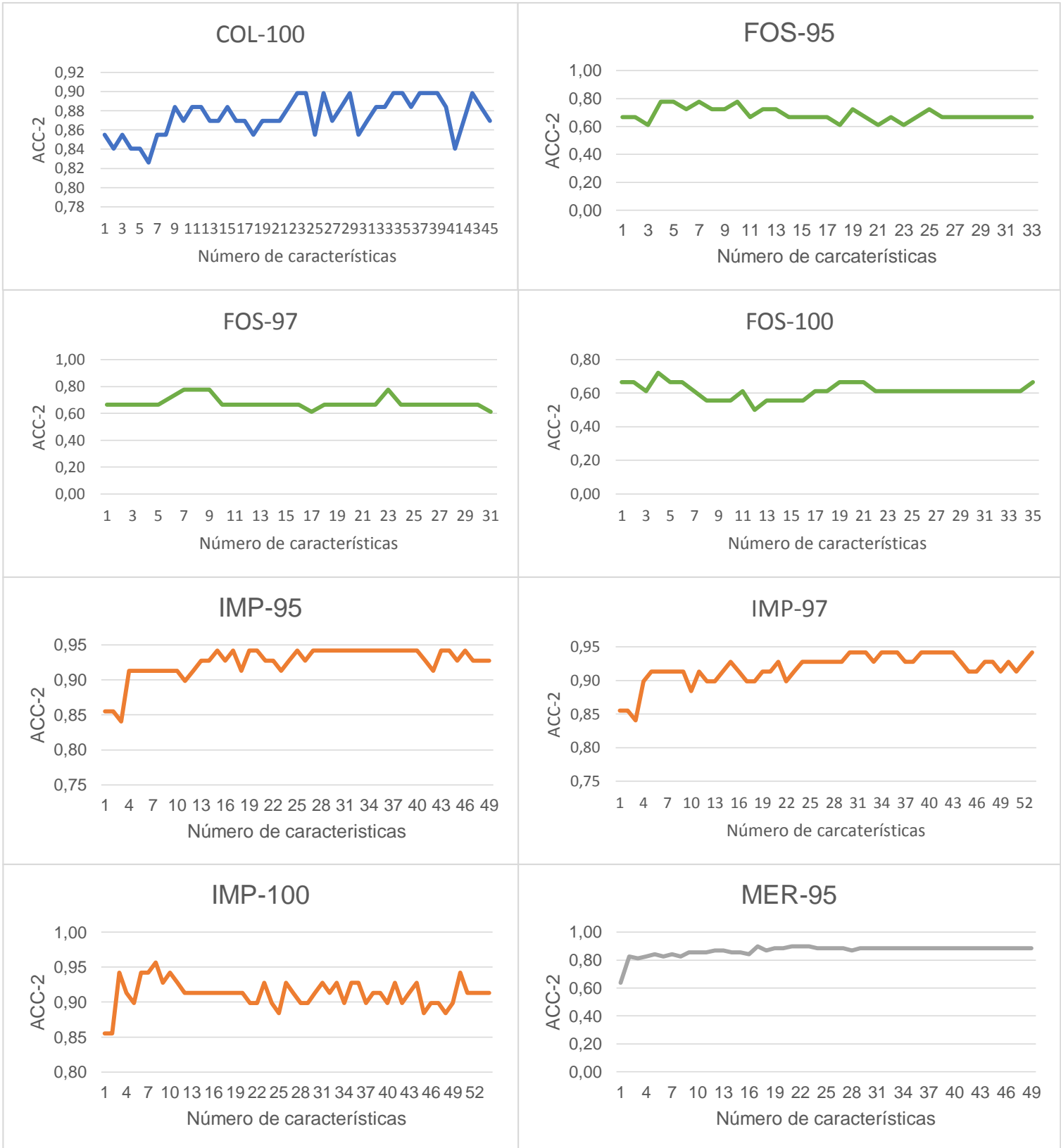


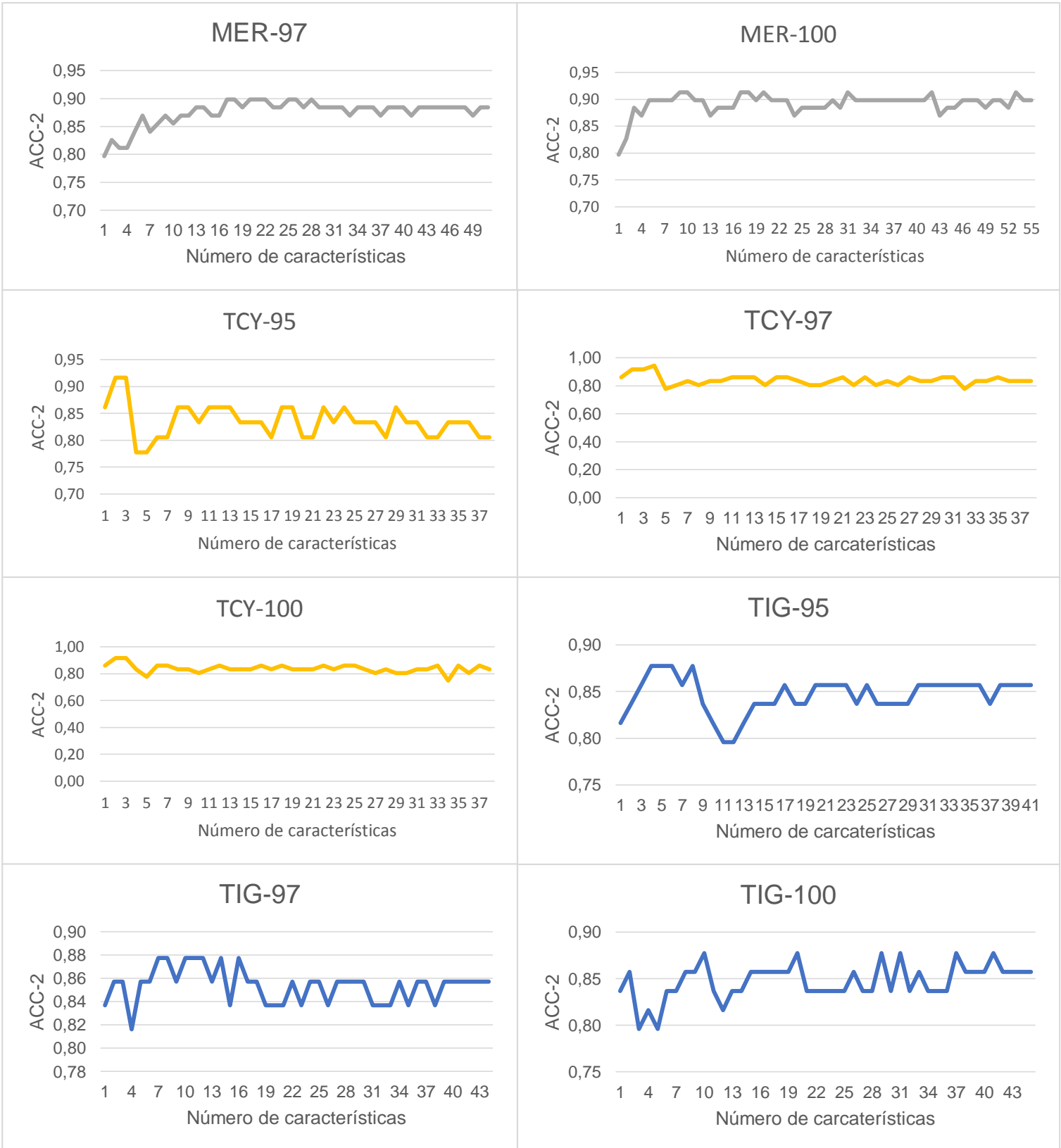
COL-95



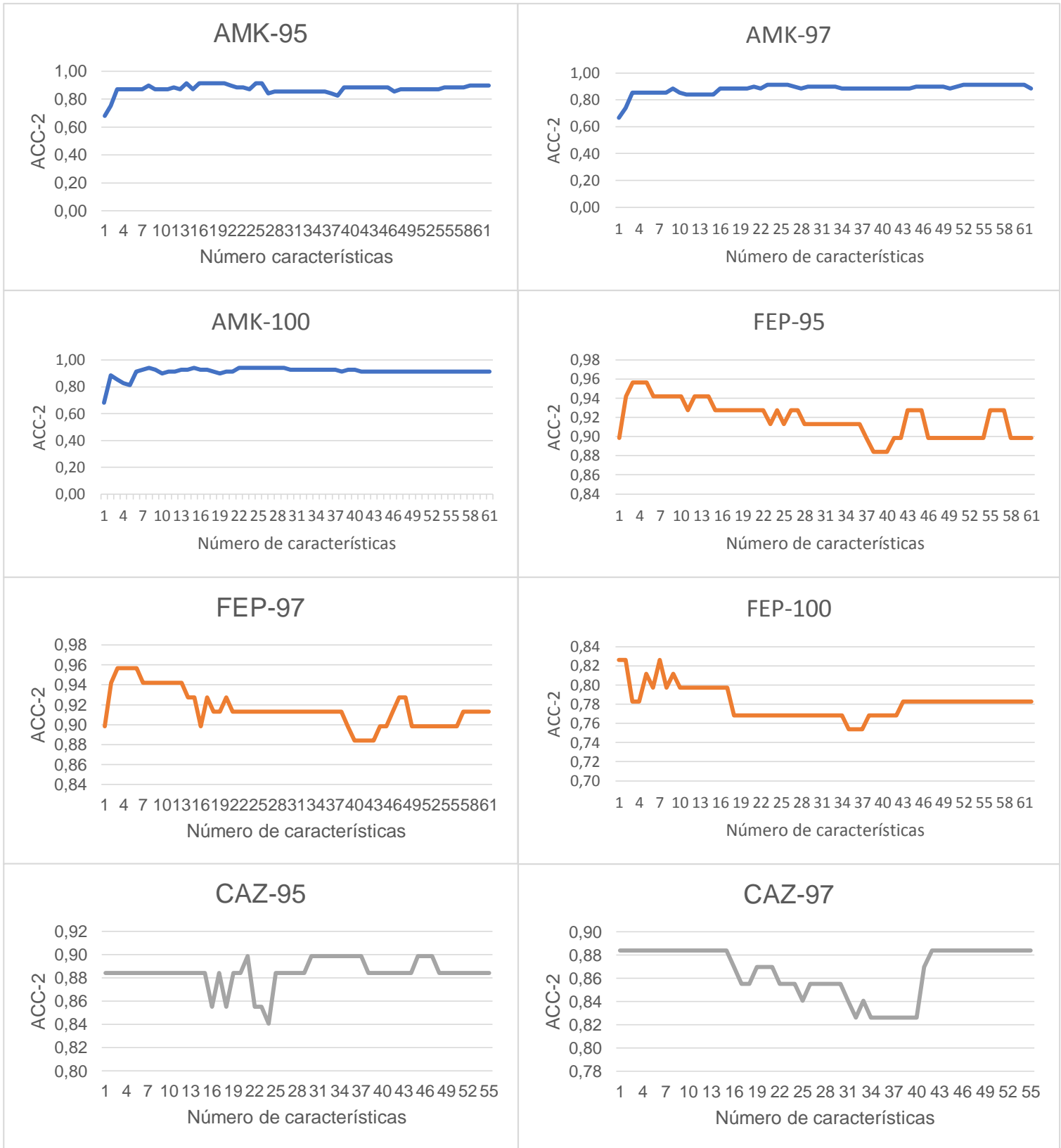
COL-97

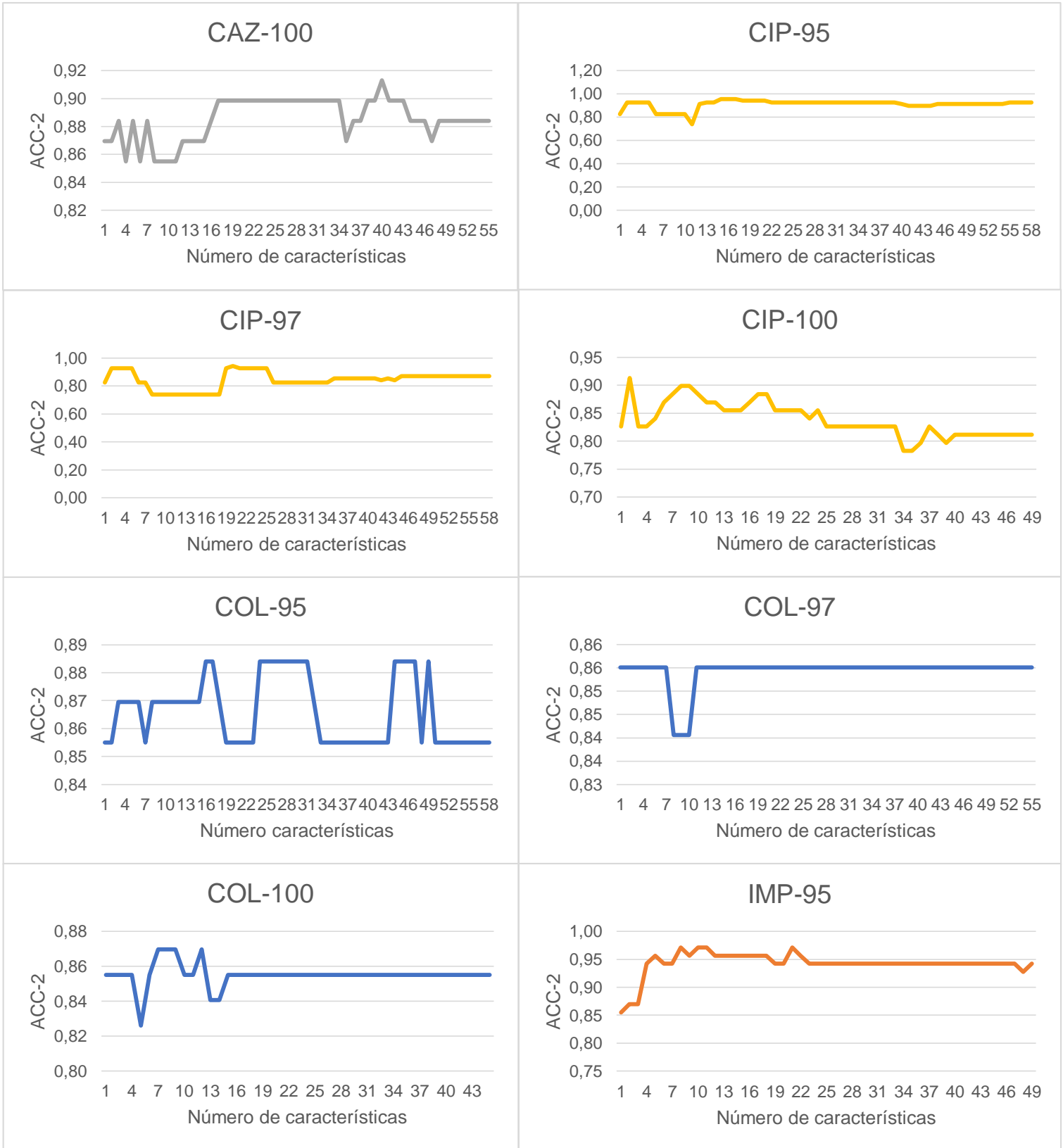


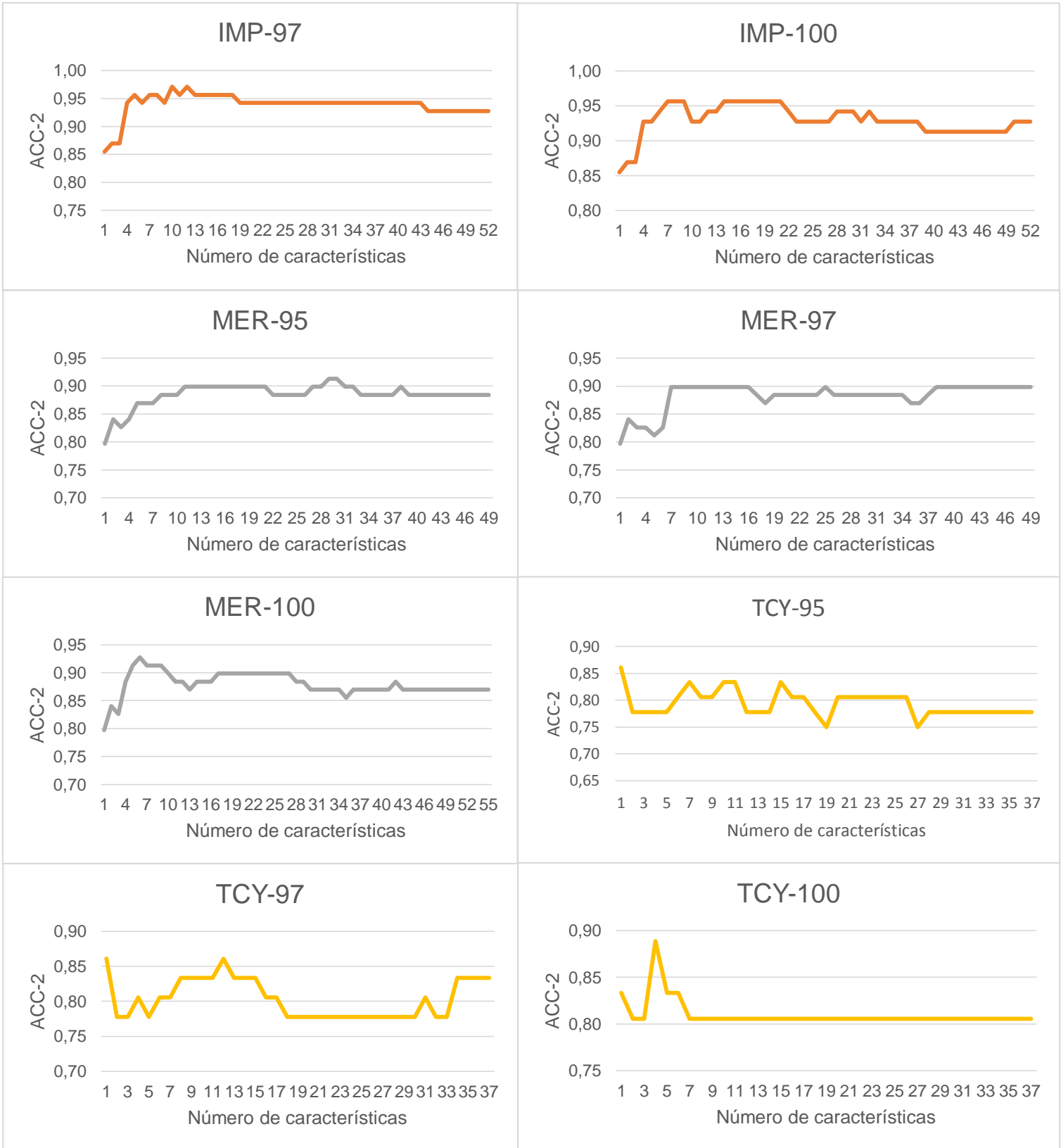


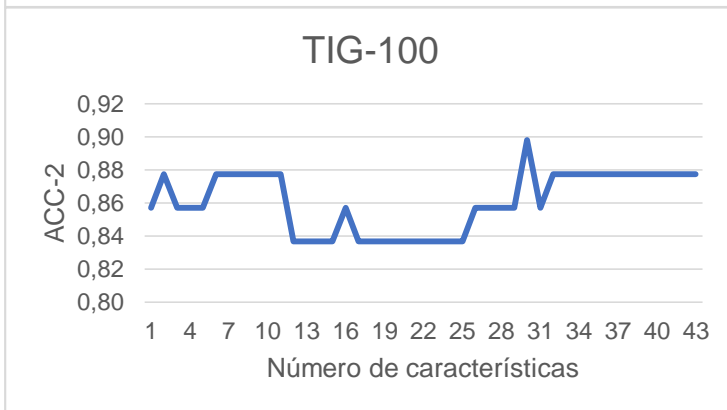
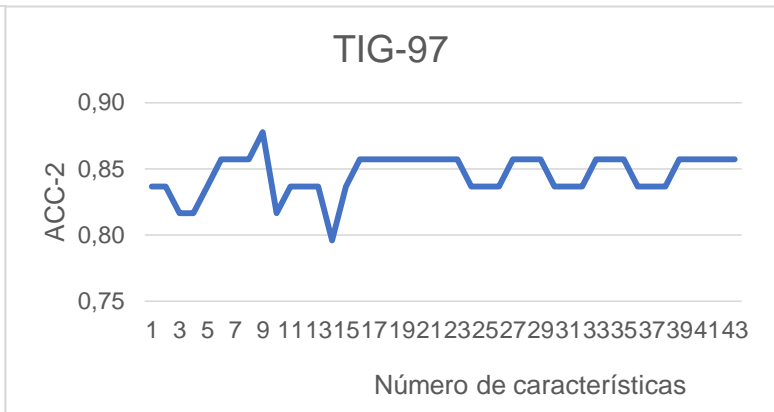
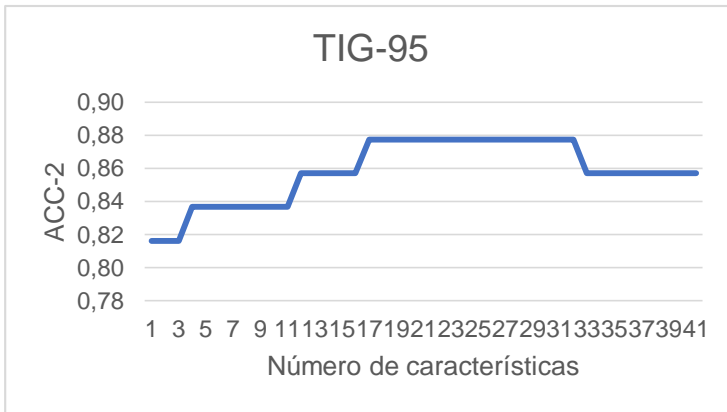


2. Precisión dentro de +/- 1 dilución doble en modelos de regresión *lasso*









Bibliografía

- Abdi, S. N., Ghotaslou, R., Ganbarov, K., Mobed, A., Tanomand, A., Yousefi, M., Asgharzadeh, M., & Kafil, H. S. (2020). Acinetobacter baumannii Efflux Pumps and Antibiotic Resistance. *Infection and Drug Resistance*, 13, 423-434. <https://doi.org/10.2147/IDR.S228089>
- Albornoz, H., Paciel, D., & Medina, J. (2018). *Tratamiento de Acinetobacter spp extremadamente resistente (XDR) y panresistente (PR)* (p. 6). Catedra de Enfermedades Infecciosas. http://www.infectologia.edu.uy/images/archivos/Esquemas_Tto_Acineto_XDR.pdf
- Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H.-K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., ... McArthur, A. G. (2020). CARD 2020: Antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 48(D1), D517-D525. <https://doi.org/10.1093/nar/gkz935>
- AL-Enawey, A. W., Saadedin, S. M. K., & Al-Khaldi, S. A.-M. (2020). ISOLATION, IDENTIFICATION AND ANTIMICROBIAL SUSCEPTIBILITY OF STAPHYLOCOCCUS AUREUS AND ACINETOBACTER BAUMANNII ISOLATED FROM BABYLON HOSPITALS. 6.
- Alzuhairi, M. A., Abdulmohsen, A. M., Falih, M. N., & Hanafiah, M. M. (2020). Genomic sequencing analysis of Acinetobacter baumannii strain ABIQM1, isolated from a meningitis patient. *Gene Reports*, 19, 100631. <https://doi.org/10.1016/j.genrep.2020.100631>
- Asif, M., Alvi, I. A., & Rehman, S. U. (2018). Insight into Acinetobacter baumannii: Pathogenesis, global resistance, mechanisms of resistance, treatment options, and alternative modalities. *Infection and Drug Resistance*, 11, 1249-1260. <https://doi.org/10.2147/IDR.S166750>

- Aun, E., Brauer, A., Kisand, V., Tenson, T., & Remm, M. (2018). A k-mer-based method for the identification of phenotype-associated genomic biomarkers and predicting phenotypes of sequenced bacteria. *PLOS Computational Biology*, *14*(10), e1006434. <https://doi.org/10.1371/journal.pcbi.1006434>
- Ayodele, T. (2010). Types of Machine Learning Algorithms. En Yagang Zhang (Ed.), *New Advances in Machine Learning*. InTech. <https://doi.org/10.5772/9385>
- Aytan-Aktug, D., Clausen, P. T. L. C., Bortolaia, V., Aarestrup, F. M., & Lund, O. (2020). Prediction of Acquired Antimicrobial Resistance for Multiple Bacterial Species Using Neural Networks. *MSystems*, *5*(1), e00774-19, /msystems/5/1/msys.00774-19.atom. <https://doi.org/10.1128/mSystems.00774-19>
- Babcock University, F.Y. O., J.E.T, A., O, A., J. O, H., O, O., & J, A. (2017). Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology*, *48*(3), 128-138. <https://doi.org/10.14445/22312803/IJCTT-V48P126>
- Barlam, T. F., Cosgrove, S. E., Abbo, L. M., MacDougall, C., Schuetz, A. N., Septimus, E. J., Srinivasan, A., Dellit, T. H., Falck-Ytter, Y. T., Fishman, N. O., Hamilton, C. W., Jenkins, T. C., Lipsett, P. A., Malani, P. N., May, L. S., Moran, G. J., Neuhauser, M. M., Newland, J. G., Ohl, C. A., ... Trivedi, K. K. (2016). Implementing an Antibiotic Stewardship Program: Guidelines by the Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America. *Clinical Infectious Diseases*, *62*(10), e51-e77. <https://doi.org/10.1093/cid/ciw118>
- Barletta Farías, R., Pérez Ponce, L., Castro Vega, G., Pujol Pérez, M., Barletta del Castillo, J., & Dueñas Pérez, Y. (2018). Acinetobacter baumannii multirresistente: Un reto para la terapéutica actual. *MediSur*, *16*(2), 322-334. http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1727-897X2018000200015&lng=es&nrm=iso&tling=es
- BioSample - NCBI. (2020). National Center for Biotechnology Information. Retrieved January 06, 2020, from <https://www.ncbi.nlm.nih.gov/biosample>
- Biswas, Raju, Panja, Anindya Sundar, & Bandopadhyay, Rajib. (2019). *Molecular Mechanism of Antibiotic Resistance: The Untouched Area of Future Hope | SpringerLink*. <https://link.springer.com/article/10.1007/s12088-019-00781-6>

- Bottou, L. (2010). Large-Scale Machine Learning with Stochastic Gradient Descent. En Y. Lechevallier & G. Saporta (Eds.), *Proceedings of COMPSTAT'2010* (pp. 177-186). Physica-Verlag HD. https://doi.org/10.1007/978-3-7908-2604-3_16
- Bradley, P., Gordon, N. C., Walker, T. M., Dunn, L., Heys, S., Huang, B., Earle, S., Pankhurst, L. J., Anson, L., Cesare, M. de, Piazza, P., Votintseva, A. A., Golubchik, T., Wilson, D. J., Wyllie, D. H., Diel, R., Niemann, S., Feuerriegel, S., Kohl, T. A., ... Iqbal, Z. (2015). Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nature Communications*, 6(1), 1-15. <https://doi.org/10.1038/ncomms10063>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Brennan, L. A., Tri, A. N., & Marcot, B. G. (2019). *Quantitative Analyses in Wildlife Science*. JHU Press.
- Cai, H. Y., McDowall, R., Parker, L., Kaufman, E. I., & Caswell, J. L. (2019). Changes in antimicrobial susceptibility profiles of *Mycoplasma bovis* over time. *Canadian Journal of Veterinary Research = Revue Canadienne de Recherche Veterinaire*, 83(1), 34-41. <https://europepmc.org/articles/PMC6318825/>
- Canovas-Segura, B., Morales, A., Lopez Martinez-Carrasco, A., Campos, M., Juarez, J. M., Lopez Rodriguez, L., & Palacios, F. (2019). Improving Interpretable Prediction Models for Antimicrobial Resistance. *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, 543-546. <https://doi.org/10.1109/CBMS.2019.00111>
- Cao, J., Wang, J., Wang, Y., Wang, L., Bi, Y., Zhu, B., & Gao, G. F. (2020). Tigecycline resistance tet(X3) gene is going wild. *Biosafety and Health*, 2(1), 9-11. <https://doi.org/10.1016/j.bsheal.2020.01.002>
- Casellas, J. M. (2011). *Resistencia a los antibacterianos en América Latina: Consecuencias para la infectología*. 16.
- Centers for Disease Control and Prevention (U.S.). (2019). *Antibiotic resistance threats in the United States, 2019*. Centers for Disease Control and Prevention (U.S.). <https://doi.org/10.15620/cdc:82532>
- Cerezales, M., Xanthopoulou, K., Wille, J., Bustamante, Z., Seifert, H., Gallego, L., & Higgins, P. G. (2019). *Acinetobacter baumannii* analysis by core genome multi-locus sequence typing in two hospitals in Bolivia: Endemicity of international clone

- 7 isolates (CC25). *International Journal of Antimicrobial Agents*, 53(6), 844-849. <https://doi.org/10.1016/j.ijantimicag.2019.03.019>
- Chakravarty, B. (2020). Genetic mechanisms of antibiotic resistance and virulence in *Acinetobacter baumannii*: Background, challenges and future prospects. *Molecular Biology Reports*, 47(5), 4037-4046. <https://doi.org/10.1007/s11033-020-05389-4>
- Chen, C., Clark, C. G., Langner, S., Boyd, D. A., Bharat, A., McCorrister, S. J., McArthur, A. G., Graham, M. R., Westmacott, G. R., & Van Domselaar, G. (2019). Detection of Antimicrobial Resistance Using Proteomics and the Comprehensive Antibiotic Resistance Database: A Case Study. *PROTEOMICS – Clinical Applications*, 1800182. <https://doi.org/10.1002/prca.201800182>
- Chen, M. L., Doddi, A., Royer, J., Freschi, L., Schito, M., Ezewudo, M., Kohane, I. S., Beam, A., & Farhat, M. (2019). Beyond multidrug resistance: Leveraging rare variants with machine and statistical learning models in *Mycobacterium tuberculosis* resistance prediction. *EBioMedicine*, 43, 356-369. <https://doi.org/10.1016/j.ebiom.2019.04.016>
- Chen, X., & Ishwaran, H. (2012). Random forests for genomic data analysis. *Genomics*, 99(6), 323-329. <https://doi.org/10.1016/j.ygeno.2012.04.003>
- Clinical and Laboratory Standards Institute, & Weinstein, M. P. (2018). *Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically*.
- Correa, A., del Campo, R., Escandón-Vargas, K., Perenguez, M., Rodríguez-Baños, M., Hernández-Gómez, C., Pallares, C., Perez, F., Arias, C. A., Cantón, R., & Villegas, M. V. (2017). Distinct Genetic Diversity of Carbapenem-Resistant *Acinetobacter baumannii* from Colombian Hospitals. *Microbial Drug Resistance*, 24(1), 48-54. <https://doi.org/10.1089/mdr.2016.0190>
- Corredor, S. M., Leal, A. L., Castillo, J. S., Gutierrez, G. B., Andrade, L. C., & Rodriguez, D. (2018). PROGRAMA DE PREVENCIÓN, VIGILANCIA Y CONTROL DE INFECCIONES ASOCIADAS A LA ATENCIÓN EN SALUD-IAAS Y LA RESISTENCIA ANTIMICROBIANA. 64.
- Cui, Z.-H., Ni, W.-N., Tang, T., He, B., Zhong, Z.-X., Fang, L.-X., Chen, L., Chen, C., Cui, C.-Y., Liu, Y.-H., Liao, X.-P., & Sun, J. (2020). Rapid detection of plasmid-mediated high-level tigecycline resistance in *Escherichia coli* and *Acinetobacter* spp. *Journal of Antimicrobial Chemotherapy*, 75(6), 1479-1483. <https://doi.org/10.1093/jac/dkaa029>

- Davis, J. J., Boisvert, S., Brettin, T., Kenyon, R. W., Mao, C., Olson, R., Overbeek, R., Santerre, J., Shukla, M., Wattam, A. R., Will, R., Xia, F., & Stevens, R. (2016). Antimicrobial Resistance Prediction in PATRIC and RAST. *Scientific Reports*, *6*, 27930. <https://doi.org/10.1038/srep27930>
- De Bruyne, S., Speeckaert, M. M., Van Biesen, W., & Delanghe, J. R. (2020). Recent evolutions of machine learning applications in clinical laboratory medicine. *Critical Reviews in Clinical Laboratory Sciences*, 1-22. <https://doi.org/10.1080/10408363.2020.1828811>
- De'ath, G., & Fabricius, K. E. (2000). Classification and Regression Trees: A Powerful yet Simple Technique for Ecological Data Analysis. *Ecology*, *81*(11), 3178-3192. [https://doi.org/10.1890/0012-9658\(2000\)081\[3178:CARTAP\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2)
- Deelder, W., Christakoudi, S., Phelan, J., Benavente, E. D., Campino, S., McNerney, R., Palla, L., & Clark, T. G. (2019). Machine Learning Predicts Accurately Mycobacterium tuberculosis Drug Resistance From Whole Genome Sequencing Data. *Frontiers in Genetics*, *10*. <https://doi.org/10.3389/fgene.2019.00922>
- Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A., & Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics*, *13*(9), 601-612. <https://doi.org/10.1038/nrg3226>
- Drouin, A., Giguère, S., Déraspe, M., Marchand, M., Tyers, M., Loo, V. G., Bourgault, A.-M., Laviolette, F., & Corbeil, J. (2016). Predictive computational phenotyping and biomarker discovery using reference-free genome comparisons. *BMC Genomics*, *17*(1), 754. <https://doi.org/10.1186/s12864-016-2889-6>
- Drouin, A., Letarte, G., Raymond, F., Marchand, M., Corbeil, J., & Laviolette, F. (2019). Interpretable genotype-to-phenotype classifiers with performance guarantees. *Scientific Reports*, *9*(1), 4071. <https://doi.org/10.1038/s41598-019-40561-2>
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460-2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Eliopoulos, G. M., Maragakis, L. L., & Perl, T. M. (2008). Acinetobacter baumannii: Epidemiology, Antimicrobial Resistance, and Treatment Options. *Clinical Infectious Diseases*, *46*(8), 1254-1263. <https://doi.org/10.1086/529198>
- Eyre, D. W., De Silva, D., Cole, K., Peters, J., Cole, M. J., Grad, Y. H., Demczuk, W., Martin, I., Mulvey, M. R., Crook, D. W., Walker, A. S., Peto, T. E. A., & Paul, J. (2017). WGS

- to predict antibiotic MICs for *Neisseria gonorrhoeae*. *Journal of Antimicrobial Chemotherapy*, 72(7), 1937-1947. <https://doi.org/10.1093/jac/dkx067>
- Fàbrega, A., Madurga, S., Giralt, E., & Vila, J. (2009). Mechanism of action of and resistance to quinolones. *Microbial Biotechnology*, 2(1), 40-61. <https://doi.org/10.1111/j.1751-7915.2008.00063.x>
- Gao, L., Lyu, Y., & Li, Y. (2017). Trends in Drug Resistance of *Acinetobacter baumannii* over a 10-year Period: Nationwide Data from the China Surveillance of Antimicrobial Resistance Program. *Chinese Medical Journal*, 130(6), 659-664. <https://doi.org/10.4103/0366-6999.201601>
- Garnacho-Montero, J., & Timsit, J.-F. (2019). Managing *Acinetobacter baumannii* infections. *Current Opinion in Infectious Diseases*, 32(1), 69–76. <https://doi.org/10.1097/QCO.0000000000000518>
- Geisinger, E., Vargas-Cuebas, G., Mortman, N. J., Syal, S., Dai, Y., Wainwright, E. L., Lazinski, D., Wood, S., Zhu, Z., Anthony, J., Opijnen, T. van, & Isberg, R. R. (2019). The Landscape of Phenotypic and Transcriptional Responses to Ciprofloxacin in *Acinetobacter baumannii*: Acquired Resistance Alleles Modulate Drug-Induced SOS Response and Prophage Replication. *MBio*, 10(3). <https://doi.org/10.1128/mBio.01127-19>
- Genome - NCBI. (2020). National Center for Biotechnology Information. Retrieved January 06, 2020, from <https://www.ncbi.nlm.nih.gov/genome>
- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc.
- Giacobbe, D. R., Mora, S., Giacomini, M., & Bassetti, M. (2020). Machine Learning and Multidrug-Resistant Gram-Negative Bacteria: An Interesting Combination for Current and Future Research. *Antibiotics*, 9(2), 54. <https://doi.org/10.3390/antibiotics9020054>
- Goldstein, E., MacFadden, D. R., & Lipsitch, M. (2018). Antimicrobial resistance and use, and rates of hospitalization associated with bacterial infections, including sepsis. *arXiv:1803.07189 [q-bio]*. <http://arxiv.org/abs/1803.07189>
- Gómez, R. F., Castillo, A., Chávez-Vivas, M., Gómez, R. F., Castillo, A., & Chávez-Vivas, M. (2017). Characterization of multidrug-resistant *Acinetobacter* ssp. Strains

- isolated from medical intensive care units in Cali—Colombia. *Colombia Médica*, 48(4), 183-190. <https://doi.org/10.25100/cm.v48i4.2858>
- Gordon, N. C., Price, J. R., Cole, K., Everitt, R., Morgan, M., Finney, J., Kearns, A. M., Pichon, B., Young, B., Wilson, D. J., Llewelyn, M. J., Paul, J., Peto, T. E. A., Crook, D. W., Walker, A. S., & Golubchik, T. (2014). Prediction of *Staphylococcus aureus* Antimicrobial Resistance by Whole-Genome Sequencing. *Journal of Clinical Microbiology*, 52(4), 1182-1191. <https://doi.org/10.1128/JCM.03117-13>
- Gwinn, M., MacCannell, D., & Armstrong, G. L. (2019). Next-Generation Sequencing of Infectious Pathogens. *JAMA*, 321(9), 893-894. <https://doi.org/10.1001/jama.2018.21669>
- Hamidian, M., Ambrose, S. J., & Hall, R. M. (2016). A large conjugative *Acinetobacter baumannii* plasmid carrying the sul2 sulphonamide and strAB streptomycin resistance genes. *Plasmid*, 87-88, 43-50. <https://doi.org/10.1016/j.plasmid.2016.09.001>
- Hamidian, M., & Nigro, S. J. (2019). Emergence, molecular mechanisms and global spread of carbapenem-resistant *Acinetobacter baumannii*. *Microbial Genomics*. <https://doi.org/10.1099/mgen.0.000306>
- Hasani, A., Sheikhalizadeh, V., Ahangarzadeh Rezaee, M., Rahmati-Yamchi, M., Hasani, A., Ghotaslou, R., & Goli, H. R. (2016). Frequency of Aminoglycoside-Modifying Enzymes and ArmA Among Different Sequence Groups of *Acinetobacter baumannii* in Iran. *Microbial Drug Resistance*, 22(5), 347-353. <https://doi.org/10.1089/mdr.2015.0254>
- Hauskrecht, M., Pelikan, R., Valko, M., & Lyons-Weiler, J. (2011). *Feature Selection and Dimensionality Reduction in Genomics and Proteomics*. 28.
- He, T., Wang, R., Liu, D., Walsh, T. R., Zhang, R., Lv, Y., Ke, Y., Ji, Q., Wei, R., Liu, Z., Shen, Y., Wang, G., Sun, L., Lei, L., Lv, Z., Li, Y., Pang, M., Wang, L., Sun, Q., ... Wang, Y. (2019). Emergence of plasmid-mediated high-level tigecycline resistance genes in animals and humans. *Nature Microbiology*, 4(9), 1450-1456. <https://doi.org/10.1038/s41564-019-0445-2>
- Health, C. for D. and R. (2020). Antimicrobial Susceptibility Test (AST) Systems—Class II Special Controls Guidance for Industry and FDA. *FDA*. <https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and->

- radiation-emitting-products/antimicrobial-susceptibility-test-ast-systems-class-ii-special-controls-guidance-industry-and-fda
- Her, H.-L., & Wu, Y.-W. (2018). A pan-genome-based machine learning approach for predicting antimicrobial resistance activities of the Escherichia coli strains. *Bioinformatics*, *34*(13), i89-i95. <https://doi.org/10.1093/bioinformatics/bty276>
- Hicks, A. L., Wheeler, N., Sánchez-Busó, L., Rakeman, J. L., Harris, S. R., & Grad, Y. H. (2019). Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data. *bioRxiv*, 607127. <https://doi.org/10.1101/607127>
- Hombach, M., Böttger, E. C., & Roos, M. (2013). The critical influence of the intermediate category on interpretation errors in revised EUCAST and CLSI antimicrobial susceptibility testing guidelines. *Clinical Microbiology and Infection*, *19*(2), E59-E71. <https://doi.org/10.1111/1469-0691.12090>
- Honarmand Jahromy, S., Ranjbar, S., & Pishkar, L. (2020). Phenotypic Activity of Efflux Pumps by Carbonyl Cyanide M-Chlorophenyl Hydrazone (CCCP) and Mutations in GyrA and ParC Genes Among Ciprofloxacin-Resistant Acinetobacter baumannii Isolates. *Jundishapur Journal of Microbiology*, *13*(4). <https://doi.org/10.5812/jjm.99435>
- Hyun, J. C., Kavvas, E. S., Monk, J. M., & Palsson, B. O. (2020). Machine learning with random subspace ensembles identifies antimicrobial resistance determinants from pan-genomes of three pathogens. *PLOS Computational Biology*, *16*(3), e1007608. <https://doi.org/10.1371/journal.pcbi.1007608>
- Instituto Nacional de Salud. (2019). *Fichas y Protocolos*. <https://www.ins.gov.co/buscador-eventos/Paginas/Fichas-y-Protocolos.aspx>
- Isler, B., Doi, Y., Bonomo, R. A., & Paterson, D. L. (2019). New Treatment Options against Carbapenem-Resistant Acinetobacter baumannii Infections. *Antimicrobial Agents and Chemotherapy*, *63*(1). <https://doi.org/10.1128/AAC.01110-18>
- Jahan, M. I., Rahaman, M. M., Hossain, M. A., & Sultana, M. (2019). Occurrence of int11-associated VIM-5 carbapenemase and co-existence of all four classes of β -lactamase in carbapenem-resistant clinical Pseudomonas aeruginosa DMC-27b. *Journal of Antimicrobial Chemotherapy*, dkz426. <https://doi.org/10.1093/jac/dkz426>

- Jenuth, J. P. (1999). The NCBI. En S. Misener & S. A. Krawetz (Eds.), *Bioinformatics Methods and Protocols* (pp. 301-312). Humana Press. <https://doi.org/10.1385/1-59259-192-2:301>
- Jeukens, J., Freschi, L., Kukavica-Ibrulj, I., Emond-Rheault, J.-G., Tucker, N. P., & Levesque, R. C. (2019). Genomics of antibiotic-resistance prediction in *Pseudomonas aeruginosa*: Predicting antibiotic resistance in *Pseudomonas*. *Annals of the New York Academy of Sciences*, 1435(1), 5-17. <https://doi.org/10.1111/nyas.13358>
- Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., Lago, B. A., Dave, B. M., Pereira, S., Sharma, A. N., Doshi, S., Courtot, M., Lo, R., Williams, L. E., Frye, J. G., Elsayegh, T., Sardar, D., Westman, E. L., Pawlowski, A. C., ... McArthur, A. G. (2017). CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 45(D1), D566-D573. <https://doi.org/10.1093/nar/gkw1004>
- Jiménez Pearson, M. A., Galas, M., Corso, A., Hormazábal, J. C., Duarte Valderrama, C., Salgado Marcano, N., Ramón-Pardo, P., & Melano, R. G. (2019). Consenso latinoamericano para definir, categorizar y notificar patógenos multirresistentes, con resistencia extendida o panresistentes. *Revista Panamericana de Salud Pública*, 43. <https://doi.org/10.26633/RPSP.2019.65>
- Kankainen, M., Ojala, T., & Holm, L. (2012). BLANNOTATOR: Enhanced homology-based function prediction of bacterial proteins. *BMC Bioinformatics*, 13(1), 33. <https://doi.org/10.1186/1471-2105-13-33>
- Kester, J. C., & Fortune, S. M. (2014). Persisters and beyond: Mechanisms of phenotypic drug resistance and drug tolerance in bacteria. *Critical Reviews in Biochemistry and Molecular Biology*, 49(2), 91-101. <https://doi.org/10.3109/10409238.2013.869543>
- Khaledi, A., Weimann, A., Schniederjans, M., Asgari, E., Kuo, T.-H., Oliver, A., Cabot, G., Kola, A., Gastmeier, P., Hogardt, M., Jonas, D., Mofrad, M. R. K., Bremges, A., McHardy, A. C., & Häussler, S. (2019). Fighting antimicrobial resistance in *Pseudomonas aeruginosa* with machine learning-enabled molecular diagnostics. *BioRxiv*, 643676. <https://doi.org/10.1101/643676>
- Khurshid, M., Rasool, M. H., Siddique, M. H., Azeem, F., Naeem, M., Sohail, M., Sarfraz, M., Saqalein, M., Taj, Z., Nisar, M. A., Qamar, M. U., & Shahzad, A. (2019). Molecular mechanisms of antibiotic co-resistance among carbapenem resistant

- Acinetobacter baumannii. *The Journal of Infection in Developing Countries*, 13(10), 899-905. <https://doi.org/10.3855/jidc.11410>
- Kim, J., Greenberg, D. E., Pifer, R., Jiang, S., Xiao, G., Shelburne, S. A., Koh, A., Xie, Y., & Zhan, X. (2020). VAMPr: VAriant Mapping and Prediction of antibiotic resistance via explainable features and machine learning. *PLOS Computational Biology*, 16(1), e1007511. <https://doi.org/10.1371/journal.pcbi.1007511>
- Köser, C. U., Ellington, M. J., Cartwright, E. J. P., Gillespie, S. H., Brown, N. M., Farrington, M., Holden, M. T. G., Dougan, G., Bentley, S. D., Parkhill, J., & Peacock, S. J. (2012). Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. *PLOS Pathogens*, 8(8), e1002824. <https://doi.org/10.1371/journal.ppat.1002824>
- Labreche, M. J., Graber, C. J., & Nguyen, H. M. (2015). Recent Updates on the Role of Pharmacokinetics-pharmacodynamics in Antimicrobial Susceptibility Testing as Applied to Clinical Practice. *Clinical Infectious Diseases*, 61(9), 1446-1452. <https://doi.org/10.1093/cid/civ498>
- Ladner, J. T. (2019). Precision epidemiology for infectious disease control. *Nature Medicine*, 25, 6.
- Lee, C.-R., Lee, J. H., Park, M., Park, K. S., Bae, I. K., Kim, Y. B., Cha, C.-J., Jeong, B. C., & Lee, S. H. (2017). Biology of Acinetobacter baumannii: Pathogenesis, Antibiotic Resistance Mechanisms, and Prospective Treatment Options. *Frontiers in Cellular and Infection Microbiology*, 7. <https://doi.org/10.3389/fcimb.2017.00055>
- Li, B., & Webster, T. J. (2018). Bacteria antibiotic resistance: New challenges and opportunities for implant-associated orthopedic infections. *Journal of Orthopaedic Research*, 36(1), 22-32. <https://doi.org/10.1002/jor.23656>
- Li, Y., Metcalf, B. J., Chochua, S., Li, Z., Gertz, R. E., Walker, H., Hawkins, P. A., Tran, T., McGee, L., Beall, B. W., & on behalf of the Active Bacterial Core surveillance team. (2017). Validation of β -lactam minimum inhibitory concentration predictions for pneumococcal isolates with newly encountered penicillin binding protein (PBP) sequences. *BMC Genomics*, 18(1), 621. <https://doi.org/10.1186/s12864-017-4017-7>
- Lin, M.-F., Lin, Y.-Y., Tu, C.-C., & Lan, C.-Y. (2017). Distribution of different efflux pump genes in clinical isolates of multidrug-resistant Acinetobacter baumannii and their

- correlation with antimicrobial resistance. *Journal of Microbiology, Immunology and Infection*, 50(2), 224-231. <https://doi.org/10.1016/j.jmii.2015.04.004>
- Lopez, P. (2019, noviembre 7). *Problema grave, la resistencia bacteriana a los antibióticos*. Gaceta UNAM. <https://www.gaceta.unam.mx/problema-grave-la-resistencia-bacteriana-a-los-antibioticos/>
- Maasdorp, S., Potgieter, S., Glover, E., & Joubert, G. (2021). Treatment outcomes of *Acinetobacter baumannii*-associated pneumonia and/or bacteraemia at the intensive care unit of Universitas Academic Hospital, Bloemfontein, South Africa. *African Journal of Thoracic and Critical Care Medicine*, 27(1), 14. <https://doi.org/10.7196/AJTCCM.2021.v27i1.122>
- Macesic, N., Polubriaginof, F., & Tatonetti, N. P. (2017a). Machine learning: Novel bioinformatics approaches for combating antimicrobial resistance. *Current Opinion in Infectious Diseases*, 30(6), 511. <https://doi.org/10.1097/QCO.0000000000000406>
- Macesic, N., Polubriaginof, F., & Tatonetti, N. P. (2017b). Machine learning: Novel bioinformatics approaches for combating antimicrobial resistance. *Current Opinion in Infectious Diseases*, 30(6), 511. <https://doi.org/10.1097/QCO.0000000000000406>
- Macesic, N., Walk, O. J. B. D., Pe'er, I., Tatonetti, N. P., Peleg, A. Y., & Uhlemann, A.-C. (2020). Predicting Phenotypic Polymyxin Resistance in *Klebsiella pneumoniae* through Machine Learning Analysis of Genomic Data. *MSystems*, 5(3). <https://doi.org/10.1128/mSystems.00656-19>
- Maglogiannis, I. G. (Ed.). (2007). *Emerging artificial intelligence applications in computer engineering: Real world AI systems with applications in eHealth, HCI, information retrieval and pervasive technologies*. IOS Press.
- Maguire, F., Rehman, M. A., Carrillo, C., Diarra, M. S., & Beiko, R. G. (2019). Identification of Primary Antimicrobial Resistance Drivers in Agricultural Nontyphoidal *Salmonella enterica* Serovars by Using Machine Learning. *MSystems*, 4(4). <https://doi.org/10.1128/mSystems.00211-19>
- Mailman, M. D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L., Popova, N., Pretel, S., Ziyabari, L., Shao, Y., Wang, Z. Y., Sirotkin, K., Ward, M., Kholodov, M., Zbicz, K., ... Sherry, S. T. (2007). The NCBI

- dbGaP database of genotypes and phenotypes. *Nature genetics*, 39(10), 1181-1186. <https://doi.org/10.1038/ng1007-1181>
- Marchand, M., Shawe-Taylor, J., Brodley, C. E., & Danyluk, A. (2003). The Set Covering Machine. *Journal of Machine Learning Research*, 3(4/5), 723-746. <http://ezproxy.unal.edu.co/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=12320303&lang=es&site=eds-live>
- Martínez-Agüero, S., Mora-Jiménez, I., Léri-da-García, J., Álvarez-Rodríguez, J., & Soguero-Ruiz, C. (2019). Machine Learning Techniques to Identify Antimicrobial Resistance in the Intensive Care Unit. *Entropy*, 21(6), 603. <https://doi.org/10.3390/e21060603>
- Mentasti, M., Prime, K., Sands, K., Khan, S., & Wootton, M. (2020). Rapid detection of OXA-23-like, OXA-24-like and OXA-58-like carbapenemases from *Acinetobacter* species by real-time PCR. *Journal of Hospital Infection*. <https://doi.org/10.1016/j.jhin.2020.06.015>
- Mobasser, P., Azimi, L., Salehi, M., Hosseini, F., & Fallah, F. (2018). Multi-drug Resistance Profiles and Expression of *AdelJK* and *AbeM* in *Acinetobacter baumannii* Collected from Humans by Real-time PCR. *Journal of Medical Bacteriology*, 7(1-2), 50-56. <http://jmb.tums.ac.ir/index.php/jmb/article/view/383>
- Moniri, R., Farahani, R. K., Shajari, G., Shirazi, M. H. N., & Ghasemi, A. (2010). Molecular Epidemiology of Aminoglycosides Resistance in *Acinetobacter* Spp. With Emergence of Multidrug-Resistant Strains. *Iranian Journal of Public Health*, 39(2), 63-68. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3481761/>
- Moradigaravand, D., Palm, M., Farewell, A., Mustonen, V., Warringer, J., & Parts, L. (2018). Prediction of antibiotic resistance in *Escherichia coli* from large-scale pan-genome data. *PLOS Computational Biology*, 14(12), e1006258. <https://doi.org/10.1371/journal.pcbi.1006258>
- Morens, D. M., Folkers, G. K., & Fauci, A. S. (2004). The challenge of emerging and re-emerging infectious diseases. *Nature*, 430(6996), 242-249. <https://doi.org/10.1038/nature02759>
- Munck, C., Gumpert, H. K., Wallin, A. I. N., Wang, H. H., & Sommer, M. O. A. (2014). Prediction of resistance development against drug combinations by collateral responses to component drugs. *Science Translational Medicine*, 6(262), 262ra156-262ra156. <https://doi.org/10.1126/scitranslmed.3009940>

- Múnera, J. M. V., Villamil, G. A. R., & Quiceno, J. N. J. (2014). Acinetobacter baumannii: Importancia clínica, mecanismos de resistencia y diagnóstico. *CES Medicina*, 28(2), 233-246. <http://revistas.ces.edu.co/index.php/medicina/article/view/2857>
- Ngiam, K. Y., & Khor, I. W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5), e262-e273. [https://doi.org/10.1016/S1470-2045\(19\)30149-4](https://doi.org/10.1016/S1470-2045(19)30149-4)
- Nguyen, M., Brettin, T., Long, S. W., Musser, J. M., Olsen, R. J., Olson, R., Shukla, M., Stevens, R. L., Xia, F., Yoo, H., & Davis, J. J. (2018). Developing an in silico minimum inhibitory concentration panel test for *Klebsiella pneumoniae*. *Scientific Reports*, 8(1), 1-11. <https://doi.org/10.1038/s41598-017-18972-w>
- Nguyen, M., Long, S. W., McDermott, P. F., Olsen, R. J., Olson, R., Stevens, R. L., Tyson, G. H., Zhao, S., & Davis, J. J. (2018). Using Machine Learning To Predict Antimicrobial MICs and Associated Genomic Features for Nontyphoidal *Salmonella*. *Journal of Clinical Microbiology*, 57(2), e01260-18, /jcm/57/2/JCM.01260-18.atom. <https://doi.org/10.1128/JCM.01260-18>
- Nigro, S. J., & Hall, R. M. (2018). Does the intrinsic oxaAb (blaOXA-51-like) gene of *Acinetobacter baumannii* confer resistance to carbapenems when activated by ISAbal? *Journal of Antimicrobial Chemotherapy*. <https://doi.org/10.1093/jac/dky334>
- Nowak, P., Paluchowska, P., & Budak, A. (2014). Co-occurrence of carbapenem and aminoglycoside resistance genes among multidrug-resistant clinical isolates of *Acinetobacter baumannii* from Cracow, Poland. *Medical Science Monitor Basic Research*, 20, 9-14. <https://doi.org/10.12659/MSMBR.889811>
- Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the Future—Big Data, Machine Learning, and Clinical Medicine. *New England Journal of Medicine*, 375(13), 1216-1219. <https://doi.org/10.1056/NEJMp1606181>
- Okser, S., Pahikkala, T., Airola, A., Salakoski, T., Ripatti, S., & Aittokallio, T. (2014). Regularized Machine Learning in the Genetic Prediction of Complex Traits. *PLoS Genetics*, 10(11), e1004754. <https://doi.org/10.1371/journal.pgen.1004754>
- Opazo, A., Sonnevend, A., Lopes, B., Hamouda, A., Ghazawi, A., Pal, T., & Amyes, S. G. B. (2012). Plasmid-encoded PER-7 β -lactamase responsible for ceftazidime resistance in *Acinetobacter baumannii* isolated in the United Arab Emirates. *Journal*

- of Antimicrobial Chemotherapy*, 67(7), 1619-1622.
<https://doi.org/10.1093/jac/dks087>
- Opazo, Andrés, Domínguez, M., Bello, H., Amyes, S. G. B., & González-Rocha, G. (2012). OXA-type carbapenemases in *Acinetobacter baumannii* in South America. *The Journal of Infection in Developing Countries*, 6(04), 311-316.
<https://doi.org/10.3855/jidc.2310>
- Organización Mundial de la Salud. (2011). *Mejorar la prevención y el control de infecciones*.
<https://www.who.int/world-health-day/2011/whd2011-fs5-prevcontr-es.pdf?ua=1>
- Organización Mundial de la Salud. (2016, septiembre 21). *Los líderes mundiales se comprometen en la ONU a abordar la resistencia a los antimicrobianos*.
<https://www.who.int/es/news-room/detail/21-09-2016-at-un-global-leaders-commit-to-act-on-antimicrobial-resistance>
- Organización Mundial de la Salud. (2018a). *Global Priority List of Antibiotic-resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics*.
<http://apps.who.int/medicinedocs/en/m/abstract/Js23171en/>
- Organización Mundial de la Salud. (2018b, enero 29). *OMS | Datos recientes revelan los altos niveles de resistencia a los antibióticos en todo el mundo*. WHO.
<http://www.who.int/mediacentre/news/releases/2018/antibiotic-resistance-found/es/>
- Organización Mundial de la Salud. (2018c, febrero 5). *Resistencia a los antibióticos*.
<https://www.who.int/es/news-room/fact-sheets/detail/resistencia-a-los-antibioticos>
- Organización Mundial de la Salud. (2018d, febrero 15). *Resistencia a los antimicrobianos*.
<https://www.who.int/es/news-room/fact-sheets/detail/resistencia-a-los-antimicrobianos>
- Park, S., Lee, K. M., Yoo, Y. S., Yoo, J. S., Yoo, J. I., Kim, H. S., Lee, Y. S., & Chung, G. T. (2011). Alterations of *gyrA*, *gyrB*, and *parC* and Activity of Efflux Pump in Fluoroquinolone-resistant *Acinetobacter baumannii*. *Osong Public Health and Research Perspectives*, 2(3), 164-170. <https://doi.org/10.1016/j.phrp.2011.11.040>
- Parquet, M. del C., Savage, K. A., Allan, D. S., Ang, M. T. C., Chen, W., Logan, S. M., & Holbein, B. E. (2019). Antibiotic-Resistant *Acinetobacter baumannii* Is Susceptible to the Novel Iron-Sequestering Anti-infective DIBI In Vitro and in Experimental Pneumonia in Mice. *Antimicrobial Agents and Chemotherapy*, 63(9), e00855-19.
<https://doi.org/10.1128/AAC.00855-19>

- Pataki, B. Á., Matamoros, S., van der Putten, B. C. L., Remondini, D., Giampieri, E., Aytan-Aktug, D., Hendriksen, R. S., Lund, O., Csabai, I., & Schultz, C. (2019). *Understanding and predicting ciprofloxacin minimum inhibitory concentration in Escherichia coli with machine learning* [Preprint]. Microbiology. <https://doi.org/10.1101/806760>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825-2830. <http://jmlr.org/papers/v12/pedregosa11a.html>
- Peseky, M. W., Hussain, T., Wallace, M., Patel, S., Andleeb, S., Burnham, C.-A. D., & Dantas, G. (2016). Evaluation of Machine Learning and Rules-Based Approaches for Predicting Antimicrobial Resistance Profiles in Gram-negative Bacilli from Whole Genome Sequence Data. *Frontiers in Microbiology*, 7. <https://doi.org/10.3389/fmicb.2016.01887>
- Potron, A., Poirel, L., & Nordmann, P. (2015). Emerging broad-spectrum resistance in *Pseudomonas aeruginosa* and *Acinetobacter baumannii*: Mechanisms and epidemiology. *International Journal of Antimicrobial Agents*, 45(6), 568-585. <https://doi.org/10.1016/j.ijantimicag.2015.03.001>
- Prada Cardozo, D. A. (2018). *Evaluación de los perfiles genómicos de resistencia a antibióticos de aislamientos colombianos de Acinetobacter baumannii mediante secuenciación del genoma completo* [Masters, Universidad Nacional de Colombia - Sede Bogotá]. <http://www.bdigital.unal.edu.co/64552/>
- Python Software Foundation. (2021, March 1). Welcome to. Python.Org. Retrieved July 6, 2020, from <https://www.python.org/>
- Qu, K., Guo, F., Liu, X., Lin, Y., & Zou, Q. (2019). Application of Machine Learning in Microbiology. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.00827>
- Ranjbar, R., & Farahani, A. (2019). Study of genetic diversity, biofilm formation, and detection of Carbapenemase, MBL, ESBL, and tetracycline resistance genes in multidrug-resistant *Acinetobacter baumannii* isolated from burn wound infections in Iran. *Antimicrobial Resistance & Infection Control*, 8(1), 172. <https://doi.org/10.1186/s13756-019-0612-5>

- Raschka, S. (2015). *Python Machine Learning* (Vol. 1). Packt Publishing Ltd.
- Ríos, R., Mattar, S., & González, M. (2011). Bibliometric analysis of publications on infectious diseases in Colombia, 2000-2009. *Revista de Salud Pública*, 13(2), 298-307. http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S0124-00642011000200011&lng=en&nrm=iso&tlng=pt
- Rolain, J.-M., Loucif, L., Al-Maslmani, M., Elmagboul, E., Al-Ansari, N., Taj-Aldeen, S., Shaukat, A., Ahmedullah, H., & Hamed, M. (2016). Emergence of multidrug-resistant *Acinetobacter baumannii* producing OXA-23 Carbapenemase in Qatar. *New Microbes and New Infections*, 11, 47-51. <https://doi.org/10.1016/j.nmni.2016.02.006>
- Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Engineering, Design and Selection*, 12(2), 85-94. <https://doi.org/10.1093/protein/12.2.85>
- Sader, H. S., Jones, R. N., Gales, A. C., Winokur, P., Kugler, K. C., Pfaller, M. A., & Doern, G. V. (1998). Antimicrobial susceptibility patterns for pathogens isolated from patients in Latin American medical centers with a diagnosis of pneumonia: Analysis of results from the SENTRY Antimicrobial Surveillance Program (1997). *Diagnostic Microbiology and Infectious Disease*, 32(4), 289-301. [https://doi.org/10.1016/S0732-8893\(98\)00124-2](https://doi.org/10.1016/S0732-8893(98)00124-2)
- Salih, T. S., & Shafeek, R. R. (2019). *In silico Detection of Acquired Antimicrobial Resistance Genes in 110 Complete Genome Sequences of Acinetobacter baumannii*. 12(5), 6.
- Samuel, S. O., Kayode, O. O., Musa, O. I., Nwigwe, G. C., Aboderin, A. O., Salami, T. a. T., & Taiwo, S. S. (2010). Nosocomial infections and the challenges of control in developing countries. *African Journal of Clinical and Experimental Microbiology*, 11(2). <https://doi.org/10.4314/ajcem.v11i2.53916>
- Sanchez, Cynthia: front-end and UI. Marjanovic, Zvezdana: graphic design. (2021). The makers' choice for sysadmins, developers and desktop users. OpenSUSE. Retrieved October 15, 2020, from <https://www.opensuse.org/>
- Schrider, D. R., & Kern, A. D. (2018). Supervised Machine Learning for Population Genetics: A New Paradigm. *Trends in Genetics*, 34(4), 301-312. <https://doi.org/10.1016/j.tig.2017.12.005>
- Sheikhalizadeh, V., Hasani, A., Ahangarzadeh Rezaee, M., Rahmati-yamchi, M., Hasani, A., Ghotaslou, R., & Goli, H. R. (2017). Comprehensive study to investigate the role

- of various aminoglycoside resistance mechanisms in clinical isolates of *Acinetobacter baumannii*. *Journal of Infection and Chemotherapy*, 23(2), 74-79. <https://doi.org/10.1016/j.jiac.2016.09.012>
- Sid Ahmed, M. A., Khan, F. A., Sultan, A. A., Söderquist, B., Ibrahim, E. B., Jass, J., & Omrani, A. S. (2020). β -lactamase-mediated resistance in MDR-*Pseudomonas aeruginosa* from Qatar. *Antimicrobial Resistance & Infection Control*, 9(1), 170. <https://doi.org/10.1186/s13756-020-00838-y>
- Su, M., Satola, S. W., & Read, T. D. (2019). Genome-Based Prediction of Bacterial Antibiotic Resistance. *Journal of Clinical Microbiology*, 57(3). <https://doi.org/10.1128/JCM.01405-18>
- Sultan, I., Rahman, S., Jan, A. T., Siddiqui, M. T., Mondal, A. H., & Haq, Q. M. R. (2018). Antibiotics, Resistome and Resistance Mechanisms: A Bacterial Perspective. *Frontiers in Microbiology*, 9. <https://doi.org/10.3389/fmicb.2018.02066>
- Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D. L., Pulcini, C., Kahlmeter, G., Kluytmans, J., Carmeli, Y., Ouellette, M., Outtersson, K., Patel, J., Cavaleri, M., Cox, E. M., Houchens, C. R., Grayson, M. L., Hansen, P., Singh, N., ... Zorzet, A. (2018). Discovery, research, and development of new antibiotics: The WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases*, 18(3), 318-327. [https://doi.org/10.1016/S1473-3099\(17\)30753-3](https://doi.org/10.1016/S1473-3099(17)30753-3)
- The Comprehensive Antibiotic Resistance Database. (2020). Resistance Gene Identifier. Retrieved October 15, 2019, from <https://card.mcmaster.ca/analyze/rgi>
- Troyano-Hernández, P., Gutiérrez-Arroyo, A., Gómez-Gil, R., Mingorance, J., & Lázaro-Perona, F. (2018). Emergence of *Klebsiella pneumoniae* Harboring the *aac(6')-Ia* Amikacin Resistance Gene. *Antimicrobial Agents and Chemotherapy*, 62(12), e01952-18. <https://doi.org/10.1128/AAC.01952-18>
- Van Camp, P.-J., Haslam, D. B., & Porollo, A. (2020). Prediction of Antimicrobial Resistance in Gram-Negative Bacteria From Whole-Genome Sequencing Data. *Frontiers in Microbiology*, 11. <https://doi.org/10.3389/fmicb.2020.01013>
- Warner, W. A., Kuang, S. N., Hernandez, R., Chong, M. C., Ewing, P. J., Fleischer, J., Meng, J., Chu, S., Terashita, D., English, L., Chen, W., & Xu, H. H. (2016). Molecular characterization and antimicrobial susceptibility of *Acinetobacter baumannii* isolates

- obtained from two hospital outbreaks in Los Angeles County, California, USA. *BMC Infectious Diseases*, 16(1), 194. <https://doi.org/10.1186/s12879-016-1526-y>
- Weinstein, M. P., & Clinical and Laboratory Standards Institute. (2019). *Performance standards for antimicrobial susceptibility testing*.
- Wiegand, I., Hilpert, K., & Hancock, R. E. W. (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nature Protocols*, 3(2), 163-175. <https://doi.org/10.1038/nprot.2007.521>
- Wilson, D. N., Hauryliuk, V., Atkinson, G. C., & O'Neill, A. J. (2020). Target protection as a key antibiotic resistance mechanism. *Nature Reviews Microbiology*, 18(11), 637-648. <https://doi.org/10.1038/s41579-020-0386-z>
- Xu, C., & Jackson, S. A. (2019). Machine learning and complex biological data. *Genome Biology*, 20(1), 76, s13059-019-1689-0. <https://doi.org/10.1186/s13059-019-1689-0>
- Yakkala, H., Samantarrai, D., Gribskov, M., & Siddavattam, D. (2019). Comparative genome analysis reveals niche-specific genome expansion in *Acinetobacter baumannii* strains. *PLOS ONE*, 14(6), e0218204. <https://doi.org/10.1371/journal.pone.0218204>
- Yamamoto, N., Hamaguchi, S., Akeda, Y., Santanirand, P., Kerdsin, A., Seki, M., Ishii, Y., Paveenkittiporn, W., Bonomo, R. A., Oishi, K., Malathum, K., & Tomono, K. (2015). Clinical Specimen-Direct LAMP: A Useful Tool for the Surveillance of blaOXA-23-Positive Carbapenem-Resistant *Acinetobacter baumannii*. *PLOS ONE*, 10(7), e0133204. <https://doi.org/10.1371/journal.pone.0133204>
- Zankari, E., Hasman, H., Kaas, R. S., Seyfarth, A. M., Agersø, Y., Lund, O., Larsen, M. V., & Aarestrup, F. M. (2013). Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *Journal of Antimicrobial Chemotherapy*, 68(4), 771-777. <https://doi.org/10.1093/jac/dks496>
- Zhang, C., Ju, Y., Tang, N., Li, Y., Zhang, G., Song, Y., Fang, H., Yang, L., & Feng, J. (2019). Systematic analysis of supervised machine learning as an effective approach to predicate β -lactam resistance phenotype in *Streptococcus pneumoniae*. *Briefings in Bioinformatics*. <https://doi.org/10.1093/bib/bbz056>
- Zhang, Yanpeng, Li, Z., He, X., Ding, F., Wu, W., Luo, Y., Fan, B., & Cao, H. (2018). Overproduction of efflux pumps caused reduced susceptibility to carbapenem under

consecutive imipenem-selected stress in *Acinetobacter baumannii*. *Infection and Drug Resistance*, 11, 457-467. <https://doi.org/10.2147/IDR.S151423>

Zheng, W., Sun, W., & Simeonov, A. (2019). Drug repurposing screens and synergistic drug-combinations for infectious diseases. *British Journal of Pharmacology*, 181-191. [https://doi.org/10.1111/bph.13895@10.1111/\(ISSN\)1476-5381.BJP-BJCP-Open-Access-Collection](https://doi.org/10.1111/bph.13895@10.1111/(ISSN)1476-5381.BJP-BJCP-Open-Access-Collection)