



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
DOCTORADO EN CIENCIAS BIOMÉDICAS
CENTRO DE CIENCIAS GENÓMICAS

CARACTERIZACIÓN GENÓMICA DEL PATÓGENO EMERGENTE
***ACINETOBACTER BAUMANNII* EN MÉXICO**

TESIS QUE PARA OPTAR POR EL GRADO DE:
DOCTORA EN CIENCIAS BIÓMICAS

PRESENTA:

LUCÍA GRAÑA MIRAGLIA

DIRECTOR DE TESIS:

DR. SANTIAGO CASTILLO RAMÍREZ

Centro de Ciencias Genómicas

COMITÉ TUTOR:

DRA. ESPERANZA MARTÍNEZ ROMERO

Centro de Ciencias Genómicas

DR. LORENZO SEGOVIA FORCELLA

Instituto de Biotecnología

CUERNAVACA, MORELOS. MARZO DE 2019



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

RESUMEN

Acinetobacter baumannii es un patógeno nosocomial con un impacto muy importante en la salud pública de la actualidad. La epidemiología de esta bacteria ha sido ampliamente estudiada y se ha determinado que existen tres linajes principales que son responsables de la mayoría de infecciones y brotes epidémicos a nivel global. Aquí realizamos una caracterización genómica de aislados mexicanos de *A. baumannii* usando secuenciación de genomas completos y encontramos que la pérdida y ganancia de genes tiene un rol preponderante en la evolución temprana de poblaciones de la especie en entornos hospitalarios.

También detectamos una incidencia muy importante en México (y el resto de América) de linajes no relacionados con las clonas internacionales, concluyendo que en los países subdesarrollados *A. baumannii* presenta características epidemiológicas particulares que han sido subestudiadas. A su vez, las cepas encontradas en estos países podrían ser reservorios de nuevos genes de resistencia que como aquí vimos, surgen y se dispersan a través de diferentes linajes muy rápidamente.

AGRADECIMIENTOS

Al Dr. Santiago Castillo por su apoyo, guía y confianza. Fue un placer formar parte de su grupo de trabajo.

A los integrantes de mi Comité Tutoral la Dra. Esperanza Martínez y el Dr. Lorenzo Segovia, quienes han seguido con gran interés mi trabajo, y cuyos aportes y consejos han sido siempre de gran valor para mí.

Al los integrantes del Laboratorio de Genómica Evolutiva, excelentes compañeros y amigos. Al Dr. Miguel Angel Cevallos por su colaboración. Al Dr. Luis Lozano por su apoyo técnico. A todo el personal del Centro de Ciencias Genómicas que hace de este un maravilloso lugar de trabajo.

Quiero agradecer profundamente a mi Mamá y mi Papá que son incansables trabajadores y que con su esfuerzo me dieron la oportunidad de estudiar y desarrollarme profesionalmente. A mis hermanas y hermano, Ceci, Mari, Sole y Santi, mi tribu. Al más reciente integrante de la familia, el increíble Jeremías.

A mi compañero Andrés, que recorrió este camino a mi lado, gracias por todo.

A mis amigos, gente importante si la hay: Clau, Charly, Noé, Andre, Miri, Kbza, Luli, Quentin, Marel, Nata, Agus y Sil, Lula, Naty, Di y Ale, Paloma y Checho, Loba, Eri, Mina y otros más!

Lucia

Índice general

RESUMEN	I
AGRADECIMIENTOS	II
LISTA DE FIGURAS	VI
LISTA DE TABLAS	IX
1 Introducción	1
1.1 Características Generales del Patógeno Emergente <i>Acinetobacter baumannii</i>	1
1.2 La clave del éxito de <i>A.baumannii</i> en las infecciones intrahospitalarias . .	2
1.2.1 RESISTENCIA	2
1.2.2 PERSISTENCIA	3
1.3 Epidemiología Global y Regional	4
1.4 Dinámica Genómica de <i>A. baumannii</i>	5
2 Justificación	7
3 Objetivos	9
3.1 Objetivo General	9
3.2 Objetivos Específicos	9
4 Metodología	10
4.1 Base de Datos	10
4.1.1 Genomas de aislados de <i>Acinetobacter baumannii</i> recientemente secuenciados.	10
4.1.2 Genomas de aislados de <i>Acinetobacter baumannii</i> públicamente disponibles.	11
4.2 Ensamble y Anotación de Genomas.	12
4.3 Determinantes de Resistencia a Antibióticos	13
4.4 Pangenoma y <i>Single Gene Families</i> (SGF)	14

4.5	Reconstrucciones filogenéticas	14
4.6	Análisis de Estructura Poblacional	15
4.7	Análisis de Fechamiento Molecular	15
4.8	Test de Topologías: Shimodaira-Hasegawa	16
4.9	Análisis de Contenido Génico	16
4.9.1	Matriz de Composición de Genes	16
4.9.2	Modelos de Pérdida/Ganacia de Genes	17
4.10	Diversidad Nucleotídica	17
4.11	Índice D de Tajima	17
4.12	Test de Tasas Relativas de Tajima (Tajima's Relative test)	18
4.13	Pruebas Estadísticas	18
5	Resultados: Expansión reciente y rápida de un linaje mexicano.	20
5.1	Expansión reciente y rápida de un linaje mexicano: ST _O 758	20
5.2	Dinámica de la Variación en el Contenido Génico a Diferentes Escalas	25
5.3	Efecto neutral o levemente deletéreo de la adquisición de nuevos genes	27
5.4	Efecto de distintas fuentes de variación genética.	28
5.5	Discusión	31
6	Resultados: El linaje ST758 es endémico de América y ha sido subestimado a nivel global.	35
6.1	Dinámica de transmisión del linaje ST758 en hospitales Mexicanos	35
6.2	El reciente linaje ST _O 758 muestra variación en los perfiles de resistencia a antibióticos.	39
6.3	Historia del alelo bla _{OXA-239}	42
6.4	Discusión	46
7	Conclusiones y Perspectivas	50
7.1	Conclusiones	50
7.2	Perspectivas	50
8	Bibliografía	52

Appendices

Apéndice A	Glosario	60
Apéndice B	Tabla Suplementaria 1	64
Apéndice C	Artículos arbitrados	67

Índice de figuras

1.1	Esquema para la tipificación de secuencias multilocus adaptado de mlst.net. MLST utiliza variaciones de secuencia en hasta siete genes constitutivos. Los números de alelos se asignan a secuencias únicas y la combinación de números de alelos da como resultado un tipo de secuencia. . . .	5
1.2	eBURST de 313 aislamientos de <i>A. baumannii</i> tipificados por el esquema MLST de Oxford. Cada círculo corresponde a un tipo de secuencia (ST). El tamaño del círculo aumenta logarítmicamente con el número de aislados de cada ST. Cada línea indica que los círculos conectados corresponden a los ST que comparten 6 de 7 de los alelos. Cada grupo de círculos unidos por al menos una línea corresponde a un complejo clonal (CC). La distancia entre círculos no corresponde al número de desajustes alélicos entre las ST correspondientes. Cada círculo o grupo de círculos resaltados en rojo corresponde a un clon. (Karah et al., 2011)	6
5.1	Complejo Clonal 636. En rojo el ST758, linaje prevalente en las muestras de los hospitales Mexicanos. La figura se generó a partir de los perfiles de MLST con el programa goeBURST. Cada punto es un ST y las líneas indican una diferencia en un alelo. Hay dos ST fundadores ST636 y ST113.	20
5.2	Reconstrucción filogenética en base a SGFs. El recuadro verde indica el clado que comprende las cepas mexicanas. Las cepas del mismo ST aparecen de igual color, los ST que están presentes solo una vez no están coloreados. La escala del árbol se ve arriba a la derecha e indica un cambio cada 100 nucleótidos.	22
5.3	Probabilidad <i>a posteriori</i> del tiempo al ancestro común más cercano del clado Mexicano. La media indica la presencia del ST758 en el hospital desde el año 2009	23

5.4	Reconstrucción filogenética por ML usando RaxML basada en el concatenado de SGF, incluyendo solamente cepas Mexicanas resuparadas en el INCAN.	24
5.5	Heatmap construido en base a una matriz de correlaciones que refleja la similitud entre los diferentes genomas en cuanto a su composición génica.	25
5.6	Reconstrucción de estados ancestrales obtenida con BadiRate	28
5.7	Comparaciones pareadas entre cepas. Mutaciones puntuales vs contenido de genes. En A) se incluyen las cepas hipermutadoras (11598 y 4113), en B) se remueven.	29
5.8	Perfiles de plásmidos. Las cepas 11598 y 4113 son las que muestran diferencias con respecto a las demás. La primera y la última línea muestran el perfil plasmídico de <i>Acinetobacter haemolyticus</i> 11616 que se utilizó como referencia de peso molecular	30
5.9	Genes asociados a MGE en genoma accesorio y core.	32
6.1	Reconstrucción filogenética en base a SGFs. La escala del árbol se ve arriba a la derecha e indica un cambio cada 100 nucleótidos.	37
6.2	Distribución de las probabilidades a posteriori de emergencia del clado ST758 (superior), y del alelo de B-lactamasa de clase d, OXA-239 (inferior).	38
6.3	Filogenia del caldo Mexicano y aislados cercanamente emparentados. Se puede observar que el clado está compuesto por cepas de los tres hospitales de Ciudad de Mexico muestreados y la presencia o ausencia del alelos OXA-239, la cual indica una adquisición reciente del alelo. . .	39
6.4	Perfiles de Resistencia a antibióticos obtenidos con la predicción in silico. Las filas no están agrupadas por clustering sino que se ordenan por clase de antibiótico. El agrupamiento de las cepas si demuestra la similitud de sus perfiles. 14 cepas tienen perfiles idénticos, todas ellas Mexicanas pero de distintos hospitales.	41

6.5	Reconstrucción filogenética por ML que refleja las relaciones de parentesco entre los diferentes alelos de la familia de β -lactamasas de clase D, OXA-23. Se incluyeron secuencias reportadas para <i>A.bauamnnii</i> pero también para otras especies. Alelos de esta familia se han descrito alrededor del mundo.	43
6.6	Orden de los genes en el vecindario del alelo OXA-239. Las cepas superior e inferior son las usadas como referencia por tener sus genomas cerrados.	45

Índice de cuadros

5.1	Aislados recuperados del INCAN	21
5.2	Test de Topologías	26
5.3	Comparación de modelos en BadiRate	27
5.4	Comparaciones Pareadas de Contenido Génico	31
6.1	Aislados recuperados del INR	36
6.2	Perfiles de Resistencia para las cepas Mexicanas en base a Antibiograma. R:Resistente,S:Suceptible,I:Intermedio,NA:No Analizado	40
B.1	Tabla suplementaria 1. Base de Datos	64
B.2	Tabla suplementaria 1. Continuación	65
B.3	Tabla suplementaria 1. Continuación	66

1. Introducción

1.1. Características Generales del Patógeno Emergente *Acinetobacter baumannii*

El género *Acinetobacter* posee en la actualidad 59 especies con nombre asignado, la mayoría de ellas han sido descritas en los últimos años gracias al desarrollo importante que han tenido las metodologías de identificación de especies bacterianas (Al Atrouni et al., 2016). El género incluye tanto especies patógenas como inocuas. Dentro de las especies patógenas, *Acinetobacter baumannii* es sin duda la que presenta la mayor relevancia clínica (Roca et al., 2012).

A. baumannii se encuentra entre los patógenos nosocomiales más importantes de la actualidad. Hasta hace pocas décadas, no era un patógeno nosocomial frecuente, su incidencia era mucho menor que la de por ejemplo, *Staphylococcus aureus* o *Pseudomonas* (Gootz and Marra, 2008) pero en los últimos 40 años las infecciones causadas por *A. baumannii* han aumentado en frecuencia, cobrando así gran importancia. Por este motivo se la considera una bacteria emergente a nivel mundial, responsable de diversos tipos de infecciones nosocomiales y causa frecuente de brotes epidémicos (Boucher, Talbot et al. 2009).

En las unidades de cuidados intensivos de los hospitales se encuentran pacientes generalmente en estado crítico, los cuales son extremadamente vulnerables a las infecciones proporcionando así un nicho para los microorganismos oportunistas (Fournier et al., 2014). Las infecciones generadas por estos organismos son muy difíciles de tratar principalmente por el alto grado de resistencia a antibióticos que presentan. Su erradicación del entorno hospitalario puede requerir el aislamiento de los pacientes o incluso el cierre temporal de las salas. Por lo tanto, su presencia, representa grandes dificultades para la salud pública (Dijkshoorn et al., 2007)

Habitualmente, *A. baumannii* no infecta personas sanas, son muy pocos los casos de infecciones producidas por *A. baumannii* adquiridas fuera del entorno hospitalario, y las que han sido descritas se detectaron en personas con dolencias crónicas (Kara-georgopoulos and Falagas, 2008). En los hospitales se ha demostrado contaminación ambiental y se ha podido rastrear el origen de brotes de infección de *A. baumannii* a

equipamiento de los centros asistenciales (Kanamori et al., 2016).

El éxito de *A. baumannii* en la infección nosocomial se debe por un lado a la extraordinaria rapidez y capacidad que presenta para desarrollar resistencia a los antimicrobianos y, por otro, a su persistencia en el medio inanimado (Karageorgopoulos and Falagas, 2008).

Estas dos importantes propiedades de *A. baumannii* se desarrollan brevemente a continuación.

1.2. La clave del éxito de *A.baumannii* en las infecciones intrahospitalarias

1.2.1. RESISTENCIA

A. baumannii parece tener la capacidad de desarrollar resistencia a nuevos antibióticos muy rápidamente, la frecuencia de aislados MDR aumentó de un 20 a un 60 % en 10 años (Giammanco et al., 2017). La resistencia de *A. baumannii* a antibióticos de amplio espectro como amikacina, cefalosporinas de espectro extendido, carbapenemos y tigeciclina va en aumento (Poirel et al., 2011). Actualmente, ciertas cepas de *A. baumannii* son altamente resistentes a la mayoría de los antibióticos disponibles en la práctica clínica, reduciendo de manera drástica las opciones de terapias efectivas (Lin and Lan, 2014). En un reciente reporte (2018) emitido por la Organización Mundial de la Salud, *A. baumannii* se cataloga como un patógeno de prioridad crítica para cuyo tratamiento se necesitan de manera urgente nuevos antibióticos (www.who.int).

En *A. baumannii* se ha identificado un extenso y diverso arsenal de genes de resistencia tanto a nivel cromosómico como plasmídico. Este organismo posee todos los mecanismos de resistencia a antibióticos conocidos para bacterias, tanto intrínsecos como adquiridos: producción de diferentes tipos de β -lactamasas, cambios en las proteínas ligadoras de penicilinas, reducción en la captura de antibióticos mediados por modificaciones en las porinas y bombas de flujo, alteración en el sitio blanco de acción farmacológica, y la producción de enzimas que alteran molecularmente a los fármacos. Estos mecanismos le confieren resistencia a penicilinas, inhibidores de β -lactamasas, cefalosporinas de tercera y cuarta generación, monobactámicos, aminoglucósidos, quinolonas, tetraciclinas, carbapenémicos y, recientemente, a las polimixinas y gliciliclinas (Poirel et al., 2011; Roca et al., 2012). Cabe destacar que muchos de estos mecanismos actúan de manera sinérgica, y otros, como las bombas de flujo pueden ofrecer resistencia a varios antibióticos (Lin and Lan, 2014).

La propagación de los determinantes de resistencia a múltiples fármacos en *A. baumannii* se debe principalmente a la conjugación de plásmidos, la adquisición de trans-

posones o la movilización de integrones para obtener grupos de genes que codifican la resistencia a varias familias de antibióticos (Esterly et al., 2011; Sahl et al., 2015). Además, las secuencias de inserción funcionales son importantes para amplificar la resistencia antimicrobiana y la plasticidad genética (Turton et al., 2006).

Los determinantes genéticos de la resistencia en *A. baumannii* pueden ser cientos, algunos como las β -lactamasas tienen un rol destacado por su relevancia clínica y epidemiológica, y en algunos casos por ser específicas de *A. baumannii* (Coyne et al., 2011; Turton et al., 2006). Debido a los altos niveles de resistencia de *A. baumannii* a muchas clases de antibióticos, los carbapenémicos representan los medicamentos de elección para el tratamiento de infecciones con esta especie. Sin embargo, la actividad de los carbapenémicos contra *A. baumannii* se ha visto amenazada por la identificación de β -lactamasas de clase D con la capacidad de hidrolizarlos (**CHDLs**: *carbapenem hydrolysing class D β lactamases*). La inactivación de los β -lactámicos constituye una parte muy importante de la resistencia a antibióticos en *A. baumannii* (Zarrilli et al., 2013).

1.2.2. PERSISTENCIA

La persistencia es otra característica de *A. baumannii* que se considera clave para su éxito como patógeno nosocomial. Esta persistencia puede durar hasta 5 meses en ambientes inanimados, confiriendo a estos organismos una gran ventaja para la re infección (Karageorgopoulos and Falagas, 2008). La supervivencia de *A. baumannii* se debe a mecanismos que le permiten resistir a la desecación, desinfección y al stress oxidativo. Cuáles son exactamente estos mecanismos no está claro en todos los casos, para la resistencia a la desecación por ejemplo hay indicios de que *A. baumannii* podría sufrir periodos de dormancia durante épocas de escasez de agua y también de que podría tener la capacidad de generar una cápsula de glicanos tal como lo hace la especie emparentada *A. baylyi* (Harding et al., 2018; Weber et al., 2015). La desecación provoca estrés oxidativo, *A. baumannii* sobre-expresa una gran cantidad de enzimas asociadas a la detoxificación causada por especies reactivas del oxígeno. En cuanto a la desinfección, en *A. baumannii* se ha observado la expulsión activa de algunos de estos antisépticos a través de la proteína AceI (Harding et al., 2018).

Por último, una característica clave de este patógeno es la capacidad de formar bio-películas o *biofilms*. Este es un mecanismo de protección que permite a *A. baumannii* sobrevivir en ambientes hostiles y durante la infección. Así, *A. baumannii* puede adherirse a superficies abióticas o a células, y sobrevivir tanto en materiales como vidrio y plástico, como en la piel, durante largos periodos (Roca et al., 2012).

1.3. Epidemiología Global y Regional

A. baumannii era un patógeno poco común en los centros de salud hasta hace pocas décadas, pero actualmente es la causa principal de neumonía e infecciones quirúrgicas y del tracto urinario, entre otras enfermedades (Peleg et al., 2008). Este aumento en la prevalencia de *A. baumannii* como causa de infecciones intra-hospitalarias coincide con un aumento en la prevalencia de cepas MDR (multi droga resistentes). 60 % de los aislados de *A. baumannii* en EEUU son MDR (Wright et al., 2016a), y en China por ejemplo, los aislados MDR (resistentes a imipenem y meropenem) aumentaron su frecuencia, de un 4.5 % en 2003 a 62.8 % en 2010 (Li et al., 2015). En Latinoamérica *Acinetobacter spp* es una de las especies más aisladas de pacientes con infecciones intra-hospitalarias como neumonía. En México, la información sobre la incidencia de infecciones causadas por *A. baumannii* es limitada pero existen datos específicos que atribuyen una mortalidad de entre un 40.7 % a un 73 % en pacientes críticos (Ñamendys-Silva et al., 2015). *A. baumannii* es responsable de aproximadamente el 2 % de la infecciones nosocomiales en EEUU y Europa y de un 4 % en Asia. Aunque otros patógenos pueden tener mayor incidencia, el porcentaje de infecciones de *A. baumannii* que presentan resistencia a múltiples drogas puede ser 45 % a nivel global y hasta un 70 % en América Latina (Harding et al., 2018). Existen varios motivos por los cuales es muy importante distinguir entre cepas de una especie bacteriana. El análisis de la variación genética en las poblaciones naturales es un requisito para estudiar los procesos moleculares subyacentes a ella (mutación, recombinación) y los procesos que determinan el destino de esa variación. Además asignar a las bacterias patógenas a diferentes “tipos” provee de información relevante a los actores del ambiente clínico, sobre las propiedades de resistencia y virulencia de las cepas, y facilita la vigilancia epidemiológica (EJ Feil, 2010).

La aplicación de varios métodos de genotipificación ha llevado al reconocimiento de que un número limitado de clones generalizados son responsables de los brotes hospitalarios en muchos países. De acuerdo al esquema de genotipificación más ampliamente usado en *A. baumannii*, el MLST (*multi locus sequence typing*) (Figura 1.1), tres linajes clonales son los principales responsables de los brotes hospitalarios a nivel global (GC1-3 ó IC1-3) (Diancourt et al., 2010) (Figura 1.2). La resistencia a múltiples fármacos a menudo se asocia con aislamientos que pertenecen a estos clones internacionales (Nemec et al., 2004).

Aunque estos linajes muestran poca diversidad genética entre genes *core*, las cepas muestran notable variabilidad a nivel de genes accesorios, incluyendo genes de resistencia a antibióticos, incluso entre aislados de un mismo brote (Wright et al., 2016a). Esta variabilidad posiblemente refleja la transmisión de plásmidos, IS, fagos, integrones, transformación natural y recombinación (Gallagher et al., 2015), y plantea un proble-

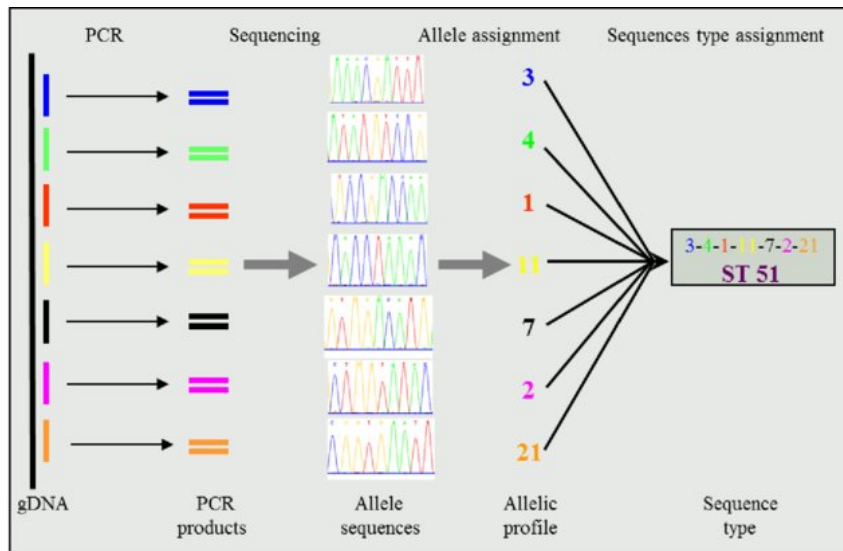


Figura 1.1 Esquema para la tipificación de secuencias multilocus adaptado de mlst.net. MLST utiliza variaciones de secuencia en hasta siete genes constitutivos. Los números de alelos se asignan a secuencias únicas y la combinación de números de alelos da como resultado un tipo de secuencia.

ma grave para estos métodos clásicos de genotipificación de *A. baumannii* como son los esquemas de MLST (Actualmente existen dos: **ST Pasteur**-ST_P y **ST Oxford**-ST_O) (Castillo-Ramírez and Graña-Miraglia, 2019).

El uso de la secuenciación genómica (**WGS**: *whole genome sequencing*) ha permitido una resolución insuperable, sobre todo en la diferenciación de aislados muy cercanamente relacionados pertenecientes a brotes epidémicos. **WGS** ha permitido identificar rutas de transmisión de *A. baumannii*, la naturaleza de diferencias genéticas entre aislados sometidos a la presión de antibióticos y el rol de los elementos genéticos móviles, entre otras muchas características genómicas (Zarrilli et al., 2013).

1.4. Dinámica Genómica de *A. baumannii*

A pesar del enorme esfuerzo que se está realizando en el estudio de la epidemiología y evolución de *A. baumannii*, siguen existiendo grandes lagunas acerca de los procesos evolutivos que contribuyen a la diversificación de cepas en los entornos hospitalarios. Los procesos a esta escala son importantes para comprender las rutas de transmisión y si las infecciones son principalmente el resultado de un contacto paciente-paciente (incluso a través de mediadores como los trabajadores de la salud o las superficies ambientales) o el resultado de múltiples eventos fundadores (Diancourt et al., 2010).

La resolución limitada de los esquemas de tipificación de *A. baumannii* (Castillo-Ramírez and Graña-Miraglia, 2019) enmascara la extensión de la transferencia lateral

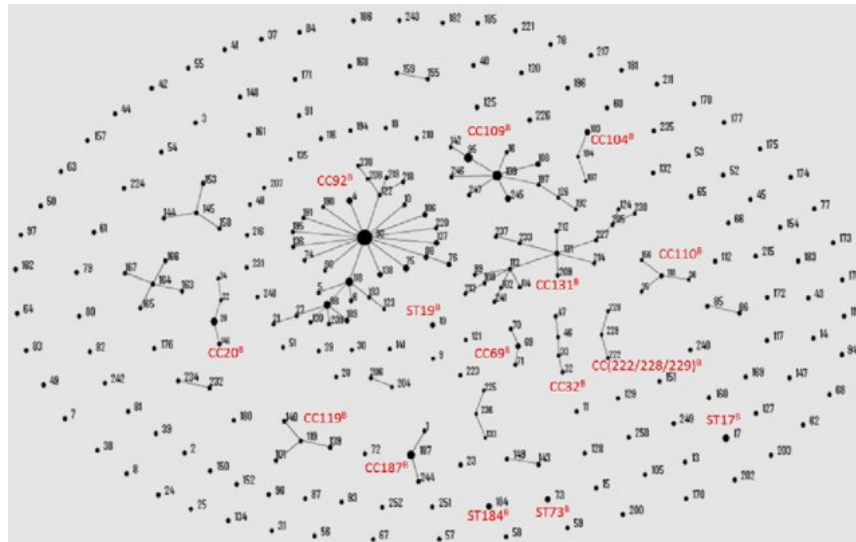


Figura 1.2 eBURST de 313 aislamientos de *A. baumannii* tipificados por el esquema MLST de Oxford. Cada círculo corresponde a un tipo de secuencia (ST). El tamaño del círculo aumenta logarítmicamente con el número de aislados de cada ST. Cada línea indica que los círculos conectados corresponden a los ST que comparten 6 de 7 de los alelos. Cada grupo de círculos unidos por al menos una línea corresponde a un complejo clonal (CC). La distancia entre círculos no corresponde al número de desajustes alélicos entre las ST correspondientes. Cada círculo o grupo de círculos resaltados en rojo corresponde a un clon. (Karah et al., 2011)

de genes (HGT) y la recombinación, que son fundamentales para impulsar la diferenciación de cepas (Wright et al., 2014).

Se ha reportado que la recombinación contribuye al cambio y evolución de los genomas de *A. baumannii* facilitando el intercambio de genes que codifican las proteínas de la superficie de la célula y sintetizan las moléculas de superficie (Snitkin et al., 2011), siendo de gran importancia para la patogenicidad de la especie.

Esto hace que esta especie de gran interés desde el punto de vista de la genómica evolutiva de bacterias dado que la contribución de diferentes fuentes de variación genética puede variar entre especies, generando diferentes dinámicas evolutivas (Dixit et al., 2017).

2. Justificación

La magnitud del problema que *A. baumannii* implica en la actualidad para la salud pública ha sido reconocido a nivel internacional y está ampliamente documentado (Dijkshoorn et al., 2007; Perez et al., 2007). Existe una necesidad urgente de comprender la epidemiología y la dinámica poblacional de los patógenos nosocomiales para mejorar los procedimientos de diagnóstico y control de infecciones. La secuenciación de genomas completos es una herramienta muy potente ya que permite estudiar las rutas de transmisión de patógenos nosocomiales, y la dinámica de genes asociados a resistencia y virulencia (Wright et al., 2016a). En México la información sobre las infecciones causadas por *A. baumannii* es escasa pero se ha registrado el aumento de la incidencia de este patógeno en hospitales mexicanos, y el problema de la resistencia a antibióticos (Ponce de Leon-Rosales, Molinar-Ramos et al. 2000, Aguirre-Avalos, Mijangos-Mendez et al. 2009, Morfin-Otero, Alcantar-Curiel et al. 2013). La mayoría de estos trabajos utilizan metodologías de tipificación como **PFGE** (Electroforesis de Campos Pulsados) o **MLST** para estudiar la epidemiología de la especie. También se analizan los fenotipos de resistencia a antibióticos de diferentes cepas y se han podido detectar nuevos alelos de resistencia (Gonzalez-Villoria et al., 2016). Pero a pesar del esfuerzo que se está llevando a cabo para comprender la dinámica evolutiva de esta especie, el primer genoma de una cepa de *A. baumannii* mexicana fue secuenciado recientemente como parte de este proyecto (Graña-Miraglia et al., 2016). Para esta especie bacteriana que tiene un genoma muy dinámico, las estrategias de genotipado comúnmente utilizadas, como **MLST**, no revelan con precisión las relaciones genéticas de las cepas en estudio (Castillo-Ramírez and Graña-Miraglia, 2019) Por lo tanto, **WGS** es la mejor estrategia para establecer las relaciones evolutivas de los clones dentro de esta especie. Además, **WGS** ha sido de suma importancia para estudiar la dinámica de transmisión de los genes de resistencia a los antibióticos en muchas poblaciones bacterianas (Holden et al., 2013). Dado el escaso conocimiento, a nivel genómico, de *A. baumannii* en México, la idea central de este proyecto fue llevar a cabo una caracterización genómica de las cepas más comunes del país para así entender cómo éstas han evolucionado y adquirido mayor resistencia a los antibióticos. Por otro lado, se sabe que un número importante de las cepas mexicanas no pertenecen a las clonas internacionales I, II, y III (**GCI-III**) y, por ello, la obtención de las secuencias genómicas de estas cepas atípicas

ayudará a entender mejor la biología y epidemiología de este patógeno emergente en México.

3. Objetivos

3.1. Objetivo General

El objetivo es contribuir a la comprensión de los procesos evolutivos y de dinámica poblacional de *A. baumannii*, así como las rutas de transmisión de este patógeno y el rol de diferentes mecanismos genéticos como recombinación, mutación y transferencia horizontal en su evolución. Usando estudios de genómica comparativa podemos evaluar los mecanismos de resistencia, patogenicidad y evolución de los patógenos bacterianos a nivel de todo su genoma.

3.2. Objetivos Específicos

- Estudiar mecanismos de variación genética en linajes mexicanos de *A. bauamnii*
- Analizar la dinámica de transferencia de genes de resistencia y la epidemiología de *A.baumannii* en México.
- Estudiar los determinantes de resistencia de los aislados mexicanos y su relación con la distribución global de genes de resistencia a antibióticos en *A. baumannii*.

4. Metodología

4.1. Base de Datos

La base de datos utilizada en este trabajo está compuesta por dos grandes grupos de genomas, los genomas de aislados secuenciados expresamente para este trabajo, y los genomas que se encuentran públicamente disponibles (NCBI, <https://www.ncbi.nlm.nih.gov/>). Todos los genomas usados junto con la información más relevante sobre ellos se encuentran en la Tabla Suplementaria 1 en el Apéndice 2.

4.1.1. Genomas de aislados de *Acinetobacter baumannii* recientemente secuenciados.

Para este trabajo se secuenciaron genomas de aislados Mexicanos de *A. baumannii*, obtenidos de dos hospitales de la ciudad de México. Se trata del Instituto Nacional de Cancerología (INCAN) y del Instituto Nacional de Rehabilitación (INR), ambos ubicados en el área de hospitales de la ciudad de México.

1. Aislados del INCAN:

Los aislados del hospital INCAN usados en este trabajo forman parte de la colección de *A. baumannii* del programa de Genómica Evolutiva. En esta colección se pueden encontrar aislados de varios puntos de la República, muestreados en un período de más de 10 años. Nosotros nos concentramos en aislados del INCAN obtenidos en un período de tiempo corto por varios motivos. Estos aislados fueron tipificados como pertenecientes al ST_O758 de acuerdo al esquema de MLST Oxford (ST_O) y obtenidos en un período de tiempo muy corto, además de que a través de comunicación personal con la Dra. Patricia Volkowf, Subdirectora de Servicios Auxiliares de Diagnóstico y Tratamiento del Hospital, se indicó que en la muestra podrían estar representados aislados de un importante brote epidémico. Se escogieron 10 de estos aislados recuperados entre 2011 y 2013 para su secuenciación. La información clínica que acompaña estos aislados es muy variable, por ejemplo las pruebas fenotípicas de resistencia llevadas a cabo en el hospital no

fueron realizadas para los mismos antibióticos en todos los casos. Todos los aislados se secuenciaron por medio de una plataforma Illumina MiSeq, con lecturas pareadas de 250 pb.

2. Aislados de INR:

Los aislados del INR fueron obtenidos a través de la colaboración con el Dr. Rafael Franco, Jefe de División del centro de Atención a Quemados del INR. Se pudieron secuenciar diez aislados, obtenidos entre 2011 y 2015, muestreados de diferentes fuentes (heridas, orina, sangre) y con perfiles de resistencia a antibióticos variables. Los antibiogramas fueron llevados a cabo en el laboratorio de Infectología del hospital así como la extracción de ADN para su secuenciación. Los genomas se secuenciaron utilizando una plataforma Illumina MiSeq con una configuración de 2 x 250 pb. La secuenciación se realizó en el Instituto Nacional de Medicina Genómica (<http://www.inmegen.gob.mx/>) en la Ciudad de México.

3. Aislados Hondureños:

Estos aislados no fueron secuenciados expresamente para este trabajo, pero fueron agregados a la colección del programa de Genómica Evolutiva hace relativamente poco y se incluyeron en nuestra base de datos dada la importancia geográfica que tienen. Son de los pocos genomas de aislados centroamericanos que hay disponibles, y algunos de ellos fueron tipificados como ST_O758, linaje de interés para nosotros. Estos aislados ya se encuentran disponibles públicamente y sus números de acceso están disponibles en la Tabla Suplementaria 1 en el Apéndice 2.

4.1.2. Genomas de aislados de *Acinetobacter baumannii* públicamente disponibles.

Existen miles de genomas de *A. baumannii* disponibles públicamente, la mayoría de ellos se encuentran en estado de *draft*. Nosotros escogimos todos los genomas completamente secuenciados y cerrados, y los descargamos. Estas bases de datos se actualizan continuamente por lo que la primera parte de este trabajo incluye 85 genomas de *A. baumannii* y en la segunda parte del trabajo la base de datos aumenta a 111. En la Tabla Suplementaria 1 del Apéndice B se explicita para cada uno de estos genomas si fue usado solo para la primera parte del trabajo o para ambas. Tomando en cuenta los 111 genomas, la base de datos incluye entonces 54 ST diferentes además de varias cepas con nuevos ST a los cuales aún no se asigna un número. En la base de datos están representados 21 países de diferentes regiones del mundo, cabe destacar que hay una importante representación de Asia (China y Corea del Sur principalmente). Además las

fechas de aislamiento cubren un período de más de 60 años. La base de datos es una de las más exhaustivas presentadas en estudios de *A. baumannii*.

4.2. Ensamble y Anotación de Genomas.

Este procedimiento se llevó a cabo con los aislados mexicanos secuenciados expresamente para este trabajo. Dado que las plataformas de secuenciación usadas arrojan el mismo tipo de lecturas, 2x250pb, el procedimiento que aquí se describe se aplicó a todos ellos.

Se analizó la calidad de las lecturas utilizando FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>), en caso de ser necesario se limpiaron las lecturas a través de un proceso conocido como trimming, en el cual se retiran las bases de baja calidad de acuerdo a el *Phred Quality Score* ($Q \leq 20$) esto se realizó con el *software* de SolexQA (Cox et al., 2010). En este mismo proceso se eliminan adaptadores en caso de estar presentes.

El siguiente paso es ensamblar los genomas, para esto se usaron dos programas, Velvet v1.2.10 (Zerbino and Birney, 2008) y SPAdes v3.11.0 (Bankevich et al., 2012). En ambos programas es posible generar ensambles a partir de lecturas cortas pareadas, los algoritmos que se utilizan son diferentes y también los tamaños de *kmer* disponibles, por lo tanto, dependiendo del aislado y del programa se pueden generar ensambles de diversa calidad. Una vez exploradas las diferentes posibilidades (ambos programas con diversos tamaños de *kmer*) se escogió el ensamble de mejor calidad para cada aislado. La calidad de los ensambles se refleja en varios índices, pero principalmente aquí se consideró el número de *contigs* finales, el **N50**, el largo total del ensamble el cual debe reflejar el tamaño total del genoma de la especie y finalmente la cobertura del ensamble. La cobertura se estima re alineando las lecturas sobre el ensamble usando Bowtie 2 (Langmead and Salzberg, 2012).

Una vez que se escogió el mejor ensamble en cada caso, se procedió a anotar los genomas con Prokka v1.12 (Seemann, 2014). Este programa realiza tanto la predicción de secuencias codificantes como la anotación funcional de los genes.

Los ensambles de todos los genomas que se secuenciaron para este trabajo están disponibles públicamente bajo el número de acceso PRJNA355850 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA355850>).

En algunos casos los aislados fueron tipificados previamente de manera experimental usando los esquemas de Oxford (**ST Oxford**-ST_O) o Pasteur (**ST Pasteur**-ST_P). Esta tipificación consiste en la amplificación de siete genes constitutivos y el posterior análisis comparativo de su secuencia. De esta manera la secuencia de cada gen puede asignarse a alelos ya descritos, o se pueden reportar nuevos alelos. Los alelos son numerados y con la asignación de cada uno se genera un perfil que se denomina el *sequence*

type (ST) de la cepa. El ST Se corroboró para todos los aislados usando la base de datos PubMLST (<https://pubmlst.org/>). Escogimos usar el esquema Oxford porque proporciona mayor resolución a la hora de distinguir aislados (Castillo-Ramírez and Graña-Miraglia, 2019). También establecimos la conexión de los ST_O en nuestra muestra con el resto de los ST_O utilizando goeBURST (Francisco et al., 2009).

4.3. Determinantes de Resistencia a Antibióticos

Como se comentó previamente, los análisis fenotípicos de resistencia de las cepas recuperadas de los hospitales son en algunos casos incompletos, esto no es algo poco habitual, normalmente en los hospitales no se hacen pruebas de resistencia a todos los antibióticos disponibles sino que generalmente se hacen pruebas dirigidas a las posibilidades de tratamiento. De todos modos, para poder identificar una cepa de *A.baumannii* como MDR se necesita que esta sea resistente a por lo menos un antibiótico en 3 clases de antibióticos. Todas las cepas Mexicanas son MDR.

Dado que tenemos los genomas de estas cepas procedimos a realizar una búsqueda *in silico* de los determinantes de resistencia. Esta búsqueda consistió en realizar un BLASTP de los proteomas contra la base de datos CARD (*The Comprehensive Antibiotic Resistance Database*) (Jia et al., 2017). CARD es una colección rigurosamente curada de determinantes de resistencia caracterizados y sus antibióticos asociados. La búsqueda se realizó a nivel proteico, todo el proteoma de cada uno de los aislados se comparó a través de un BLASTP (evalúe 1e-100) con la base de datos. Dado que muchos de los determinantes de resistencia tienen variación genética considerable, presentando un gen varios alelos, se escogió de la salida de BLASTP el mejor hit para cada proteína y se corroboró que un alto porcentaje de la proteína haya sido alineada con el gen de resistencia (mínimo 80 %). De esta manera obtuvimos un perfil de genes de resistencia para cada aislado, estableciendo la presencia (y en qué cantidad) o ausencia de los determinantes de resistencia. Estos perfiles se visualizaron a través de *heatmaps* elaborados en R (librerías: pheatmap y ComplexHeatmap).

Familia OXA-23: Las oxacilinasas del género *Acinetobacter* tienen una característica muy especial, pueden hidrolizar carbapenémicos (meropenem, imipenem). Esto es muy importante porque este tipo de antibióticos tienen un amplio espectro de acción bactericida y son por lo tanto una herramienta clave en el tratamiento de infecciones causadas por *A. baumannii*. Hay 5 familias de oxacilinasas identificadas en *A. baumannii*. La OXA-23 está ampliamente distribuída en varios linajes y presenta más de 20 alelos (Nigro and Hall, 2016). Se realizó un análisis en profundidad de la familia de CHDL OXA-23 y las relaciones filogenéticas entre sus diferentes alelos. Se utilizó un gen del alelo bla_{OXA-23} perteneciente a la cepa A85 como semilla para la búsqueda a través de BLASTN de alelos de esta familia reportados en la base de datos nucleotídica

de NCBI. Se eligieron los hits con más del 90 % de identidad nucleotídica (y más del 80 % del gen alineado), ya que con criterios menos estrictos se obtienen hits con otras familias de β -lactamasas de clase D. Obtuvimos 213 secuencias y 29 alelos de la familia OXA-23 diferentes (consulte la Tabla Suplementaria 2 en Apéndice 3). Las secuencias se alinearon con PRANK (Löytynoja, 2014) y utilizamos RaxML (Stamatakis, 2014) para la reconstrucción filogenética. La visualización y anotación del árbol se realizó con iTOL (Letunic and Bork, 2016).

4.4. Pangenoma y *Single Gene Families* (SGF)

Para la construcción de los pangenomas analizados en este trabajo se utilizaron dos aproximaciones alternativas, PanOCT (Fouts et al., 2012) y ROARY (Page et al., 2015). Ambas permiten obtener el conjunto de genes *Genoma core* y *Genoma accesorio* de los genomas en estudio, el tamaño de estos conjuntos va a depender ligeramente de los criterios usados para definirlos, entre ellos el porcentaje de identidad necesario para considerar a dos genes cualquiera como homólogos. Dado que trabajamos con diferentes cepas pero siempre dentro de una misma especie bacteriana, utilizamos un mínimo de 90 % de identidad nucleotídica y un 80 % del largo del gen conservado para considerar a dos secuencias como homólogas.

El *pipeline* de estos programas incluye el procedimiento de BLAST y la aplicación de criterios como BRH (Best Reciprocal Hit) y sintenia de genes para definir ortólogos. A partir de la salida de cada uno de estos programas se pueden identificar genes *ortólogos* y *parálogos*, con esta información definimos lo que llamamos genes ortólogos de copia única, o familias de genes de copia única (*SGFs: single gene families*). Estos genes son los que cumplen con la condición de estar presentes en todos los genomas con los criterios de identidad que propusimos y además no tener genes parálogos en un mismo genoma, o sea estar presentes en una sola copia en cada genoma. Se pueden entender como un subconjunto del genoma *Genoma core*, el cual puede contener parálogos.

4.5. Reconstrucciones filogenéticas

Utilizamos los *SGF* para obtener una aproximación robusta al árbol de las especies (Castillo-Ramírez and González, 2008). Los alineamientos de genes se realizaron con Muscle y tranalign (Resultados, Primera parte), y con PRANK (Resultados Segunda parte). Se descartaron para la inferencia filogenética los genes con evidencia de recombinación, para esto usamos el test Phi disponible en el paquete PhiPack (Bruen et al., 2006). Para realizar la inferencia filogenética los *SGF* no recombinantes

fueron concatenados respetando el marco de lectura, el mejor modelo evolutivo para el alineamiento se estimó con ModelTest (<https://github.com/ddarriba/modeltest>)(Posada, 2008) y posteriormente la reconstrucción filogenética se realizó con RaxML (Stamatakis, 2014). Cuando estimamos aproximaciones al árbol de las especies, realizamos 20 inferencias independientes con 20 topologías iniciales diferentes estimadas por parsimonia, cuando se infieren árboles de genes este número baja a 10. Las réplicas de bootstrap (# 100) también se estimaron con RaxML (Stamatakis, 2014). La visualización de los árboles (Figuras en Resultados) se realizó con iTOL (Letunic and Bork, 2016).

4.6. Análisis de Estructura Poblacional

Para establecer el grado de estructuración de la población de *A.baumannii* que conforma la muestra a analizar se usó el programa BAPS (*Bayesian Analysis of Population Structure*) (Corander et al., 2008). Se supone que la población objetivo está potencialmente estructurada genéticamente, de modo que existen límites (o han existido) que limitan el flujo de genes. La extensión y la forma de dicha subestructura es típicamente desconocida para las poblaciones naturales.

Para este análisis empleamos el concatenado de 1383 familias de genes únicos sin excluir recombinantes (Resultados Primera parte). Ejecutamos un agrupamiento jerárquico de las cepas usando la versión tándem de BAPS (Cheng et al., 2013). Este agrupamiento se llevó a cabo estableciendo cuatro niveles en la jerarquía. Para obtener una medida del número de grupos divergentes genéticamente, realizamos un análisis preliminar que establece el número máximo de grupos genéticamente diferenciados en 35, como se hizo anteriormente para otras especies bacterianas (Joseph et al., 2015). El primer nivel de agrupamiento produjo cinco grupos, el segundo produjo 15, el tercero proporcionó 22 y el último dio 32.

4.7. Análisis de Fechamiento Molecular

Las estimaciones precisas de los tiempos de divergencia entre especies son información vital para comprender la biogeografía histórica, estimar las tasas de diversificación e identificar causas en la variación de tasas de evolución molecular.

Muchos factores pueden influenciar la tasa de sustitución de una población, como la tasa de mutación, el tamaño de la población, el tiempo generacional y la selección natural. Por este motivo se han desarrollado muchos modelos que describen cómo varían las tasas de sustitución en un árbol, el reloj molecular es uno de ellos.

Muchos de estos modelos pueden aplicarse como *priors* en métodos de inferencia

bayesianos. Para el fechamiento que se realizó a través de BEAST2 (Bouckaert et al., 2014), usamos **SNPs** (*single Nucleotide Polimorphisms*) extraídos de los concatenados de **SGFs**. Solo se consideró un subconjunto de cepas, para las cuales teníamos información confiable sobre la fecha de aislamiento (ver Tabla Suplementaria 1, apéndice 2), de modo de poder calibrar el reloj molecular relajado con confianza. Establecimos un reloj relajado log-normal, empleando el modelo evolutivo **GTR** y utilizando la corrección para la variación entre sitios; este modelo fue elegido a través de jModelTest2 (Posada, 2009; Darriba et al., 2012). Este análisis se realizó para 200,000,000 generaciones, muestreando cada 10,000 generaciones y descartando las primeras 20,000,000 generaciones.

4.8. Test de Topologías: Shimodaira-Hasegawa

Este test no paramétrico compara simultáneamente todas las topologías en determinado conjunto M. Es necesario que M contenga la topología “verdadera” para que los niveles de significancia computados sean precisos. La hipótesis nula es que todas las topologías en M explican igual de bien los datos, y la alternativa, que algunas de estas topologías no son explicaciones igual de buenas de los datos (Shimodaira and Hasegawa, 1999). Este test se implementó con RaxML (Stamatakis, 2014).

4.9. Análisis de Contenido Génico

4.9.1. Matriz de Composición de Genes

Para comparar la composición génica de los diferentes genomas de la base de datos, se utilizaron los grupos de genes homólogos (GH) generados durante la búsqueda de SGF. Con esta información se generó una matriz compuesta de columnas que corresponden a los genomas, y filas que corresponden a los grupos de homólogos. De esta manera, para cada cepa tenemos un perfil de composición de genes, que consiste en la presencia (y en qué cantidad) o ausencia de genes en determinado GH. La matriz se estandarizó para poder generar a partir de ella una matriz de correlaciones, la cual se visualizó a través de un *heatmap*. Las correlaciones se realizaron en R así como la visualización de las matrices (librerías: *ade4*, *ComplexHeatmap*).

Además de la matriz de correlaciones, generamos una matriz de distancias. Usando esta matriz construimos topologías por el método de *Neighbor Joining* (NJ) que reflejaran las relaciones entre las cepas basadas en su contenido génico.

4.9.2. Modelos de Pérdida/Ganacia de Genes

Para comprender el rol de la selección y la adaptación en el moldeado de las familias de genes, es necesario llevar a cabo análisis comparativos de tasas de pérdida y ganancia de genes (Librado et al., 2012). La mayoría de las familias de genes son muy dinámicas y evolucionan de acuerdo a un proceso de muerte y nacimiento (BD, *birth-and-death*). En este modelo, los nuevos genes se crean mediante duplicación y algunos de ellos permanecen en el genoma durante mucho tiempo, mientras que otros se desactivan o eliminan del genoma (Nei and Rooney, 2005).

Implementamos un modelo estocástico de BD mediante el *software* BadiRate (Librado et al., 2012). Este análisis requiere un árbol ultramétrico de los taxones considerados, por lo tanto, utilizamos la biblioteca APE (Paradis et al., 2004) para convertir la filogenia del clado mexicano construida por máxima verosimilitud en un árbol ultramétrico. Consideramos dos modelos: uno fue el modelo de tasas globales (GD-GR-ML) y el otro fue el modelo de tasas libres (GD-FR-ML). GD-GR-ML incluye un modelo de ganancia-pérdida de genes (GD, *gain-and-death*), que todas las ramas tienen la misma tasa de recambio de genes (GR) y el método estadístico para estimar esta tasa es máxima verosimilitud (ML). En el segundo modelo la única diferencia es que la tasa de cambio del número de genes puede variar en cada rama del árbol.

Para evaluar la bondad de ajuste de estos dos modelos, utilizamos el *log-likelihood* de los modelos y el Criterio de Información de Akaike (Akaike, 1974). También establecimos la opción *anc* de BadiRate que realiza una reconstrucción de caracteres ancestrales. Se puede obtener el número de genes en las ramas internas para cada familia de genes así como el número mínimo de pérdidas y ganancias en ramas internas y externas del árbol.

4.10. Diversidad Nucleotídica

La diversidad nucleotídica se utiliza para estimar el grado de polimorfismos presente en una población. Es por lo tanto una medida de variación genética. Una manera de determinar la diversidad nucleotídica es calculando el número promedio de diferencias de nucleótidos por sitio entre dos secuencias de DNA en todos los pares posibles de secuencias en la población de la muestra, y se denota con la letra π . Este estadístico fue calculado con la librería *pegas* en R.

4.11. Índice D de Tajima

Es un estadístico que compara el número promedio de diferencias nucleotídicas pareadas con el número de sitios segregantes. Esperamos que la selección direccional

(o barridas selectivas) nos proporcione una D negativa en una población que no tiene ningún cambio demográfico en curso (expansión, reducción poblacional, migración, etc.). Después de una barrida selectiva, la mayoría de los haplotipos en una población serán los mismos. Por lo tanto, cuando ocurren mutaciones serán raras. Cuando tienes muchas mutaciones raras, obtienes una D de Tajima negativa. Esto también se observa cuando ocurren cambios demográficos como incrementos súbitos en el tamaño poblacional posteriores a un cuello de botella.

En el caso de selección balanceadora, los alelos se mantienen en frecuencias intermedias. Esto produce una D de Tajima positiva porque habrá más diferencias entre pares que sitios de segregación. Este índice se estimó usando MEGA6 (Tamura et al., 2013).

4.12. Test de Tasas Relativas de Tajima (Tajima's Relative test)

Este es un test comparativo entre dos grupos internos (especies estrechamente relacionadas) y una especie de referencia o grupo externo. Se compara la mutación y las tasas de evolución entre las especies o taxa. Cada grupo interno se compara independientemente con el grupo externo para determinar qué tan relacionadas están las dos especies sin saber el momento exacto de la divergencia de su ancestro común más cercano. Si se han producido más cambios en un linaje en relación con otro linaje desde su ancestro común compartido, entonces las especies del grupo externo serán más diferentes.

Esta es una prueba que permite evaluar la hipótesis del reloj molecular ya que este no necesita ser (y nunca debería ser) asumido: puede evaluarse directamente desde los propios datos. Empleamos MEGA6 (Tamura et al., 2013) para realizar la prueba relativa de Tajima y evaluar si las posibles cepas de hipermutadores tienen mayores tasas de evolución. Un resultado frecuente de la selección en pacientes sometidos a terapias con antibióticos y de la evolución en laboratorio es la generación de cepas hipermutadoras. Estas hipermutadoras a menudo tienen tasas de mutación de 10 a 100 veces más altas que sus antecesoras y pueden proporcionar una ventaja adaptativa significativa sobre las cepas de evolución más lenta (Hammerstrom et al., 2015). Realizamos dos análisis: uno considerando la cepa Ab4113 y otro sobre la cepa Ab11598.

4.13. Pruebas Estadísticas

Se usaron dos pruebas estadísticas, el Test de Wilcoxon y el de Chi cuadrado. Ambos se implementaron en R con la librería stats (wilcox.test, chisq.test).

NOTA

A continuación se presentan los Resultados, los cuales fueron divididos en dos capítulos. El trabajo del capítulo 5 se realizó con una base de datos más pequeña. El capítulo 6 incluye más genomas de *A. baumannii* Mexicanos y también centroamericanos que pudimos ir colectando durante el período del doctorado. A pesar de tener cosas en común, las secciones tienen diferentes objetivos por lo tanto se discuten por separado en cada capítulo, y se incluye una conclusión general al final. Además los resultados presentados en el capítulo 5 están publicados (se anexa artículo), mientras que los del capítulo 6 están en proceso de revisión aunque aún el artículo no ha sido aceptado.

5. Resultados: Expansión reciente y rápida de un linaje mexicano.

5.1. Expansión reciente y rápida de un linaje mexicano: ST_O758

En el programa de Genómica Evolutiva contamos con una amplia colección de aislados de *A.baumannii* de varios estados de la República, y diferentes hospitales, la muestra abarca más de 10 años, y muchos ST diferentes.

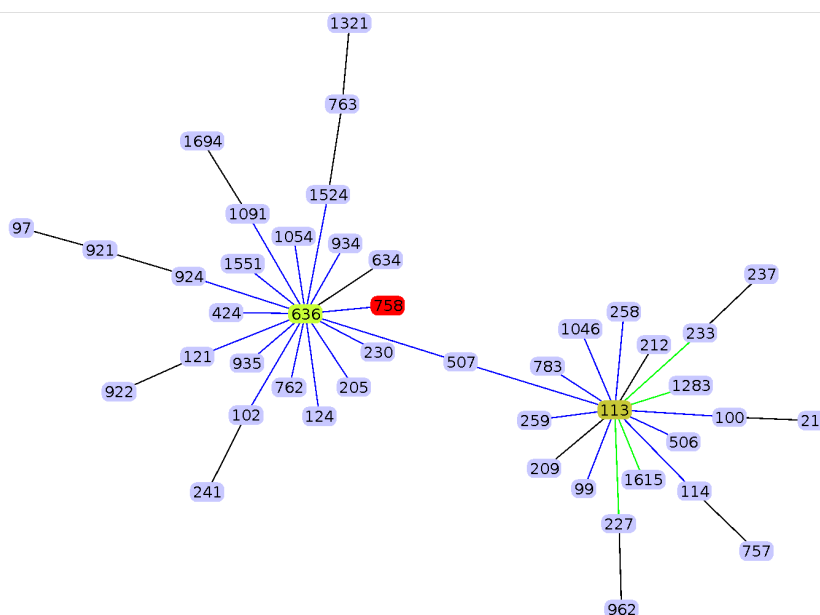


Figura 5.1 Complejo Clonal 636. En rojo el ST758, linaje prevalente en las muestras de los hospitales Mexicanos. La figura se generó a partir de los perfiles de MLST con el programa goeBURST. Cada punto es un ST y las líneas indican una diferencia en un alelo. Hay dos ST fundadores ST636 y ST113.

En esta colección, se pudo identificar un ST particularmente frecuente, el ST_O758. El ST_O758 se reportó por primera vez en México y se ha aislado frecuentemente en muchos hospitales, principalmente en el sur y el centro de la República (Gonzalez-

Villoria et al., 2016). Se encuentra especialmente en hospitales de la ciudad de México, y es muy interesante el hecho de que no forma parte de los grupos clonales globales más frecuentemente encontrados en Europa y Asia (Figura 5.1).

Tabla 5.1 Aislados recuperados del INCAN

Cepa	Hospital	Año	ST _O	Genoma (pb)	Cobertura	Contigs	N50	Genes
Ab11536	INCAN	2011	758	4117800	49.5	46	194147	3866
Ab11598	INCAN	2011	758	4004456	80	86	137359	3763
Ab4113	INCAN	2013	758	4032501	101.6	80	11820	3807
Ab11547	INCAN	2013	758	4027479	128.1	85	47382	3804
Ab55151	INCAN	2011	758	4061335	146.8	77	52744	3771
Ab11502	INCAN	2012	758	3929478	157.6	57	68938	3762
Ab11606	INCAN	2011	758	3997583	68.6	79	198285	3697
Ab11510	INCAN	2012	758	4155504	375	2	4154130	3955

Secuenciamos aislados de *A. baumannii* pertenecientes a este linaje obtenidos del Instituto Nacional de Cancerología (INCAN) de la ciudad de México, entre el año 2011 y el 2013 (Tabla 5.1). Estos datos se complementaron con todos los genomas de *A. baumannii* completamente secuenciados y cerrados disponibles al momento (2015) en la base de datos de NCBI (85 genomas que incluyen 38 ST diferentes) (Tabla Suplementaria 1, Apéndice B -B.1-). El primer paso fue contextualizar nuestras cepas en el escenario global construyendo una filogenia robusta. Utilizando PanOCT identificamos un total de 14194 grupos homólogos, de los cuales 1462 pertenecen al genoma *core*. Entre estos encontramos 1383 familias SGF, grupos de genes homólogos que tienen un solo gen por genoma. Se realizó una reconstrucción filogenética por el método de máxima verosimilitud (ML: *maximum likelihood*) basada en el concatenado de los alineamientos de SGF sin evidencia de recombinación (574 genes)(Figura 5.2).

Las cepas con el mismo ST se colorearon en la Figura 5.2, y se observa que hay casos en los que la asignación de ST no refleja las relaciones verdaderas entre los aislados de esta especie. La mayoría de los ST sí se agrupan y forman clados monofiléticos, pero hay casos en que cepas con un mismo ST forman grupos polifiléticos (ST_O455, ST_O369). Esto puede deberse a que algunos de los genes usados en el esquema Oxford de MLST tienen señales de recombinación. Esto muestra que las relaciones correctas entre aislados no se pueden recuperar utilizando solamente los esquemas MLST (Castillo-Ramírez and Graña-Miraglia, 2019).

Los aislados Mexicanos forman un grupo monofilético bien diferenciado del resto, la cohesividad de la población se pudo comprobar también a través de un análisis de estructura poblacional (ver métodos), en el cual uno de los grupos recuperados es el clado Mexicano (en verde en Figura 5.2).

En el clado Mexicano se observan ramas extremadamente cortas indicando que el grupo tiene muy poca variación genética, esto se corroboró estimando la diversidad

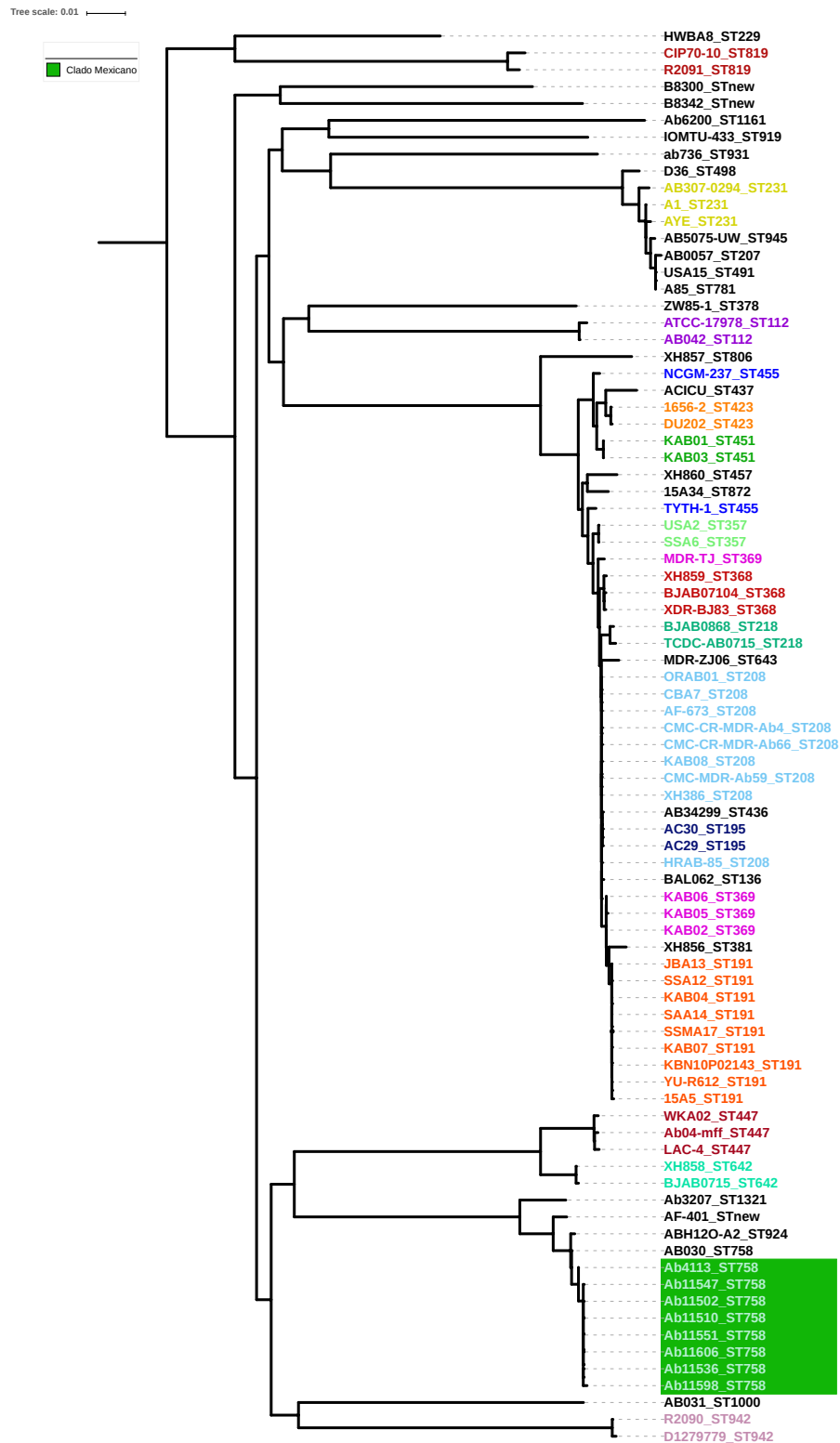


Figura 5.2 Reconstrucción filogenética en base a SGFs. El recuadro verde indica el clado que comprende las cepas mexicanas. Las cepas del mismo ST aparecen de igual color, los ST que están presentes solo una vez no están coloreados. La escala del árbol se ve arriba a la derecha e indica un cambio cada 100 nucleótidos.

nucleotídica en el grupo, la cual resultó ser muy baja, ($\pi=0.000532$) y se encontraron solamente 147 sitios segregantes. Ambos datos apuntan a una introducción reciente de este linaje al hospital Mexicano. Un análisis de dataje molecular nos permitió explorar más a fondo esta posibilidad. Se estimó que el tiempo al ancestro común más cercano del clado es el año 2009 (Figura 5.3).

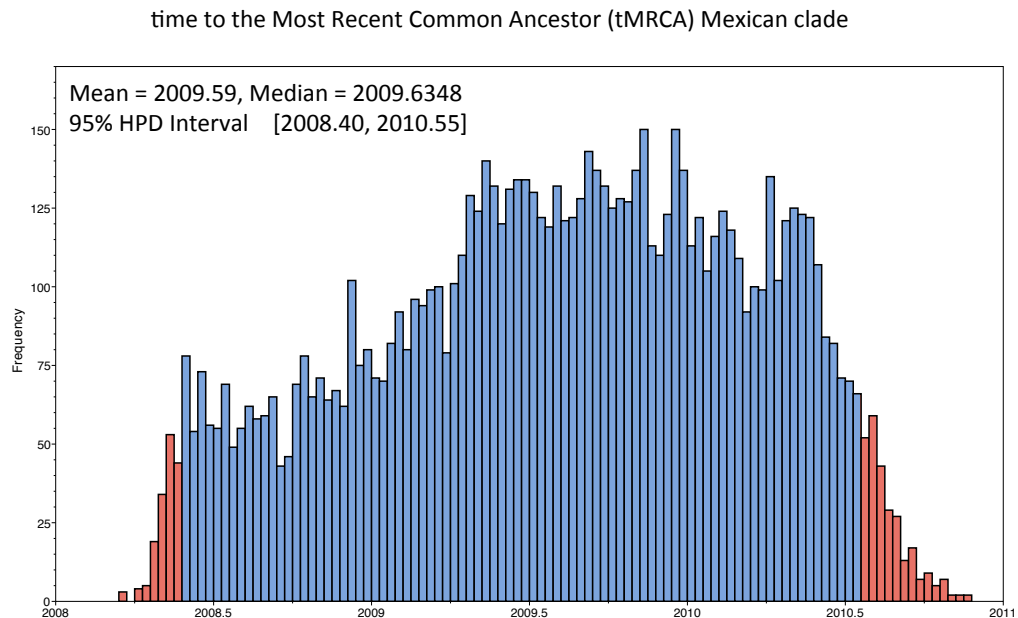


Figura 5.3 Probabilidad *a posteriori* del tiempo al ancestro común más cercano del clado Mexicano. La media indica la presencia del ST758 en el hospital desde el año 2009

En la Figura 5.2 es difícil apreciar claramente las relaciones filogenéticas entre las cepas mexicanas debido a la escala del árbol, por lo tanto construimos una filogenia que incluye solamente los aislados de este grupo (Figura 5.4). Se puede observar entonces que no hay politomías y que la topología del clado tiene una forma de “estrella”, similar a las topologías observadas en las radiaciones evolutivas, lo que sugiere una expansión reciente y muy rápida del linaje.

Este tipo de hipótesis puede probarse a través del índice D de Tajima que permite evaluar los niveles de variación de una población con respecto a lo esperado por neutralidad (Tajima, 1989). Este índice resultó tener un valor de -1.922, su valor negativo proviene de una sobre abundancia de polimorfismos de baja frecuencia, consistentes con un rápido crecimiento de la población luego de un cuello de botella. Cabe destacar que la mayoría de los cambios que se observan entre las cepas mexicanas, se localizan en las ramas de las cepas Ab11598 y Ab4113, mientras que el resto de las cepas parecen

prácticamente idénticas. Esto indica una gran variación en la tasa de mutación.

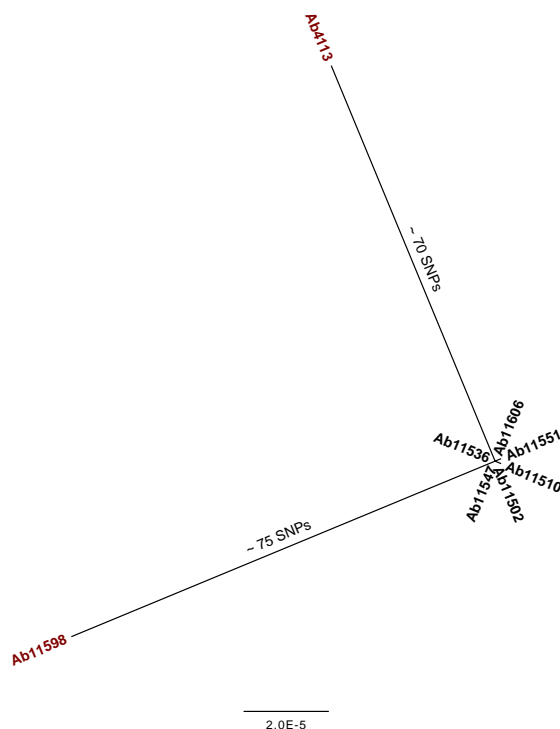


Figura 5.4 Reconstrucción filogenética por ML usando RaxML basada en el concatenado de SGF, incluyendo solamente cepas Mexicanas resuparadas en el INCAN.

Llevamos a cabo una prueba relativa de Tajima para determinar si esas dos cepas han acumulado significativamente más SNPs que las otras cepas mexicanas. Se pudo corroborar que este es el caso, ya que para ambas cepas, Ab4113 y Ab11598, se rechazó la hipótesis nula de tasas iguales en comparación con las otras cepas mexicanas. Además, buscamos la ubicación de estos cambios puntuales, observando que las sustituciones que se encuentran en ambas ramas no están agrupadas en una pequeña región del genoma, lo cual podría indicar un evento de recombinación no identificado. Por lo tanto, estas dos cepas pueden ser cepas hipermutadoras, recientemente se han encontrado cepas hipermutadoras en *A. baumannii* (Hammerstrom et al., 2015; Komp Lindgren et al., 2016). Sin embargo, reconocemos que se requieren más evidencias para asegurar esto. Por ejemplo, los estudios previos muestran que las cepas se convirtieron en hipermutadores a través del movimiento de un elemento móvil para inactivar mutS (proteína involucrada en la reparación de mutaciones puntuales en el ADN.), en este trabajo no encontramos evidencia de modificaciones en este gen.

Esta evidencia en conjunto, indica que los aislados Mexicanos forman parte de una joven población, la cual ha sido introducida recientemente al INCAN, y está experimentando una rápida expansión clonal.

5.2. Dinámica de la Variación en el Contenido Génico a Diferentes Escalas

El genoma de *A. baumannii* es un genoma muy dinámico, hay estudios que indican que *A. baumannii* puede fácilmente tomar DNA del ambiente y que esto sería ventajoso bajo la fuerte presión selectiva de los antibióticos (Wright et al., 2016b; Zhu et al., 2013). Pero cómo afecta esto al resto del genoma no es tan claro. Aquí se compara la composición de genes de las cepas en nuestro conjunto de datos el cual contiene ejemplares de diversos linajes los cuales han sido aislados alrededor del mundo en diferentes fechas. Las diferencias en el número de genes entre cepas se estimó utilizando grupos de genes homólogos cercanamente emparentados (obtenidos con PanOCT). Estos grupos se formaron bajo el criterio de un 90 % de identidad aminoacídica compartida por todos los genomas, y un 80 % del largo del gen alineado.

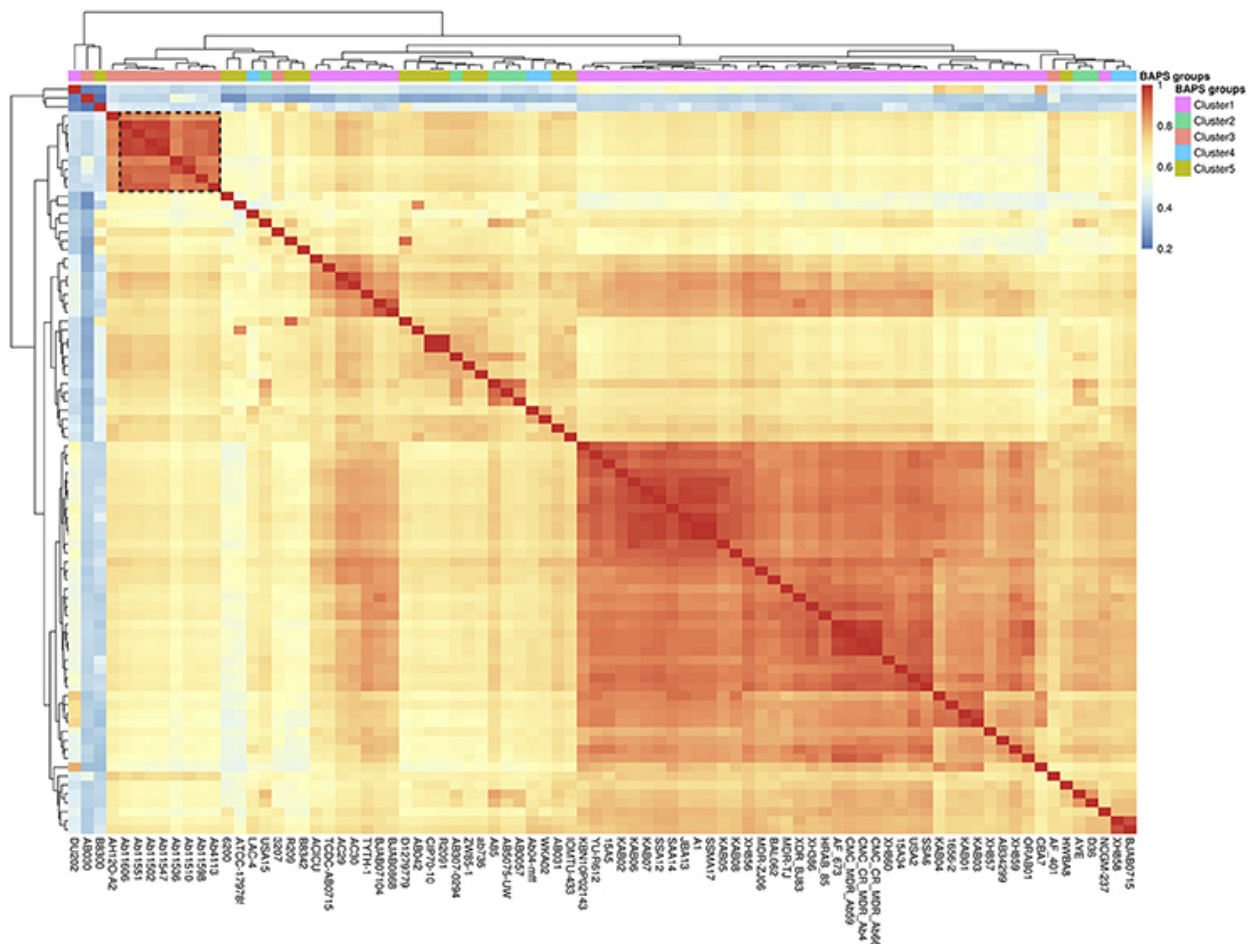


Figura 5.5 Heatmap construido en base a una matriz de correlaciones que refleja la similitud entre los diferentes genomas en cuanto a su composición génica.

A partir de esta información, se pudo generar una matriz de correlación, la cual puede visualizarse usando un heatmap (Figura 5.5). En la Figura 5.5 la diagonal en rojo

oscuro corresponde a un valor de correlación de 1, que corresponde a su vez a la comparación de la cepa consigo misma. Se pueden observar también dos regiones con altos valores de correlación (tonos rojizos), una de ellas corresponde a las comparaciones entre las cepas mexicanas (recuadro negro). La otra región que muestra alto grado de similitud corresponde en su mayoría a cepas del GC2. La similitud entre las cepas se refleja también en la topología de la parte superior de la figura, la cual está basada en la matriz. La barra de colores en la parte superior de la figura corresponde a los grupos recuperados con el análisis de estructura poblacional (BAPS).

La variación en contenido génico observada es considerable, además el agrupamiento generado por el análisis de *clustering* jerárquico (topología superior en la Figura 5.5) no arroja a simple vista, las mismas relaciones que las establecidas en la filogenia, con excepción del clado mexicano, el cual forma también aquí un grupo bien definido. Para saber si esta topología es significativamente distinta a la topología de la reconstrucción filogenética por ML, se llevó a cabo un test de Shimodaira-Hasegawa (SH).

Se construyeron dos topologías a través del método de *Neighbor Joining* (NJ) basadas en una matriz de distancia estimada a partir de la matriz de contenido génico, una para todo el conjunto de datos y otra solamente para el clado mexicano. Las topologías se compararon con sus contrapartes generadas por el método de ML basado en el alineamiento de SGF. Los resultados del test SH se muestran en la Tabla 5.2. Cuando se compara todo el conjunto de datos, el test indica que las diferencias topológicas son significativas, sin embargo cuando se compara el reducido grupo del clado mexicano, las diferencias no son significativas. Esto indica que a cortas escalas de tiempo el patrón de composición génica es congruente con las relaciones filogenéticas, pero que a medida que las cepas divergen, la variación en el contenido génico deja de reflejar las relaciones de parentesco.

Tabla 5.2 Test de Topologías

	Diff -lnL*	pKH ⁺	pSH ⁼	pRELL [@]
NJ todo el Data Set	-658053.472	0.000	0.000	0.000
NJ clado Mexicano	-0.172	0.336	0.352	0.29

* Diferencia respecto a filogenia ML, + Kishino-Hasegawa p-val, = Shimodaira-Hasegawa p-val, @ test de proporciones de bootstrap p-val

De todos modos, si observamos la cantidad de genes predichos para cada genoma del clado mexicano (Tabla 5.1), podemos observar diferencias que parecen sustanciales (cientos de genes), por lo tanto utilizamos otras aproximaciones para profundizar en la dinámica de la variación de contenido génico en cortas escalas de tiempo.

5.3. Efecto neutral o levemente deletéreo de la adquisición de nuevos genes

Para explorar más en detalle la variación del contenido de genes en el clado mexicano, comparamos el ajuste de dos modelos evolutivos que describen la ganancia y pérdida de genes a través de una filogenia. Esto se realizó con el programa BadiRate. El primer modelo que se ajustó es el modelo de tasas globales (GD-GR-ML) el cual asume que las tasas de ganancia y pérdida de genes se mantienen constantes en el tiempo (o sea la tasa es la misma en todas las ramas del árbol). El segundo modelo, asume que existen diferentes tasas para cada rama del árbol, permitiendo así variación en la ganancia y pérdida de genes en el tiempo (y linajes) (GD-FR-ML). GD-FR-ML, o modelo de tasas variables, resultó ser una mejor explicación para nuestros datos debido a que presentó un menor valor de verosimilitud (*Log-Likelihood*) y también de acuerdo al Criterio de Información de Akaike (AIC) (Tabla 5.3).

Tabla 5.3 Comparación de modelos en BadiRate

Modelo	-lnL*	K ⁺	AIC	$\Delta AIC^{\textcircled{a}}$
GD-FR-ML	-7500.2956	29	15058.59	0.00
GD-GR-ML	-8290.2042	3	16586.41	15015.68

* *Logaritmo de la Verosimilitud*, + *número de parámetros*, @ *Diferencia de AIC entre cada modelo y el mejor modelo*

Pero con BadiRate se puede investigar un poco más y obtener datos muy útiles para esta pregunta. Se realizó una reconstrucción de estados ancestrales, siendo en este caso los estados la cantidad de genes en el genoma (por grupo homólogo). Se estimó también el número mínimo de ganancias/pérdidas de genes para cada nodo de la filogenia (Figura 5.6). De acuerdo a este análisis, los genomas han estado ganando genes, o sea, encontramos mayor cantidad de genes en el presente (tips del árbol) (=3770) que en los nodos ancestrales (=3677). La significancia de esta diferencia se confirmó con un test de Wilcoxon ($W=7, p=0.020$).

Lo que aquí se propone es que la disminución en el número de genes en los nodos ancestrales se debe a la remoción de genes a través de selección purificadora. La emergencia y subsecuente remoción de la variación levemente deletérea resulta en un aparente aumento de la tasa de variación en las puntas del árbol. El efecto de la mayoría de estos genes será levemente deletéreo de modo que en las hojas del árbol (dado el poco tiempo que separa a las cepas) aún no vemos el efecto de la selección.

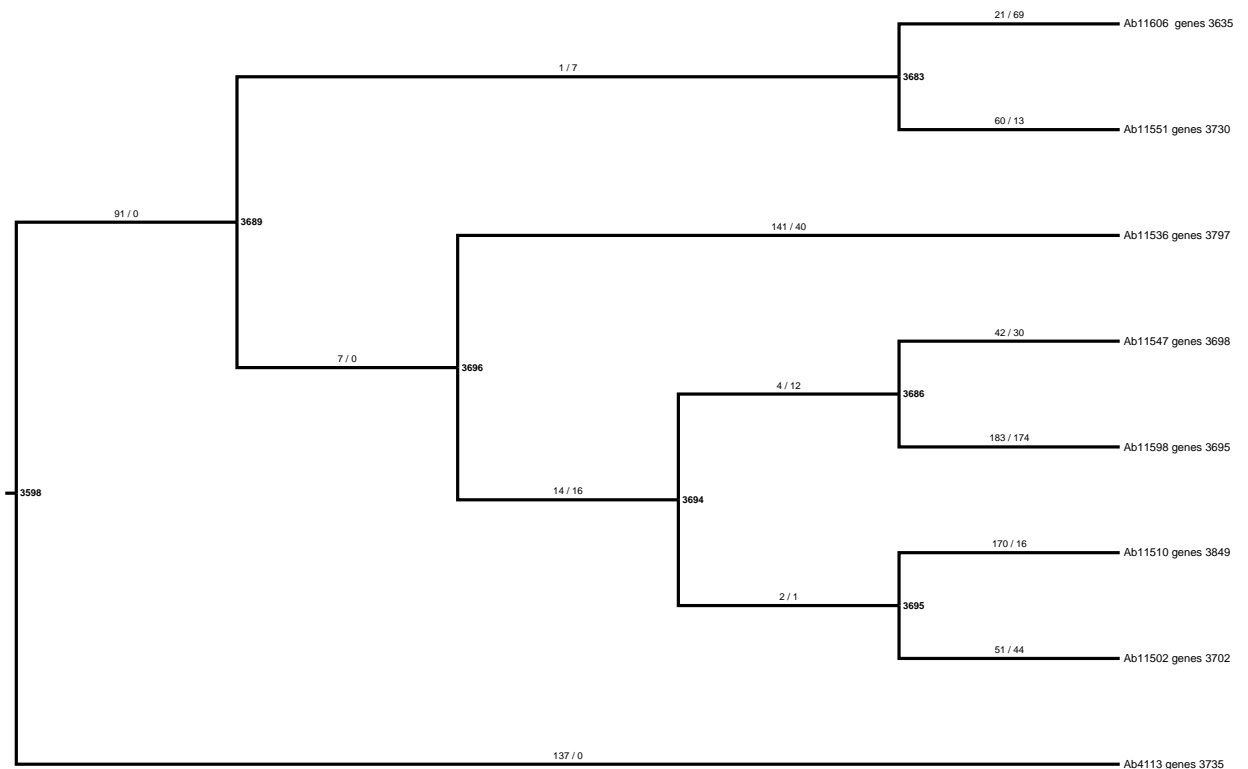


Figura 5.6 Reconstrucción de estados ancestrales obtenida con BadiRate

5.4. Efecto de distintas fuentes de variación genética.

Debido a que los cambios en la composición de genes se pueden observar a nivel de la población Mexicana, la cual se demostró que tiene niveles de divergencia muy bajos, se decidió comparar el efecto de dos fuentes de variabilidad sobre estos genomas: las mutaciones puntuales y la ganancia/pérdida de genes.

Para esto se llevaron a cabo comparaciones pareadas entre las cepas. Por un lado se calculó el número de diferencias entre ellas, y por otro la cantidad de genes en que se diferencian. Entre la mayoría de las cepas se encuentran muy pocas mutaciones (Figura 5.7), la diferencia media se encuentra en 1.5, realmente muy baja considerando que miles de sitios se están teniendo en cuenta. Sin embargo, la media de la diferencia en cantidad de genes se eleva a cientos (≈ 294.5). Cabe destacar que ningún par de cepas mostró un contenido de genes idéntico. Este análisis muestra que la tasa de recambio de genes es mucho mayor que la acumulación de mutaciones puntuales (Test de Wilcoxon: $p=1.57e-10$).

Hay que tener en cuenta, que las diferencias de varios genes entre dos cepas cualquiera puede deberse a un solo evento de transferencia, o sea, el número de genes en los que difieren los genomas no necesariamente se traduce en el mismo número de eventos de pérdida o ganancia de material genético. Muchos genes pueden introducirse en un solo evento, los elementos genéticos móviles (MGE: *Mobile Genetic Elements*) como por ejemplo los plásmidos, pueden mover cientos de genes. Debido a esto, se llevaron a

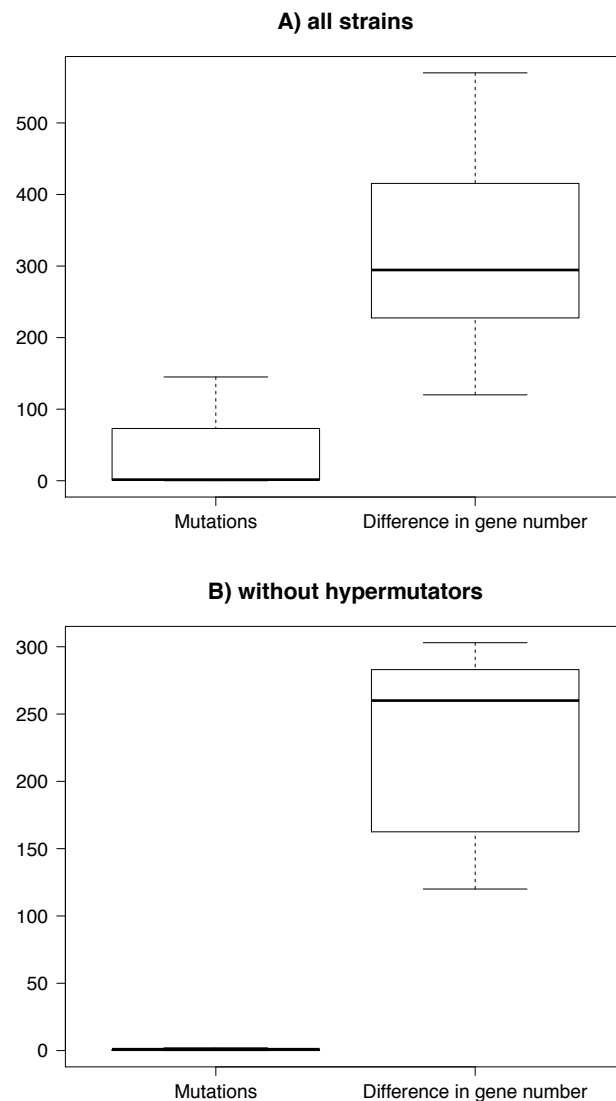


Figura 5.7 Comparaciones pareadas entre cepas. Mutaciones puntuales vs contenido de genes. En A) se incluyen las cepas hipermutadoras (11598 y 4113), en B) se remueven.

cabo estudios para determinar si las diferencias observadas pueden reducirse en realidad a unos pocos eventos de transferencia.

Primero se realizaron perfiles de plásmidos de las cepas Mexicanas a través del método Eckardt. Seis de estos aislados presentan el mismo perfil (Figura 5.8) de modo que en estos casos, la variación de la composición de genes no se debe a la pérdida o ganancia de un plásmido entero. Los dos aislados restantes comparten parte del perfil, pero presentan además en un caso una banda más, y en el otro tres. En estos dos casos parte de la variación puede entonces deberse a este tipo de elementos, pero las diferencias de genes encontradas entre cepas con perfiles idénticos indican que la variabilidad observada no se explica exclusivamente por esto. A su vez, esta variación afecta de

manera significativa al cromosoma como se verá a continuación.

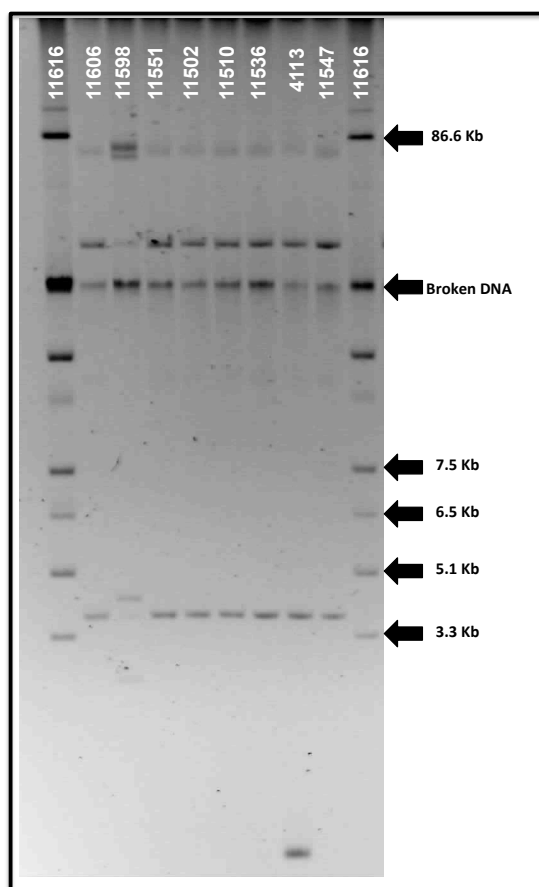


Figura 5.8 Perfiles de plásmidos. Las cepas 11598 y 4113 son las que muestran diferencias con respecto a las demás. La primera y la última línea muestran el perfil plasmídico de *Acinetobacter haemolyticus* 11616 que se utilizó como referencia de peso molecular

Posteriormente, se compararon cuatro grupos de dos cepas, escogidas por haber sido aisladas en el mismo año. Durante estas comparaciones, asumimos que dos o más genes provienen del mismo evento de transferencia, si están contiguos en el genoma. En los cuatro casos, de acuerdo a este criterio, el número de eventos estimados es considerablemente menor a la cantidad de genes en que difieren las cepas. Por ejemplo, en la comparación de las cepas Ab11502 y Ab11510, a pesar de que difieren en 311 genes (82 en Ab11502 ausentes en Ab11510 y 229 en Ab11510 ausentes en Ab11502), se estimó que por lo menos 112 eventos generaron la diferencia. El mismo patrón aplica para el resto de las comparaciones (Tabla 5.4) lo que implica que muchos genes se ganan o pierden en un solo evento de transferencia.

Claramente muchos de estos genes, aproximadamente el 52 %, pudieron ser asociados con MGEs como ISs, fagos o plásmidos. Encontramos que el genoma accesorio de

Tabla 5.4 Comparaciones Pareadas de Contenido Génico

Pares de Cepas	Genes Únicos	Eventos	% Genes Potencial Localización en Cromosoma	% Genes Potenciales MGE
Ab11510	229	72	70.74	58.95
Ab11502	82	40	59.76	42.68
Totales	311	112	67.85	54.66
Ab11598	193	120	87.05	46.63
Ab11551	215	96	80.0	45.58
Totales	408	216	78.92	44.12
Ab11536	132	54	60.61	59.85
Ab11606	47	27	59.57	59.57
Totales	179	81	60.34	59.78
Ab11547	137	48	71.53	49.64
Ab4113	165	65	52.73	48.48
Totales	302	113	61.26	49.01

los aislados mexicanos está enriquecido en este tipo de elementos comparado con el genoma core (Test de Chi², $p; 2.2e-16$) (Figura 5.9). Además de la información obtenida de los perfiles de Eckardt, un 60 % de los genes únicos en las comparaciones pareadas tienen hits significativos con genes localizados en cromosoma (en los genomas completos obtenidos de NCBI) (Tabla Suplementaria 1, Apéndice B- B.1-). Entonces el recambio de genes observado no se debe solamente a la pérdida y ganancia de plásmidos sino que ha afectado también al cromosoma.

En conjunto estos datos indican que el recambio de genes ha ocurrido mucho más rápido que la acumulación de mutaciones puntuales y que es una importante fuente de variación genética tanto a nivel de plásmidos como de cromosoma.

5.5. Discusión

Utilizamos una aproximación de genómica de poblaciones para estudiar la dinámica evolutiva del contenido génico en un período de tiempo muy corto en una población de *A.baumannii* bien definida. Esto se logró muestreando varias cepas de un mismo linaje, dentro de un hospital Mexicano. Todas las cepas secuenciadas entre 2011 y 2013 en el Instituto Nacional de Cancerología pertenecen al ST_O758. Si se consulta la página de PubMLST (<https://pubmlst.org/>), solo se reportan 3 cepas con este esquema, una Canadiense, una Mexicana y una Colombiana, pero la única de ellas que ha sido secuenciada es la Canadiense, AB030. Entonces, hasta ahora este linaje ha sido reportado principalmente en América del Norte, podría tratarse de un linaje endémico pero para

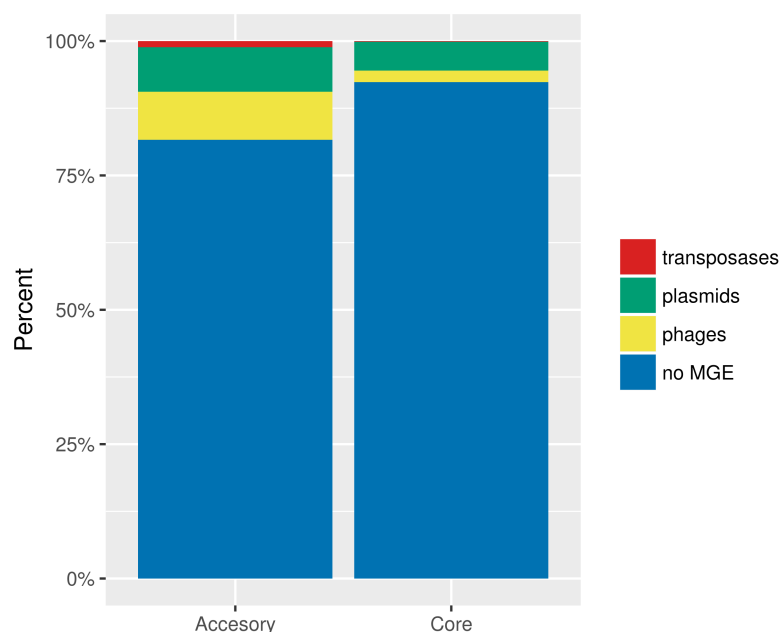


Figura 5.9 Genes asociados a MGE en genoma accesorio y core.

decir esto es necesario muestrear otras regiones que claramente están sub muestreadas como América del Sur. De todos modos, el linaje (ST_O758) pertenece a un complejo clonal, el CC636, no relacionado con los IC I-III, por lo tanto es poco probable que el linaje se encuentre globalmente distribuido. Realizamos una filogenia con estas cepas y con cepas que se encuentran públicamente disponibles, para ubicar este linaje en el contexto global dado que *A.baumannii* es un patógeno que se encuentra globalmente distribuido. Escogimos cepas aisladas de diferentes partes del mundo y que representan 38 ST diferentes.

Los análisis filogenéticos y de dataje molecular indican que las cepas Mexicanas comprenden una población reciente que está experimentando una expansión poblacional muy rápida. El ingreso de el linaje al hospital se estimó para el año 2009, la alta tasa de variantes (SNPs) únicas indica un aumento en el tamaño poblacional en un período ente 2 y 4 años (los aislados aquí analizados fueron recuperados entre 2011 y 2013). Este tipo de expansiones clonales han sido descritas para otros patógenos nosocomiales como *S.aureus* (Aanensen et al., 2016), indicando que esta podría ser una tendencia común para poblaciones bacterianas establecidas en los hospitales. Hasta donde sabemos, este es el primer análisis de variación en contenido génico que toma en cuenta la estructura poblacional para estudiar apropiadamente la dinámica de pérdida/ganancia de genes entre cepas, pero también compara esta fuente de variación con la acumulada por mutaciones puntuales. De todos modos tendencias similares a las aquí observadas se han descrito recientemente en *Vibrio cholerae* (Levade et al., 2017).

En este trabajo, demostramos que el recambio de genes ha sido un factor principal en la generación de variación genética en los pocos años de existencia del clado

Mexicano, y que ha introducido 8 veces más variación que la acumulada a través de mutaciones puntuales. Evidentemente, la ganancia o pérdida de genes no son siempre eventos independientes, ya que muchos genes pueden ser introducidos en un solo evento a través de algún MGE. De acuerdo con esto, nosotros encontramos que el genoma accesorio de *A.baumannii* está enriquecido en MGEs (Figura 5.9). Sin embargo, regiones cromosómicas también están implicadas en este tipo de variación, a pesar de esto a muy pequeñas escalas de tiempo, el patrón de contenido génico aún refleja las relaciones filogenéticas entre las cepas.

Acá se plantea que los patrones de contenido génico ancestral que se encontraron son compatibles con la naturaleza levemente deletérea de muchas de las adquisiciones de genes. Aunque algunos de estos genes pueden ser beneficiosos para el organismo, la mayoría de ellos representan un costo y eventualmente serán removidos por selección natural (Mira et al., 2001). Dado que el costo de la adquisición de estos genes puede no ser tan alto (levemente deletéreo) el proceso de remoción de los mismos no es necesariamente instantáneo, de modo que estas diferencias entre cepas pueden ser muestreadas con una estrategia que implique aislados muy cercanamente emparentados (poca divergencia) (Kuo et al., 2009). Bajo este escenario, las ramas externas se encontrarían enriquecidas debido a las adquisiciones de genes ya que no ha pasado tiempo suficiente tiempo para que actúe la selección natural, mientras tanto las ramas internas muestran menor cantidad de genes ya que las adquisiciones de carácter levemente deletéreo ya fueron purgadas.

La idea preponderante es que los procesos de reducción genómica, selectivos o neutrales, son el principal modo de evolución de los genomas bacterianos (Wolf and Koonin, 2013). Estos argumentos se basan en estudios que consideran escalas macro y no microevolutivas. En la misma línea de las evidencias aquí presentadas, Gogarten and Townsend, 2005 (Gogarten and Townsend, 2005) observaron que la mayoría de los genes que son transferidos horizontalmente tienen efectos neutrales o levemente neutrales sobre el genoma de destino y solo unos pocos presentan una ventaja selectiva. En conclusión, este trabajo demuestra que la variación en la composición génica puede ser una fuerza evolutiva de gran importancia y que puede generar mayor variabilidad que la acumulada por mutaciones puntuales en cortas escalas de tiempo, en etapas tempranas de diversificación de un linaje.

Además se sugiere que la disminución en el número de genes en nodos internos de la filogenia se debe a la remoción no inmediata de genes cuya adquisición tiene un efecto levemente deletéreo. Este no es el primer trabajo que analiza aislados Mexicanos, hay varios estudios que han sido de gran importancia epidemiológica pero ninguno de ellos se enfocó en la diversidad genómica de las cepas que circulan en México, y hasta donde sabemos este es el primer estudio que analiza la variación genética de una población de *A.baumannii* Mexicana a nivel genómico.

Aquí también se evidencia la utilidad de la secuenciación de genomas completos para estudiar la dinámica micro evolutiva de los genomas bacterianos, ya sea para explicar la variación en composición génica, así como para observar mutaciones puntuales.

6. Resultados: El linaje ST758 es endémico de América y ha sido subestimado a nivel global.

6.1. Dinámica de transmisión del linaje ST758 en hospitales Mexicanos

Se decidió explorar la situación de los linajes de *A. baumannii* circulando en otros hospitales de la ciudad de México. Se pudo acceder a aislados de *A. baumannii* del Centro Nacional de Investigación y Atención a Quemados del Instituto Nacional de Rehabilitación, obtenidos entre 2011 y 2015. En este centro se tratan pacientes quemados de diferentes estados de la República, muchos de ellos con períodos previos de hospitalización en otros estados. Aquí, la incidencia de *A. baumannii* MDR se ha convertido en un problema grave, ya que es el segundo microorganismo más frecuentemente aislado de pacientes infectados (comunicación personal con el Dr. Rafael Franco). Se secuenciaron diez aislados de *A. baumannii* que mostraron un grado importante de variación en perfiles de resistencia y en otros datos clínicos como la fuente de aislamiento (sange, heridas, vías respiratorias) (Tabla 6.1). Los aislados pertenecen a solo cuatro ST_O diferentes. El aislado H350 se asignó al ST_O208 en el esquema de Oxford, que es equivalente al ST_P2 del esquema de Pasteur y pertenece al complejo clonal 2 (CC2), también conocido como International Clone II (IC II), que comprende cepas encontradas en todo el mundo. El aislado H008 se asignó al ST_O690, un ST previamente reportado en Río de Janeiro, Brasil, y para el cual no conocemos otro genoma secuenciado reportado hasta el momento, ni tampoco reportes de aislamiento en otras partes del mundo. El ST_O más frecuente en la muestra es el ST_O758, seguido de un single locus variant (SLV), ST_O1091 (Ver Figura 5.1). El ST_O1091, al igual que ST_O758 se reportó por primera vez en México pero se han encontrado también aislados en los EE. UU (PubMLST, <https://pubmlst.org/>) (Wright et al., 2016).

El primer paso fue poner en contexto estos nuevos aislados en relación a los aislados Mexicanos del mismo linaje que ya se han analizado (Tabla 6.1), pero también

con respecto a aislados provenientes de Honduras que recientemente pasaron a formar parte de la colección del Programa de Genómica Evolutiva, y cepas con genomas completamente secuenciados y cerrados, públicamente disponibles (NCBI) que al inicio de este trabajo alcanzaban un total de 111. La base de datos generada es muy exhaustiva, además de representar una gran parte de los ST más incidentes a nivel global (más de 50), incluye cepas provenientes de 21 países aisladas en un período de 66 años (Tabla suplementaria 1, Apéndice B-B.1). Esta base de datos incluye también un aislado Mexicano del ST_O758 recuperado del Hospital Infantil Federico Gómez de la ciudad de México (810CP), el cual nos permite, a pesar de ser uno solo, ampliar el análisis a tres hospitales Mexicanos.

Tabla 6.1 Aislados recuperados del INR

Cepa	ST _O	Contigs	N50	L50	Genoma (pb)	Cobertura
A023	1091	220	46224	27	3974954	138.7
Ab158	758	397	21352	48	3733923	193.2
A229	1091	241	31792	37	3806034	192.6
A278	758	166	76762	15	4013639	169.1
A465	758	328	41960	28	3995467	175.3
A580	758	328	36471	34	3942644	139.3
H008	690	271	50945	26	4149770	153.7
H170	1091	327	21382	50	3665579	145.2
H350	208	203	55522	19	3882957	150.2

Se construyó una filogenia basada en el concatenado de los alineamientos de los SGF (766 SGF), sin evidencia de recombinación (388 SGF) (Figura 6.1). Esta filogenia es congruente con la presentada en la primera parte de Resultados, por lo tanto se pueden observar nuevamente los clados en los que se agrupan los aislados de las clonas internacionales (IC1 en naranja, IC2 en azul) y también el clado que contiene a la mayoría de los aislados Mexicanos. Este clado se ve claramente que contiene los ST_O pertenecientes al complejo clonal 636 (CC636) (en violeta) (Figura 5.1), el cual incluye no solamente a los ST_O1091 y ST_O758, sino también otros ST relacionados como el ST_O924 y ST_O1321. Los aislados ST_O758 están agrupados en la filogenia, tanto los Mexicanos como los Hondureños y la cepa Canadiense.

La mayoría de aislados Hondureños e incluso un par de aislados Mexicanos no agrupan en el clado C636, sino que se encuentran dispersos en el árbol, revelando un bajo grado de parentesco con el linaje ST_O758. Esto indica que varios clones coexisten en esos países. Por ejemplo, con respecto a los aislados recientemente secuenciados del INR (puntos anaranjados en la Figura 6.1), encontramos no solo los asignados a ST_O758 sino también un aislado (H008) que pertenece al ST_O690 y otro aislado (H350) asignado a el ST_O208, el primero agrupa con cepas no relacionadas con los complejos internacionales, y el segundo se ubica en el clado IC II. Se observa que un patrón simi-

lar se aplica a los aislamientos Hondureños, donde dos de los hospitales muestreados (puntos azules y verdes) tienen varios ST (consulte la Tabla Suplementaria 1, Apéndice A). Indicando que, linajes muy distantes son capaces de coexistir en un mismo hospital, pero también que un mismo linaje puede ser transmitido fácilmente entre hospitales.

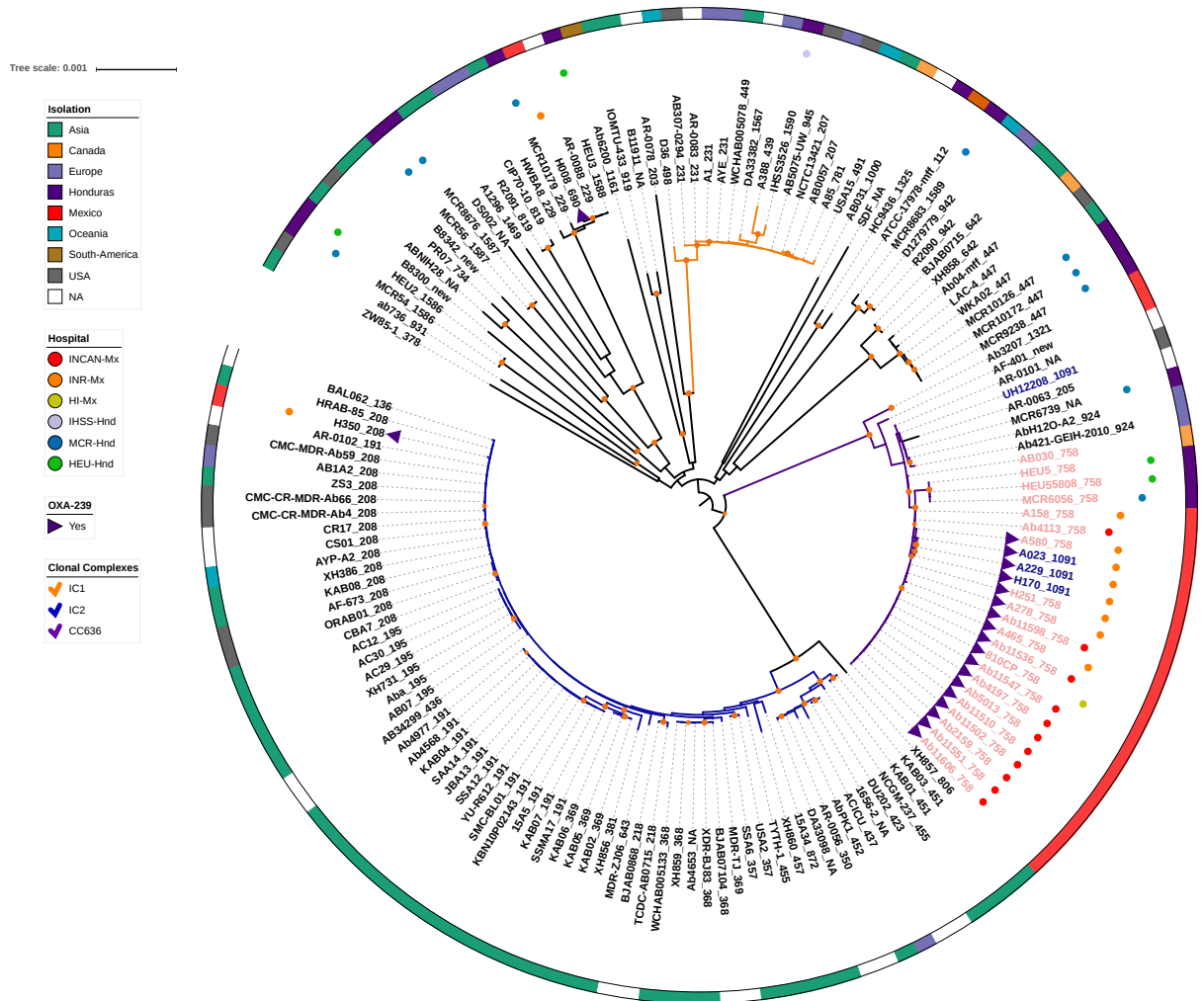


Figura 6.1 Reconstrucción filogenética en base a SGFs. La escala del árbol se ve arriba a la derecha e indica un cambio cada 100 nucleótidos.

El linaje ST_O758, que incluye cepas Mexicanas, Hondureñas y una Canadiense parece tener muy poca variación genética juzgando por las ramas cortas en la filogenia, lo que podría implicar un surgimiento muy reciente de este clado, se realizó un análisis de dataje molecular para obtener una estimación de su aparición. De acuerdo con este análisis, el clado surgió muy recientemente a principios de 2008, con un intervalo de HPD del 95 % (finales de 2007, finales de 2008) (Figura 6.2, panel superior).

De igual forma que en la filogenia realizada anteriormente, en esta, las relaciones entre los aislados dentro del clado Mexicano son difíciles de observar debido a los bajos niveles de variación que presentan, los cuales se traducen en ramas muy cortas en la escala del árbol. Por lo tanto, se construyó una filogenia que incluye solamente los

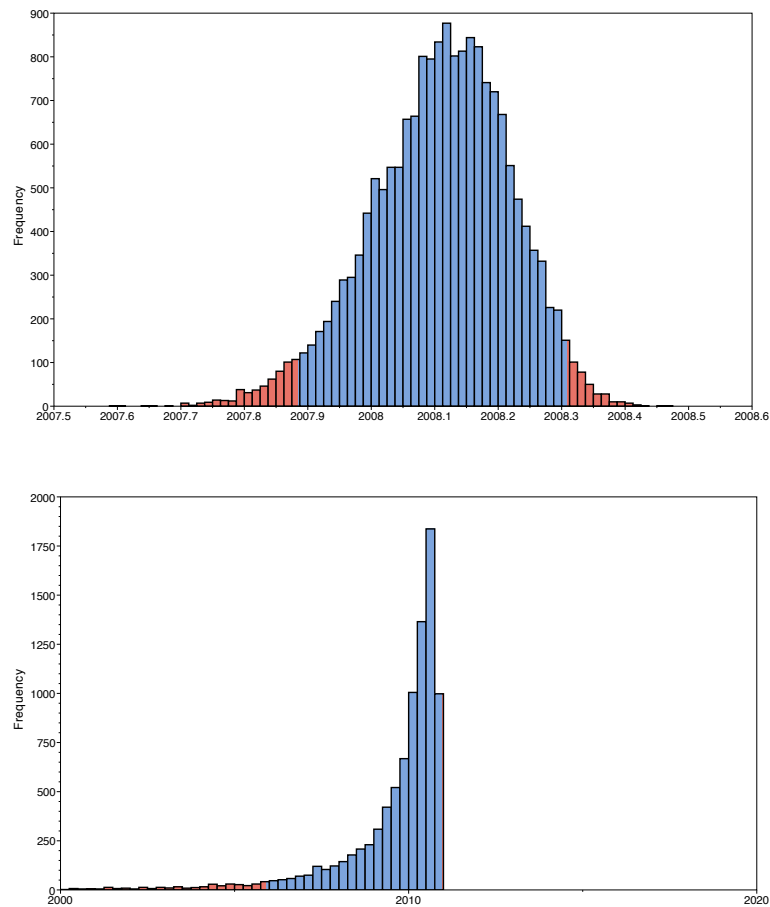


Figura 6.2 Distribución de las probabilidades a posteriori de emergencia del clado ST758 (superior), y del alelo de B-lactamasa de clase d, OXA-239 (inferior).

aislados Mexicanos de este clado y otros cercanamente emparentados (Figura 6.3). Lo más interesante de esta filogenia es que al revelarse las relaciones entre estos aislados Mexicanos, se nota que no hay un agrupamiento por hospital, o sea, los aislados de cada hospital no forman grupos monofiléticos, sino que el clado es un grupo polifilético con respecto a los sitios de aislamiento, sugiriendo que posiblemente ocurrieron diferentes eventos de intercambio entre los hospitales durante este corto período. Cabe destacar el hecho de que los aislados de ST_O1091 del INR agrupan con aislados del ST_O758 Mexicanos y no con el genoma UH12208 (ST_O1091 de EEUU), esto representa un evento de diversificación independiente y no una introducción exógena del linaje ST_O1091 al hospital.

En conjunto, estos resultados implican que varios linajes están circulando en los hospitales de México y Honduras analizados aquí y que el clado ST_O758 parece haber surgido hace unos 10 años y en ese período de tiempo se ha extendido por América

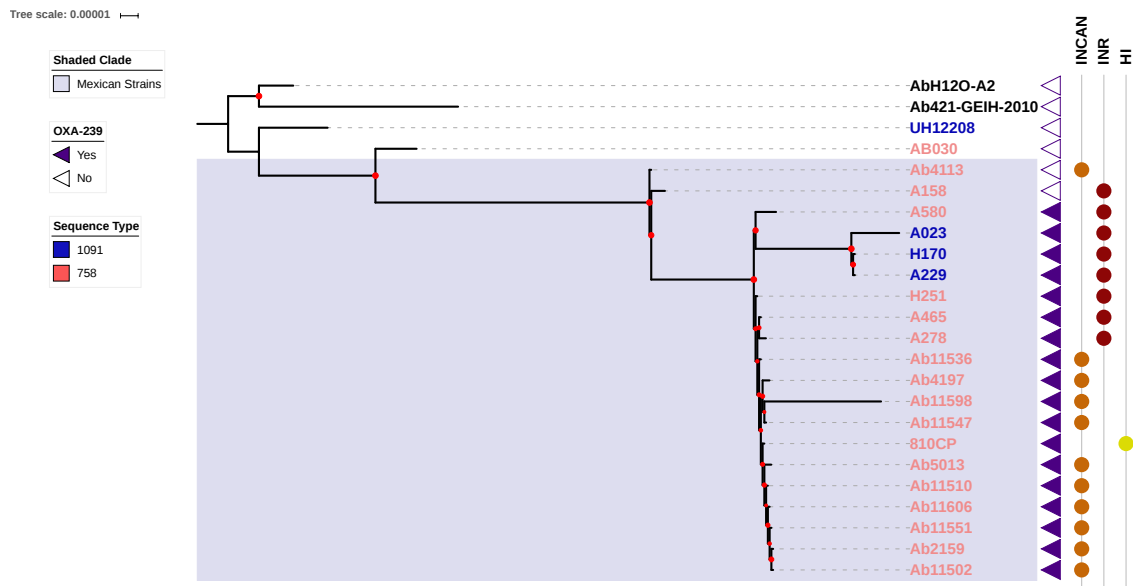


Figura 6.3 Filogenia del caldo Mexicano y aislados cercanamente emparentados. Se puede observar que el clado está compuesto por cepas de los tres hospitales de Ciudad de Mexico muestreados y la presencia o ausencia del alelos OXA-239, la cual indica una adquisición reciente del alelo.

Central (Honduras) y América del Norte (México y Canadá). Queda también demostrada la importancia del uso de WGS para analizar los aislados presentes en un hospital, ya que la resolución limitada de métodos de tipificación como MLST pueden enmascarar el rol de la transferencia horizontal y la recombinación en los procesos de diversificación. Además, aparentemente el ST_{O758} y otros linajes han sido bastante descuidados en cuanto a su estudio, ya que casi no hay información sobre ellos en el PubMLST.

6.2. El reciente linaje ST_{O758} muestra variación en los perfiles de resistencia a antibióticos.

La variabilidad fenotípica en cuanto a la resistencia a antibióticos de las cepas del clado mexicano se estudió a través de antibiogramas (Tabla 6.2). En el INR se realizaron estos experimentos específicamente para estos aislados, mientras que en el caso de las cepas del INCAN, la información fenotípica se encontraba dispersa en diferentes documentos, y las pruebas no fueron realizadas de manera sistemática. Por lo tanto, la tabla 6.2 muestra también la escasez de información que tenemos en algunos casos.

Se realizó también la predicción *in silico* de la resistencia a los antibióticos considerando todos los aislamientos dentro del clado ST_{O758} , o sea incluyendo AB030, UH1228, 810CP, cepas hondureñas, y mexicanas del INCAN y del INR (ver Figura 6.4). La información fenotípica que poseemos para contrastar estos resultados se limita

a cepas del INR (Tabla 6.1) y del INCAN (Tabla 5.1).

Los aislados del INR mostraron marcadas diferencias en cuanto a la resistencia a diferentes antibióticos. A nivel fenotípico, se puede observar variación en los perfiles de resistencia a antibióticos entre los aislados, pero también altas tasas de resistencia a muchos de los antibióticos analizados.

Parece claro que existe una variación considerable en el contenido de genes de resistencia a los antibióticos entre los aislados del ST_O758. Aunque este es un clado que emergió recientemente, encontramos 9 perfiles de determinantes de resistencia diferentes considerando la predicción *in silico*. Para los aminoglucósidos gentamicina y amikacina, encontramos fenotipos sensibles, intermedios e incluso resistentes sin bases genéticas claras que expliquen las diferencias. La mayoría de los aislados fueron resistentes contra doripenem, imipenem y meropenem, ya que todas excepto dos cepas tenían CIM superiores a 16 para los tres antibióticos. No se encontró ningún aislamiento resistente a colistina ni las mutaciones asociadas a este fenotipo (Moffatt et al., 2010).

Tabla 6.2 Perfiles de Resistencia para las cepas Mexicanas en base a Antibiograma. R:Resistente,S:Suceptible,I:Intermedio,NA:No Analizado

Cepa	IPM	MEM	DOR	TZP	CRO	CAZ	CTX	FEP	AMK	GEN	MIN	TOB	FOX	CIP	LVX	SXT	TET	CST	AMP/S
INCAN																			
11502	R	R	NA	R	R	NA	R	R	R	R	NA	R	NA	R	R	R	S	NA	NA
11510	R	R	NA	R	R	NA	R	R	R	R	NA	R	NA	R	R	R	I	NA	NA
11536	R	R	NA	R	R	NA	R	R	R	R	NA	R	NA	R	R	R	R	NA	NA
11547	R	R	NA	R	R	NA	R	R	R	R	NA	R	NA	R	R	R	I	NA	NA
11551	R	R	NA	R	R	NA	R	R	R	R	NA	R	NA	R	R	R	I	NA	NA
11598	R	R	NA	NA	R	R	R	NA	R	R	NA	R	NA	R	R	R	NA	NA	NA
11606	R	R	NA	NA	R	R	R	NA	R	R	NA	R	R	R	NA	NA	NA	NA	NA
4113	S	S	NA	I	R	R	R	I	R	R	NA	R	R	R	R	R	S	NA	NA
INR																			
H350	R	R	R	R	NA	R	NA	R	R	R	S	NA	NA	R	R	NA	NA	S	I
H008	R	R	R	R	NA	R	NA	R	S	R	S	NA	NA	R	R	NA	NA	S	I
H170	R	R	R	R	NA	R	NA	R	S	S	S	NA	NA	R	R	NA	NA	S	R
H251	R	R	R	I	NA	I	NA	R	R	R	S	NA	NA	R	R	NA	NA	S	S
A023	R	R	R	R	NA	R	NA	R	S	R	S	NA	NA	R	R	NA	NA	S	I
A229	R	R	R	R	NA	R	NA	R	R	I	S	NA	NA	R	R	NA	NA	S	R
A580	R	R	R	R	NA	R	NA	R	S	I	S	NA	NA	R	R	NA	NA	S	R
A465	R	R	R	R	NA	R	NA	R	S	I	S	NA	NA	R	R	NA	NA	S	I
A158	R	R	R	R	NA	R	NA	I	S	I	S	NA	NA	R	R	NA	NA	S	S
A278	R	R	R	R	NA	R	NA	R	S	R	S	NA	NA	R	I	NA	NA	S	I

Amikacina(AMK);Gentamicin(GEN);Ceftazidima(CAZ);Cefepime(CPE);Ciprofloxacino(CIP); Levofloxacino(LVX);Doripenem(DOR);Imipenem(IPM);Meropenem(MEM);Colistin(CST); Piperacillin/tazobactam(TZP);AmpicilinaSulbactam(AMP/s);Minocycline(MIN);Ceftriaxone(CRO); Ceftazidime(CAZ);Cefotaximo(CTX);Cefepime(FEP);Tobramycin(TOB);Cefoxitin(FOX); Trimethoprim-sulfamethoxazole(SXT);tetracycline(TET)

Tampoco se observó resistencia a minociclina lo cual es concordante con la ausencia en todos los aislados de tetB (Martí et al.). Finalmente, todos los aislados mostraron resistencia contra piperacilina-tazobactam, pero se encontró una gran variabilidad en la respuesta a la ampicilina/sulbactam. No se encontró una base genética clara para es-

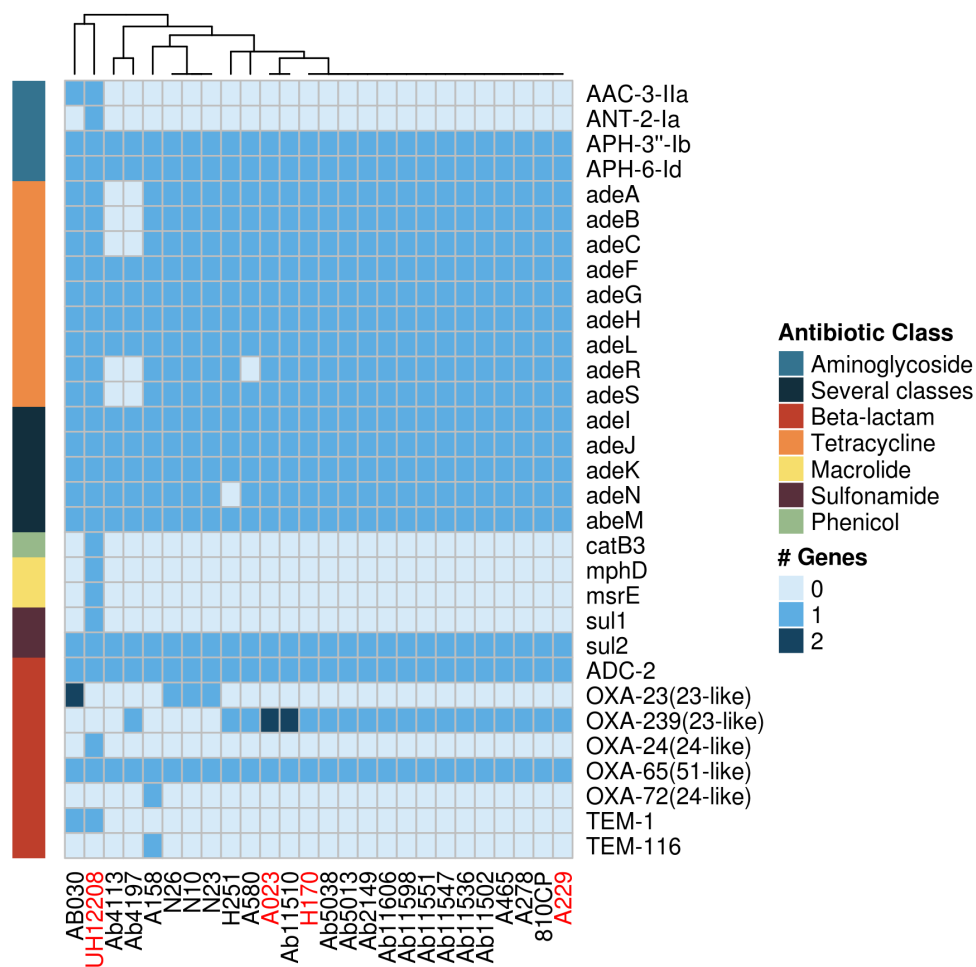


Figura 6.4 Perfiles de Resistencia a antibióticos obtenidos con la predicción in silico. Las filas no están agrupadas por clustering sino que se ordenan por clase de antibiótico. El agrupamiento de las cepas si demuestra la similitud de sus perfiles. 14 cepas tienen perfiles idénticos, todas ellas Mexicanas pero de distintos hospitales.

ta variación, aislados con perfiles de resistencia idénticos mostraron tanto resistencia como susceptibilidad. La cepa A158 muestra susceptibilidad pero se ha informado de resistencia en una cepa con un perfil de β -lactamasas muy similar (TEM-1, OXA-72, OXA- 66) (Li et al., 2015). El nivel de expresión de *adeB* se ha asociado con la resistencia a la ampicilina/sulbactam (Chiu et al., 2010), lo que podría aumentar el fenotipo resistente y no podría ser determinado con este análisis. Como era de esperar, todos los aislados comparten el alelo correspondiente al gen *bla_{OXA-51}* (OXA-65), dada la asociación de estas variantes con los diferentes linajes epidémicos (Evans et al., 2008), esto explica también a la resistencia observada a ciproflaxina y cefepime.

Con respecto a las β -lactamasas de clase D adquiridas, un par de aislados del clado poseen alelos de la familia *bla_{OXA-24}* pero la mayoría de los aislados (todos menos tres) tienen genes que pertenecen a la familia *bla_{OXA-23}*. Curiosamente, el alelo predominante de la familia de genes tipo *bla_{OXA-23}* fue *bla_{OXA-239}*, una variante reciente-

mente descrita en México (Tamayo-Legorreta, Garza-Ramos et al. 2014). De las cepas Mexicanas, solo Ab4113 es sensible a meropenem, esto está de acuerdo con la falta de OXA-239 en Ab4113. A158 también carece de OXA-239 pero es resistente a meropenem e imipenem, en lugar de OXA-239, A158 posee una β -lactamasa de la familia OXA-72 (también encontrada en ST_O758 (Gonzalez-Villoria et al., 2016)) y una ESBL (extended spectrum -lactamase), TEM-116. Los aislamientos A158 y Ab4113 son los que tienen las secuencias más divergentes dentro del grupo INR + INCAN (diferencias pareadas de SNPs) y la falta del alelo OXA-239 en ambas es una característica interesante porque significa que la adquisición del alelo bla_{OXA-239} posiblemente es muy reciente.

El gen bla_{OXA-239} se encuentra en los aislados mexicanos solamente, los aislados del clado ST_O758 que fueron recuperados de otros países (Honduras, Canadá, EEUU) poseen el alelo bla_{OXA-23} o carbapenemasas de la familia bla_{OXA-24}.

En conjunto, nuestros datos sugieren que estos aislamientos no solo son resistentes a carbapenémicos sino también MDR y que a pesar de pertenecer a un clado de origen reciente, muestran variación en sus perfiles de resistencia. Además, parece que los genes de la familia bla_{OXA-23} funcionan como el principal determinante de resistencia a carbapenémicos y en el caso de las cepas mexicanas esta familia está representada por un alelo distinto recientemente descrito y posiblemente recientemente adquirido.

6.3. Historia del alelo bla_{OXA-239}

Alelos de la familia de β -lactamasas de clase D OXA-23 se han encontrado en aislados de *A.baumannii* alrededor del mundo (Mugnier et al., 2010), el alelo bla_{OXA-239} fue descrito recientemente en aislados Mexicanos del ST_O758 y ST_O1091 (Tamayo-Legorreta et al., 2014) y hasta ahora no se ha reportado en otras ubicaciones. Realizamos una búsqueda de BLASTP (ver métodos) y no encontramos ninguna coincidencia exacta (100 % de identidad de aminoácidos) con OXA-239 en la base de datos de NCBI, excepto los aislados reportados y descritos anteriormente en México. Notablemente, ninguno de los aislados que tienen el alelo bla_{OXA-239} se recolectó antes de 2010, esto sugiere una aparición muy reciente de esta variante. Para corroborar esto, realizamos otro análisis de dataje molecular utilizando las secuencias bla_{OXA-239} y sus genes bla_{OXA} más cercanamente relacionados (ver métodos).

De acuerdo con este análisis, bla_{OXA-239} surgió a principios de 2010 (ver Figura 6.2, panel inferior), lo cual se ajusta completamente a las fechas de recolección de los aislados que poseen este gen. Luego, para explorar las relaciones de parentesco de bla_{OXA-239} con otros alelos de la familia OXA-23, construimos una filogenia (ver Figura 6.5). Notamos que la secuencia más cercana a bla_{OXA-239} fue bla_{OXA-469}, un gen que también se describió solo en México en 2012 (Gonzalez-Villoria et al., 2016).

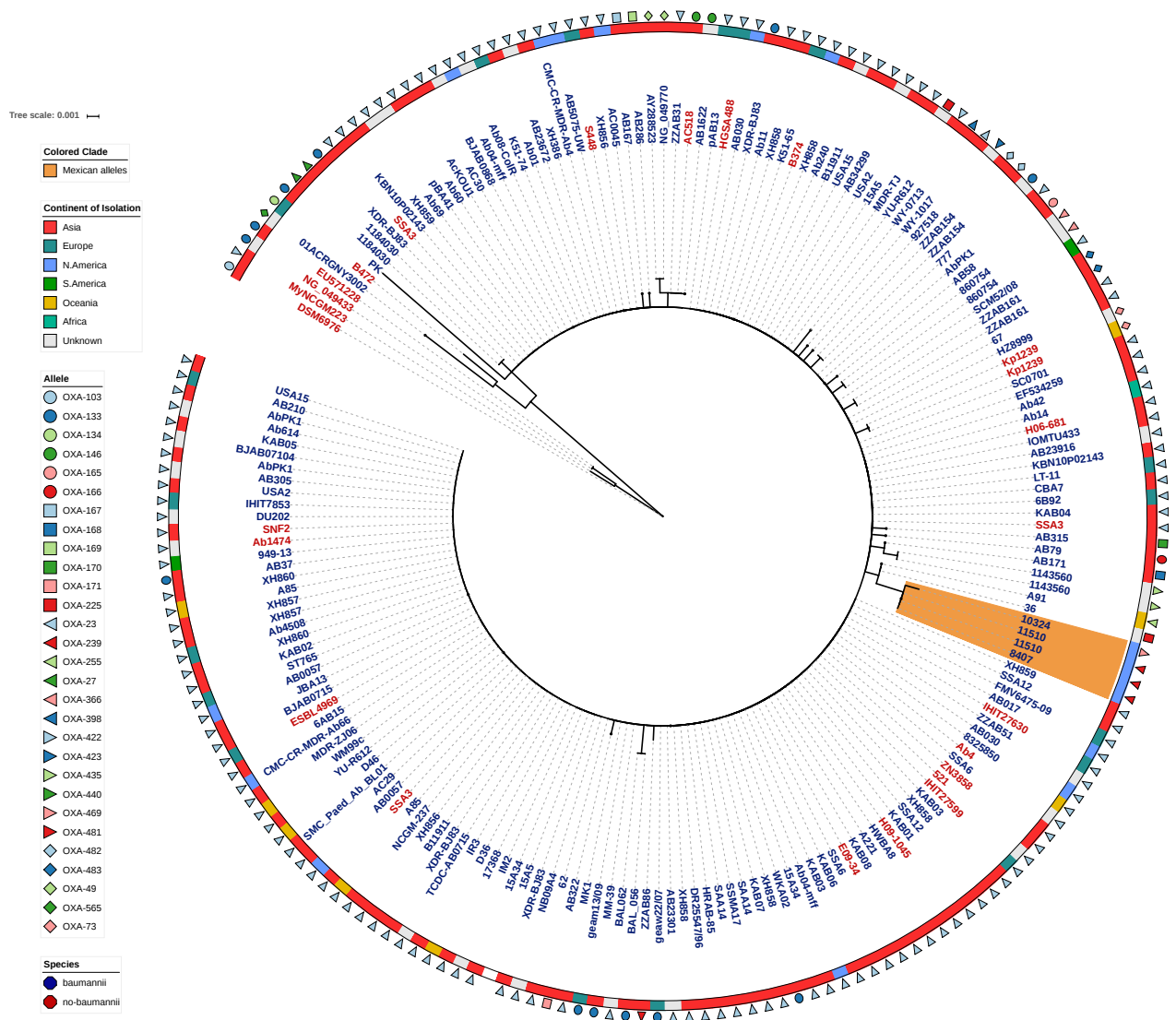


Figura 6.5 Reconstrucción filogenética por ML que refleja las relaciones de parentesco entre los diferentes alelos de la familia de β -lactamasas de clase D, OXA-23. Se incluyeron secuencias reportadas para *A. baumannii* pero también para otras especies. Alelos de esta familia se han descrito alrededor del mundo.

Las secuencias de aminoácidos de OXA-239 y OXA-469 son casi idénticas, difieren en un solo aminoácido en la posición 222 (M en OXA-239 e I en OXA-469). Además, en términos de la secuencia aminoacídica relativa a OXA-23, OXA-239 tiene las siguientes sustituciones: S109L, D222N y P225S. La posición 109 está en el bucle α 3- α 4, mientras que 222 y 225 están en el bucle β 6- β 7. En OXA-23, Phe110 en el bucle α 3- α 4 y Met221 en el bucle β 6- β 7 forman el túnel que se cree que es importante para la actividad de la carbapenemasa (Smith, Antunes et al. 2013). En particular, OXA-239 parece ser el primer caso de un alelo con dos sustituciones en comparación con OXA-23, en el bucle 6-7 (consulte la Figura 1, Apéndice 3). Dos de los aminoácidos que difieren entre OXA-23 y OXA-239 (L109 y N222) están inmediatamente adyacentes al F110 y se cree que

M221 desempeña un papel importante en la actividad de la enzima carbapenemasa.

Se ha planteado la hipótesis de que los cambios en los residuos de aminoácidos en el bucle 6-7 son responsables de la especificidad alterada al sustrato enzimático (Kaitany, Klinger et al. 2013). Por lo tanto, parece que las sustituciones encontradas en OXA-239 han modificado la actividad enzimática de la OXA-23 original. Sorprendentemente, no solo encontramos $bla_{OXA-239}$ en los aislados ST_O758 y ST_O1091 del CC636, sino también en otros dos aislados no relacionados, el H008 tipificado como ST_O690 y el H350 tipificado como ST_O208 (ver triángulos morados en la 6.1); esto implica que el alelo $bla_{OXA-239}$ ya ha sido sujeto a transferencia horizontal entre linajes. Cabe destacar que los tres linajes implicados en estos posibles eventos de HGT se aislaron en el INR (ver la Figura 6.1 y Tabla 6.1). Para dos de las cepas mexicanas registramos dos copias de este gen (ver Figura 6.4, A023 y Ab11510); en el caso de A023, una de las copias está ubicada en el cromosoma y la otra está en un plásmido putativo, mientras que para el Ab11510 ambas copias residen en el cromosoma.

Desde una perspectiva epidemiológica es importante conocer la incidencia de genes de resistencia a carbapenémicos en hospitales, pero más relevante aún puede ser conocer el contexto de estos genes, lo que provee información adicional sobre la dinámica de su dispersión. Hasta la fecha, los genes de la familia OXA-23 se han encontrado en por lo menos 4 contextos diferentes; Tn2006, Tn2008, Tn2008B y Tn2009 (Nigro and Hall, 2016). El origen de OXA-23 se ha rastreado a la especie *Acinetobacter radioresistens*, donde se localiza en el cromosoma y se considera una carbapenemasa intrínseca, su salto a *A. baumannii* se ha estudiado y se cree que puede haber sido transferida en varias ocasiones (Poirel et al., 2008). Intentamos establecer el contexto genómico del alelo $bla_{OXA-239}$ en esta muestra (Figura 6.6). En algunos casos es extremadamente difícil conocer el contexto debido a la fragmentación de los ensamblados (ver Tabla 6.1).

Cuando usamos lecturas cortas pareadas para ensamblar (ver métodos), las secuencias repetidas que se encuentran en los genomas, pueden confundir a los programas de ensamblado generando quiebres en los grafos, lo que resulta en genomas altamente fragmentados. Los transposones que movilizan a los genes de la familia OXA-23 están compuestos por una o más secuencias de inserción (ISs) (que presentan generalmente muchas copias en *A. baumannii*), de modo que en muchos de nuestros aislados no se pudo establecer con exactitud la estructura de dispersión que acompaña a $bla_{OXA-239}$. Afortunadamente, algunos de estos genomas pueden ser usados de referencia debido a que fueron secuenciados con tecnología PacBIO (AB030, Ab11510 y 810CP), la cual usa lecturas largas que resuelven el problema mencionado anteriormente.

Encontramos que la cepa canadiense AB030 posee el Tn2006 con el alelo bla_{OXA-23} , mientras que Ab11510 y 810CP, ambas mexicanas, poseen el Tn2008 transportando el alelo $bla_{OXA-239}$. A pesar de no poder establecer con precisión esto en el resto de las cepas recientemente secuenciadas, estudiamos el contexto de $bla_{OXA-239}$ en la medida de

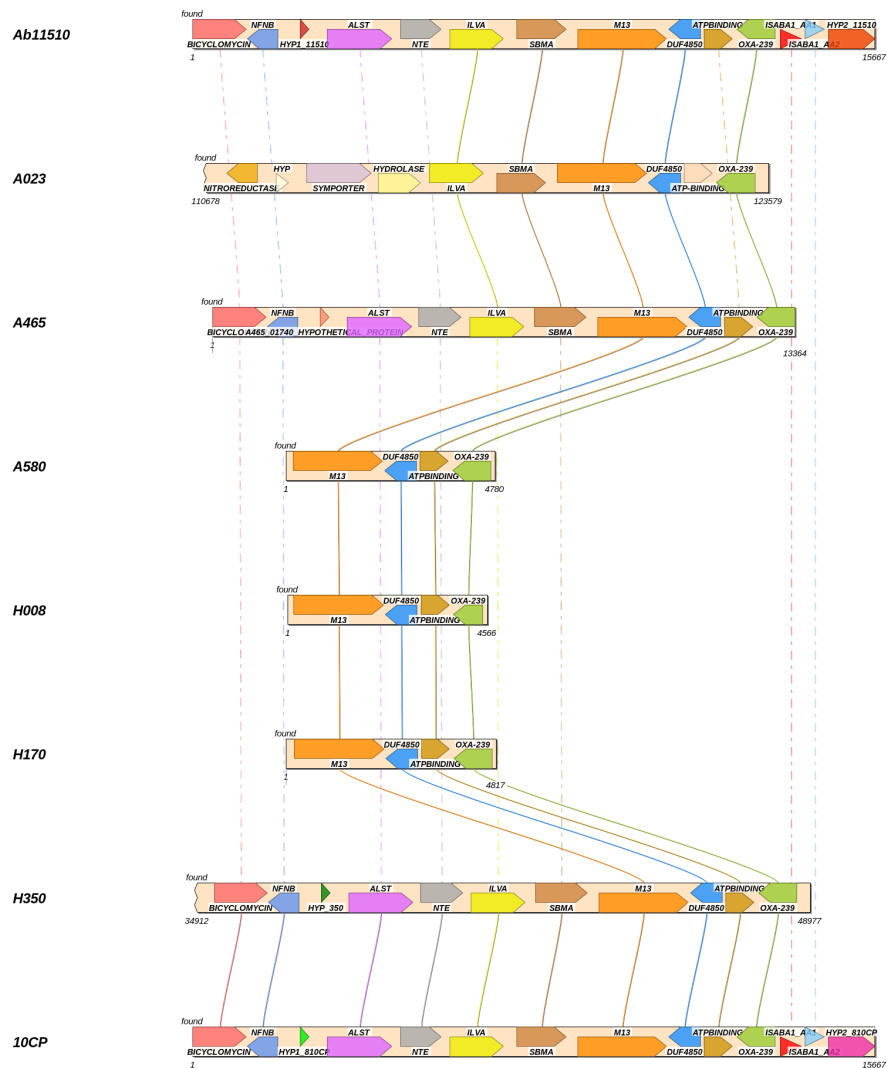


Figura 6.6 Orden de los genes en el vecindario del alelo OXA-239. Las cepas superior e inferior son las usadas como referencia por tener sus genomas cerrados.

lo posible, y lo comparamos con las cepas de genomas cerrados. Encontramos que posiblemente, las cepas mexicanas compartan el contexto genómico de $bla_{OXA-239}$ dado que aunque no podemos identificar la secuencia de inserción (ISAbA1) que conforma el Tn2008, sí podemos identificar el fragmento de ATPasa derivado de *A. radioresistens*, y el orden de los genes subsiguientes el cual es compartido con Ab11510 y 810CP

en la mayoría de los aislados mexicanos (Figura 6.6). Como se observa en la Figura 6.6, hay aislados con un grado de fragmentación del ensamble tal, que poseen el alelo $bla_{OXA-239}$ en un pequeño *contig* (H008, H170 y A580), el cual además de este alelo, solo posee la secuencia ATPasa y una proteína más, M13. Otros aislados presentan mejor calidad de ensamble, y el orden de muchos genes más se puede comparar con las cepas de referencia, permitiendo corroborar que el contexto genómico de OXA-239 se conserva en aislados de INR, INCAN y HI.

Hay una herramienta más que podemos usar, que es buscar los repetidos directos que se generan en el genoma debido a la inserción del transposón. Los repetidos directos pueden tener entre 2 y 15 pares de bases y son específicos del sitio de inserción del transposón en el genoma. Buscamos estos repetidos directos muchos de los cuales han sido reportados para diferentes sitios de inserción del Tn2008 (Nigro and Hall, 2016). Encontramos los mismos repetidos directos en varias de las cepas, lo que confirma que el sitio de inserción en el genoma es el mismo, y que posiblemente los aislados de linajes no relacionados con ST_{O758}, obtuvieron el alelo partir de este linaje. A su vez, la mayoría de las cepas del linaje ST_{O758} podrían haber obtenido el alelo por transferencia vertical. Sin embargo, además de este contexto genómico existen otros, en el caso de los genomas con más de una copia, y en el genoma del aislado A278.

En general, nuestros resultados sugieren que el alelo $bla_{OXA-239}$ surgió recientemente en el linaje ST_{O758} y que rápidamente ha sido transferido a otros linajes e incluso se ha duplicado en dos cepas.

6.4. Discusión

Hoy en día, los aislados de *A. baumannii* que poseen carbapenemasas, son una amenaza importante en muchos hospitales de todo el mundo. No obstante, se sabe muy poco acerca de la diversidad genómica de los aislamientos que circulan en América Latina. Aquí utilizamos WGS para revelar la dinámica espacial y temporal de los linajes (y sus genes de resistencia a los antibióticos) de *A. baumannii* en Honduras y México.

En particular, encontramos que muchos de los aislados de estos dos países no forman parte de los clones internacionales I-III, de hecho no están ni cercanamente emparentados; además, la mayoría de los aislados mexicanos y algunos hondureños pertenecen al linaje ST_{O758}, confirmando una vez más la importante presencia de este ST en América. Según la base de datos PubMLST y varios estudios, se han reportado aislados de ST_{O758} en Canadá (Loewen et al., 2014), México (Graña-Miraglia et al., 2017), Honduras (López-Leal et al., 2019), y en Colombia (Vanegas et al., 2015). El análisis de esta información y nuestros propios estudios, sugieren que este es un linaje endémico de las Américas.

El análisis de dataje molecular indica una aparición muy reciente de este clado y las

fechas de recolección informadas en los estudios mencionados anteriormente para los aislamientos ST_O758 son congruentes, ya que ninguno de esos aislamientos se recolectó antes de 2010. En conjunto, estos datos sugieren una muy rápida propagación de este clon: en menos de 12 años se ha extendido a varios países en el sur, centro y norte de América. Sin embargo, se requieren más aislamientos, especialmente de América del Sur, para obtener una estimación más precisa del origen de este clado.

Aquí es importante mencionar que el ST_P79 bajo el esquema Pasteur MLST, es equivalente a varios de los ST_O que conforman el CC636. La presencia de ST_P79, se informó recientemente en América del Sur. Por ejemplo, un informe que analiza aislados del sur de Brasil reportó que el ST más frecuente a finales de la década del 2000 en el sur de Brasil era ST_P79 (Pagano et al., 2018). Además, otro estudio reciente que considera aislados peruanos, también encontró el ST_P79 (Levy-Blitchtein et al., 2018). La secuenciación del genoma de estas cepas de Perú y Brasil podría proporcionar nuevas perspectivas sobre el momento y el lugar de aparición del clado.

La dificultad de contrastar ciertas hipótesis solamente con información de MLST se evidencia nuevamente, a pesar de que los estudios mencionados aportan indicios de un linaje Americano, lo que podemos decir es que tanto el ST_O758 como ST relacionados están circulando en América y parecen ser más relevantes en cuanto a su incidencia de lo considerado hasta ahora.

Es importante destacar también, que estudios recientes han encontrado STs que no están relacionados con los clones internacionales en otras partes de América Latina (López-Leal et al., 2019; Vanegas et al., 2015). Un estudio muy reciente sobre la diversidad genética de aislamientos de Colombia encontró 7 nuevos ST no descritos anteriormente, algunos de ellos pertenecientes al Complejo Clonal (CC) 636, que es el CC que contiene el ST_O758. Teniendo en cuenta todo esto, es muy probable que América Latina pueda ser un reservorio de STs desconocidos no relacionadas con los clones internacionales más estudiados.

Notamos que varios linajes distintos, con variados perfiles de resistencia a antibióticos pueden coexistir en el mismo hospital. Pensamos que algunos de estos linajes podrían ser linajes endémicos firmemente establecidos, mientras que otros son representantes de linajes que probablemente se introducen periódicamente en a los hospitales. Sin embargo, dada la alta cantidad de HGT en esta especie (Graña-Miraglia et al., 2017; Snitkin et al., 2011), estos linajes transitorios pueden estar introduciendo constantemente nuevos genes a los linajes establecidos (resistentes a los antibióticos y otros), siendo esto una de las razones del éxito de *A. baumannii* como patógeno nosocomial. Del mismo modo, los linajes pasajeros pueden adquirir genes a partir de los linajes establecidos, homogenizando así la población del hospital, por lo menos en cuanto a fenotipos de resistencia como vimos aquí con el alelo bla_{OXA-239}.

Las pruebas de susceptibilidad antimicrobiana para los aislados secuenciados re-

cientemente y los perfiles de resistencia *in silico* para el clado ST_O758 muestran una cantidad considerable de resistencia, ya que en ambos análisis el conjunto completo de cepas fue MDR. Además, aunque este es un clado muy reciente, encontramos 9 perfiles de resistencia diferentes considerando la predicción *in silico* que implica que la adquisición y pérdida de genes de resistencia a antibióticos es altamente dinámica.

Es importante destacar que observamos la presencia de varias familias de oxacilinasas en ST_O758; en promedio, cada aislamiento posee dos tipos de OXA, una de ellas es la bla_{OXA-51} la cual es intrínseca en *A. baumannii* y cuyos alelos están relacionados con linajes epidémicos. La segunda puede pertenecer a varias familias pero la familia OXA-23 es la más frecuente y está ausente en solo tres aislados. Por otro lado, alelos de la familia OXA-40 se encontraron en dos aislados, uno de ellos, el bla_{OXA-72}, se ha reportado tanto en México (Gonzalez-Villoria et al., 2016) como en Sudamérica (Rodríguez et al., 2016). Por lo tanto, parece muy probable que para este clado la expresión de una variedad de oxacilinasas sea el determinante más común de la resistencia al carbapenémicos.

El enfoque de genómica poblacional utilizado nos permitió estudiar la reciente aparición del nuevo alelo de la enzima OXA-23, el alelo bla_{OXA-239}; el cual parece haberse originado en 2010. Esta variante particular solo se ha reportado en México y, en todos los casos, los aislamientos que tienen este alelo se muestrearon después de 2010. Aquí vale la pena mencionar que, a pesar de los cientos de genomas disponibles públicamente en NCBI, no se encontró ninguna coincidencia exacta con el bla_{OXA-239} que no fuera el alelo ya descrito en aislados mexicanos. Las sustituciones que posee este alelo en comparación con el OXA-23 han alterado la especificidad del sustrato enzimático. Un estudio reciente que analiza las propiedades cinéticas y estructurales de bla_{OXA-239} concluyó que esta enzima ha expandido su especificidad de sustrato a monobactámicos y cefalosporinas de última generación, aunque se detectó una menor actividad contra los carbapenémicos (Harper et al., 2018).

Aunque un estudio anterior solo ha encontrado este gen asociado al linaje ST_O758 (incluye ST_O1091) (Gonzalez-Villoria et al., 2016), encontramos este gen también en otros dos linajes (ST_O690 y ST_O208). En particular, estos tres linajes han coexistido en el Instituto Nacional de Rehabilitación, lo que sugiere que a pesar de que este alelo (bla_{OXA-239}) surgió recientemente, ya ha sido transferido de su fuente original (ST_O758) a otros dos linajes distantes. Por lo tanto, OXA-239 es una enzima que tiene poco tiempo circulando, aparentemente endémica de México, que muestra una historia muy dinámica, ya que ya ha sido sometida a HGT e incluso a duplicaciones internas dentro de algunas cepas.

En términos epidemiológicos este estudio es muy relevante, dado que refleja que es necesaria mayor vigilancia sobre los países no desarrollados, ya que podrían ser no solo reservorios de nuevos linajes sino también de nuevos genes de resistencia. En ese

sentido, la secuenciación de todo el genoma debería ser una herramienta mas frecuentemente utilizada, ya que otras estrategias de genotipificación no representan adecuadamente las relaciones filogenéticas de las cepas de esta especie (Castillo-Ramírez and Graña-Miraglia, 2019b).

7. Conclusiones y Perspectivas

7.1. Conclusiones

En este trabajo se presenta una caracterización genómica robusta de los aislados de *A. baumannii* más frecuentes circulando en hospitales mexicanos. Además se estudian aislados relacionados recuperados en la región, y el análisis se lleva a cabo en un marco global al incluir cepas de todo el mundo.

Utilizando una aproximación de genómica de poblaciones, fuimos capaces de definir una población del patógeno nosocomial *A. baumannii* y estudiar el rol de diferentes fuentes de variación genética sobre su genoma. Pudimos concluir que en etapas tempranas de diversificación de un linaje, la pérdida y ganancia de genes tiene un rol preponderante en la generación de diversidad genética.

Establecimos la importancia de un linaje de *A. baumannii*, el ST_O758 del CC636, posiblemente endémico de las Américas, y claramente subestimado en su estudio, este ST y otros relacionados son frecuentes en América y no están emparentados a los Complejos Internacionales I-III.

Datamos el surgimiento del alelo bla_{OXA-239} en 2010. Esta alelo fue descrito en México y no se ha reportado en otros lugares del mundo, a su vez detectamos la transferencia horizontal de bla_{OXA-239} a linajes divergentes del linaje en el que tuvo origen, en un período muy corto de tiempo.

En este trabajo se hace mucho énfasis en la utilidad de la WGS para estudios de vigilancia epidemiológica, especialmente los enfocados en *A. baumannii*. Hemos visto que la dinámica de pérdida y ganancia de genes de la especie es muy rápida y que esto afecta también a los genes de resistencia, impidiendo la asociación de determinados perfiles de resistencia con determinados linajes, esto es relevante para el diagnóstico y tratamiento de infecciones.

7.2. Perspectivas

Este trabajo evidencia que aún queda mucho por hacer en México en cuanto a vigilancia epidemiológica de *A. baumannii*. Uno de los posibles caminos a explorar serían

los estudios epidemiológicos prospectivos, a diferencia de los datos recabados en este trabajo los cuales son retrospectivos y dependen de las colecciones de cepas guardadas en los hospitales. Diseñar el propio muestreo afecta el tipo de preguntas que pueden hacerse.

También se demostró aquí que el uso de WGS para el estudio de la evolución de bacterias patógenas es clave, por ejemplo para estudiar la dinámica del contenido génico de las cepas. Sería relevante en este sentido evaluar que familias génicas se ven mayormente afectadas por esto. La acumulación de genomas de poblaciones de *A. baumannii* de diferentes estados de la República seguramente permita develar procesos evolutivos de la especie importantes no sólo para México, sino para el saber general.

Quizás el enfoque principal sobre *A. baumannii* a futuro sea el de explorar los determinantes de patogenicidad y virulencia de la especie, los cuáles aún son en gran medida un misterio. Para esto sería de gran importancia incluir en los análisis comparativos cepas de *A. baumannii* obtenidas de ambientes no hospitalarios. Estas cepas son escasas pero pueden ser de crítica importancia en el estudio de la evolución de la patogenicidad y virulencia.

La crisis de salud pública que se está atravesando en cuanto a resistencia a antibióticos y a las pesimistas predicciones a futuro, implica que la mayoría de los esfuerzos estén en la búsqueda de nuevos antibióticos. Pero la realidad es que sabemos menos sobre los procesos que llevan a la infección y enfermedad causadas por esta bacteria, y siendo que la disponibilidad de antibióticos es limitada, la prevención de las infecciones a través de información epidemiológica se torna cada vez más importante.

8. Bibliografía

- Aanensen, D.M., Feil, E.J., Holden, M.T.G., Dordel, J., Yeats, C.A., Fedosejev, A., Goater, R., Castillo-Ramírez, S., Corander, J., Colijn, C., et al. (2016). Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: a Population Snapshot of Invasive *Staphylococcus aureus* in Europe. *MBio* 7, e00444-16.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19, 455–477.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Comput. Biol.* 10, e1003537.
- Bruen, T.C., Philippe, H., and Bryant, D. (2006). A Simple and Robust Statistical Test for Detecting the Presence of Recombination. *Genetics* 172, 2665–2681.
- Castillo-Ramírez, S., and González, V. (2008). Factors affecting the concordance between orthologous gene trees and species tree in bacteria. *BMC Evol. Biol.* 8, 300.
- Castillo-Ramírez, S., and Graña-Miraglia, L. (2019). Inaccurate Multilocus Sequence Typing of *Acinetobacter baumannii*. *Emerg. Infect. Dis.* 25, 186–187.
- Chiu, C.-H., Lee, H.-Y., Tseng, L.-Y., Chen, C.-L., Chia, J.-H., Su, L.-H., and Liu, S.-Y. (2010). Mechanisms of resistance to ciprofloxacin, ampicillin/sulbactam and imipenem in *Acinetobacter baumannii* clinical isolates in Taiwan. *Int. J. Antimicrob. Agents* 35, 382–386.
- Corander, J., Marttinen, P., Sirén, J., and Tang, J. (2008). Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* 9, 539.

- Cox, M.P., Peterson, D.A., and Biggs, P.J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11, 485.
- Coyne, S., Courvalin, P., and Périchon, B. (2011). Efflux-Mediated Antibiotic Resistance in *Acinetobacter* spp. *Antimicrob. Agents Chemother.* 55, 947–953.
- Diancourt, L., Passet, V., Nemeč, A., Dijkshoorn, L., and Brisse, S. (2010). The Population Structure of *Acinetobacter baumannii*: Expanding Multiresistant Clones from an Ancestral Susceptible Genetic Pool. *PLOS ONE* 5, e10034.
- Dijkshoorn, L., Nemeč, A., and Seifert, H. (2007). An increasing threat in hospitals: multidrug-resistant *Acinetobacter baumannii*. *Nat. Rev. Microbiol.* 5, 939–951.
- Dixit, P. D., Pang, T. Y., Maslov, S. (2017). Recombination-Driven Genome Evolution and Stability of Bacterial Species. *Genetics*, 207(1), 281–295.
- EJ Feil, R.D., Robinson DA, Falush D. and Feil (2010). *Bacterial Population Genetics in Infectious Disease.* (Hoboken, NJ.).
- Esterly, J.S., Richardson, C.L., Eltoukhy, N.S., Qi, C., and Scheetz, M.H. (2011). Genetic Mechanisms of Antimicrobial Resistance of *Acinetobacter baumannii*. *Ann. Pharmacother.* 45, 218–228.
- Evans, B.A., Hamouda, A., Towner, K.J., and Amyes, S.G.B. (2008). OXA-51-like -lactamases and their association with particular epidemic lineages of *Acinetobacter baumannii*. *Clin. Microbiol. Infect.* 14, 268–275.
- Fournier, P.-E., Dubourg, G., and Raoult, D. (2014). Clinical detection and characterization of bacterial pathogens in the genomics era. *Genome Med.* 6, 114.
- Fouts, D.E., Brinkac, L., Beck, E., Inman, J., and Sutton, G. (2012). PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. *Nucleic Acids Res.* 40, e172.
- Francisco, A.P., Bugalho, M., Ramirez, M., and Carriço, J.A. (2009). Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* 10, 152.
- Gallagher, L.A., Ramage, E., Weiss, E.J., Radey, M., Hayden, H.S., Held, K.G., Huse, H.K., Zurawski, D.V., Brittnacher, M.J., and Manoil, C. (2015). Resources for Genetic and Genomic Analysis of Emerging Pathogen *Acinetobacter baumannii*. *J. Bacteriol.* 197, 2027–2035.

- Gogarten, J.P., and Townsend, J.P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 3, 679–687.
- Gonzalez-Villoria, A.M., Tamayo-Legorreta, E., Garza-Ramos, U., Barrios, H., Sanchez-Pérez, A., Rodríguez-Medina, N., Uribe-Aviña, N., Cevallos, M.A., CRAB Study Group, and Silva-Sanchez, J. (2016). A Multicenter Study in Mexico Finds *Acinetobacter baumannii* Clinical Isolates Belonging to Clonal Complexes 636B (113B) and 92B Harboring OXA-72, OXA-239, and OXA-469. *Antimicrob. Agents Chemother.* 60, 2587–2588.
- Gootz, T.D., and Marra, A. (2008). *Acinetobacter baumannii*: an emerging multidrug-resistant threat. *Expert Rev. Anti Infect. Ther.* 6, 309–325.
- Graña-Miraglia, L., Lozano, L., Castro-Jaimes, S., Cevallos, M.A., Volkow, P., and Castillo-Ramírez, S. (2016). First Genome Sequence of a Mexican Multidrug-Resistant *Acinetobacter baumannii* Isolate. *Genome Announc.* 4.
- Graña-Miraglia, L., Lozano, L.F., Velázquez, C., Volkow-Fernández, P., Pérez-Oseguera, Á., Cevallos, M.A., and Castillo-Ramírez, S. (2017). Rapid Gene Turnover as a Significant Source of Genetic Variation in a Recently Seeded Population of a Healthcare-Associated Pathogen. *Front. Microbiol.* 8, 1817.
- Hammerstrom, T.G., Beabout, K., Clements, T.P., Saxer, G., and Shamoo, Y. (2015). *Acinetobacter baumannii* Repeatedly Evolves a Hypermutator Phenotype in Response to Tigecycline That Effectively Surveys Evolutionary Trajectories to Resistance. *PLOS ONE* 10, e0140489.
- Harding, C.M., Hennon, S.W., and Feldman, M.F. (2018). Uncovering the mechanisms of *Acinetobacter baumannii* virulence. *Nat. Rev. Microbiol.* 16, 91–102.
- Harper, T.M., June, C.M., Taracila, M.A., Bonomo, R.A., Powers, R.A., and Leonard, D.A. (2018). Multiple substitutions lead to increased loop flexibility and expanded specificity in *Acinetobacter baumannii* carbapenemase OXA-239. *Biochem. J.* 475, 273–288.
- Holden, M.T.G., Hsu, L.-Y., Kurt, K., Weinert, L.A., Mather, A.E., Harris, S.R., Strommenger, B., Layer, F., Witte, W., de Lencastre, H., et al. (2013). A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.* 23, 653–664.
- 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., et al. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573.

- Kanamori, H., Parobek, C.M., Weber, D.J., van Duin, D., Rutala, W.A., Cairns, B.A., and Juliano, J.J. (2016). Next-Generation Sequencing and Comparative Analysis of Sequential Outbreaks Caused by Multidrug-Resistant *Acinetobacter baumannii* at a Large Academic Burn Center. *Antimicrob. Agents Chemother.* 60, 1249–1257.
- Karageorgopoulos, D.E., and Falagas, M.E. (2008). Current control and treatment of multidrug-resistant *Acinetobacter baumannii* infections. *Lancet Infect. Dis.* 8, 751–762.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242-245.
- Levade, I., Terrat, Y., Leducq, J.-B., Weil, A.A., Mayo-Smith, L.M., Chowdhury, F., Khan, A.I., Boncy, J., Buteau, J., Ivers, L.C., et al. (2017). *Vibrio cholerae* genomic diversity within and between patients. *Microb. Genomics* 3.
- Levy-Blitchein, S., Roca, I., Plasencia-Rebata, S., Vicente-Taboada, W., Velásquez-Pomar, J., Muñoz, L., Moreno-Morales, J., Pons, M.J., Del Valle-Mendoza, J., and Vila, J. (2018). Emergence and spread of carbapenem-resistant *Acinetobacter baumannii* international clones II and III in Lima, Peru. *Emerg. Microbes Infect.* 7, 119.
- Li, H., Liu, F., Zhang, Y., Wang, X., Zhao, C., Chen, H., Zhang, F., Zhu, B., Hu, Y., and Wang, H. (2015). Evolution of carbapenem-resistant *Acinetobacter baumannii* revealed through whole-genome sequencing and comparative genomic analysis. *Antimicrob. Agents Chemother.* 59, 1168–1176.
- Librado, P., Vieira, F.G., and Rozas, J. (2012). BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinforma. Oxf. Engl.* 28, 279–281. Lin, M.-F., and Lan, C.-Y. (2014). Antimicrobial resistance in *Acinetobacter baumannii*: From bench to bedside. *World J. Clin. Cases WJCC* 2, 787–814.
- Loewen, P.C., Alsaadi, Y., Fernando, D., and Kumar, A. (2014). Genome Sequence of an Extremely Drug-Resistant Clinical Isolate of *Acinetobacter baumannii* Strain AB030. *Genome Announc.* 2.
- López-Leal, G., Zuniga-Moya, J.C., Castro-Jaimes, S., Graña-Miraglia, L., Pérez-Oseguera, Á., Reyes-García, H.S., Gough-Coto, S.D., Pavón-Madrid, R., Bejarano, S.A., Ferrera, A., et al. (2019). Unexplored Genetic Diversity of Multidrug-

and Extremely Drug-Resistant *Acinetobacter baumannii* Isolates from Tertiary Hospitals in Honduras. *Microb. Drug Resist. Larchmt. N.*

- Löytynoja, A. (2014). Phylogeny-aware alignment with PRANK. *Methods Mol. Biol. Clifton NJ* 1079, 155–170.
- Martí, S., Fernández-Cuenca, F., Pascual, Á., Ribera, A., Rodríguez-Baño, J., Bou, G., Miguel Cisneros, J., Pachón, J., Martínez-Martínez, L., Vila, J., et al. Prevalencia de los genes tetA y tetB como mecanismo de resistencia a tetraciclina y minociclina en aislamientos clínicos de *Acinetobacter baumannii*. *Enfermedades Infecc. Microbiol. Clínica* 77–80.
- Moffatt, J.H., Harper, M., Harrison, P., Hale, J.D.F., Vinogradov, E., Seemann, T., Henry, R., Crane, B., Michael, F.S., Cox, A.D., et al. (2010). Colistin Resistance in *Acinetobacter baumannii* Is Mediated by Complete Loss of Lipopolysaccharide Production. *Antimicrob. Agents Chemother.* 54, 4971–4977.
- Mugnier, P.D., Poirel, L., Naas, T., and Nordmann, P. (2010). Worldwide Dissemination of the blaOXA-23 Carbapenemase Gene of *Acinetobacter baumannii*. *Emerg. Infect. Dis.* 16, 35–40.
- Ñamendys-Silva, S.A., Correa-García, P., García-Guillén, F.J., González-Herrera, M.O., Pérez-Alonso, A., Texcocano-Becerra, J., Herrera-Gómez, A., Cornejo-Juárez, P., and Meneses-García, A. (2015). Outcomes of critically ill cancer patients with *Acinetobacter baumannii* infection. *World J. Crit. Care Med.* 4, 258–264.
- Nemeč, A., Dijkshoorn, L., and van der Reijden, T.J.K. (2004). Long-term predominance of two pan-European clones among multi-resistant *Acinetobacter baumannii* strains in the Czech Republic. *J. Med. Microbiol.* 53, 147–153.
- Nigro, S.J., and Hall, R.M. (2016). Structure and context of *Acinetobacter* transposons carrying the oxa23 carbapenemase gene. *J. Antimicrob. Chemother.* 71, 1135–1147.
- Pagano, M., Nunes, L.S., Niada, M., Barth, A.L., and Martins, A.F. (2018). Comparative Analysis of Carbapenem-Resistant *Acinetobacter baumannii* Sequence Types in Southern Brazil: From the First Outbreak (2007-2008) to the Endemic Period (2013-2014). *Microb. Drug Resist. Larchmt. N.*
- Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., and Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693.

- Peleg, A.Y., Seifert, H., and Paterson, D.L. (2008). *Acinetobacter baumannii*: Emergence of a Successful Pathogen. *Clin. Microbiol. Rev.* 21, 538–582.
- Poirel, L., Figueiredo, S., Cattoir, V., Carattoli, A., and Nordmann, P. (2008). *Acinetobacter radioresistens* as a Silent Source of Carbapenem Resistance for *Acinetobacter spp.* *Antimicrob. Agents Chemother.* 52, 1252–1256.
- Poirel, L., Bonnin, R.A., and Nordmann, P. (2011). Genetic basis of antibiotic resistance in pathogenic *Acinetobacter* species. *IUBMB Life* 63, 1061–1067.
- Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256.
- Roca, I., Espinal, P., Vila-Farrés, X., and Vila, J. (2012). The *Acinetobacter baumannii* Oxymoron: Commensal Hospital Dweller Turned Pan-Drug-Resistant Menace. *Front. Microbiol.* 3, 148.
- Rodríguez, C.H., Balderrama Yarhui, N., Nastro, M., Nuñez Quezada, T., Castro Cañarte, G., Magne Ventura, R., Ugarte Cuba, T., Valenzuela, N., Roach, F., Mota, M.I., et al. (2016). Molecular epidemiology of carbapenem-resistant *Acinetobacter baumannii* in South America. *J. Med. Microbiol.* 65, 1088–1091.
- Sahl, J.W., Del Franco, M., Pournaras, S., Colman, R.E., Karah, N., Dijkshoorn, L., and Zarrilli, R. (2015). Phylogenetic and genomic diversity in isolates from the globally distributed *Acinetobacter baumannii* ST25 lineage. *Sci. Rep.* 5, 15188.
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinforma. Oxf. Engl.* 30, 2068–2069.
- Shimodaira, H., and Hasegawa, M. (1999). Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Mol. Biol. Evol.* 16, 1114.
- Snitkin, E.S., Zelazny, A.M., Montero, C.I., Stock, F., Mijares, L., NISC Comparative Sequence Program, Murray, P.R., and Segre, J.A. (2011). Genome-wide recombination drives diversification of epidemic strains of *Acinetobacter baumannii*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 13758–13763.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Tajima, F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* 123, 585–595.

- Tamayo-Legorreta, E.M., Garza-Ramos, U., Barrios-Camacho, H., Sanchez-Perez, A., Galicia-Paredes, A., Meza-Chavez, A., and Silva-Sanchez, J. (2014). Identification of OXA-23 carbapenemases: novel variant OXA-239 in *Acinetobacter baumannii* ST758 clinical isolates in Mexico. *New Microbes New Infect.* 2, 173–174.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). ME-GA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729.
- Turton, J.F., Ward, M.E., Woodford, N., Kaufmann, M.E., Pike, R., Livermore, D.M., and Pitt, T.L. (2006). The role of ISAbal in expression of OXA carbapenemase genes in *Acinetobacter baumannii*. *FEMS Microbiol. Lett.* 258, 72–77.
- Vanegas, J.M., Higuera, L.F., Vargas, C.A., Cienfuegos, A.V., Rodríguez, É.A., Roncancio, G.E., and Jiménez, J.N. (2015). [Carbapenem-resistant *Acinetobacter baumannii* causing osteomyelitis and infections of skin and soft tissues in hospitals of Medellín, Colombia]. *Biomed. Rev. Inst. Nac. Salud* 35, 522–530.
- Weber, B.S., Harding, C.M., and Feldman, M.F. (2015). Pathogenic *Acinetobacter*: from the Cell Surface to Infinity and Beyond. *J. Bacteriol.* 198, 880–887.
- Wolf, Y.I., and Koonin, E.V. (2013). Genome reduction as the dominant mode of evolution. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 35, 829–837.
- Wright, M.S., Iovleva, A., Jacobs, M.R., Bonomo, R.A., and Adams, M.D. (2016a). Genome dynamics of multidrug-resistant *Acinetobacter baumannii* during infection and treatment. *Genome Med.* 8, 26.
- Wright, M.S., Iovleva, A., Jacobs, M.R., Bonomo, R.A., and Adams, M.D. (2016b). Genome dynamics of multidrug-resistant *Acinetobacter baumannii* during infection and treatment. *Genome Med.* 8.
- Zarrilli, R., Pournaras, S., Giannouli, M., and Tsakris, A. (2013). Global evolution of multidrug-resistant *Acinetobacter baumannii* clonal lineages. *Int. J. Antimicrob. Agents* 41, 11–19.
- Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829.
- Zhu, L., Yan, Z., Zhang, Z., Zhou, Q., Zhou, J., Wakeland, E.K., Fang, X., Xuan, Z., Shen, D., and Li, Q.-Z. (2013). Complete Genome Analysis of Three *Acinetobacter baumannii* Clinical Isolates in China for Insight into the Diversification of Drug Resistance Elements. *PLOS ONE* 8, e66584.

Appendices

A. Glosario

Glosario

CHDL Carbapenem Hydrolising class D B-lactamases. Enzimas capaces de hidrolizar carbapenémicos (Ej: Doripenem, meropenem). 3

draft Estado de completitud de un ensamble bacteriano en el cual a pesar de haber sido secuenciado todo el genoma, no se ha logrado cerrar y determinar cromosoma y número de plásmidos.. 11

GC Global Complex, Complejos Clonales o Complejos Internacionales. Los sequence types se agrupan en complejos clonales por su similitud con un perfil alélico central (genotipo). Estos genotipos centrales se identifican mediante una serie de medios heurísticos, incluidos BURST y la descomposición dividida, junto con los comentarios de los laboratorios de salud pública y los epidemiólogos.. 4, 7

Genoma accesorio incluye los genes ausentes de uno o más aislados o exclusivos de un aislado dado, en el grupo considerado.. 14

Genoma core Convencionalmente definido como aquellos genes presentes en todos los aislados considerados.. 14

GTR El modelo general reversible en el tiempo de Tavaré 1986, es el modelo más general, neutral, independiente, de sitios finitos y reversible en el tiempo posible. 16

lecturas una lectura es una secuencia inferida de pares de bases (o probabilidades de pares de bases) correspondientes a un fragmento de ADN. Un experimento de secuenciación típico implica la fragmentación del genoma en millones de moléculas, que se seleccionan por tamaño y se ligan a adaptadores que en conjunto se conocen como una biblioteca de secuenciación. La secuenciación de los mismos genera lecturas. 12

MDR Multi Drogo Resistente. Se denomina de esta forma a cepas de bacterias patógenas que son resistentes a por lo menos un antibiótico en por lo menos tres clases de ellos (Ej: aminoglucósidos, sulfonamidas y beta-lactámicos). 4, 13

MLST Multi Locus Sequence Typing. Método de genotipificación de cepas de bacterias patógenas. Se basa en la secuenciación de fragmentos de 7 genes housekeeping. A cada fragmento único se asigna un identificador que corresponde al alelo. Estos identificadores suelen ser números. Se asigna un número a cada loci, creando así un perfil (un alelo por cada uno de los 7 genes), el cual es denominado *sequence type* de la cepa. En *A. baumannii* se usan dos esquemas diferentes con combinaciones de genes diferentes, el esquema Oxford y el esquema Pasteur.. 4, 7

N50 La longitud de los contigs tal que usando contigs de igual o mayor tamaño produce la mitad de las bases del genoma. El tamaño N50 se calcula ordenando todos los contigs de mayor a menor, y determinando el conjunto mínimo de contigs cuyo tamaño total sea el 50 % de todo el genoma.. 12

ortólogos relación entre secuencias en la que la divergencia acontece tras un evento de especiación. El ancestro común es el cen ancestro. La filogenia recuperada de estas secuencias refleja la filogenia de las especies.. 14

parálogos condición evolutiva en la que la divergencia observada acontece tras un evento de duplicación génica. La mezcla de ortólogos y parálogos en un mismo análisis filogenético recupera la filogenia correcta de los genes pero no necesariamente la de los organismos o taxa.. 14

PFGE Pulse Gel Field Electrophoresis. Es una electroforesis en gel estándar excepto que en lugar de correr constantemente el voltaje en una dirección el voltaje se cambia periódicamente entre tres direcciones. PFGE se puede utilizar para el genotipado. Es comúnmente considerado un estándar en los estudios epidemiológicos de organismos patógenos.. 7

prior el *prior* de una cantidad incierta, es la distribución de probabilidad que expresaría el conocimiento previo de uno sobre esta cantidad antes de que se tome en cuenta alguna evidencia. Es una parte clave de la inferencia bayesiana.. 15

SGF Grupos de genes ortólogos que presentan una sola copia en el genoma. 14, 16

SNP es una variación en la secuencia de ADN que afecta a una sola base (adenina (A), timina (T), citosina (C) o guanina (G)) de una secuencia del genoma.. 16

ST Perfil alélico de una cepa patógena.. 11

ST Oxford En este esquema se usan 7 genes: *gltA*, *gyrB*, *gdhB*, *recA*, *cpn60*, *gpi*, *rpoD*. 5, 12

ST Pasteur En este esquema se usan 7 genes: cpn60, fusA, gltA, pyrG, recA, rplB, rpoB. 5, 12

WGS Es el proceso de determinar la secuencia completa de ADN del genoma de un organismo. Esto implica la secuenciación de todo el ADN cromosómico, así como el ADN plasmídico en caso de bacterias.. 5, 7

B. Tabla Suplementaria 1

Tabla B.1 Tabla suplementaria 1. Base de Datos

Strain Name	Isolation Date	Assembly ID NCBI	Oxford (MLST)	Location
15A34	2014	GCA.002082685.1	872	South-Korea
15A5	2013	GCA.002082705.1	191	South-Korea
1656-2	NA	GCA.000188215.1	423	NA
A1	1982	GCA.000830055.1	231	United Kingdom
A1296	2010	GCA.002504145.1	1469	China
A388	2002	GCA.002741415.1	439	Greece
A85	2003	GCA.002210065.1	781	Australia
AB0057	2001	GCA.000021245.1	207	USA
AB030	2010	GCA.000746645.1	758	Canadá
AB031	2010	GCA.000746605.1	1000	Canadá
Ab04-mff	2012	GCA.001077655.1	447	Canada
AB07	NA	GCA.000576535.1	1816, 195	India
AB1A2	2017	GCA.002948475.1	1806, 208	USA
AB307-0294	1994	GCA.000021145.1	231	USA
Ab207	2008	GCA.001636235.1	1321	Mexico
AB34299	2015	GCA.002009115.1	1834, 436	NA
Ab421-GEIH-2010	2010	GCA.001628795.1	924	Spain
Ab4568	2011	GCA.002762115.1	191	China
Ab4653	2011	GCA.002762155.1	NA	China
Ab4977	2011	GCA.002762095.1	191	China
AB5075-UW	2008	GCA.000963815.1	945	USA
Ab6200	2012	GCA.000814345.1	1161	Colombia
ab736	2015	GCA.002116925.1	931	USA
Aba	2016	GCA.003288775.1	1816, 195	China
ABH120-A2	2006-2008	GCA.000761175.1	924	Spain
ABNIH28	2016	GCA.002902885.1	NA	USA
AbPK1	2012	GCA.002753915.1	452	Pakistan
AC12	2011	GCA.000282795.3	1816, 195	Malasya
AC29	2011	GCA.000695855.2	195	Malasya
AC30	2011	GCA.000307975.2	195	Malasya
ACICU	2005	GCA.000018445.1	437	Italy
AF-401	2009	GCA.001896005.1	new	Mexico
AF-673	2008	GCA.001895985.1	1806, 208	USA
AR-0056	NA	GCA.003029495.1	350	NA
AR-0063	NA	GCA.002947845.1	205	NA
AR-0078	NA	GCA.002948925.1	203	NA
AR-0083	NA	GCA.002996805.1	1604, 231	NA
AR-0088	NA	GCA.003006035.1	229	NA
AR-0101	NA	GCA.003010675.1	NA	NA
AR-0102	NA	GCA.003010655.1	191	NA
ATCC-17978	1951	GCA.000015425.1	112	Canada
AYE	2001	GCA.000069245.1	231	France

Tabla B.2 Tabla suplementaria 1. Continuación

Strain Name	Isolation Date	Assembly ID NCBI	Oxford (MLST)	Location
AYP-A2	2013	GCA.002761575.1	1806, 208	Australia
B11911	2014	GCA.001077565.2	NA	India
B8300	2015	GCA.001077965.2	new	India
B8342	2015	GCA.001077555.2	new	India
BAL062	2014	GCA.900088705.1	136, 1851	NA
BJAB07104	2007-2008	GCA.000419385.1	368	NA
BJAB0715	2007	GCA.000419405.1	642	China
BJAB0868	NA	GCA.000419425.1	218	NA
CBA7	2013	GCA.002082645.1	1806, 208	South-Korea
CIP70-10	1970	GCA.001457535.1	819	France
CMC-CR-MDR-Ab4	NA	GCA.001922205.1	1806, 208	USA
CMC-CR-MDR-Ab66	NA	GCA.001922245.1	1806, 208	USA
CMC-MDR-Ab59	NA	GCA.001922225.1	1806, 208	USA
CR17	NA	GCA.000820685.1	1806, 208	NA
CS01	NA	GCA.000820665.1	1806, 208	NA
D1279779	2009	GCA.000186665.4	942	Australia
D36	2008	GCA.001399655.1	498	Australia
DA33098	NA	GCA.003181015.1	350	Spain
DA33382	NA	GCA.003296225.1	1567	Germany
DS002	2005	GCA.003012895.1	new	India
DU202	NA	GCA.000498375.2	423	NA
HRAB-85	2014	GCA.001887305.1	208	China
HWBA8	2013	GCA.002082785.1	229	South-Korea
IOMTU-433	2013	GCA.000828935.1	919	Nepal
JBA13	2013	GCA.002082625.1	191	South-Korea
KAB01	2015	GCA.001806345.1	1809, 451	South-Korea
KAB02	2015	GCA.001806365.1	369	South-Korea
KAB03	2015	GCA.001806385.1	1809, 451	South-Korea
KAB04	2015	GCA.001806405.1	191	South-Korea
KAB05	2015	GCA.001806425.1	1837, 369	South-Korea
KAB06	2015	GCA.001806445.1	1837, 369	South-Korea
KAB07	2015	GCA.001806465.1	191	South-Korea
KAB08	2015	GCA.001806485.1	1806, 208	South-Korea
KBN10P02143	2012	GCA.001514375.1	191	South-Korea
LAC-4	1997	GCA.000786735.1	447	USA
MDR-TJ	NA	GCA.000187205.4	1837, 369	NA
MDR-ZJ06	2006	GCA.000226275.1	643	China
NCGM-237	2012	GCA.000828795.1	455	Japan
NCTC13421	2004	GCA.900478145.1	1840, 207	United Kingdom
ORAB01	2012	GCA.000939415.2	208	USA
PR07	2012	GCA.000372585.2	734	Malasya
R2090	2013	GCA.001261895.2	942	France
R2091	2014	GCA.001517645.1	819	France
SA A14	2013	GCA.002082845.1	191	South-Korea
SDF	NA	GCA.000069205.1	NA	NA
SMC-BL01	2012	GCA.002843665.1	191	South-Korea
SSA12	2013	GCA.002082865.1	191	South-Korea
SSA6	2013	GCA.002082745.1	357	South-Korea
SSMA17	2013	GCA.002082885.1	191	South-Korea
TCDC-AB0715	NA	GCA.000189735.2	218	NA
TYTH-1	2008	GCA.000302575.1	455	Taiwan
UH12208	2008	GCA.000515575.1	1091	USA
USA15	2013	GCA.002082825.1	491	South-Korea
USA2	2013	GCA.002082725.1	357	South-Korea
WCHAB005078	NA	GCA.002999195.1	449	China
WCHAB005133	NA	GCA.002950495.1	368	China
WKA02	2013	GCA.002082805.1	447	South-Korea
XDR-BJ83	2007	GCA.001902375.1	368	China
XH386	2017	GCA.003264275.1	1806, 208	China
XH731	2014	GCA.003264295.1	1816, 195	China
XH856	2010	GCA.001573125.1	381	China
XH857	2010	GCA.001573105.1	806	China
XH858	2010	GCA.001578145.1	642	China
XH859	2009	GCA.001573085.1	368	China
XH860	2009	GCA.001573065.1	457	China
YU-R612	2014	GCA.001543995.1	191	South-Korea
ZS3	2009	GCA.003202135.1	1806, 208	China

Tabla B.3 Tabla suplementaria 1. Continuación

Strain Name	Isolation Date	Assembly ID NCBI	Oxford (MLST)	Location	Hospital	Sequencing
H350	2015	SGSF00000000	208	Mexico	INR	Newly sequenced
H008	2011	SGSG00000000	690	Mexico	INR	Newly sequenced
A158	2012	SGSH00000000	758	Mexico	INR	Newly sequenced
A580	2015	SGSI00000000	758	Mexico	INR	Newly sequenced
H251	2014	SGSJ00000000	758	Mexico	INR	Newly sequenced
A465	2014	GCA.004209035.1	758	Mexico	INR	Newly sequenced
A278	2013	GCA.004209005.1	758	Mexico	INR	Newly sequenced
A023	2011	GCA.004209075.1	1091	Mexico	INR	Newly sequenced
A229	2013	GCA.004208995.1	1091	Mexico	INR	Newly sequenced
H170	2013	SGSO00000000	1091	Mexico	INR	Newly sequenced
810CP	2015	GCA.003345235.1	758	Mexico	HI	Recently sequenced by us
Ab2159	2011	SIDE00000000	758	Mexico	INCAN	Recently sequenced by us
Ab11606	2011	GCA.001922755.1	758	Mexico	INCAN	Recently sequenced by us
Ab11598	2011	GCA.001923005.1	758	Mexico	INCAN	Recently sequenced by us
Ab11551	2011	GCA.001922745.1	758	Mexico	INCAN	Recently sequenced by us
Ab11547	2013	GCA.001922695.1	758	Mexico	INCAN	Recently sequenced by us
Ab11536	2011	GCA.001922705.1	758	Mexico	INCAN	Recently sequenced by us
Ab11502	2012	GCA.001922635.1	758	Mexico	INCAN	Recently sequenced by us
Ab11510	2012	GCA.001922425.2	758	Mexico	INCAN	Recently sequenced by us
Ab4113	2013	GCA.001922685.1	758	Mexico	INCAN	Recently sequenced by us
Ab4197	2013	SIDD00000000	758	Mexico	INCAN	Recently sequenced by us
Ab5013	2014	SIDC00000000	758	Mexico	INCAN	Recently sequenced by us
HEU5	2016	GCA.002927915.1	758	Honduras	University School Hospital	Recently sequenced by us
HEU55808	2016	GCA.002928095.1	758	Honduras	University School Hospital	Recently sequenced by us
MCR6056	2015	GCA.002928135.1	758	Honduras	Mario Catarino Rivas Hospital	Recently sequenced by us
MCR10126	2015	GCA.002927995.1	447	Honduras	Mario Catarino Rivas Hospital	Recently sequenced by us
MCR10172	2015	GCA.002927805.1	447	Honduras	Mario Catarino Rivas Hospital	Recently sequenced by us
MCR10179	2015	GCA.002927965.1	229	Honduras	Mario Catarino Rivas Hospital	Recently sequenced by us
MCR54	2015	GCA.002928045.1	1586	Honduras	Mario Catarino Rivas Hospital	Recently sequenced by us
MCR56	2015	GCA.002928195.1	1587	Honduras	Mario Catarino Rivas Hospital	Recently sequenced by us
MCR6739	2015	GCA.002928035.1	NA	Honduras	Mario Catarino Rivas Hospital	Recently sequenced by us
MCR8676	2015	GCA.002928015.1	1587	Honduras	Mario Catarino Rivas Hospital	Recently sequenced by us
MCR8683	2015	GCA.002928155.1	1589	Honduras	Mario Catarino Rivas Hospital	Recently sequenced by us
MCR9238	2015	GCA.002927775.1	447	Honduras	Mario Catarino Rivas Hospital	Recently sequenced by us
HC9436	2015	GCA.002928145.1	1325	Honduras	Cemes Hospital	Recently sequenced by us
HEU2	2015	GCA.002928075.1	1586	Honduras	University School Hospital	Recently sequenced by us
HEU3	2016	GCA.002927855.1	1588	Honduras	University School Hospital	Recently sequenced by us
IHSS3526	2016	GCA.002927895.1	1590	Honduras	Honduran Institute of Social Security	Recently sequenced by us

C. Artículos arbitrados

A continuación se anexan los artículos publicados en el marco del trabajo de Doctorado

References

1. Chenais E, Ståhl K, Guberti V, Depner K. Identification of wild boar–habitat epidemiologic cycle in African swine fever epizootic. *Emerg Infect Dis.* 2018;24:810–2. <http://dx.doi.org/10.3201/eid2404.172127>
2. Gallardo C, Fernández-Pinero J, Pelayo V, Gzaev I, Markowska-Daniel I, Pridotkas G, et al. Genetic variation among African swine fever genotype II viruses, eastern and central Europe. *Emerg Infect Dis.* 2014;20:1544–7. <http://dx.doi.org/10.3201/eid2009.140554>
3. Ge S, Li J, Fan X, Liu F, Li L, Wang Q, et al. Molecular characterization of African swine fever virus, China, 2018. *Emerg Infect Dis.* 2018;24:2131–3. <http://dx.doi.org/10.3201/eid2411.181274>
4. Kolbasov D, Titov I, Tsybanov S, Gogin A, Malogolovkin A. African swine fever virus, Siberia, Russia, 2017. *Emerg Infect Dis.* 2018;24:796–8. <http://dx.doi.org/10.3201/eid2404.171238>

Address for correspondence: Daniel Desmecht, University of Liège, Pathology, Faculty of Veterinary Medicine, Sart Tilman B43, Liège 4000, Belgium; email: daniel.desmecht@ulg.ac.be

Inaccurate Multilocus Sequence Typing of *Acinetobacter baumannii*

Santiago Castillo-Ramírez, Lucía Graña-Miraglia

Author affiliation: Programa de Genómica Evolutiva, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Mexico

DOI: <https://doi.org/10.3201/eid2412.180374>

Multilocus sequence typing has been useful for genotyping pathogens in surveillance and epidemiologic studies. However, it cannot reflect the true relationships of isolates for species with very dynamic genomes. Using a robust genome phylogeny, we demonstrated the limitations of this method for typing *Acinetobacter baumannii*.

An adequate genotyping system is of paramount importance for infectious disease epidemiology. Two decades ago, the multilocus sequence typing (MLST) scheme was proposed as a genotyping method (1), and today, because of its reproducibility and portability, MLST schemes are available for many human pathogens (2). MLST has been instrumental in increasing understanding of the epidemiology and population structure of many bacteria.

Acinetobacter baumannii, a major source of nosocomial infections, is no exception, and 2 MLST schemes (Oxford and Pasteur) have been established for this species (3,4). Each scheme uses just 7 loci and, therefore, only samples a small fraction of the chromosome, which could be a serious issue for genotyping species with highly variable genomes. Some studies have shown that *A. baumannii* has both high gene content variation (5) and substantial levels of recombination (6).

We revisited one of the most comprehensive genome datasets of *A. baumannii* (5) to construct a robust phylogeny to show that sequence type (ST) assignment in both MLST schemes does not reflect true relationships among isolates of this species. This dataset of >80 genomes covers 36 different STs according to the Oxford scheme (STox) and 19 different STs according to the Pasteur scheme (STp) (Appendix Table 1, <https://wwwnc.cdc.gov/EID/article/25/1/18-0374-App1.pdf>). We constructed a concatenated alignment of 574 orthologous genes and conducted statistical model selection as in a previous study (5) and, on that alignment, constructed a maximum-likelihood phylogeny by PhyML (7).

The 2 schemes showed different levels of resolution. Although in many instances a single STp had just 1 equivalent STox, 2 STs exist in the Pasteur scheme that encompass many Oxford STs (Appendix Figure, red branches). For instance, under the Pasteur scheme, STp2 represents >15 STs in the Oxford scheme and STp1 encompasses 5 STox. Thus, the Pasteur scheme seems to have considerably less resolution than the Oxford scheme to distinguish isolates. The Pasteur scheme's lack of resolution was not insignificant, however. STp2 comprises 43 isolates (approximately half of our dataset) showing considerable levels of genetic variation according to our phylogeny, but according to this MLST scheme, they constitute just 1 genotype.

Many of the STs in either scheme formed coherent (monophyletic) groups in our phylogeny. However, we recorded some clear exceptions in which isolates from some STs did not form monophyletic groups, that is, isolates with the same ST did not cluster. The most striking case is STox208 (orange tips in the phylogeny), where there are 2 well-defined groups with several isolates each and an extra isolate not close to either of those well-defined groups. We also noted that the 2 STox455 isolates did not cluster and are located far apart on the tree (green tips). Additionally, 1 of the STox369 isolates did not fall within the ST369 group (blue tips). These 3 examples show that the Oxford MLST does not accurately reflect the relationships among the isolates. Also notable is that, although for the Oxford scheme 36 STs are represented in this dataset, only 16 of them have ≥ 2 isolates and therefore only in these STs could we detect problems with the clustering within any given ST. Thus, 3 of these 16 STs did not cluster the isolates properly inasmuch as these STs were polyphyletic. In summary, for the Oxford scheme we demonstrated that some STs form polyphyletic groups

because 4 of the 7 loci have signals of recombination (Appendix Table 2), whereas for the Pasteur scheme, we noted a serious lack of resolution for some STs because the loci used only by this scheme have the lowest levels of genetic diversity (Appendix Table 2). Two previous studies noted problems with the MLST schemes for this species (8,9); nonetheless, neither was as extensive as our study, nor did they benchmark both schemes against a genome-based phylogeny.

In conclusion, we showed that the correct relationships among isolates cannot be recovered using either of the MLST schemes for *A. baumannii*. In addition, we highlighted the importance of using more powerful genotyping strategies when analyzing bacteria with highly dynamic genomes; in this regard, the ever-decreasing cost of genome sequencing will make this technology the perfect tool for genotyping bacterial species.

Acknowledgments

S.C.-R. thanks Timothy Read for his valuable comments on the manuscript.

This work was supported by “Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT” (grant no. IA201317) and CONACyT Ciencia Básica 2016 (grant no. 284276) to S.C.-R. L.G.-M. received a CONACyT doctoral fellowship (no. 585414).

About the Authors

Dr. Castillo-Ramírez is an associate professor in the Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México. His primary research interests include evolutionary genomics and phylogeography of bacterial pathogens.

Ms. Graña-Miraglia is a doctoral student at the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México. Her primary research interest is the comparative genomics of bacteria.

References

- Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A*. 1998;95:3140–5. <http://dx.doi.org/10.1073/pnas.95.6.3140>
- Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*. 2010;11:595. <http://dx.doi.org/10.1186/1471-2105-11-595>
- Bartual SG, Seifert H, Hippler C, Luzon MA, Wisplinghoff H, Rodríguez-Valera F. Development of a multilocus sequence typing scheme for characterization of clinical isolates of *Acinetobacter baumannii*. *J Clin Microbiol*. 2005;43:4382–90. <http://dx.doi.org/10.1128/JCM.43.9.4382-4390.2005>
- Diancourt L, Passet V, Nemeč A, Dijkshoorn L, Brisse S. The population structure of *Acinetobacter baumannii*: expanding multiresistant clones from an ancestral susceptible genetic pool. *PLoS One*. 2010;5:e10034. <http://dx.doi.org/10.1371/journal.pone.0010034>
- Graña-Miraglia L, Lozano LF, Velázquez C, Volkow-Fernández P, Pérez-Oseguera Á, Cevallos MA, et al. Rapid gene turnover as a significant source of genetic variation in a recently seeded population of a healthcare-associated pathogen. *Front Microbiol*. 2017;8:1817. <http://dx.doi.org/10.3389/fmicb.2017.01817>
- Snitkin ES, Zelazny AM, Montero CI, Stock F, Mijares L, Murray PR, et al.; NISC Comparative Sequence Program. Genome-wide recombination drives diversification of epidemic strains of *Acinetobacter baumannii*. *Proc Natl Acad Sci U S A*. 2011;108:13758–63. <http://dx.doi.org/10.1073/pnas.1104404108>
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–21. <http://dx.doi.org/10.1093/sysbio/syq010>
- Hamidian M, Nigro SJ, Hall RM. Problems with the Oxford multilocus sequence typing scheme for *Acinetobacter baumannii*: do sequence type 92 (ST92) and ST109 exist? *J Clin Microbiol*. 2017;55:2287–9. <http://dx.doi.org/10.1128/JCM.00533-17>
- Tomaschek F, Higgins PG, Stefanik D, Wisplinghoff H, Seifert H. Head-to-head comparison of two multi-locus sequence typing (MLST) schemes for characterization of *Acinetobacter baumannii* outbreak and sporadic isolates. *PLoS One*. 2016; 11:e0153014. <http://dx.doi.org/10.1371/journal.pone.0153014>

Address for correspondence: Santiago Castillo-Ramírez, Centro de Ciencias Genómicas, Av Universidad S/N, Cuernavaca 62210, Mexico; email: iago@ccg.unam.mx

Severe Disseminated Infection with Emerging Lineage of Methicillin-Sensitive *Staphylococcus aureus*

Paul Jewell, Luke Dixon, Aran Singanayagam, Rohma Ghani, Ernie Wong, Meg Coleman, Bruno Pichon, Angela Kearns, Georgina Russell, James Hatcher

Author affiliations: Imperial College Healthcare NHS Trust, London, United Kingdom (P. Jewell, L. Dixon, A. Singanayagam, R. Ghani, E. Wong, M. Coleman, G. Russell, J. Hatcher); Public Health England, London (B. Pichon, A. Kearns)

DOI: <https://doi.org/10.3201/eid2501.180684>

We report a case of severe disseminated infection in an immunocompetent man caused by an emerging lineage of methicillin-sensitive *Staphylococcus aureus* clonal complex 398. Genes encoding classic virulence factors were



Rapid Gene Turnover as a Significant Source of Genetic Variation in a Recently Seeded Population of a Healthcare-Associated Pathogen

Lucía Graña-Miraglia¹, Luis F. Lozano¹, Consuelo Velázquez², Patricia Volkow-Fernández², Ángeles Pérez-Oseguera¹, Miguel A. Cevallos¹ and Santiago Castillo-Ramírez^{1*}

¹ Programa de Genómica Evolutiva, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Mexico, ² Departamento de Enfermedades Infecciosas, Instituto Nacional de Cancerología, Mexico, Mexico

OPEN ACCESS

Edited by:

Yasir Muhammad,
King Abdulaziz University, Saudi Arabia

Reviewed by:

Govindan Rajamohan,
Institute of Microbial Technology
(CSIR), India
Benjamin Andrew Evans,
University of East Anglia,
United Kingdom

*Correspondence:

Santiago Castillo-Ramírez
iago@ccg.unam.mx

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 05 May 2017

Accepted: 06 September 2017

Published: 20 September 2017

Citation:

Graña-Miraglia L, Lozano LF, Velázquez C, Volkow-Fernández P, Pérez-Oseguera Á, Cevallos MA and Castillo-Ramírez S (2017) Rapid Gene Turnover as a Significant Source of Genetic Variation in a Recently Seeded Population of a Healthcare-Associated Pathogen. *Front. Microbiol.* 8:1817. doi: 10.3389/fmicb.2017.01817

Genome sequencing has been useful to gain an understanding of bacterial evolution. It has been used for studying the phylogeography and/or the impact of mutation and recombination on bacterial populations. However, it has rarely been used to study gene turnover at microevolutionary scales. Here, we sequenced Mexican strains of the human pathogen *Acinetobacter baumannii* sampled from the same locale over a 3 year period to obtain insights into the microevolutionary dynamics of gene content variability. We found that the Mexican *A. baumannii* population was recently founded and has been emerging due to a rapid clonal expansion. Furthermore, we noticed that on average the Mexican strains differed from each other by over 300 genes and, notably, this gene content variation has accrued more frequently and faster than the accumulation of mutations. Moreover, due to its rapid pace, gene content variation reflects the phylogeny only at very short periods of time. Additionally, we found that the external branches of the phylogeny had almost 100 more genes than the internal branches. All in all, these results show that rapid gene turnover has been of paramount importance in producing genetic variation within this population and demonstrate the utility of genome sequencing to study alternative forms of genetic variation.

Keywords: population genomics, microevolution, genetic variation, pathogen, phylogeography, *A. baumannii*, gene content

INTRODUCTION

Genome sequencing of bacteria has drastically transformed our view of how bacteria change. For example, the use of genome sequencing in microbial experimental evolution has been of paramount importance to better understand the mechanisms that generate genetic variation in bacterial populations. However, these types of experiments are far from what really happens in nature—as they are simplified versions of natural populations. On the other hand, natural populations of bacterial pathogens themselves provide a copious source to address the question of how bacteria evolve. Over the last two decades genome sequencing has allowed us to better understand not only the microevolution but also the epidemiology of many bacterial pathogens. For instance, this tool has been used to study the phylogeography and population structure of several human and animal

bacterial pathogens (Brynildsrud et al., 2014; Ezewudo et al., 2015; Joseph et al., 2015; Castillo-Ramirez et al., 2016). It has also been employed to study the impact of recombination on bacterial clones (Castillo-Ramirez et al., 2011, 2012) and even to analyse the intra-host evolution of several human pathogens (Diaz Caballero et al., 2015; Azarian et al., 2016). Although pan-genome analyses have been used to indirectly study gene content variation in bacteria, these studies have rarely been used to study gene turnover over very short periods of time and, more importantly, explicitly taking into account the population structure of the bacteria involved. Furthermore, only a few studies have tried to compare the genetic variation generated by point mutation vs. that generated by gene gains and losses within a well-defined population.

We now know much about the microevolution of a few pathogens, such as *Neisseria gonorrhoea* or the species from the genus *Chlamydia* (Ezewudo et al., 2015; Joseph et al., 2015), however for most bacterial species we do not know very much. While *Staphylococcus aureus* has been extensively studied (Holden et al., 2010; Nubel et al., 2010; Castillo-Ramirez et al., 2011, 2012), much less is known about the evolutionary dynamics of *A. baumannii*, especially in developing countries (Castro-Jaimes et al., 2016; Grana-Miraglia et al., 2016; Silva et al., 2016; Zenati et al., 2016). Importantly, this bacterial species has emerged as one of the main causes of nosocomial infections over the last decades (Zarrilli et al., 2013; Antunes et al., 2014). One of the factors contributing to this pattern is the ability of this species to show multi-drug (MDR) and extreme (XDR) phenotypes; these phenotypes have been on the rise the last decade. Remarkably, the infections caused by MDR and/or XDR isolates have been linked with higher mortality rates and longer hospitalization (Sunenshine et al., 2007; Metan et al., 2009).

We propose that *A. baumannii* is a good model to study the microevolutionary dynamics of gene content variation for several reasons. First, given that this is a hospital-associated pathogen, it should be rather easy to define single populations. Secondly, considering the highly dynamic genome of this species (Chan et al., 2015), one would expect that gene content should be an important factor as far as the genetic variation is concerned. Here, we use genome sequencing to characterize the microevolution of a single lineage (Sequence Type [ST] 758) within a single locale—a tertiary hospital in a developing country (Mexico). To gain further insights into the evolution of this lineage, we have also incorporated publicly available sequenced strains of this species to create one of the most inclusive data sets to date with 85 genomes and representing 38 STs. We found that gene content variation is of chief importance to generate genetic variation within the population, as it occurs much faster than *de novo* mutations. Furthermore, we also note that more genes are found in the external branches of the phylogeny, which is consistent with the slightly deleterious nature of gene acquisitions.

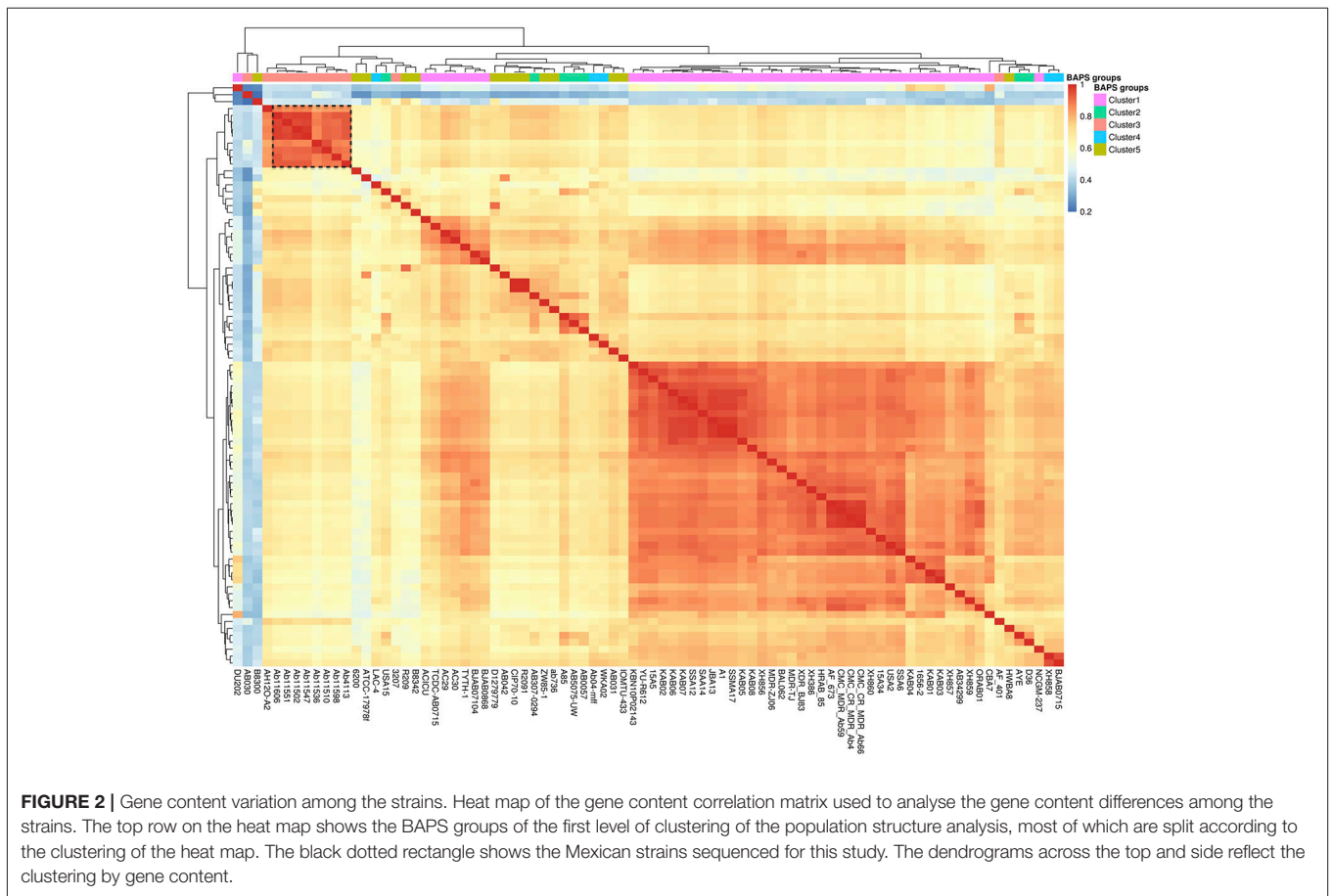
RESULTS

A Very Recent Expanding Population

In order to have a proper data set we first conduct genome sequencing of eight Mexican isolates (all recovered from the

same tertiary hospital in Mexico city during 2011–2013, see **Supplementary Table 1**). Of note, all these isolates belong to the same Sequence Type (ST758). This set was supplemented with a global data set of publicly available genomes (see **Supplementary Table 2**). Our combined data set has 85 genomes and it is one of the most comprehensive collections in terms of *A. baumannii* lineages (38 STs). Using the combined data set, we initially conducted a Maximum Likelihood (ML) phylogeny on a concatenated alignment of single gene families not affected by recombination to see how the newly sequenced Mexican strains relate to the rest of *A. baumannii* strains. The phylogenetic tree reveals that the Mexican isolates form a tight monophyletic cluster (see green square, **Figure 1**) and that these are well-differentiated from the remaining strains. A hierarchical population structure analysis at the deepest level (see methods for details) further reinforces this picture, as one of the 32 clusters recovered (green square, **Figure 1**) has all the newly sequenced Mexican isolates. The very short branches of the Mexican isolates suggest that this population has low genetic variation, to corroborate this we computed the nucleotide diversity for the Mexican isolates and noted that the value is very low ($\pi = 0.000532$). Furthermore, there were only 147 segregating sites in this population. Among other things, the low genetic variation in this population may point to a recent introduction of this lineage in to the Mexican hospital. To prove this we conducted a molecular dating analysis (see methods). Our molecular dating analysis confirms this as the time to the most recent common ancestor was found to be mid 2009 (see **Supplementary Figure 1**). This is consistent with the collection dates of all the Mexican isolates between mid 2011 and late 2013 (see **Supplementary Table 1**). From **Figure 1** is very difficult to appreciate how the Mexican isolates relate to one and another; hence, to examine this in more detail, we constructed a ML phylogeny just considering the Mexican isolates (see **Supplementary Figure 2**). This phylogeny has a star like topology, which suggests a recent and rapid expansion of this Mexican lineage. To corroborate these findings, we used an independent line of evidence: a Tajima's D analysis on the alignment just containing the Mexican isolates. The clear negative value of Tajima's D, -1.922 , implies an overabundance of low frequency polymorphisms and this is consistent with a population size expansion as suggested by the ML phylogeny. Taken together, these results suggest that the Mexican isolates constitute a very young population, which was very recently introduced in Mexico's National Institute of Oncology, and since then has been undergoing a clonal expansion.

It is worth paying attention to the fact that two strains (Ab4113 and Ab11598) from the star like phylogeny of the Mexican lineage show two very long branches, having each one an average distance of over 70 SNPs to the rest of the Mexican strains (see **Supplementary Figure 2**). We carried out a Tajima's relative test to establish whether those two strains have accumulated significantly more SNPs than the other Mexican strains; this was clearly the case as for both strains, Ab4113 and Ab11598, the null hypothesis of equal rates was rejected when compared to the other Mexican strains. Furthermore, we determined that substitutions defining those two branches



gene content and the minimum number of gains/losses in each internal node of the phylogeny again by means of BadiRate. The results are shown in **Figure 3**. This analysis suggests that there are more genes toward the present than in the past. For instance, the most basal node is the one that shows the least number of genes, whereas the external nodes have many more genes. To establish this more formally, we compared the ancestral gene content of the internal branches of the phylogeny vs. the gene content of the external branches. We noted that considerably more genes are present in the external branches (mean = 3,770) than at the internal nodes (mean = 3,677, Wilcoxon rank sum test $W = 7.5$, p -value = 0.02052). This supports the idea that the presence of fewer genes in the internal branches might be due to the selective removal of genes, which could have deleterious effects. We think that given sufficient time purifying selection will remove most of the genes but at the tips of the tree there has not been enough time for that to happen. To sum up this part, this array of analyses shows that gain and loss of genes do not remain constant over time and there seems to be more genes toward the present time and this could be due to natural selection (purifying selection).

Comparison of Gene Turnover vs. Accumulation of *De novo* Mutation

Thus far we have analyzed the dynamics of gene turnover but we have not compared this to a more typical source of variation, i.e., mutation. Given the very recent emergence of the Mexican

population it is a suitable scenario to compare the contribution of gene content variation and the accrual of *de novo* mutations to genetic variation. We carried out pairwise comparisons of Mexican strains as a means to conduct such analysis. On the one hand, we computed the number of mutations that differentiate any two Mexican strains and, on the other hand, we calculated the number of genes in which any two Mexican strains differ. Most of the strains differ in just a few mutations from one another (boxplots on the left, **Figure 4**), the median difference being just 1.5 mutations. There were six pairwise comparisons that involved strains that did not have any mutations to differentiate them. However, these strains differed from each other by hundreds of genes (boxplots on the right, **Figure 4**), with a median value of 294.5 genes and, remarkably, no two strains were equal to each other in terms of their gene content. Irrespective of the inclusion of hypermutator strains (see Discussion) or not, this analysis clearly showed that the rate of gene turnover is much higher than the accumulation of mutations (see **Figure 4**); this difference was statistically significant either way (Wilcoxon rank sum test with hypermutators, p -value = 1.573e-10 and without hypermutators, p -value = 2.498e-06). For instance, if the hypermutators are included in the comparisons the mean number of mutations is 36.75, whereas the mean number in gene differences is 314.2; this is also evident when hypermutators are left out, as the mean number of mutations is 0.6667, whereas the mean number of gene differences is 224. However, the difference in gene

TABLE 1 | Topology tests.

	*Diff $-\ln L$	pKH ⁺	pSH ⁺	pRELL ⁺
NJ whole data set	-658053.472	0.000	0.000	0.000
NJ Mexican clade	-0.172	0.336	0.352	0.29

Kishino-Hasegawa (KH), Shimodaira-Hasegawa (SH), and REll bootstrap proportion tests to determine whether the NJ topologies, based on the gene content matrix, differ significantly from the ML phylogenies.

⁺Difference in log-likelihood to the ML phylogeny.

⁺p-values under the different tests.

TABLE 2 | Branch models of gene family turnover.

Branch model	$-\ln L^+$	K*	AIC	$\Delta AIC^{\&}$
GD-FR-ML	-7500.2956	29	15058.59	0.00
GD-GR-ML	-8290.2042	3	16586.41	15015.68

The GD-GR-ML model implies that all the branches have the same turnover rates, whereas GD-FR-ML model assumes that each branch has its own turnover rates. In both models, turnover rates were estimated by Maximum Likelihood (ML).

⁺Log-Likelihood scores; *Number of parameters; [&] ΔAIC is the difference in the Akaike Information Criterion (AIC) for each model to the best model. We used *BadRate* to implement the Gain-and-Death stochastic population model to estimate the gene family turnover rates.

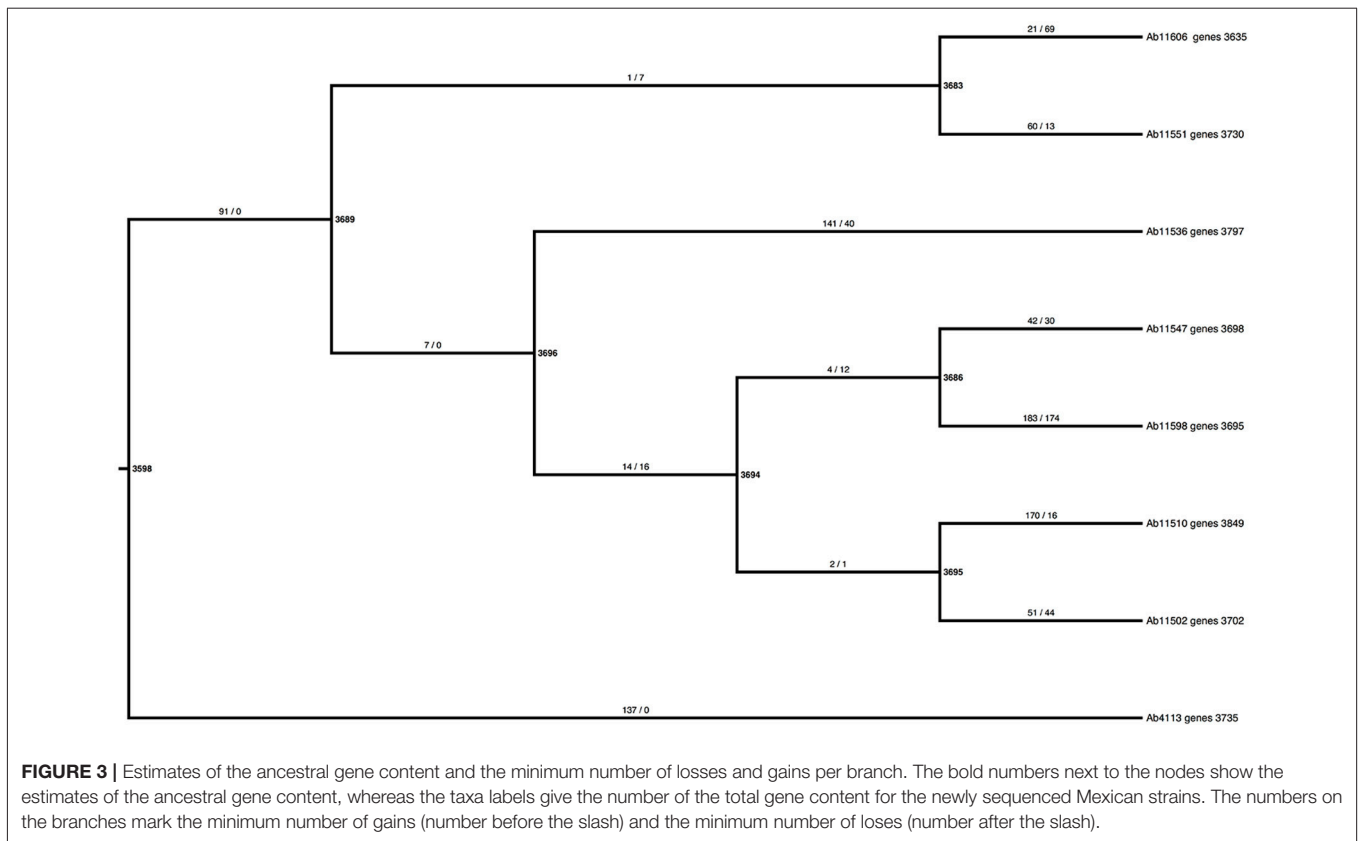
content between any two strains might have been due to a small number of events—just because several genes can be introduced at once thanks to Mobile Genetic Elements (MGEs), for example Insertion Sequences (ISs) or phage. In order to explore this, we compared four sets of two strains that both were sampled the same year (see **Supplementary Table 4**); we assumed that two or more genes were introduced in the same event if they are contiguous to one another. In all the four pairwise comparisons the number of events is considerably lower than the number of gene differences. For instance, in the first comparison (involving Ab11502 and Ab11510), although these two strains differ in 311 genes (82 present in Ab11502 but not in Ab11510 and 229 in Ab11510 but absent in Ab11502), we estimated that around 112 events have produced this difference in genes (72 events in Ab11510 and 40 in Ab11502). Clearly, the same pattern applies to the rest of the comparisons and implies that several genes are introduced/lost simultaneously (see **Supplementary Table 4**).

Notably, many of these genes, ~52% (see **Supplementary Table 4** for the value for each comparison), could be associated with MGEs, such as ISs, phage or plasmids. Along these lines, we found that the accessory genome of the Mexican strains is enriched in MGEs (Chi-squared test, p -value $< 2.2e-16$) compared to the core genome. Finally, it is worth mentioning that this gene turnover has affected both plasmids and chromosomes and it is not only due to the loss and gain of plasmids. First, we noted that a significant number (more than 60% in all comparisons) of the unique genes in the pairwise comparisons have significant hits in chromosomes from complete genome sequences (see third column, **Supplementary Table 4**). Second, we conducted a plasmid profile analysis of the Mexican strains (see

Supplementary Figure 3) and noted that six of the eight strains had identical plasmid profiles, which suggests that the difference in gene content is not due to the loss or gain of complete plasmids for these six strains. Whereas, the two remaining strains have the same set of bands as the other six strains plus 1 and 3 extra bands. These last two analyses imply that gene content variation cannot be exclusively due to gain and loss of plasmids and that this variation has considerably affected the chromosome. Collectively, these data indicate that gene turnover (mediated by MGEs) has happened considerably more rapid than the accumulation of mutations and that this rapid turnover is an important force in generating genomic variation in both plasmid and chromosomes.

DISCUSSION

In this study we used a population-genomics approach to analyse the evolutionary dynamics of gene content over a very short period of time and from a single population. In order to do this, we sampled several strains (all from the same lineage) from a tertiary hospital in a developing country (Mexico). To put this lineage in the context of the global population of *A. baumannii*, we have also incorporated publicly available genomes to create one of the most inclusive data sets for this bacterium, as far as different lineages (STs) are concerned—more than 35 STs are represented in our data set. All the newly sequenced Mexican strains belong to the ST758 and, notably, the PubMLST webpage (<https://pubmlst.org>) only reports two other strains with this ST. One is the Canadian strain AB030, which was included in our data set, and the other is a Mexican strain, isolate 6463. Thus far, this ST has mainly been reported in North America and it might be that this ST is endemic in this region; however a recent report found this ST also in a hospital in Pretoria, South Africa (Lowings et al., 2015). Many more isolates from different parts of the world are needed to properly establish the distribution of this ST; nonetheless this ST belongs to the clonal complex 636 (as per the Oxford MLST scheme), which is not within the global clones (GCs), and, therefore, it is very unlikely that ST758 would be as widely distributed as STs within the GCs. With respect to the *A. baumannii* populations in Mexico, we are adamant that our study could be of paramount importance for future studies tackling the population genomics of this species in this country. Clearly this is not the first study analyzing *A. baumannii* isolates from Mexico; in fact, over the last decade there have been several studies (Ares et al., 2013; Morfin-Otero et al., 2013; Alcantar-Curiel et al., 2014; Bocanegra-Ibarias et al., 2015; Cornejo-Juarez et al., 2015; Gonzalez-Villoria et al., 2016; Tamayo-Legorreta et al., 2016), which have been very useful from a clinical point of view, as they focused on the molecular epidemiology (Morfin-Otero et al., 2013; Gonzalez-Villoria et al., 2016; Tamayo-Legorreta et al., 2016) or the antibiotic resistance profiles of hospital isolates (Ares et al., 2013; Alcantar-Curiel et al., 2014; Bocanegra-Ibarias et al., 2015). However, none of those studies have analyzed the genomic diversity of the isolates in question and, to the best of our knowledge, this is the first study that has addressed the genetic variation within a population of *A. baumannii* isolates



in Mexico at a genome level. Our phylogenetic and molecular dating analyses indicated that the Mexican strains comprise a very recently founded population undergoing a rapid expansion. Notably, similar rapid clonal expansions have been also described for some STs of *S. aureus* (Aanensen et al., 2016) and, very likely, this is a common trend exhibited by many bacterial populations within the hospital setting.

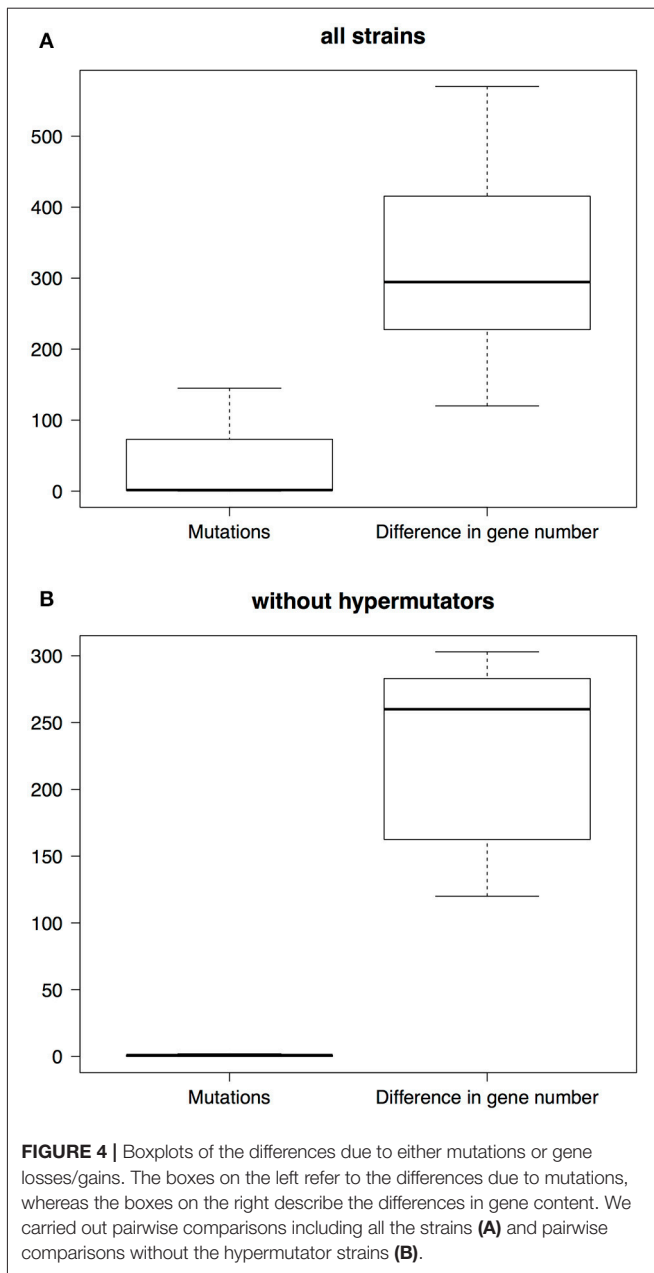
Although pan-genome analyses have been extremely useful, it should be remembered that they could also be misleading. First, core and accessory genomes are bound to be artifacts of the sampling criteria employed and, importantly, the genes falling in each of these two categories totally depend on the taxonomic level chosen. Here, we avoided these distinctions and we considered all the genes for the gene content analysis, whereas for defining the global population and the Mexican clade we used single gene families (without signals of recombination) as a proxy for orthologous genes. Therefore, our strategy considerably improves on previous studies that have used pan-genome analyses.

Although there have been some studies that have gone beyond the use of pan genome analysis, and have even used proper statistical approaches, these have considered broader timescales and did not take the population structure into account (Librado et al., 2014; Nowell et al., 2014). Furthermore, none of those studies compared the rate of change due to gene content variability to that attributable to *de novo* mutation. To the best of our knowledge, this is the first study that not only takes

into account the population structure to properly understand gene content variation among strains but also compares this gene content variability to a better known source of genetic variation, that is mutation. However, we note that similar trends have been very recently described in a work studying within-patient genomic diversity of *Vibrio cholera* isolates (Levade et al., 2017).

In this study, we were not only able to show that rapid gene turnover has been a major factor in the generation of genomic variation in the few years of existence of this expanding Mexican population but also to show that, on average, gene turnover has introduced more than 8 times more variation than mutation and this is taking into account the hypermutator strains. However, if hypermutator strains are not included in the analysis, on average, gene turnover has added ~ 335 more variation than mutation. Notably many of these gene differences are not independent, as the genes were introduced in the same event via some MGE; in conformity with this, we found that the accessory genome is enriched in MGEs. Nonetheless, chromosome regions were also implicated in this type of variation. Notably at this short time scale gene content variation still reflects the common ancestry relationships of the strains (i.e., the topology of the ML likelihood) and, given the increased resolution, gene content variability could be very useful for studying outbreaks or other events that happen at very short periods of time.

We reason that patterns found in the ancestral gene content (external branches showing more genes) are compatible with the slightly deleterious nature of many gene acquisitions. Although



some gene acquisitions might be beneficial initially, most of them have a fitness cost and therefore will be eliminated by natural selection. Nevertheless, if the fitness cost of these acquisitions is not high (i.e., slightly deleterious) the removal process will not be happening instantaneously—giving time for acquisitions to be sampled if the sampling strategy considers very closely related isolates. Under this scenario, external branches would be enriched in gene acquisitions, as selection has not had enough time to remove them; whereas, internal branches would have less genes, as natural selection has had more time to purge those slightly deleterious acquisitions. Furthermore, the slightly deleterious nature of the gene content variability is further reinforced by the fact that the global rates model (GD-FR-ML in Table 2) assuming different gain and death

turnover rates over time was a considerably better fit for our data. Interestingly, Wolf and Koonin (2013) have discussed the idea that genome reduction—either by neutral or selective processes—is the principal mode of evolution, although their arguments were based on studies that had considered macro evolutionary scales rather than micro evolutionary scales. In a slightly different view, Gogarten and Townsend have also discussed that most of the genes horizontally transferred have neutral or nearly neutral effects in the receiver genome and just a few of those genes have a positive effect (increase the fitness) in the recipient genome (Gogarten and Townsend, 2005). Along these lines, several studies have proposed that many gene gains/losses have minor effects on the fitness of bacteria and behave rather neutral (Baumdicker et al., 2012; Knoppel et al., 2014; Andreani et al., 2017).

In conclusion, our study shows that gene content variability could be a major source of genetic variation that takes place much faster than the accumulation of *de novo* mutations in bacterial populations at very early stages of diversification. Furthermore, it suggests that the loss of genes in the internal branches is due to the gradual removal of the slightly deleterious gene acquisitions. This study clearly demonstrates the utility of NGS for studying the micro evolutionary dynamics of gene content variation over very short periods of time and for comparing this to more typical forms of genetic variation such as mutations.

MATERIALS AND METHODS

Genomes used and Homologous Groups

Because we wanted to know the evolutionary dynamics of a single lineage in a single location over a very short period of time, we selected eight *A. baumannii* isolates (Supplementary Table 1), all of which belong to ST758 (according to the Oxford MLST scheme), from the Instituto Nacional de Cancerología (Mexico's National Institute of Oncology) that is a tertiary hospital located in Mexico City. The antibiograms and the source of the Mexican isolates are provided in Supplementary Table 3. This study was carried out with isolated strains, confidentiality of the patients is preserved and there is no possible way to link the information here provided to any individual; thus, the Ethics Committee approval of the Instituto Nacional de Cancerología was not required. All the isolates were sequenced by means of an Illumina MiSeq platform, considering 250-bp paired-end reads. We used the SolexQA software (Cox et al., 2010) to trim the reads prior to assemble the genomes. These draft genomes were assembled using Velvet version 1.2.09 (Zerbino and Birney, 2008) and Spades v3.9.0 (Bankevich et al., 2012) and contigs smaller than 300 base pairs were not taken into account. We manually edited our assembly for gap closure and error correction. The whole genome sequences have been deposited at DDBJ/ENA/GenBank under the accession numbers MSCX00000000, MSCY00000000, MSCZ00000000, MSDA00000000, MSDB00000000, MSDC00000000, MSDD00000000 (BioProject PRJNA355850). One of the isolates (Ab11510) was also sequenced using Pacific Biosciences technology and assembled with SMRT (accession numbers CP018861, CP018862 and CP023300). Although the quality

of these assemblies varied to some extent, all these draft genomes have good values considering the coverage (range 49.5–375) and the number of contigs was below 90 in all cases (**Supplementary Table 1**). We annotated these Mexican isolates through PROKKA v1.11 (Seemann, 2014). We also included 77 *A. baumannii* complete genomes from the NCBI (see **Supplementary Table 2**); these include isolates from Europe, Asia, South and North America. For consistency, we also annotated these genomes using PROKKA v1.11 (Seemann, 2014). Then, in order to construct the homologous groups, we used the program PanOCT (Fouts et al., 2012). We first carried out BLAST searches between all genomes, considering an e-value of 1.0×10^{-30} , we then fed these to PanOCT to be able to create the homologous groups. Because we wanted to define closely related homologous groups (CRHGs), we required that sequences aligned $\geq 90\%$ of their lengths and were $\geq 80\%$ identical when running PanOCT—other than that, default parameters were used. We identified a total of 14194 homologous groups, of which 1462 belong to the core genome according to PanOCT. We identified 1383 single gene families, homologous groups that have only one gene per genome. We also assigned, when possible, the CRHGs to their potential MGEs, for doing that we used two databases: ISfinder (Siguier et al., 2006) and ACLAME (Leplae et al., 2010). We conducted BLAST searches for each gene of each CRHG against those two databases with an e-value of 1.0×10^{-30} and requiring that the query sequence aligned $\geq 60\%$ of its length and was $\geq 40\%$ identical. Finally, we also ran another pangenome analysis, again via PanOCT, but this time just considering the Mexican strains.

Phylogenies, Molecular Dating, and Population Genetics Analysis

Identifying true orthologous genes is not trivial (Castillo-Ramirez and Gonzalez, 2008), however as proxy for orthologous genes we used single gene families. We employed a concatenated alignment of 1383 single gene families to conduct a mixture analysis via the Bayesian Analysis of Population Structure Analysis (BAPS) program version 6 (Tang et al., 2009). We ran a hierarchical model-based clustering of the strains implementing the tandem version of BAPS (Cheng et al., 2013). This clustering was carried out setting four levels in the hierarchy. To get a measure of the number of genetically diverged groups we conduct a preliminary analysis setting the maximum number of genetically differentiated groups to 35, as it has been done previously for other bacterial species (Joseph et al., 2015), and to 35 in the final analysis. The first level of clustering yielded five clusters, the second produced 15, the third provided 22, and the last one gave 32. We constructed a species phylogeny for this data set using a concatenated alignment of 574 single gene families that did not show evidence of recombination, as inferred by the test for detecting recombination implemented via the PhiPack program (Bruen et al., 2006). We ran a Maximum Likelihood phylogeny on the concatenated alignment through PhyML (Guindon et al., 2010) and setting the model described below. We conducted statistical model selection, as in Lopez-Leal et al. (2016), to find the most adequate model. This analysis

was done by means of jModelTest (Abascal et al., 2005) and the model selected was GTR+R+I. We also made another phylogeny just considering the Mexican isolates also via PhyML. We created a SNPs alignment, from the concatenated alignment of the 574 single genes families not affected by recombination by keeping only the variable sites. We used the SNPs alignment to carry out a dating analysis, which was conducted via BEAST2 (Bouckaert et al., 2014). For the SNP alignment only a subset of strains, for which we had reliable information on the date of isolation, were considered (see **Supplementary Table 6**) so to be able to calibrate the relaxed molecular clock confidently. We set a log-normal relaxed clock, employing the GTR DNA model and using the correction for among site variation; this model was chosen as the statistical model selection analysis, implemented via jModelTest2 (Posada, 2009; Darriba et al., 2012), indicated that was the most suitable model for this data set. This analysis was run for 200,000,000 generations, sampling every 10,000 generations and discarding the first 20,000,000 generations as burn-in. We used VariScan version 2.0 (Vilella et al., 2005) to compute some population genetics summary statistics (nucleotide diversity [π] and Tajima's D) setting the runmode to 12. We employed MEGA6 (Tamura et al., 2013) to carry out Tajima's relative test and evaluate whether the potential hypermutators strains have higher evolutionary rates. We conducted two analyses: one considering the strain Ab4113 and other involving the strain Ab11598. In both analyses, the out group was Ab11551 and the taxon B was Ab11510.

Gene Content Analysis and Pairwise Comparisons

We have created a matrix that contains the number of genes per homologous groups and the genomes considered. We normalized the matrix by dividing each value by the sum of all the values. Using this matrix we have created two additional matrices used for downstream analysis. One was a correlation matrix obtained using the `cor()` function in R and setting a Pearson correlation. The correlation matrix was visualized using a heat map employing the `heatmap()` function also in R. We also constructed a distance matrix, for which we utilize the `dist()` function and the distance measure used was "Euclidean." Then, by means of the APE library in R (Paradis et al., 2004) we used the Neighbor-joining algorithm to construct two phylogenies based on the distance matrix. We conducted several topology tests (see **Table 1**) to determine whether the NJ phylogenies based on the gene content matrix differ significantly from the ML phylogenies; the topology tests were implemented via PAML 4 (Yang, 2007). We implemented a gain-death model stochastic by means of the BadiRate software (Librado et al., 2012). This analysis requires an ultrametric tree of the taxa considered, therefore we used the APE library (Paradis et al., 2004) to convert the ML phylogeny of the Mexican clade into an ultrametric tree. We considered two branch models: one was the global rates model (GD-GR-ML) and the other was the free-rates model (GD-FR-ML). To assess the goodness of fit of these two models we used the log-likelihood of the models and the Akaike Information Criterion (Akaike, 1974). We also set the option "anc" that gives the number of

genes at the internal branches for each gene family and the total number of genes; it also reports the minimum number of losses and gains per internal and external branch. We also carried out a set of pairwise comparisons (see **Supplementary Table 4**); for each comparison we determine the number of genes present in one but not the other strain and the sum of those unique genes per strain gives the total difference in gene number between the two strains under consideration. We assigned a potential chromosomal location for the unique genes per strain using a set of chromosomes from complete genomes, which are listed in **Supplementary Table 5**. For doing this, we carried out BLAST searches for each gene present in one but not the other strain against a database containing the chromosomes mentioned above with an e -value of 1.0×10^{-30} and requiring that the query sequence aligned $\geq 60\%$ of its length and was $\geq 40\%$ identical.

AUTHOR CONTRIBUTIONS

SC conceived, designed, and coordinated the study. LG assembled and annotated the genomes, constructed the homologous groups, ran the ML phylogeny and created the gene content matrices. SC conducted the population structure and molecular dating analyses, ran the gene turnover analysis, and conducted the population genetic parameters. LL helped with the genome assemblies and genome annotations. AP conducted the plasmid profiles. PV, CV, and MC contributed the Mexican isolates and participated in the general discussion. SC and LG wrote the manuscript. All the authors revised and approved the manuscript.

ACKNOWLEDGMENTS

This work was supported by “Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT” (grant numbers IA200515 and IA201317) to SC. LG is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), and received a fellowship (number 585414) from CONACYT. The genome sequencing of all the Mexican strains was conducted at Macrogen (<http://www.macrogen.com/eng/>). SC is grateful to his

colleague Timothy Read for helpful discussions on this project and his comments on the manuscript. SC is also indebted to Gabrielle Margos and Jesse Shapiro for critical reading of the manuscript. SC and LG are thankful to the reviewers as their comments have significantly improved this manuscript. Finally, SC and LG would like to extend warm thanks to Esperanza Martinez and Lorenzo Segovia for their helpful comments on LG’s PhD project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.01817/full#supplementary-material>

Supplementary Table 1 | Newly sequenced Mexican isolates on this study.

Supplementary Table 2 | Publically available complete genomes downloaded from the NCBI.

Supplementary Table 3 | Antibiograms and source of the Mexican isolates.

Supplementary Table 4 | Pairwise comparisons of a set of strains.

Supplementary Table 5 | List of the complete genomes from which chromosomes were extracted to assign potential chromosomal localization for the gene differences regarding the pairwise comparisons of the set of strains used in **Supplementary Table 4**.

Supplementary Table 6 | Strains, and their dates of isolation, used for the molecular dating analysis.

Supplementary Figure 1 | Molecular dating Mexican clade. Marginal posterior distribution for the time to the Most Recent Common Ancestor of the (tMRCA) Mexican strains. Blue bars denote the 95% highest posterior density interval. The mean and median values for the tMRCA are also provided.

Supplementary Figure 2 | Phylogeny Mexican clade. The phylogeny is based on the concatenated alignment of all the single gene families not affected by recombination and was constructed via PhyML. This phylogeny includes only the newly Mexican strains. The red labels denote the hypermutators strains, the number on the branches denote the average SNP distance to the non-hypermutators strains. The scale bar represents substitution per sites.

Supplementary Figure 3 | Plasmid profile analysis of the Mexican strains. Plasmid profiles of the 8 Mexican isolates analyzed by agarose gel electrophoresis (0.8%). Names of strains are listed at the top of the figure. The first and last lanes show the plasmid profile of *Acinetobacter haemolyticus* 11616 that was utilized as molecular weight standard. Arrows at the right indicate the plasmid molecular weights of *A. haemolyticus* 11616 and the migration point of broken DNA.

REFERENCES

- Aanensen, D. M., Feil, E. J., Holden, M. T., Dordel, J., Yeats, C. A., Fedosejev, A., et al. (2016). Whole-genome sequencing for routine pathogen surveillance in public health: a population snapshot of invasive *Staphylococcus aureus* in Europe. *MBio* 7:e00444-16. doi: 10.1128/mBio.00444-16
- Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104–2105. doi: 10.1093/bioinformatics/bti263
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control* 19, 716–723. doi: 10.1109/TAC.1974.1100705
- Alcantar-Curiel, M. D., Garcia-Torres, L. F., Gonzalez-Chavez, M. I., Morfin-Otero, R., Gayosso-Vazquez, C., Jarillo-Quijada, M. D., et al. (2014). Molecular mechanisms associated with nosocomial carbapenem-resistant *Acinetobacter baumannii* in Mexico. *Arch. Med. Res.* 45, 553–560. doi: 10.1016/j.arcmed.2014.10.006
- Andreani, N. A., Hesse, E., and Vos, M. (2017). Prokaryote genome fluidity is dependent on effective population size. *ISME J.* 11, 1719–1721. doi: 10.1038/ismej.2017.36
- Antunes, L. C., Visca, P., and Towner, K. J. (2014). *Acinetobacter baumannii*: evolution of a global pathogen. *Pathog. Dis.* 71, 292–301. doi: 10.1111/2049-632X.12125
- Ares, M. A., Alcantar-Curiel, M. D., Jimenez-Galicia, C., Rios-Sarabia, N., Pacheco, S., and De la Cruz, M. A. (2013). Antibiotic resistance of gram-negative bacilli isolated from pediatric patients with nosocomial bloodstream infections in a Mexican tertiary care hospital. *Chemotherapy* 59, 361–368. doi: 10.1159/000362085
- Azarian, T., Daum, R. S., Petty, L. A., Steinbeck, J. L., Yin, Z., Nolan, D., et al. (2016). Intrahost evolution of methicillin-resistant *Staphylococcus aureus* USA300 among individuals with reoccurring skin and soft-tissue infections. *J. Infect. Dis.* 214, 895–905. doi: 10.1093/infdis/jiw242

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Baumdicker, F., Hess, W. R., and Pfaffelhuber, P. (2012). The infinitely many genes model for the distributed genome of bacteria. *Genome Biol. Evol.* 4, 443–456. doi: 10.1093/gbe/evs016
- Bocanegra-Ibarias, P., Pena-Lopez, C., Camacho-Ortiz, A., Llaca-Diaz, J., Silva-Sanchez, J., Barrios, H., et al. (2015). Genetic characterisation of drug resistance and clonal dynamics of *Acinetobacter baumannii* in a hospital setting in Mexico. *Int. J. Antimicrob. Agents* 45, 309–313. doi: 10.1016/j.ijantimicag.2014.10.022
- Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537
- Bruen, T. C., Philippe, H., and Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172, 2665–2681. doi: 10.1534/genetics.105.048975
- Brynildsrud, O., Feil, E. J., Bohlin, J., Castillo-Ramirez, S., Colquhoun, D., McCarthy, U., et al. (2014). Microevolution of *Renibacterium salmoninarum*: evidence for intercontinental dissemination associated with fish movements. *ISME J.* 8, 746–756. doi: 10.1038/ismej.2013.186
- Castillo-Ramirez, S., and Gonzalez, V. (2008). Factors affecting the concordance between orthologous gene trees and species tree in bacteria. *BMC Evol. Biol.* 8:300. doi: 10.1186/1471-2148-8-300
- Castillo-Ramirez, S., Corander, J., Marttinen, P., Aldeljawi, M., Hanage, W. P., Westh, H., et al. (2012). Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus*. *Genome Biol.* 13:R126. doi: 10.1186/gb-2012-13-12-r126
- Castillo-Ramirez, S., Fingerle, V., Jungnick, S., Straubinger, R. K., Krebs, S., Blum, H., et al. (2016). Trans-Atlantic exchanges have shaped the population structure of the Lyme disease agent *Borrelia burgdorferi* sensu stricto. *Sci. Rep.* 6:22794. doi: 10.1038/srep22794
- Castillo-Ramirez, S., Harris, S. R., Holden, M. T., He, M., Parkhill, J., Bentley, S. D., et al. (2011). The impact of recombination on dN/dS within recently emerged bacterial clones. *PLoS Pathog.* 7:e1002129. doi: 10.1371/journal.ppat.1002129
- Castro-Jaimes, S., Salgado-Camargo, A. D., Grana-Miraglia, L., Lozano, L., Bocanegra-Ibarias, P., Volkow-Fernandez, P., et al. (2016). Complete genome sequence of a multidrug-resistant *Acinetobacter baumannii* isolate obtained from a Mexican hospital (Sequence Type 422). *Genome Announc.* 4:e00583-16. doi: 10.1128/genomeA.00583-16
- Chan, A. P., Sutton, G., DePew, J., Krishnakumar, R., Choi, Y., Huang, X. Z., et al. (2015). A novel method of consensus pan-chromosome assembly and large-scale comparative analysis reveal the highly flexible pan-genome of *Acinetobacter baumannii*. *Genome Biol.* 16, 143. doi: 10.1186/s13059-015-0701-6
- Cheng, L., Connor, T. R., Siren, J., Aanensen, D. M., and Corander, J. (2013). Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* 30, 1224–1228. doi: 10.1093/molbev/mst028
- Cornejo-Juarez, P., Vilar-Compte, D., Perez-Jimenez, C., Namendys-Silva, S. A., Sandoval-Hernandez, S., and Volkow-Fernandez, P. (2015). The impact of hospital-acquired infections with multidrug-resistant bacteria in an oncology intensive care unit. *Int. J. Infect. Dis.* 31, 31–34. doi: 10.1016/j.ijid.2014.12.022
- Cox, M. P., Peterson, D. A., and Biggs, P. J. (2010). SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC bioinformatics* 11:485. doi: 10.1186/1471-2105-11-485
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772. doi: 10.1038/nmeth.2109
- Diaz Caballero, J., Clark, S. T., Coburn, B., Zhang, Y., Wang, P. W., Donaldson, S. L., et al. (2015). Selective sweeps and parallel pathoadaptation drive *Pseudomonas aeruginosa* evolution in the cystic fibrosis lung. *MBio* 6, e00981–e00915. doi: 10.1128/mBio.00981-15
- Ezewudo, M. N., Joseph, S. J., Castillo-Ramirez, S., Dean, D., Del Rio, C., Didelot, X., et al. (2015). Population structure of *Neisseria gonorrhoeae* based on whole genome data and its relationship with antibiotic resistance. *PeerJ.* 3:e806. doi: 10.7717/peerj.806
- Fouts, D. E., Brinkac, L., Beck, E., Inman, J., and Sutton, G. (2012). PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. *Nucleic Acids Res.* 40:e172. doi: 10.1093/nar/gks757
- Gogarten, J. P., and Townsend, J. P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 3, 679–687. doi: 10.1038/nrmicro1204
- Gonzalez-Villoria, A. M., Tamayo-Legorreta, E., Garza-Ramos, U., Barrios, H., Sanchez-Perez, A., Rodriguez-Medina, N., et al. (2016). A multicenter study in Mexico finds *Acinetobacter baumannii* clinical isolates belonging to clonal complexes 636B (113B) and 92B Harboring OXA-72, OXA-239, and OXA-469. *Antimicrob. Agents Chemother.* 60, 2587–2588. doi: 10.1128/AAC.02042-15
- Grana-Miraglia, L., Lozano, L., Castro-Jaimes, S., Cevallos, M. A., Volkow, P., and Castillo-Ramirez, S. (2016). First genome sequence of a Mexican multidrug-resistant *Acinetobacter baumannii* isolate. *Genome Announc.* 4:e00156-16. doi: 10.1128/genomeA.00156-16
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Hammerstrom, T. G., Beabout, K., Clements, T. P., Saxer, G., and Shamoo, Y. (2015). *Acinetobacter baumannii* repeatedly evolves a hypermutator phenotype in response to tigecycline that effectively surveys evolutionary trajectories to resistance. *PLoS ONE* 10:e0140489. doi: 10.1371/journal.pone.0140489
- Holden, M. T., Lindsay, J. A., Corton, C., Quail, M. A., Cockfield, J. D., Pathak, S., et al. (2010). Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant *Staphylococcus aureus*, sequence type 239 (TW). *J. Bacteriol.* 192, 888–892. doi: 10.1128/JB.01255-09
- Joseph, S. J., Marti, H., Didelot, X., Castillo-Ramirez, S., Read, T. D., and Dean, D. (2015). Chlamydiaceae genomics reveals interspecies admixture and the recent evolution of *Chlamydia abortus* infecting lower mammalian species and humans. *Genome Biol. Evol.* 7, 3070–3084. doi: 10.1093/gbe/evv201
- Knoppel, A., Lind, P. A., Lustig, U., Nasvall, J., and Andersson, D. I. (2014). Minor fitness costs in an experimental model of horizontal gene transfer in bacteria. *Mol. Biol. Evol.* 31, 1220–1227. doi: 10.1093/molbev/msu076
- Komp Lindgren, P., Higgins, P. G., Seifert, H., and Cars, O. (2016). Prevalence of hypermutators among clinical *Acinetobacter baumannii* isolates. *J. Antimicrob. Chemother.* 71, 661–665. doi: 10.1093/jac/dkv378
- Lepiae, R., Lima-Mendez, G., and Toussaint, A. (2010). ACLAME: a CLAssification of mobile genetic elements, update 2010. *Nucleic Acids Res.* 38, D57–D61. doi: 10.1093/nar/gkp938
- Levade, I., Terrat, Y., Leducq, J. B., Weil, A. A., Mayo-Smith, L. M., Chowdhury, F., et al. (2017). *Vibrio cholerae* genomic diversity within and between patients. *BioRxiv.* doi: 10.1101/169292
- Librado, P., Vieira, F. G., and Rozas, J. (2012). BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics* 28, 279–281. doi: 10.1093/bioinformatics/btr623
- Librado, P., Vieira, F. G., Sanchez-Gracia, A., Kolokotronis, S. O., and Rozas, J. (2014). Mycobacterial phylogenomics: an enhanced method for gene turnover analysis reveals uneven levels of gene gain and loss among species and gene families. *Genome Biol. Evol.* 6, 1454–1465. doi: 10.1093/gbe/evu117
- Lopez-Leal, G., Cevallos, M. A., and Castillo-Ramirez, S. (2016). Evolution of a sigma factor: an all-in-one of gene duplication, horizontal gene transfer, purifying selection, and promoter differentiation. *Front. Microbiol.* 7:581. doi: 10.3389/fmicb.2016.00581
- Lowings, M., Ehlers, M. M., Dreyer, A. W., and Kock, M. M. (2015). High prevalence of oxacillinases in clinical multidrug-resistant *Acinetobacter baumannii* isolates from the Tshwane region, South Africa - an update. *BMC Infect. Dis.* 15:521. doi: 10.1186/s12879-015-1246-8
- Metan, G., Sariguzel, F., and Sumerkan, B. (2009). Factors influencing survival in patients with multi-drug-resistant *Acinetobacter bacteraemia*. *Eur. J. Intern. Med.* 20, 540–544. doi: 10.1016/j.ejim.2009.05.005
- Morfin-Otero, R., Alcantar-Curiel, M. D., Rocha, M. J., Alpuche-Aranda, C. M., Santos-Preciado, J. I., Gayosso-Vazquez, C., et al. (2013). *Acinetobacter baumannii* infections in a tertiary care hospital in Mexico over the past 13 years. *Chemotherapy* 59, 57–65. doi: 10.1159/000351098

- Nowell, R. W., Green, S., Laue, B. E., and Sharp, P. M. (2014). The extent of genome flux and its role in the differentiation of bacterial lineages. *Genome Biol. Evol.* 6, 1514–1529. doi: 10.1093/gbe/evu123
- Nubel, U., Dordel, J., Kurt, K., Strommenger, B., Westh, H., Shukla, S. K., et al. (2010). A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. *PLoS Pathog.* 6:e1000855. doi: 10.1371/journal.ppat.1000855
- Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412
- Posada, D. (2009). Selection of models of DNA evolution with jModelTest. *Methods Mol. Biol.* 537, 93–112. doi: 10.1007/978-1-59745-251-9_5
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36. doi: 10.1093/nar/gkj014
- Silva, L., Grosso, F., Branquinho, R., Ribeiro, T. G., Sousa, C., and Peixe, L. (2016). Exploring non-hospital-related settings in Angola reveals new *Acinetobacter* reservoirs for blaOXA-23 and blaOXA-58. *Int. J. Antimicrob. Agents* 48, 228–230. doi: 10.1016/j.ijantimicag.2016.06.003
- Sunenshine, R. H., Wright, M. O., Maragakis, L. L., Harris, A. D., Song, X., Hebden, J., et al. (2007). Multidrug-resistant *Acinetobacter* infection mortality rate and length of hospitalization. *Emerg. Infect. Dis.* 13, 97–103. doi: 10.3201/eid1301.060716
- Tamayo-Legorreta, E., Turrubiarres-Martinez, E., Garza-Ramos, U., Nino-Moreno, P., Barrios, H., Sanchez-Perez, A., et al. (2016). Outbreak caused by blaOXA-72-producing *Acinetobacter baumannii* ST417 detected in clinical and environmental isolates. *Microb. Drug Res.* 22, 129–133. doi: 10.1089/mdr.2015.0157
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tang, J., Hanage, W. P., Fraser, C., and Corander, J. (2009). Identifying currents in the gene pool for bacterial populations using an integrative approach. *PLoS Comput. Biol.* 5:e1000455. doi: 10.1371/journal.pcbi.1000455
- Vilella, A. J., Blanco-Garcia, A., Hutter, S., and Rozas, J. (2005). VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics* 21, 2791–2793. doi: 10.1093/bioinformatics/bti403
- Wolf, Y. I., and Koonin, E. V. (2013). Genome reduction as the dominant mode of evolution. *Bioessays* 35, 829–837. doi: 10.1002/bies.201300037
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Zarrilli, R., Pournaras, S., Giannouli, M., and Tsakris, A. (2013). Global evolution of multidrug-resistant *Acinetobacter baumannii* clonal lineages. *Int. J. Antimicrob. Agents* 41, 11–19. doi: 10.1016/j.ijantimicag.2012.09.008
- Zenati, K., Touati, A., Bakour, S., Sahli, F., and Rolain, J. M. (2016). Characterization of NDM-1- and OXA-23-producing *Acinetobacter baumannii* isolates from inanimate surfaces in a hospital environment in Algeria. *J. Hosp. Infect.* 92, 19–26. doi: 10.1016/j.jhin.2015.09.020
- Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Graña-Miraglia, Lozano, Velázquez, Volkow-Fernández, Pérez-Oseguera, Cevallos and Castillo-Ramírez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

First Genome Sequence of a Mexican Multidrug-Resistant *Acinetobacter baumannii* Isolate

Lucía Graña-Miraglia,^a Luis Lozano,^a Semiramis Castro-Jaimes,^a Miguel A. Cevallos,^a Patricia Volkow,^b Santiago Castillo-Ramírez^a

Programa de Genómica Evolutiva, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México^a; Instituto Nacional de Cancerología, Ciudad de México, D.F., México^b

L.G.-M. and L.L. contributed equally to this article.

***Acinetobacter baumannii* has emerged as an important nosocomial pathogen worldwide. Here, we present the draft genome of the first multidrug-resistant *A. baumannii* isolate, sampled from a tertiary hospital in Mexico City. This genome will provide a starting point for studying the genomic diversity of this species in Mexico.**

Received 2 February 2016 Accepted 6 February 2016 Published 24 March 2016

Citation Graña-Miraglia L, Lozano L, Castro-Jaimes S, Cevallos MA, Volkow P, Castillo-Ramírez S. 2016. First genome sequence of a Mexican multidrug-resistant *Acinetobacter baumannii* isolate. *Genome Announc* 4(2):e00156-16. doi:10.1128/genomeA.00156-16.

Copyright © 2016 Graña-Miraglia et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Santiago Castillo-Ramírez, iago@ccg.unam.mx.

Acinetobacter baumannii, a Gram-negative coccobacillus, has surfaced as one of the leading causes of hospital-acquired infections in recent years (1, 2). Remarkably, many strains of this species have multidrug-resistant phenotypes showing resistance against most antibiotic classes (2). Ever since the publication of the first genome sequence of an *A. baumannii* isolate a considerable amount of genome sequences have been published. As of 23 November 2015, there are slightly over 30 complete genomes and more than a thousand draft genomes publically available. However, none of them has come from Mexico. Here, in order to start exploring the genome diversity of this species in Mexico, we report the first genome sequence of the Mexican isolate *A. baumannii* Mex11594, hereinafter referred to as Mex11594.

Mex11594 was isolated on 17 June 2011 at the Instituto Nacional de Cancerología (Mexico's National Institute of Oncology), a tertiary hospital located in Mexico City, from pleural fluid of a male patient. Multilocus sequence typing (the Oxford scheme) on this strain revealed that it belongs to the sequence type (ST) 231 (3). According to the PubMLST database (4), isolates sharing this ST have been found in Argentina, Brazil, France, Germany, the Netherlands, and United States, implying that this ST is not exclusive to Latin America. The Mex11594 genome was sequenced using an Illumina MiSeq platform with a 300-bp library (250-bp paired-end reads). The sequencing run produced 1,631,060 reads that were assembled into a preliminary assembly using Velvet version 1.2.10 (5) and SPAdes (6). In order to improve our assembly, we used Consed version 23.0 (7) for gap closure, assembly editing, and error correction. The improved assembly yielded 12 contigs and had a total size of 3,745,499 bp, with an N_{50} of 684,685 bp, an average coverage of 41 \times and a G+C content of 39.03%. Given the number of contigs, the N_{50} , and the average coverage, we are confident that this is a good assembly that should be helpful for future comparative genome analysis. Our improved assembly suggested the existence of several plasmids, which was corroborated via an Eckhardt gel analysis (8). The chromosome was composed of 8 contigs and had an estimated size of 3,700,249 bp.

The draft genome was annotated by means of the Rapid Annotations Using Subsystems Technology (RAST) (9) server, which predicted 3,488 coding sequences, 772 of them hypothetical proteins. Ten rRNAs and 51 tRNAs were identified. This is a multidrug-resistant isolate (10), as it has molecular determinants conferring resistance to at least 3 different types of antibiotics. For instance, this isolate is nonsusceptible to meropenem and imipenem (carbapenems), tobramycin (aminoglycosides), levofloxacin (fluoroquinolones), and ceftazidime and ceftriaxone (cephalosporins) to name but a few.

Further and more extensive comparative analysis of this and other Mexican isolates will help to unveil the genomic diversity (as well as the genetic components conferring the multidrug-resistant phenotypes) of this bacterial pathogen in Mexico and Latin America and, hence, to better understand its worldwide diversity.

Nucleotide sequence accession number. The *A. baumannii* MEX11594 genome sequence was deposited in GenBank under the accession number [LQXZ00000000](https://www.ncbi.nlm.nih.gov/nuclseq/LQXZ00000000).

ACKNOWLEDGMENTS

This work was supported by "Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT," grant number IA200515, to S.C.-R. Lucía Graña-Miraglia is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), and received a fellowship (number 585414) from CONACYT.

FUNDING INFORMATION

This work, including the efforts of Santiago Castillo-Ramírez, was funded by Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT (IA200515).

REFERENCES

1. Antunes LC, Visca P, Towner KJ. 2014. *Acinetobacter baumannii*: evolution of a global pathogen. *Pathog Dis* 71:292–301. <http://dx.doi.org/10.1111/2049-632X.12125>.
2. Zarrilli R, Pournaras S, Giannouli M, Tsakris A. 2013. Global evolution of multidrug-resistant *Acinetobacter baumannii* clonal lineages. *Int J An-*

- timicrob Agents 41:11–19. <http://dx.doi.org/10.1016/j.ijantimicag.2012.09.008>.
3. Bartual SG, Seifert H, Hippler C, Luzon MA, Wisplinghoff H, Rodríguez-Valera F. 2005. Development of a multilocus sequence typing scheme for characterization of clinical isolates of *Acinetobacter baumannii*. J Clin Microbiol 43:4382–4390. <http://dx.doi.org/10.1128/JCM.43.9.4382-4390.2005>.
 4. Jolley KA, Maiden MC. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595. <http://dx.doi.org/10.1186/1471-2105-11-595>.
 5. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de bruijn graphs. Genome Res 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
 6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
 7. Gordon D, Green P. 2013. *Consed*: a graphical editor for next-generation sequencing. Bioinformatics 29:2936–2937. <http://dx.doi.org/10.1093/bioinformatics/btt515>.
 8. Eckhardt T. 1978. A rapid method for the identification of plasmid deoxyribonucleic acid in bacteria. Plasmid 1:584–588. [http://dx.doi.org/10.1016/0147-619X\(78\)90016-1](http://dx.doi.org/10.1016/0147-619X(78)90016-1).
 9. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42: D206–D214. <http://dx.doi.org/10.1093/nar/gkt1226>.
 10. Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, Giske CG, Harbarth S, Hindler JF, Kahlmeter G, Olsson-Liljequist B, Paterson DL, Rice LB, Stelling J, Struelens MJ, Vatopoulos A, Weber JT, Monnet DL. 2012. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin Microbiol Infect 18: 268–281. <http://dx.doi.org/10.1111/j.1469-0691.2011.03570.x>.

NOTA

A continuación se anexan artículos publicados en el marco de colaboraciones tanto externas como dentro del laboratorio de Genómica Evolutiva

Phylogenomics picks out the par excellence markers for species phylogeny in the genus *Staphylococcus*

Lucia Graña-Miraglia¹, César Arreguín-Pérez², Gamaliel López-Leal³, Alan Muñoz¹, Angeles Pérez-Oseguera¹, Estefan Miranda-Miranda², Raquel Cossío-Bayúgar² and Santiago Castillo-Ramírez¹

¹ Programa de Genómica Evolutiva, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

² Centro Nacional de Investigación Disciplinaria en Parasitología Veterinaria del Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Jiutepec, Morelos, Mexico

³ Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

ABSTRACT

Although genome sequencing has become a very promising approach to conduct microbial taxonomy, few labs have the resources to afford this especially when dealing with data sets of hundreds to thousands of isolates. The goal of this study was to identify the most adequate loci for inferring the phylogeny of the species within the genus *Staphylococcus*; with the idea that those who cannot afford whole genome sequencing can use these loci to carry out species assignment confidently. We retrieved 177 orthologous groups (OGs) by using a genome-based phylogeny and an average nucleotide identity analysis. The top 26 OGs showed topologies similar to the species tree and the concatenation of them yielded a topology almost identical to that of the species tree. Furthermore, a phylogeny of just the top seven OGs could be used for species assignment. We sequenced four *staphylococcus* isolates to test the 26 OGs and found that these OGs were far superior to commonly used markers for this genus. On the whole, our procedure allowed identification of the most adequate markers for inferring the phylogeny within the genus *Staphylococcus*. We anticipate that this approach will be employed for the identification of the most suitable markers for other bacterial genera and can be very helpful to sort out poorly classified genera.

Submitted 4 June 2018
Accepted 28 September 2018
Published 24 October 2018

Corresponding author
Santiago Castillo-Ramírez,
iago@ccg.unam.mx

Academic editor
Lesley Hoyles

Additional Information and
Declarations can be found on
page 13

DOI 10.7717/peerj.5839

© Copyright
2018 Graña-Miraglia et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Ecology, Microbiology, Taxonomy

Keywords *Staphylococcus*, Phylogenomics, Bacterial species, Infectious diseases, Evolutionary biology

INTRODUCTION

The study of ecology and evolution has been substantially transformed by the omics technologies. Remarkably, the use of these technologies has allowed essential questions in evolutionary biology to be addressed and has advanced our knowledge in many biological processes (López-Leal et al., 2014; Joseph et al., 2015; Joseph et al., 2016). The precise identification of the different species within any given genus is highly valuable for many branches of microbiology. For instance, as far as clinical microbiology is concerned, this is instrumental in establishing which species are human/animal pathogens and which

are just regular commensal organisms; whereas in microbial ecology species assignment is very helpful in defining the niche-range of the different species ([Becker et al., 2016](#)).

Although whole-genome sequencing is the best method for genotyping bacterial isolates and, therefore, to conduct species assignment/identification, this approach is not yet affordable for routine use in many laboratories all over the world. This is especially true for low and middle-income countries in which the amount of money invested in science is not as much as in developed countries. For this particular purpose, and from a theoretical point of view, orthologous genes are the perfect candidates for inferring the phylogeny of species, as they should reflect the species tree of the taxa considered. As per definition, orthologous genes are the ideal candidates to track the sequence of past of speciation events within a given lineage ([Fitch, 2000](#)). Therefore, in order to infer the phylogeny of the species, a clear-cut ascertainment of orthologous genes is of paramount importance and phylogenomic pipelines can be implemented to try to identify the potential orthologous genes. Orthologous relationships are context-sensitive and a gene family could be mainly composed of orthologous genes at one taxonomic level but if one considers a higher taxonomic level many more non-orthologous genes will appear. Furthermore, orthologous relationships could involve one to many relationships ([Gabaldón & Koonin, 2013](#)). The word co-orthologue was coined to describe that exact situation in which a genome has more than one orthologous gene ([Gabaldón & Koonin, 2013](#)). Hence, from a practical point of view (i.e., operational implementation), orthologous genes with just one gene per genome among the taxa considered should be the best markers to delineate the history of the species.

The genus *Staphylococcus* has a few dozens of species of Gram-positive bacteria, which are commensals colonizing the skin and mucous membranes of some mammals and birds. However, some of these species have a clear clinical and economic relevance, as they are a frequent cause of infection in humans, livestock, and domestic animals. Although the genus is very well known for the human opportunist pathogen *S. aureus*, which is famous worldwide as major source of nosocomial infections ([Challagundla et al., 2018](#); [Frisch et al., 2018](#)), there are some other species such as *S. epidermidis*, *S. lugdunensis*, *S. saprophyticus* and *S. schleiferi* that have been associated with human infections. Several molecular-based methods have been introduced to carry out species identification within the genus *Staphylococcus* ([Kwok & Chow, 2003](#); [Ghebremedhin et al., 2008](#); [Sasaki et al., 2010](#); [Lamers et al., 2012](#)). PCR along with sequence analysis of the genes *dnaJ*, *tuf*, *sodA*, *rpoB*, *hsp60* and *nuc* have been used to differentiate *Staphylococcus* species ([Kwok & Chow, 2003](#); [Ghebremedhin et al., 2008](#); [Sasaki et al., 2010](#); [Lamers et al., 2012](#)). Although they have been very useful and demonstrate a better resolution than the 16S *rRNA* gene ([Ghebremedhin et al., 2008](#)), these genes show different amounts of genetic diversity and, therefore, varying levels of discriminatory power for the different species. Thus, depending on the gene and the species, low-level resolution or even mis-identification can occur. Clearly, an ideal solution would be to conduct whole-genome sequencing of all the isolates considered to tell apart the different species. In terms of genome sequences, *Staphylococcus* is one of the few bacterial genera for which many genomes are available.

In this study, we address the question of which genes are the best inter-species markers (*bona fide* orthologous genes) for the genus *Staphylococcus* and, by applying a phylogenomic approach, we provide a list of the top candidates for species assignment within this genus. Underfunded groups could use these top candidates to accurately conduct species assignment without the necessity of using whole-genome sequencing. Furthermore, this approach could be used for many other species/genera to identify the most suitable markers for inferring the phylogeny of the taxa under investigation.

METHODS

Genomes and homologous groups

We downloaded 265 publically available complete genomes (see [Table S1](#)). These cover 46 different species from the genus *Staphylococcus* and are a good representation of the host range within this genus. We ran CheckM ([Parks et al., 2015](#)) on these genomes to discard poorly sequenced genomes (i.e., with contamination or incomplete) and only one of them (marked in red in [Table S1](#)) did not pass the criteria ($\geq 95\%$ complete and with $\leq 5\%$ contamination) and thus was not included in the rest of the analyses. We also sequenced the genomes of four bacterial isolates (see [Table S2](#)) obtained from hemolymph or hypostome exudates from adult ticks showing signs of bacterial infection and reared over experimentally infested bovines in the Centro Nacional de Investigación Disciplinaria en Parasitología Veterinaria (CENID-PAVET-INIFAP), Jiutepec, Morelos, México. Of note the four isolates sequenced were confirmed to be *Staphylococcus spp* by biochemical tests and 16S rRNA gene sequence analysis. The isolate INIFAP 005-08 was initially identified as *S. saprophyticus* by its physical and biochemical characteristics (Gram-positive coccus bacteria, catalase positive; novobiocin resistant, absence of coagulase, gelatinase and caseinase activity). This isolate was also able to produce acids by fermenting glycerol, lactose, D(+) mannose, sucrose and turanose; however it was unable to use L(+) arabinose or L(+) lactose. This isolate was classified as *S. xylosus* as per 16S rRNA gene analysis. The isolates INIFAP 002-16 and INIFAP 004-15 were identified as Gram-positive coccus bacteria, catalase positive and ribotyped by the 16S rRNA gene, which exhibited positive identity for *S. xylosus*. INIFAP 009-16 was identified as Gram-positive coccus bacteria, catalase-positive and identified as *S. succinus* using the 16S rRNA gene. This same isolate was identified as *S. carnosus* using the API20E (bioMérieux 2010) system. The Nextera XT DNA Library Prep Kit was used for the sequencing libraries and Agilent High Sensitivity DNA Kit was employed for quality control. The isolates were sequenced using an Illumina MiSeq platform, with a 2×250 bp configuration; the genome sequencing was conducted at Instituto Nacional de Medicina Genómica (<http://www.inmegen.gob.mx/>) in Mexico City. Prior to assembling the genomes, we employed the SolexQA v.3.7.1 program ([Cox, Peterson & Biggs, 2010](#)) to trim the reads, employing the default parameters. We used Velvet version 1.2.09 ([Zerbino & Birney, 2008](#)) and Spades v3.11.0 ([Bankevich et al., 2012](#)) to carry out *de novo* assembly setting the option careful and we tested the following k-mer sizes: 51,61,71,81,91,101,111,121,127. We did not consider contigs smaller than 300 bp in the final assemblies. In all cases, Spades outperformed Velvet in terms of the number of

contigs and the N50 statistics. Again, we used CheckM (Parks *et al.*, 2015) to evaluate the quality of the genomes sequenced by us; notably all the 4 isolates were shown to be $\geq 95\%$ complete and with $\leq 5\%$ contamination and therefore reliable for downstream analyses. The details of the genomes assemblies are provided in Table S2. These seem to be good genome assemblies as judged by the high median coverage and the low number of contigs for each one of the newly sequenced isolates. These four genomes have been submitted to the GenBank and have the following accession numbers: PIZQ01000000, PIZN00000000, PIZP00000000, PIZO00000000 (BioProject PRJNA421192). Then, we employed PROKKA v1.11 (we set the genus option to *Staphylococcus*) to annotate all the genomes sequences (newly sequenced ones and the publically available) to have a consistent annotation for all of them. To get the orthologous groups (OGs) with one-to-one relationships among the species, we only considered single gene families (SGF); these were constructed running BLASTP searches with an e-value of 1.0×10^{-30} between the genome of *S. aureus* TW20 and the rest of the genomes. We kept all the cases where there was only one hit per genome and requiring that the seed from TW20 and the hit aligned $\geq 60\%$ of their lengths and were $\geq 45\%$ identical—the rationale behind these two criteria was to make sure that we had whole genes and not just domains. Then, for the all the SGF we constructed DNA alignments in frame via the program Fast Statistical Alignment version 1.5.9 (Bradley *et al.*, 2009), setting the option `-nucprot` that align nucleotide sequences taking into account the protein space. We conducted recombination analysis on each of the SGFs using PhiTest (Bruen, Philippe & Bryant, 2006) that was implemented via the PhiPack program, using a window size of 50 bp. Additionally, we also determined the nucleotide diversity for each one of the SGFs using R pegas library function `nuc.div()` with the default parameters. We carried out a function enrichment analysis as follows: first, we conducted a Gene Ontology (GO) annotation via InterProScan version 5 using the genes from the *S. aureus* TW20 genome as a reference. Then, we conducted a GO enrichment analysis via Blast2GO PRO using a Fisher's Exact Test (<https://www.blast2go.com/>), the analysis was conducted at three different GO levels: Molecular Function (MF), Biological process (BP) and Cellular Component (CC). Then to evaluate the results, we used a False Discovery Rate of 0.05 and the Benjamini–Hochberg correction was used to account for multiple testing.

Phylogenetic reconstructions, average nucleotide identity analysis and neighbour nets

We performed phylogenetic reconstructions for every SGF that did not show recombination signals as per PhiTest (Bruen, Philippe & Bryant, 2006) and for the super alignment (read below). All the gene trees were constructed with RaxML version 8.2.11 (Stamatakis, 2014) executing 10 inferences on the alignment, using 10 distinct randomized Maximum Parsimony (MP) trees and with the GTR+G+I model. Both the Shimodaira-Hasegawa topology test and the Robinson-Foulds distance were also conducted via RAXML with the default settings. It is known that species tree estimation is not a trivial matter and we employed a previous approach that gave trustable results (Castillo-Ramirez & Gonzalez, 2008). We created a super alignment concatenating all the SGFs that did not have signals for recombination and on this alignment a ML phylogeny was constructed also through

RAXML, this time executing 20 independent inferences starting from 20 different MP trees and with GTR+G+I model. For this ML phylogeny bootstrap replicates were generated again employing RaxML, using the -x option that implements a fast algorithm for bootstrapping. The average nucleotide identity analysis (ANI) was run via the python module pyani (<http://widdowquinn.github.io/pyani/>), specifically we employed the ANIm method (Richter & Rosselló-Móra, 2009). To determine the GC content and the proportion of variable sites for each SGF, we used the function summary from the program AMAS (Borowiec, 2016). In order to visualize the conflicting phylogenetic signals in the genes *rpoB* and *tuf* we used SplitTree4 (Huson & Bryant, 2005) to construct Neighbour nets with uncorrected *P* distances.

RESULTS

Defining the species tree, confirming genome affiliations and the set of orthologous genes

We focused on the genus *Staphylococcus* as it has clear clinical and veterinary relevance and, due to that, it has been extensively covered in terms of genome sequences. The data set employed for this study represents 46 species, incorporating 40 type strains, and included a total of 269 genomes for the analyses (see Tables S1 and S2). Notably, this data set has species with a wide host-range covering many of the niches described for this genus. First, we determined the single gene families (SGFs), as these are potential candidates to be orthologous genes with one-to-one relationships, and found 208 SGFs. However recombination could have affected them and, therefore, we conducted recombination tests on them and around 15% of them (31 SGFs) showed signals of recombination and were discarded, which left 177 SGFs as good candidates to be orthologous groups (OG); the list of these 177 SGFs is provided in Table S3. We employed a previously used strategy (Castillo-Ramírez & Gonzalez, 2008) to approximate the species tree. This is the total evidence approach, in which all the 177 SGFs without signals of recombination are concatenated and treated as if they were a single marker on which a Maximum Likelihood (ML) phylogeny was constructed (see Fig. 1)—this ML phylogeny was our Proxy for the Species Tree Topology (PSTT). From Fig. 1 one can see that type strains are scattered throughout the tree and that most of the isolates from individual species tend to form monophyletic groups and these groups are very well-supported as most of them have bootstrap values higher than 80 (see orange dots in the phylogeny). However, we also noted a region on the tree (shaded areas), where different species intermingle together. This is caused by two species, namely *S. warneri* and *S. saccharolyticus*, which do not form monophyletic groups and clearly some strains from these species have been mislabelled. Additionally, as an independent strategy, we also carried out an ANI analysis to calculate the relatedness of the strains to the type strains included in this study (see Fig. 2). This analysis shows that in terms of their taxonomy most of the genomes have been properly labelled. For instance, all the strains designated as *S. lugdunensis* clustered with the type strain, *S. lugdunensis* NCTC 12217, with identity percentages well above 95% and thus could be assigned to this species (Chun et al., 2018). The same applies to 78.26% of the

other species; see for example *S. gallinarum* or *S. simulans* where genomes labelled as belonging to each of these species clustered tightly (again above the 95%) with the type strains. On the other hand, the mis-classification previously noted in the phylogeny is also evident in this analysis, as the strains of both *S. warneri* and *S. saccharolyticus* do not cluster all together in each case and their identity percentages are well below 95% (see red labels in Fig. 2). We then carried out a GO enrichment analysis to have an idea of the overrepresented functions of the 177 SGFs (see Table S4), as expected many of the enriched biological processes and molecular functions have to do with housekeeping functions; the top two GO molecular functions were ATP binding and GTP binding, whereas the top two biological processes were DNA repair and fatty acid biosynthesis (see Table S4 for more details). Taken together, these results demonstrate that the 177 SGFs seem to be real OGs and that most of the genomes have a proper affiliation regarding their taxonomy.

Ranking the orthologous groups

Then we analysed which of the 177 SGFs, from now on OGs, could be the best markers to infer the evolutionary relationships among the species. For that end, we established how similar each one of the single gene trees (from the 177 OGs) was compared to the PSTT; this was done employing the Robinson and Foulds (RF) distance between the single gene trees and the PSTT—this distance gives the number of bipartitions that are not shared by the two trees under consideration. We also used π , a common measure of genetic diversity, to establish the amount of genetic variation for each OG. Figure 3 gives the percentage of similarity of the 177 OG trees to the PSTT and the nucleotide diversity for each of the 177 OGs; this figure shows that no single OG yielded the same topology as the PSTT. Furthermore, we also noted that every single OG has its own topology—not shared by any other OG—as none of all the pairwise comparisons of the OG trees gave a RF equal to 0 (a RF distance of 0 implies that the two topologies in comparison are the same, see Fig. S2). Figure 3 also shows that 167 OGs (94%) have nucleotide diversity values higher than 0.15, which make them good candidates for phylogenetic markers. Of note, for two species (*S. aureus* and *S. epidermidis*) we also computed the intra-species diversity and it seems that these OGs even at this level have a good amount of genetic diversity (see Fig. S3). We found that the top 26 OGs were similar to the PSTT, as all of them have RF distances below 214, indicating that they are more than 60% identical to the PSTT. Notably, these 26 OGs present good values of nucleotide diversity (all but two showing values higher than 0.2 nucleotide changes per nucleotide site and the average for the 26 being 0.26 nucleotide changes per nucleotide site), which is very convenient for their use as to phylogenetic markers. Table 1 provides details about these OGs such as RF distance to the PSTT, nucleotide diversity, function and the ID in the *S. aureus* TW20 genome. Remarkably, when we concatenated these 26 OGs and constructed a phylogeny, this tree showed a topology pretty similar to the PSTT (see Fig. 4A), the RF distance was 74 being 86.1% identical to the PSTT. Furthermore, using only the best seven OGs (the bold candidates from Table 1), we got a similar result—that is the concatenated ML phylogeny of these seven OGs is almost 80% identical to the PSTT (see Fig. 4B). Here it is worth mentioning that although these two phylogenies (Fig. 4) are not 100% identical to the PSTT, the two phylogenies recovered

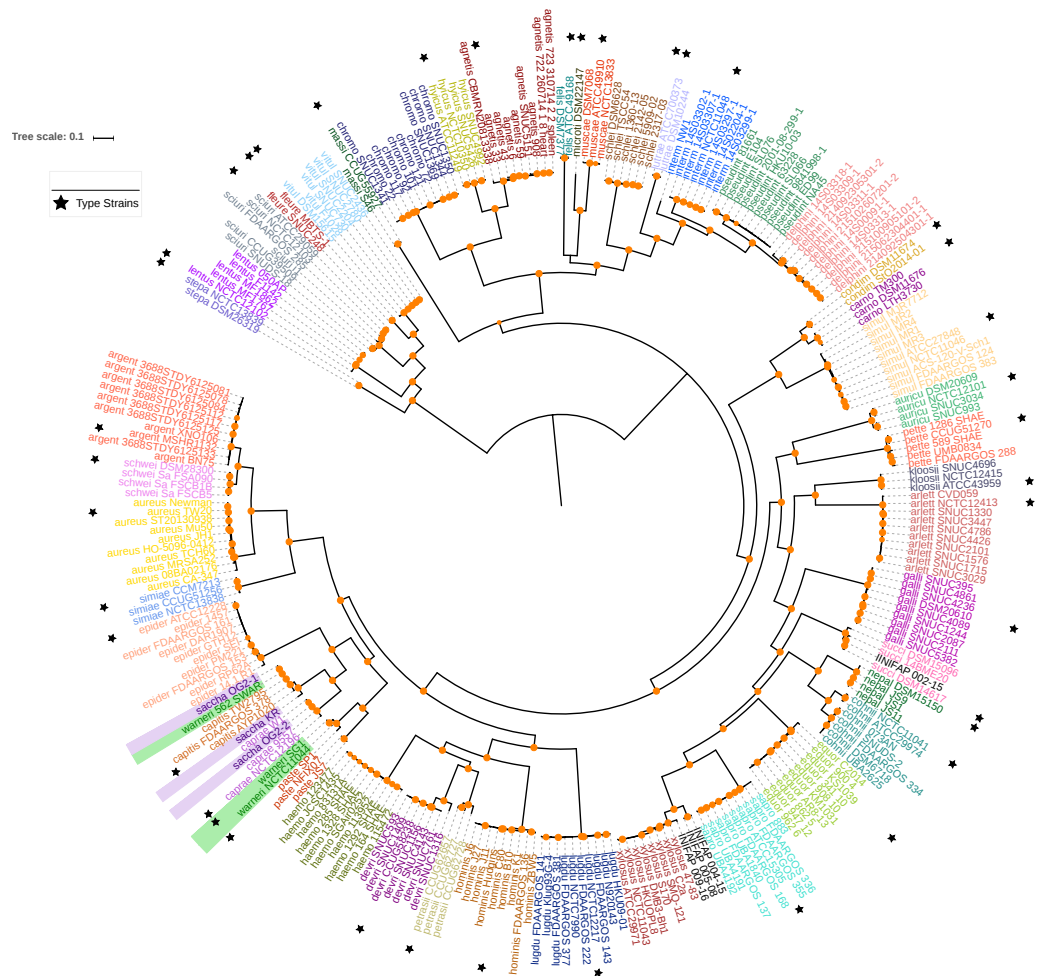


Figure 1 Maximum likelihood phylogenetic tree based on the super alignment of the orthologous groups. Maximum likelihood phylogenetic tree based on the super alignment of the 177 non-recombinant SGFs. Strains of the same species are labelled with the same colour and black stars indicate type strain for some species. Green and violet rectangles highlight strains with conflicting clade assignment, *S. warneri* and *S. saccharolyticus* respectively. Bootstrap values above 80 are shown in orange dots. The scale bar shows the number substitutions per site. The code for the strain identifiers is in Table S1.

Full-size [DOI: 10.7717/peerj.5839/fig-1](https://doi.org/10.7717/peerj.5839/fig-1)

almost all the species as monophyletic groups with very good bootstrap values (>80%), see blue dots in both phylogenies—the only exceptions being the miss-classifications noted also in the PSTT and ANI analysis. In considering this part, although no single gene tree of the 177 OGs yielded the PSTT, just using the best 26 (and even just the top seven) OGs we were able to recover a topology pretty similar to the PSTT.

The inadequacy of usual markers and the soundness of our strategy

Then, we wanted to know how a set of commonly used markers for inferring the phylogenetic relationships within this genus compared to our list. We chose the genes *dnaJ*, *tuf*, *sodA*, *rpoB* and *hsp60* as these are the set of genes most used in previous studies

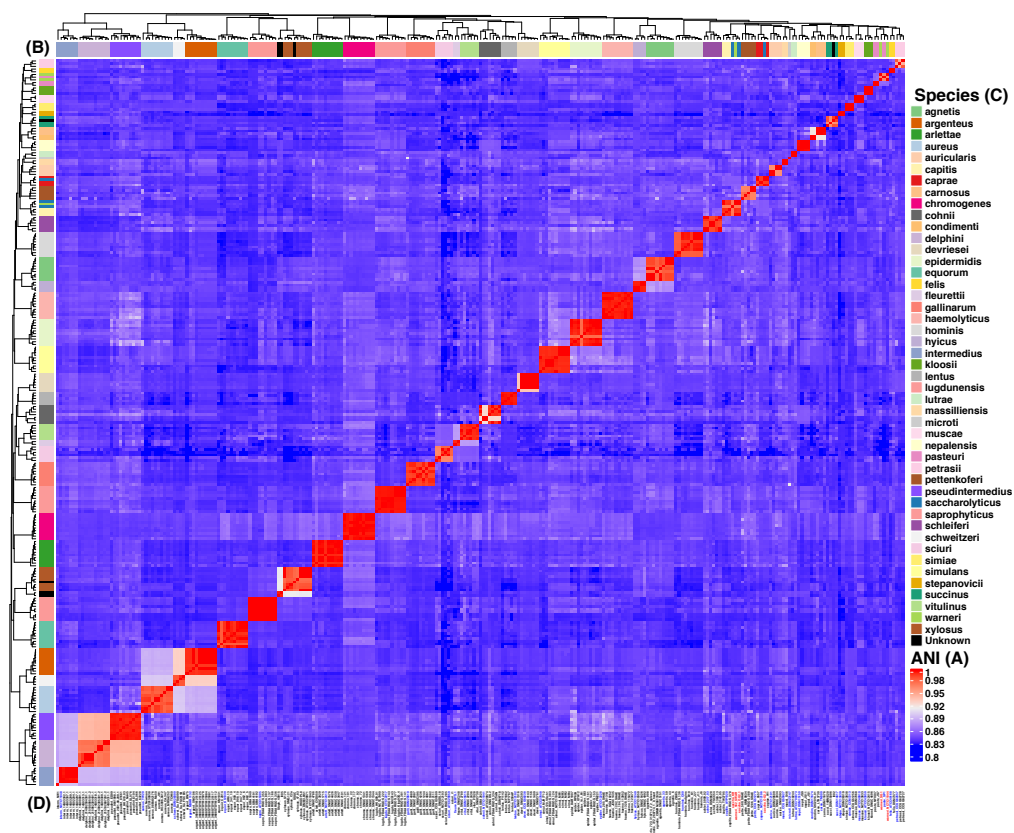


Figure 2 Heat map of the Average Nucleotide Identity (ANI) analysis of the 269 strains from *Staphylococcus spp.* The reddish cells show identity percentages above 95% implying that the strains belong to the same species, whereas non-reddish colors denote identity percentages below 95% (see ANI key, A). The rows on and by the heat map (B) show the species assignment (see Species key, C) and the dendrograms show the clustering of the strains. Strain identifiers (D) are as in Fig. 1 and the color-coding is as follows: blue gives the type strains, whereas red shows the strains with issues of mis-classification. This analysis indicates 44 clearly discernible *Staphylococcus spp.*

Full-size [DOI: 10.7717/peerj.5839/fig-2](https://doi.org/10.7717/peerj.5839/fig-2)

(Kwok & Chow, 2003; Ghebremedhin et al., 2008). Most of them had clear issues to be considered suitable markers given the criteria we employed to define our genuine OGs. Notably two of them, *tuf* and *rpoB*, did have signals for recombination as shown by their Neighbour nets and the PhiTest (see Fig. S1) and their nucleotide diversity was very low (below 0.15, see Fig. 3 orange dots). On the other hand, *sodA* and *hsp60* were not single gene families, as they did have more than one gene per genome in some species. Among these usual markers, only *dnaJ* seems to be a good candidate in as much as it is a SGF, did not have signals of recombination and has a nucleotide diversity value above 0.2 (see Fig. 3). The shortcoming of most of these genes is not totally unexpected, as these genes were not selected based on phylogenetic criteria; however, it is clear that most of these genes do not seem to be a good option to infer the history of the species.

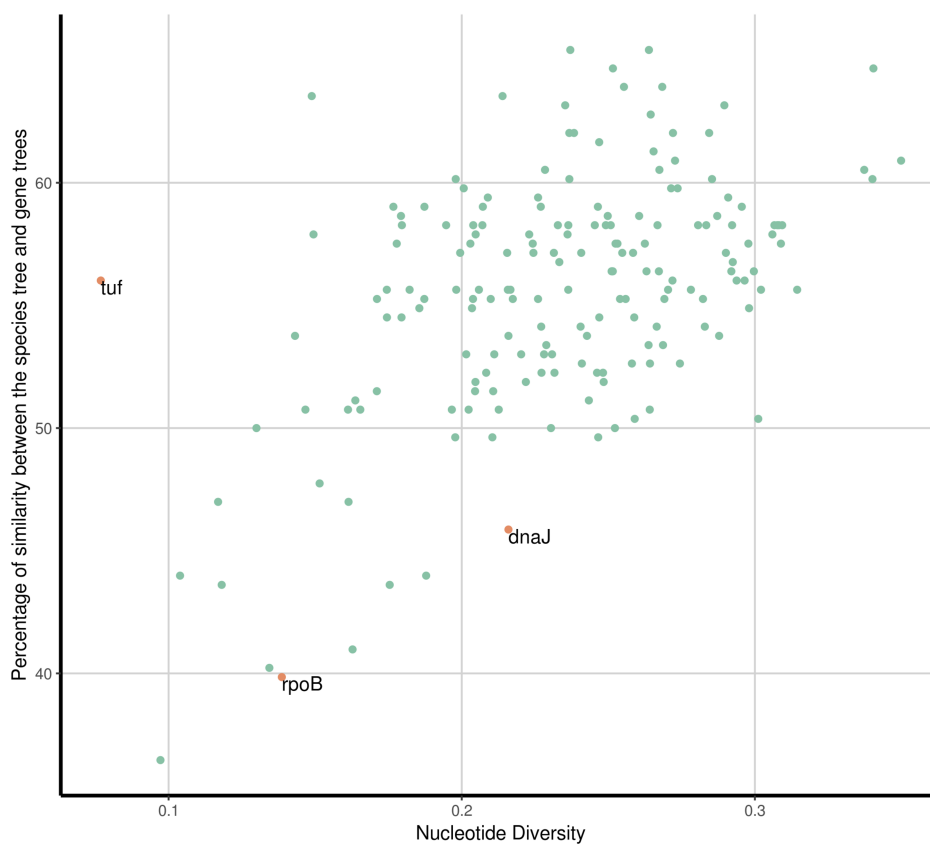


Figure 3 Similarity to the species tree and nucleotide diversity. Percentage of similarity between gene trees and the PSTT and the nucleotide diversity for each of the 177 SGF (green dots). The most similar gene tree topologies correspond to genes with high levels of nucleotide diversity. The commonly used marker genes *dnaJ*, *rpoB* and *tuf* are highlighted (orange dots), they all show low to moderate nucleotide diversity values and none of them show a similarity percentage to the PSTT above 60%.

Full-size  DOI: [10.7717/peerj.5839/fig-3](https://doi.org/10.7717/peerj.5839/fig-3)

Finally, to prove the validity of our phylogenomic approach, in the data set here employed we included the genome sequences of four bacterial isolates collected by us from cattle ticks and phenotypically classified as members of the genus *Staphylococcus* (see [Table S2](#))—these isolates were sequenced using an Illumina MiSeq platform (see ‘Methods’). Initially these isolates were classified using the 16S *rRNA* gene sequence (see [Table S2](#)); however, in two of these cases, namely INIFAP 009-16 and INIFAP 002-15, the initial classification was incorrect. For instance, INIFAP 009-16 was classified as *S. succinus* by its 16S *rRNA* gene sequence but the PSTT placed this isolate with *S. xylosus*. In the case of INIFAP 002-15, whereas PSTT assigned it to *S. succinus*, the starting classification via the 16S *rRNA* gene was *S. xylosus*. Importantly, in the concatenate alignments using our top 26 OGs ([Table 1](#)) the isolates were properly classified (see [Fig. 4A](#)). Furthermore, the ML phylogeny based on the concatenated alignment of the top seven OGs also recovers the true affiliation of these isolates (see [Fig. 4B](#)). To sum up, our genuine OGs are a much better option than the usual markers to establish the phylogeny of the species and even just a few of these (the

Table 1 Top 26 genuine orthologous genes. These are the most adequate markers for inferring the species phylogeny.

Description ^a	NCBI reference sequence	GC content	Nucleotide diversity	Proportion of variable sites	Similarity with species tree (Robinson-fould distance)
DNA mismatch repair protein (<i>MutS</i>)	WP_000073352.1	0.342	0.2370335	0.412	65.4135 (184)
ATP-dependent RecD-like DNA helicase (<i>recD2</i>)	WP_001283311.1	0.345	0.2638535	0.588	65.4135 (184)
Cytosol aminopeptidase (<i>pepA</i>)	WP_001009697.1	0.361	0.3404309	0.760	64.6617 (188)
putative ABC transporter ATP-binding protein (<i>YheS</i>)	WP_000602071.1	0.339	0.2515735	0.688	64.6617 (188)
Phenylalanine-tRNA ligase beta subunit (<i>pheT</i>)	WP_000908982.1	0.366	0.2683997	0.696	63.9098 (192)
Penicillin-binding protein H (<i>pbp3</i>)	WP_000919772	0.340	0.2553319	0.623	63.9098 (192)
DNA-directed RNA polymerase subunit beta' (<i>rpoC</i>)	CBI48492.1	0.381	0.1488365	0.442	63.5338 (194)
Pyruvate kinase (<i>pyk</i>)	WP_001232648.1	0.361	0.2139191	0.584	63.5338 (194)
DNA polymerase III subunit alpha (<i>DnaE</i>)	WP_000226911.1	0.336	0.2895834	0.737	63.1579 (196)
UDP-N-acetylmuramate-L-alanine ligase (<i>murC</i>)	WP_000150163.1	0.335	0.2352931	0.622	63.1579 (196)
Fibronectin-binding domain-containing protein (<i>YloA</i>)	WP_000312763.1	0.338	0.2644582	0.667	62.782 (198)
UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-L-lysine ligase (<i>murE</i>)	WP_000340119.1	0.374	0.2366963	0.638	62.0301 (202)
Homoserine dehydrogenase (<i>dhoM</i>)	WP_000735864.1	0.344	0.2720552	0.579	62.0301 (202)
Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha (<i>accA</i>)	WP_000883645.1	0.357	0.2383029	0.584	62.0301 (202)
Molybdopterin molybdenumtransferase (<i>MoeA</i>)	WP_000259718.1	0.386	0.2844009	0.679	62.0301 (202)
S-adenosylmethionine:tRNA ribosyltransferase-isomerase (<i>queA</i>)	WP_001019171.1	0.357	0.2468871	0.620	61.6541 (204)
ATP-dependent DNA helicase (RecQ)	WP_000983677.1	0.345	0.2654353	0.659	61.2782 (206)
hypothetical protein (<i>YibE/F</i> -like)	WP_001794550.1	0.363	0.3498756	0.723	60.9023 (208)
DNA polymerase III subunit tau (<i>dnaX</i>)	WP_001109047.1	0.368	0.2727606	0.645	60.9023 (208)
Ribonuclease R (<i>rnr</i>)	WP_001050064.1	0.367	0.2284088	0.626	60.5263 (210)
Ktr system potassium uptake protein B (<i>ktrB</i>)	WP_000021864.1	0.342	0.2674265	0.646	60.5263 (210)
Dihydrolipoyl dehydrogenase (<i>pdhD</i>)	WP_001291535.1	0.353	0.3373007	0.759	60.5263 (210)
3-oxoacyl-[acyl-carrier-protein] synthase 3 (<i>fabH</i>)	WP_001100524.1	0.377	0.2367141	0.601	60.1504 (212)
Signal recognition particle receptor (<i>FtsY</i>)	WP_000007682.1	0.366	0.1979450	0.506	60.1504 (212)
6-phosphogluconolactonase (hypothetical)	WP_000181322.1	0.358	0.3401093	0.737	60.1504 (212)
Aspartate aminotransferase (<i>AAT</i> -like)	WP_001068542.1	0.342	0.2853785	0.678	60.1504 (212)

Notes.

^aGene names were taken from the UniProt (<http://www.uniprot.org>) database when possible, otherwise the proteins were blasted against NCBI protein database (<http://www.ncbi.nlm.nih.gov>) and a well-annotated *Staphylococcus* species was used for annotation.

top seven) are able to adequately carry out taxonomy classification of the newly sequenced bacterial isolates.

DISCUSSION

The main goal of our study was to establish the best markers for species identification in the genus *Staphylococcus*. To try to be as comprehensive as possible, we used a data set with 46 species and a total of 269 genomes from this genus that broadly represents most if not all the niches cover by this genus. Here we identified 177 OGs and ranked them according to their potential as phylogenetic markers; clearly, this set of markers should be useful

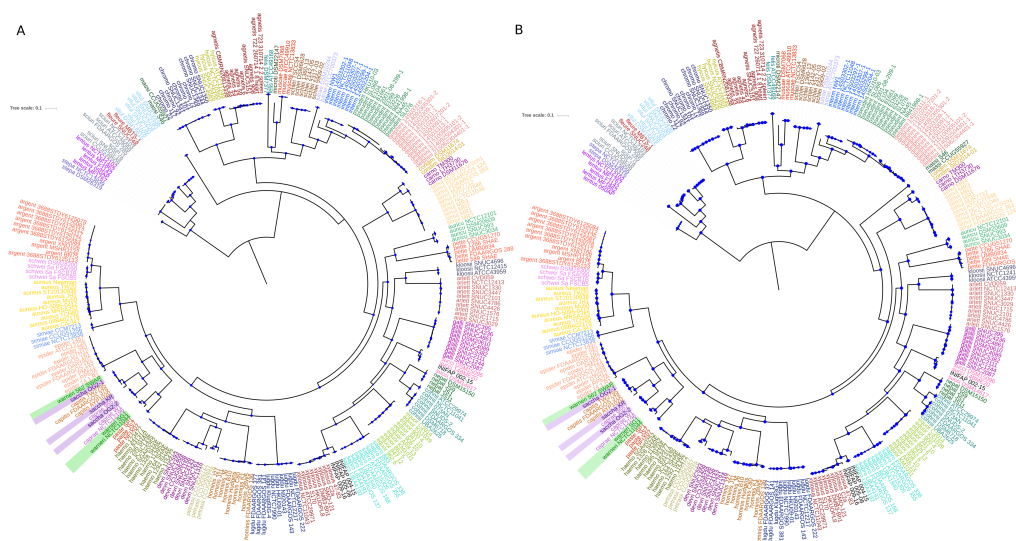


Figure 4 **Phylogenies of the top orthologous groups.** Phylogenetic trees based on the concatenated alignments of the 26 orthologous groups listed in Table 1 (A) and on the concatenated alignments of the top seven orthologous groups (B). The tree in A has a percentage of similarity with the PSTT of 86% and the tree in B of 79%. Bootstrap values above 80 are shown in blue dots. Scale bar shows the number of substitutions per site.

Full-size  DOI: [10.7717/peerj.5839/fig-4](https://doi.org/10.7717/peerj.5839/fig-4)

not only for ecological and evolutionary studies but also for clinical and biotechnological purposes. In addition, our phylogenomic approach allowed us to identify two species (*S. warneri* and *S. saccharolyticus*) in which mislabelling of deposited genomes has occurred. This issue of misclassification is not that rare, as genome sequences submitted in public databases many times do not undergo a proper inspection in terms of microbial taxonomy. However, the combination of these phylogenomic approaches (genome-based phylogeny and ANI analysis) can be very useful to pinpoint those cases of misclassification.

We want to emphasise that our study has two major contributions: one is the list of adequate markers for inferring the phylogeny of the species within the genus *Staphylococcus* and the other is the phylogenomic strategy that we used to identify such markers. Regarding the first major contribution, given that the average gene content of the species here analysed is 2,413, our 177 OGs represent barely the 7% of the genome of these species. However, we need to emphasise here that, for merely practical reasons, we focused on orthologous genes with one-to-one relationships and, very likely, there are many more orthologous genes within these species that did not pass our criteria given that they are not SGFs. Despite the fact that all these OGs are potential good markers, in as much as their histories do not deviate significantly from the history of the species, they do differ among them in the amount of phylogenetic signal they each present. Importantly, no single OG was able to exactly represent the PSTT and no two OGs yielded the same topology; clearly, we found a cloud of topologies. This is in agreement with a previous study that found that only one of the several hundreds of OGs reflected the species tree (*Castillo-Ramirez & Gonzalez, 2008*), although this focused on different evolutionary scales and different bacteria. It is worth

mentioning, that the top 26 OGs have good values of nucleotide diversity (all but two have values higher than 0.21), and are localized in different parts of the chromosome and present different functions. We think these 26 OGs represent a really good set of markers for inferring the PSTT; actually concatenating the best seven it was possible to get a very good approximation of the PSTT. Here, we want to highlight that using the phylogeny of the top 7 OGs all but the two species with issues of mis-classification were recovered as monophyletic groups and thus it seems that just using these seven markers species assignment can be carried out. Clearly, this set of markers could be very useful for underfunded laboratories working with *Staphylococcus* isolates. Although from a genotyping perspective, ideally one would want to use whole-genome sequencing for species identification, many laboratories in the world (specially in developing countries) still cannot afford the sequencing of tens to hundreds of isolates. Thus, this short list of markers should be extremely useful for those who only can sequence a few loci, which is often the case for clinical, environmental and evolutionary biology microbiologists in developing countries. Clearly, the markers highlighted in this study are a much better option than the commonly used markers, as the latter seem to exhibit obvious flaws (more than one copy per genome, signals of recombination, low nucleotide diversity values) for inferring the history of the species. Our strategy, and in turn the list of genes found, proved to be factually sound, even when using just a few markers for the taxonomic assignment of newly sequenced bacterial isolates.

Maybe more important than the first major contribution is the fact that our phylogenomic approach will allow the identification of adequate markers in many different genera—not only from bacteria but also from archaea. This is very important, as gene families showing only orthologous relationships very likely do not extend all over the tree of life—or just for a few gene families. Along these lines, it is very likely that many of the markers found here will not work for other genera (i.e., they will not show one-to-one orthologous relationships) as the homologous genes within those genera might have been affected by horizontal gene transfer or duplication and differential loss, or some other molecular event that prevent the history of the gene to reflect the speciation events. We want to highlight that the biology of the species under consideration could have a very important effect on the number of orthologous genes found. For instance, highly recombinogenic species, such as *Neisseria gonorrhoeae* (Ezewudo et al., 2015) and *Acinetobacter baumannii* (Grana-Miraglia et al., 2017), would have fewer orthologous genes than more clonal species. Nonetheless, the strategy employed by us should work even in these species, although the number of potential orthologous genes should be much less.

CONCLUSIONS

In summary, here we devised and applied a phylogenomic approach that allowed us to define the most suitable markers for inferring the species phylogeny within the genus *Staphylococcus*. We acknowledge that the effectiveness of these markers for other bacterial genera remains open to investigation. However, we are confident that the phylogenomic approach here implemented can be employed to identify the most suitable markers for other genera. On a broader level, this study has very practical implications, as it provides

a general framework for mapping out the best markers to construct the phylogeny of the taxa considered not only at the genus level but also at other taxonomic levels.

ACKNOWLEDGEMENTS

We are extremely grateful to the editor and the reviewers, as their suggestions have drastically improved our manuscript.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by CONACYT PDCPN (grant number 248049) and INIFAP SIGI (grant number 1322633028) to Raquel Cossío-Bayúgar and by “Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT” (grant number IA201317) to Santiago Castillo-Ramirez. The genome sequencing of the newly sequence isolates was conducted at Instituto Nacional de Medicina Genómica (<http://www.inmegen.gob.mx>). Lucía Graña-Miraglia is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), and received a fellowship (number 585414) from CONACYT. Gamaliel López-Leal was supported by CONACyT postdoctoral fellowship CB-2013/223279. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

CONACYT PDCPN: 248049.

INIFAP SIGI: 1322633028.

Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT: IA201317.

CONACYT: 585414.

CONACyT postdoctoral fellowship: CB-2013/223279.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Lucia Graña-Miraglia performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- César Arreguín-Pérez performed the experiments, contributed reagents/materials/analysis tools, approved the final draft.
- Gamaliel López-Leal performed the experiments, analyzed the data, approved the final draft.
- Alan Muñoz performed the experiments, approved the final draft.
- Angeles Pérez-Oseguera performed the experiments, contributed reagents/materials/analysis tools, approved the final draft.

- Estefan Miranda-Miranda and Raquel Cossío-Bayúgar contributed reagents/materials/-analysis tools, approved the final draft, contributed bacterial isolates.
- Santiago Castillo-Ramírez conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The 4 newly sequenced genomes have been submitted to the GenBank and have the following accession numbers: [PIZQ01000000](#), [PIZN00000000](#), [PIZP00000000](#), [PIZO00000000](#) (BioProject [PRJNA421192](#)).

Data Availability

The following information was supplied regarding data availability:

The four newly sequenced genomes are available under BioProject [PRJNA421192](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.5839#supplemental-information>.

REFERENCES

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19(5):455–477 DOI 10.1089/cmb.2012.0021.
- Becker NS, Margos G, Blum H, Krebs S, Graf A, Lane RS, Castillo-Ramírez S, Sing A, Fingerle V. 2016. Recurrent evolution of host and vector association in bacteria of the *Borrelia burgdorferi* sensu lato species complex. *BMC Genomics* 17(1):734 DOI 10.1186/s12864-016-3016-4.
- Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660 DOI 10.7717/peerj.1660.
- Bradley RK, Roberts A, Smoot M, Juvekar S, Do J, Dewey C, Holmes I, Pachter L. 2009. Fast statistical alignment. *PLOS Computational Biology* 5(5):e1000392 DOI 10.1371/journal.pcbi.1000392.
- Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172(4):2665–2681.
- Castillo-Ramirez S, Gonzalez V. 2008. Factors affecting the concordance between orthologous gene trees and species tree in bacteria. *BMC Evolutionary Biology* 8:300 DOI 10.1186/1471-2148-8-300.
- Challagundla L, Reyes J, Rafiqullah I, Sordelli DO, Echaniz-Aviles G, Velazquez-Meza ME, Castillo-Ramírez S, Fittipaldi N, Feldgarden M, Chapman S. 2018. Phylogenomic classification and the evolution of clonal complex 5 Methicillin-Resistant *Staphylococcus aureus* in the Western Hemisphere. *Frontiers in Microbiology* 9:Article 1901 DOI 10.3389/fmicb.2018.01901.

- Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, Da Costa MS, Rooney AP, Yi H, Xu XW, De Meyer S, Trujillo ME. 2018.** Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology* **68**(1):461–466 DOI [10.1099/ijsem.0.002516](https://doi.org/10.1099/ijsem.0.002516).
- Cox MP, Peterson DA, Biggs PJ. 2010.** SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC bioinformatics* **11**(1):485 DOI [10.1186/1471-2105-11-485](https://doi.org/10.1186/1471-2105-11-485).
- Ezewudo MN, Joseph SJ, Castillo-Ramirez S, Dean D, Del Rio C, Didelot X, Dillon J-A, Selden RF, Shafer WM, Turingan RS. 2015.** Population structure of *Neisseria gonorrhoeae* based on whole genome data and its relationship with antibiotic resistance. *PeerJ* **3**:e806 DOI [10.7717/peerj.806](https://doi.org/10.7717/peerj.806).
- Fitch WM. 2000.** Homology a personal view on some of the problems. *Trends in Genetics* **16**(5):227–231 DOI [10.1016/S0168-9525\(00\)02005-9](https://doi.org/10.1016/S0168-9525(00)02005-9).
- Frisch M, Castillo-Ramirez S, Petit R, Farley M, Ray S, Albrecht V, Limbago B, Hernandez J, See I, Satola S. 2018.** Invasive methicillin-resistant *Staphylococcus aureus* USA500 strains from the US Emerging Infections Program constitute three geographically distinct lineages. *mSphere* **3**(3):e00571-00517.
- Gabaldón T, Koonin EV. 2013.** Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics* **14**(5):360–366 DOI [10.1038/nrg3456](https://doi.org/10.1038/nrg3456).
- Ghebremedhin B, Layer F, König W, König B. 2008.** Genetic classification and distinguishing of *Staphylococcus* species based on different partial gap, 16S rRNA, hsp60, rpoB, sodA, and tuf gene sequences. *Journal of Clinical Microbiology* **46**(3):1019–1025 DOI [10.1128/JCM.02058-07](https://doi.org/10.1128/JCM.02058-07).
- Grana-Miraglia L, Lozano LF, Velazquez C, Volkow-Fernandez P, Perez-Oseguera A, Cevallos MA, Castillo-Ramirez S. 2017.** Rapid gene turnover as a significant source of genetic variation in a recently seeded population of a healthcare-associated pathogen. *Frontiers in Microbiology* **8**:Article 1817 DOI [10.3389/fmicb.2017.01817](https://doi.org/10.3389/fmicb.2017.01817).
- Huson DH, Bryant D. 2005.** Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**(2):254–267.
- Joseph SJ, Cox D, Wolff B, Morrison SS, Kozak-Muiznieks NA, Frace M, Didelot X, Castillo-Ramirez S, Winchell J, Read TD. 2016.** Dynamics of genome change among *Legionella* species. *Scientific reports* **6**:33442 DOI [10.1038/srep33442](https://doi.org/10.1038/srep33442).
- Joseph SJ, Marti H, Didelot X, Castillo-Ramirez S, Read TD, Dean D. 2015.** Chlamydiaceae genomics reveals interspecies admixture and the recent evolution of *Chlamydia abortus* infecting lower mammalian species and humans. *Genome Biology and Evolution* **7**(11):3070–3084 DOI [10.1093/gbe/evv201](https://doi.org/10.1093/gbe/evv201).
- Kwok AY, Chow AW. 2003.** Phylogenetic study of *Staphylococcus* and *Micrococcus* species based on partial hsp60 gene sequences. *International Journal of Systematic and Evolutionary Microbiology* **53**(1):87–92 DOI [10.1099/ijse.0.02210-0](https://doi.org/10.1099/ijse.0.02210-0).
- Lamers RP, Muthukrishnan G, Castoe TA, Tafur S, Cole AM, Parkinson CL. 2012.** Phylogenetic relationships among *Staphylococcus* species and refinement of cluster groups based on multilocus data. *BMC Evolutionary Biology* **12**(1):171 DOI [10.1186/1471-2148-12-171](https://doi.org/10.1186/1471-2148-12-171).

- López-Leal G, Tabche ML, Castillo-Ramírez S, Mendoza-Vargas A, Ramírez-Romero MA, Dávila G. 2014.** RNA-Seq analysis of the multipartite genome of *Rhizobium etli* CE3 shows different replicon contributions under heat and saline shock. *BMC Genomics* **15**(1):770 DOI [10.1186/1471-2164-15-770](https://doi.org/10.1186/1471-2164-15-770).
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015.** CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25**:1043–1055 DOI [10.1101/gr.186072.114](https://doi.org/10.1101/gr.186072.114).
- Richter M, Rosselló-Móra R. 2009.** Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* **106**(45):19126–19131 DOI [10.1073/pnas.0906412106](https://doi.org/10.1073/pnas.0906412106).
- Sasaki T, Tsubakishita S, Tanaka Y, Sakusabe A, Ohtsuka M, Hirotaki S, Kawakami T, Fukata T, Hiramatsu K. 2010.** Multiplex-PCR method for species identification of coagulase-positive staphylococci. *Journal of Clinical Microbiology* **48**(3):765–769 DOI [10.1128/JCM.01232-09](https://doi.org/10.1128/JCM.01232-09).
- Stamatakis A. 2014.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9):1312–1313 DOI [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033).
- Zerbino DR, Birney E. 2008.** Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**:821–829 DOI [10.1101/gr.074492.107](https://doi.org/10.1101/gr.074492.107).

Unexplored Genetic Diversity of Multidrug- and Extremely Drug-Resistant *Acinetobacter baumannii* Isolates from Tertiary Hospitals in Honduras

Gamaliel López-Leal,^{1,*} Julio César Zuniga-Moya,² Semiramis Castro-Jaimes,¹ Lucía Graña-Miraglia,¹ Ángeles Pérez-Oseguera,¹ Hilda Sarahi Reyes-García,³ Simmons David Gough-Coto,² René Pavón-Madrid,² Suyapa Aurora Bejarano,^{2,3} Annabelle Ferrera,⁴ Santiago Castillo-Ramírez,^{1,*} and Miguel A. Cevallos¹

Although *Acinetobacter baumannii* has become one of the most important nosocomial pathogens worldwide, very little is known about the genetic identity of isolates from less developed countries in Latin America. To alleviate this, we sequenced the genomes of 16 *A. baumannii* isolates from Honduras. Whole-genome sequencing was conducted on 16 isolates from five Honduran Hospitals. With the sequences of these Honduran isolates and other 42 publically available genomes, a maximum likelihood phylogeny was constructed to establish the relationship between the Honduran isolates and those belonging to the International Clones (ICs). In addition, sequence type (ST) assignment was conducted by the PubMLST, and antibiotic resistance genes were identified using ResFinder. The Honduran isolates are highly diverse and contain new allele combinations under the Bartual multilocus sequence typing scheme. The most common STs were ST_B447/ST_P10 and ST_B758/ST_P156. Furthermore, none of these isolates belongs to clonal complexes related to the ICs. Antibiotic susceptibility profiles of these isolates showed that they are multidrug resistant (MDR) or extensively drug resistant (XDR). In addition, the Honduran isolates had genes involved in resistance to seven antibiotic families. For instance, several *bla*OXA alleles were found, including *bla*OXA-23 and a gene encoding the metallo-beta-lactamase NDM-1. Notably, nine of the Honduran isolates have antibiotic resistance genes to three or more antibiotic families. In summary, in this study, we unveiled an untapped source of genetic diversity of MDR and XDR isolates; notably, these isolates did not belong to the well-known ICs.

Keywords: antibiotic resistance genes, *bla*OXA-23, NDM-1, International clonal complex

Introduction

OVER THE LAST DECADE, *Acinetobacter baumannii* has become one of the most prominent bacterial pathogens worldwide, as it has turned into one of the most problematic nosocomial pathogens in intensive care units in many parts of the world and a constant source of nosocomial infections, and a frequent cause of outbreaks. Actually, *A. baumannii* is at the very top of the 2017 World Health Organization Priority List for Research and Discovery of New Antibiotics. Although much information about *A. baumannii* clinical isolates has been gathered in developed countries, clinical isolates from many developing countries have received very little attention.

The limited amount of information available about *A. baumannii* isolates in Latin America shows that, although the International Clones (ICs) are present in this region of the world, they are not the most predominant.¹⁻³ For instance, even though Mexico shares a large border with the United States of America, the most common sequence type (ST) in Mexico is ST758, which belongs to clonal complex 636 (CC636, previously CC113) and is not related to CC92 (IC-II), the most prevalent in the United States of America.⁴ Although some of us recently started to analyze the genomic diversity of *A. baumannii* isolates in Mexico,⁵⁻⁷ the situation in Central America is worst and there is hardly any information about isolates for this region. To mitigate this lack of information, we carried out genome sequencing of 16 *A.*

¹Programa de Genómica Evolutiva, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, México.

²Campus San Pedro y San Pablo, Universidad Católica de Honduras, San Pedro Sula, Honduras.

³Liga Contra el Cáncer, San Pedro Sula, Honduras.

⁴Universidad Nacional Autónoma de Honduras, Escuela de Microbiología, Edificio E1, 2da planta, Ciudad Universitaria Tegucigalpa, Tegucigalpa M.D.C., Honduras.

*These two authors have contributed equally to this work.

baumannii isolates sampled from five different hospitals in Honduras located in two different departments. Multilocus sequence typing (MLST) analysis indicated that five isolates had new allele combinations and none of the isolates was related to the ICs. Moreover, a phylogenomic analysis showed that the isolates were highly diverse and spread over the entire phylogenetic tree. In addition, most of the *A. baumannii* isolates possessed a wide variety of antibiotic resistance genes, including several *bla*OXA alleles and *bla*NDM1, a metallo-beta-lactamase capable of conferring resistance to a wide variety of beta-lactam antibiotics, including carbapenems. Our study shows that Honduras is an untapped source of genetic diversity of multidrug-resistant and extremely drug-resistant isolates of *A. baumannii*.

Methods

Isolate collection

Clinical isolates were obtained from patients with *A. baumannii* infections in five tertiary care hospitals located in two Honduran departments: Cortés and Francisco Morazán. Initial identification of isolates was performed using VITEK 2 (bioMérieux, France). Age, sex, and source of the samples are listed in Supplementary Table S1 (Supplementary Data are available online at www.liebertpub.com/mdr). Antimicrobial susceptibility profiles for all strains were determined using BD EpiCenter™ Microbiology Data Management System V7.22A (BD Diagnostics, France) at Department of Microbiology, Instituto Nacional de Cancerología (Supplementary Table S2). Multidrug resistant (MDR) and extensively drug resistant (XDR) were defined according to the criteria used by Magiorakos *et al.*⁸

Plasmid profiles

Plasmid profiles of the isolates were visualized on agarose gels according to the protocol described by Hynes and McGregor.⁹ Briefly, cell were harvested in mid-logarithmic phase and lysed in the agarose gel well with lysozyme and sodium dodecyl sulfate before electrophoresis. Images were obtained staining the gel with ethidium bromide and photographed in a U.V. transilluminator.

Genome sequencing

DNA was obtained with the Genomic DNA Purification Kit (Thermo Scientific), using the manufacturer's instructions. Sequencing was performed at the Unidad de Secuenciación e Identificación de Polimorfismos, Instituto Nacional de Medicina Genómica (INMEGEN, México). DNA libraries were constructed with a TruSeq DNA Library Preparation Kit v3 and sequenced on an Illumina MiSeq platform with 2×300 bp paired-end reads. Genome assemblies were carried out using ABySS 2.0.1¹⁰ and VELVET 1.2.10,¹¹ and the best assembly of each genome was deposited in GenBank and annotated with the NCBI annotation pipeline. The ST assignment of each *A. baumannii* isolate was obtained from the PubMLST database (<https://pubmlst.org>). Supplementary Table S3 gives the accession numbers and ST assignment of these isolates. eBURST analysis was performed with goeBURST¹² and the identification of acquired antibiotic resistance genes was done with ResFinder.¹³

Selection of homologous groups, model selection, and phylogenetic inferences

For the phylogenetic analysis, we used the genome assemblies of our *A. baumannii* collection and 41 genomes from NCBI (Supplementary Table S4). To obtain the homologous groups, we first blasted the proteome of each genome against the rest of the proteomes with a cutoff e-value of $1.0e - 30$. Homologous protein groups were identified with PanOCT,¹⁴ requiring that proteins within a homologous group show an aligned region of $\geq 90\%$ of their lengths and $\geq 80\%$ identity. For our phylogenetic analysis, we only considered single gene families (SGFs), which have only one gene per genome, and these were aligned with MUSCLE,¹⁵ specifying 50 iterations. To create a DNA alignment in frame, we used the program TRANALING,¹⁶ and to detect and discard SGFs with recombination signals, we used PhiPack¹⁷ setting a *p*-value cutoff of 0.05. Then, the SGFs that did not show recombination signals were concatenated to form a superalignment. On this alignment, we constructed a maximum likelihood (ML) phylogeny using the generalized time-reversible model (GTR) gamma distributed rate variation among sites (+G) model, which was the most adequate model according to the jModelTest 2 program.¹⁸ We ran a nonparametric bootstrap analysis (100 replicates) on the ML phylogeny to establish the support for the clades. To determine the relatedness between the Honduran strains, the number of single-nucleotide polymorphisms (SNPs) was determined from the concatenated alignment in a pairwise manner by MEGA 7.¹⁹

Results

The collection

Previous studies have suggested that the worldwide increase in *A. baumannii* infections is the result of the rapid expansion of a limited number of clones. The prevalence of the ICs and specifically, IC-II in Europe, North America, the Middle East, Asia, and Oceania suggests that this situation could also occur in the rest of the world. To explore which clones are circulating in Honduras and which antibiotic resistance genes they carry, we gathered a small collection of nosocomial isolates (16 isolates in total), from March 2015 to June 2016, sampled from five tertiary care health institutions located in two departments in Honduras. It is important to note that our collection was made without any specific criteria regarding the antibiotic resistance profiles of the isolates. The isolates were obtained from several patients and all were initially identified as *A. baumannii* as per the automated system VITEK 2. Antimicrobial susceptibility profiles revealed that 75% of the strains were XDR and the rest were MDR. All isolates were nonsusceptible to ampicillin, cefazolin, cefoxitin, ertapenem, and fosfomycin c/G6P. Half of the isolates were nonsusceptible to imipenem and 56.25% to meropenem. However, all were susceptible to colistin (Supplementary Table S2).

The isolates came from different sources (Supplementary Table S1); where the most abundant source was endotracheal tubes (36.4%), followed by blood (18.1%) and wounds (13.6%). To avoid working with the same or very closely related clones, the plasmid profiles of all strains were obtained by *in situ* lysis gel electrophoresis (Supplementary Fig. S1). This analysis showed that with the exception of

two isolates (read below), which had the same profile, the plasmid profiles of the isolates varied widely in terms of the number and size of plasmids. Therefore, almost each isolate possessed a distinctive plasmid profile, suggesting that, in general, they did not come from the same nosocomial outbreak. The exceptions were HEU2 and MCR54, collected in different hospitals located in distinct Honduran departments.

Phylogenetic relationships of the isolates

Then, to evaluate the relationships between the Honduran isolates and isolates obtained from other parts of the world, we sequenced the genome of all members of our collection with an Illumina MiSeq platform. We first used the genome sequences to define their STs under the MLST Bartual scheme.²⁰ We found that five isolates have new allele combinations and, in consequence, curators of the *A. baumannii* database (PubMLST) provided us with new ST numbers (Supplementary Table S2). Unfortunately, we did not obtain enough sequence information for isolate MCR6739 to establish its ST. In total, nine STs were found in these isolates and the most common were ST_B447 and ST_B758, each one with three isolates. It is important to point out that isolates from the ST_B447 have been recorded in the Czech Republic, United States of America, and Canada, whereas isolates belonging to ST_B758 have been described in Mexico and in Canada.^{7,21} Another isolate (MCR10179) belongs to ST_B229, an ST that has been reported in Mexico, Brazil, and the United States of America (PubMLST database). Furthermore, the eBURST analysis showed that the Honduran isolates belong to nine different clonal complexes, none of them related with the clonal complexes embracing the ICs. The most common STs, ST_B447 and ST_B758, belonged to clonal complexes CC447 and CC636, respectively. These observations suggest the existence of new ICs preferentially circulating in North and South America. On the other hand, our collection also contains five isolates with new allele combinations, suggesting that they could be Honduran endemic clones not previously described. We also constructed a phylogenetic tree to establish more accurately the relationships between these isolates. The ML phylogenetic tree (Fig. 1) showed that the Honduran isolates are highly diverse as they are located, forming small clusters, on different parts of the tree. In line with this, we found that the average difference between any two Honduran strains was 1,456 SNPs. Furthermore, from this tree, it is clear that the Honduran isolates do not have a close association with isolates from the ICs. These results are congruent with eBURST analysis, considering that Honduran isolates sharing the same ST were located within the same clade. It is important to point out that despite the small sample studied here, these results indicate an unexpected diversity of *A. baumannii* in Honduras.

Antibiotic resistance genes

The second major goal of our study was to identify the antibiotic resistance genes present in the Honduran isolates and for that, we analyzed the draft genomes using the ResFinder database. A summary of these results is described in Table 1. The most common resistance genes were those

related with beta-lactam antibiotics. As expected, all the isolates from our collection have two intrinsic encoded beta-lactamases: *bla*ADC-25 and *bla*OXA-51-like alleles (*bla*OXA-64, *bla*OXA-65, *bla*OXA-68, *bla*OXA-69, *bla*OXA-132, or *bla*OXA-180). In two isolates, the *bla*ADC-25 gene is closely associated with an ISAbal element (HEU2 and MCR54) and in other with an ISAbal27 element (MCR26739). To the best of our knowledge, an association between ISAbal27 and a *bla*ADC allele has not been reported before. Four isolates HEU55808, HEU5, MCR6056, and MCR10179, obtained from two hospitals, contain a chromosomal encoded *bla*OXA-23 gene in close proximity with an ISAbal element, suggesting that this element influences *bla*OXA-23 expression and potentially increases the minimal inhibitory concentrations for carbapenems.

Two different isolates obtained from the same hospital possess a *bla*NDM-1, a gene encoding a class B metallo-lactamase (MCR9238 and MCR10172, see Table 1). This gene is usually carried in plasmids; however, considering that we have only draft genomes of these isolates, it was not possible to determine whether *bla*NDM-1 is located on a plasmid; nonetheless, in both cases, this gene is linked to a Tn125-like element. We also found three isolates (HEU2, MCR54, and IHSS3526) harboring a *bla*CTX-M15 gene that encodes an extended-spectrum beta-lactamase, and in the HEU2 isolate, a carbenicillin-hydrolyzing beta-lactamase allele (*bla*CARB-8) was found; notably, to the best of our knowledge, this has not previously been reported in *A. baumannii*. Finally, we identified in isolate MCR6739 a gene linked to sulbactam resistance (*bla*TEM-1B).

In addition, we identified several genes related to aminoglycoside resistance and the most common were the *strAB* genes, which are involved in streptomycin resistance; these genes were found in 5 of the 16 isolates studied here. Another gene involved in aminoglycoside resistance was *armA*, a 16S-methyltransferase gene that confers resistance to 4,6-disubstituted deoxystreptamines. This gene was found in isolates HEU2 and MCR54. Even more, we found three types of aminoglycoside modifying enzymes: first, *aadA1* and *aadA2* both encoding aminoglycoside adenylyltransferases involved in streptomycin and spectinomycin resistance. Second, two genes encoding *O*-phosphotransferases: *aph(3')-Ic* and *aph(3')-Via*. Both genes linked to kanamycin, neomycin, and puromycin resistance, and the last one also to amikacin and gentamicin B resistance. Finally, three genes encoding *N*-acetyltransferases, involved in different types of aminoglycoside resistance, were found: *aac(3)-Ia* (gentamicin resistance), *aac(3)-IId* (gentamicin and tobramycin resistance), and *aac(3)-IIa*.

Besides the genes involved in resistance to beta-lactam and aminoglycoside antibiotics, we identified genes linked to resistance to other antibiotic families: in three isolates, we found the macrolide resistance genes *msrE* and *mphE*. Seven isolates possess a *sul2* gene and three possess a *sul1* gene, both related to sulfonamide resistance. Four isolates have genes related to tetracycline resistance; two possess a *tetB* gene and the others two possess *tetA* or *tetG*. A couple of isolates have a chloramphenicol resistance-related gene, *catA1*, and one of these also has a florphenicol resistance gene (*floR*). Finally, in three isolates, we identified genes involved in trimethoprim resistance: *dfrA1* in one isolate and *dfrA12* in the other two.

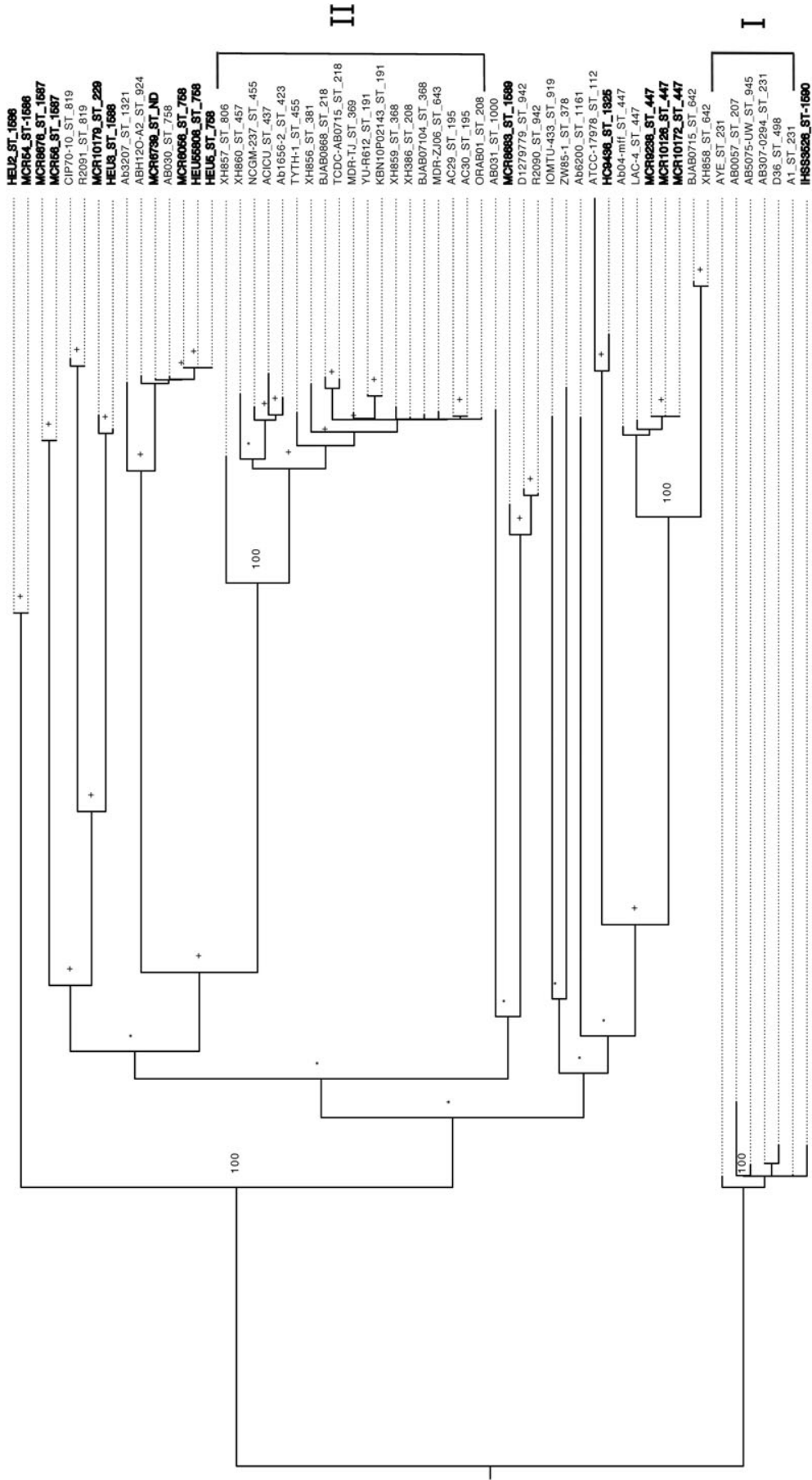


FIG. 1. Maximum likelihood phylogenetic tree. The tree was constructed on the concatenated alignment of all the single gene families not showing recombination signals. For each isolate the sequence type (Bartual scheme) is shown next to its name. Honduran isolates are indicated in *bold letters*. Isolates belonging to the International Complex II and I are marked with the two big *brackets*. For sake of clarity, the bootstrap values are shown just for some nodes; the rest of the nodes are marked with *plus signs* for those nodes with 70 or higher bootstrap support values, whereas *asterisks* denote nodes with <70 bootstrap support.

TABLE 1. ANTIBIOTIC RESISTANCE GENE PROFILES BASED ON THE RESFINDER RESULTS

Isolate	Aminoglycoside	Beta-lactam	Macrolide	Phenicol	Sulfonamide	Tetracycline	Trimethoprim
HEU2	<i>aadA2</i> <i>aac(3)-IId</i> <i>armA</i>	<i>bla</i> CTX-M-15 <i>bla</i> ADC-25 <i>bla</i> OXA-132 <i>bla</i> CARB-8	<i>msrE</i> <i>mphE</i>	—	<i>sul1</i> <i>sul2</i>	—	<i>dfrA12</i>
MCR56		<i>bla</i> ADC-25 <i>bla</i> OXA-65	—	—	—	—	—
MCR54	<i>aadA2</i> <i>aac(3)-IId</i> <i>armA</i>	<i>bla</i> OXA-132 <i>bla</i> CTX-M-15 <i>bla</i> ADC-25	<i>msrE</i> <i>mphE</i>	—	<i>sul1</i>	—	<i>dfrA12</i>
MCR6739	<i>aac(3)-IId</i> <i>strAB</i>	<i>bla</i> TEM-1B <i>bla</i> OXA-65 <i>bla</i> ADC-25	—	—	<i>sul2</i>	—	—
HC9436		<i>bla</i> ADC-25 <i>bla</i> OXA-180	—	—	—	—	—
MCR6056	<i>strAB</i>	<i>bla</i> ADC-25 <i>bla</i> OXA-65 <i>bla</i> OXA-23	—	—	<i>sul2</i>	—	—
MCR8676		<i>bla</i> ADC-25 <i>bla</i> OXA-65	—	—	—	—	—
MCR8683	<i>aadA1</i> <i>aph(3')-Ic</i>	<i>bla</i> ADC-25 <i>bla</i> OXA-180	—	<i>catA1</i> <i>floR</i>	<i>sul2</i>	<i>tetG</i>	<i>dfrA1</i>
MCR9238	<i>strAB</i> <i>aac(3)-IIa</i>	<i>bla</i> OXA-68 <i>bla</i> ADC-25 <i>bla</i> OXA-180	—	—	<i>sul2</i>	<i>tetB</i>	—
MCR10126		<i>bla</i> ADC-25 <i>bla</i> OXA-68	<i>msrE</i> <i>mphE</i>	—	—	—	—
MCR10172		<i>bla</i> ADC-25 <i>bla</i> OXA-68	—	—	—	—	—
MCR10179		<i>bla</i> NDM-1 <i>bla</i> ADC-25 <i>bla</i> OXA-23	—	—	—	<i>tetB</i>	—
IHSS3526	<i>aadA1</i> <i>aac(3)-Ia</i> <i>aph(3')-Via</i>	<i>bla</i> OXA-64 <i>bla</i> OXA-69 <i>bla</i> ADC-25	—	<i>catA1</i>	<i>sul1</i>	<i>tetA</i>	—
HEU55808	<i>strAB</i>	<i>bla</i> CTX-M-15 <i>bla</i> OXA-65 <i>bla</i> ADC-25	—	—	<i>sul2</i>	—	—
HEU3		<i>bla</i> OXA-23 <i>bla</i> OXA-64	—	—	—	—	—
HEU5	<i>strAB</i>	<i>bla</i> ADC-25 <i>bla</i> OXA-23 <i>bla</i> OXA-65	—	—	<i>sul2</i>	—	—

Discussion

Given that the Honduran isolates were scattered over the phylogeny, the considerable number of STs for just 16 isolates and the fact that almost every Honduran isolate had its own plasmid profile, we conclude that the Honduran isolates are highly diverse and that they might have had multiple origins.

These observations suggest the existence of potential new ICs preferentially circulating in North and South America. Furthermore, the five isolates with new allele combinations could be Honduran endemic clones, but more data are needed to corroborate this, nonetheless. Importantly, all isolates studied here were MDR and even XDR, and contain a wide variety of acquired antibiotic-resistant genes; a worrisome situation for the clinicians and infection control teams of the hospitals sampled—if one considers that these

isolates were collected regardless of antibiotic resistance profiles. Furthermore, this also highlights the urgent need to reinforce all measures required to detect, destroy, and prevent the further spread of these lineages. On that account, these Honduran isolates not only possess a high genetic diversity but also show MDR and XDR phenotypes.

In summary, these Honduran isolates show an untapped source of genetic diversity of this pathogen and, worryingly, some of them are MDR and XDR. All in all, this study highlights the urgent need to study isolates of this pathogen from other understudied geographic areas, as there may yet be a considerable amount of unexplored genetic diversity.

Acknowledgments

We thank Consuelo Velázquez-Acosta from the Instituto Nacional de Cancerología for helping us to determine

antimicrobial susceptibility tests. We are also thankful to all people who collected the isolates and their associated clinical data: Carlos Eduardo Brizio, Eduardo José Borjas, Carmen Elisa Chinchilla López, Javier Villegas, Iliana Tinoco, Giovanna Pazzetti, Sthephany Hernández, Said Leiva, Marlen Servellon, Ledy Brizzio, José Samaras, Dr. Gonzalo Pineda, Carmen Lean, Ana Valle, Isis Padilla, Yolanda Gonzáles Dr. Belinda Zapata, and Claudio Galo. This work was partially supported by “Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT” (grant numbers: IN200318 and IA201317).

Disclosure Statement

No competing financial interests exist.

References

- Coelho-Souza, T., J.N. Reis, N. Martins, I.S. Martins, A.O. Menezes, M.G. Reis, N.O. Silva, R.C. Dias, L.W. Riley, and B.M. Moreira. 2013. Longitudinal surveillance for meningitis by *Acinetobacter* in a large urban setting in Brazil. *Clin. Microbiol. Infect.* 19:E241–E244.
- Grosso, F., K.R. Carvalho, S. Quintera, A. Ramos, A.P. Carvalho-Assef, M.D. Asensi, and L. Peixe. 2011. OXA-23-producing *Acinetobacter baumannii*: a new hotspot of diversity in Rio de Janeiro? *J. Antimicrob. Chemother.* 66: 62–65.
- Stietz, M.S., M.S. Ramirez, E. Vilacoba, A.K. Merckier, A.S. Limansky, D. Centron, and M. Catalano. 2013. *Acinetobacter baumannii* extensively drug resistant lineages in Buenos Aires hospitals differ from the international clones I–III. *Infect. Genet. Evol.* 14:294–301.
- Gonzalez-Villoria, A.M., E. Tamayo-Legorreta, U. Garza-Ramos, H. Barrios, A. Sanchez-Perez, N. Rodriguez-Medina, N. Uribe-Avina, M.A. Cevallos, C.S. Group, and J. Silva-Sanchez. 2016. A multicenter study in Mexico finds *Acinetobacter baumannii* clinical isolates belonging to clonal complexes 636B (113B) and 92B harboring OXA-72, OXA-239, and OXA-469. *Antimicrob. Agents Chemother.* 60:2587–2588.
- Castro-Jaimes, S., A.D. Salgado-Camargo, L. Grana-Miraglia, L. Lozano, P. Bocanegra-Ibarias, P. Volkow-Fernandez, J. Silva-Sanchez, S. Castillo-Ramirez, and M.A. Cevallos. 2016. Complete genome sequence of a multidrug-resistant *Acinetobacter baumannii* isolate obtained from a Mexican hospital (sequence type 422). *Genome Announc.* 4:e00583-16.
- Grana-Miraglia, L., L. Lozano, S. Castro-Jaimes, M.A. Cevallos, P. Volkow, and S. Castillo-Ramirez. 2016. First genome sequence of a Mexican multidrug-resistant *Acinetobacter baumannii* isolate. *Genome Announc.* 4:e00156-16.
- Grana-Miraglia, L., L.F. Lozano, C. Velazquez, P. Volkow-Fernandez, A. Perez-Oseguera, M.A. Cevallos, and S. Castillo-Ramirez. 2017. Rapid gene turnover as a significant source of genetic variation in a recently seeded population of a healthcare-associated pathogen. *Front. Microbiol.* 8:1817.
- Magiorakos, A.P., A. Srinivasan, R.B. Carey, Y. Carmeli, M.E. Falagas, C.G. Giske, S. Harbarth, J.F. Hindler, G. Kahlmeter, B. Olsson-Liljequist, D.L. Paterson, L.B. Rice, J. Stelling, M.J. Struelens, A. Vatopoulos, J.T. Weber, and D.L. Monnet. 2012. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* 18:268–281.
- Hynes, M.F., and N.F. McGregor. 1990. Two plasmids other than the nodulation plasmid are necessary for formation of nitrogen-fixing nodules by *Rhizobium leguminosarum*. *Mol. Microbiol.* 4:567–574.
- Jackman, S.D., B.P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo, S.A. Hammond, G. Jahesh, H. Khan, L. Coombe, R.L. Warren, and I. Birol. 2017. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* 27:768–777.
- Zerbino, D.R., and E. Birney. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–829.
- Francisco, A.P., M. Bugalho, M. Ramirez, and J.A. Carrico. 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* 10:152.
- Zankari, E., H. Hasman, S. Cosentino, M. Vestergaard, S. Rasmussen, O. Lund, F.M. Aarestrup, and M.V. Larsen. 2012. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67:2640–2644.
- Fouts, D.E., L. Brinkac, E. Beck, J. Inman, and G. Sutton. 2012. PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. *Nucleic Acids Res.* 40:e172.
- Edgar, R.C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
- Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16:276–277.
- Bruen, T.C., H. Philippe, and D. Bryant. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172:2665–2681.
- Darriba, D., G.L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772.
- Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33:1870–1874.
- Bartual, S.G., H. Seifert, C. Hippler, M.A. Luzon, H. Wisplinghoff, and F. Rodriguez-Valera. 2005. Development of a multilocus sequence typing scheme for characterization of clinical isolates of *Acinetobacter baumannii*. *J. Clin. Microbiol.* 43:4382–4390.
- Loewen, P.C., Y. Alsaadi, D. Fernando, and A. Kumar. 2014. Genome sequence of an extremely drug-resistant clinical isolate of *Acinetobacter baumannii* strain AB030. *Genome Announc.* 2:e01035-14.

Address correspondence to:

Miguel A. Cevallos, PhD
 Programa de Genómica Evolutiva
 Centro de Ciencias Genómicas
 Universidad Nacional Autónoma de México
 A.P. 565-A
 Cuernavaca 62210
 México

E-mail: mac@ccg.unam.mx

Complete Genome Sequence of a Multidrug-Resistant *Acinetobacter baumannii* Isolate Obtained from a Mexican Hospital (Sequence Type 422)

Semiramis Castro-Jaimes,^a Abraham David Salgado-Camargo,^a Lucía Graña-Miraglia,^a Luis Lozano,^a Paola Bocanegra-Ibarias,^d Patricia Volkow-Fernández,^b Jesus Silva-Sanchez,^c Santiago Castillo-Ramírez,^a Miguel A. Cevallos^a

Centro de Ciencias Genómicas, Programa de Genómica Evolutiva, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico^a; Departamento de Enfermedades Infecciosas, Laboratorio de Microbiología, Instituto Nacional de Cancerología, Secretaría de Salud (SSA), Cuernavaca, Morelos, Mexico^b; Departamento de Diagnóstico Epidemiológico, Centro de Investigaciones sobre Enfermedades Infecciosas, Instituto Nacional de Salud Pública, Cuernavaca, Morelos, Mexico^c; Hospital Universitario de Nuevo León, Monterrey, Nuevo León, Mexico^d

Acinetobacter baumannii has emerged as a dangerous nosocomial pathogen, particularly for severely ill patients in intensive care units and patients with hematologic malignancies. Here, we present the complete genome sequence of a multidrug-resistant *A. baumannii* isolate, recovered from a Mexican hospital and classified as sequence type 422 according to the multilocus sequence typing Pasteur scheme.

Received 5 May 2016 Accepted 10 May 2016 Published 23 June 2016

Citation Castro-Jaimes S, Salgado-Camargo AD, Graña-Miraglia L, Lozano L, Bocanegra-Ibarias P, Volkow-Fernández P, Silva-Sanchez J, Castillo-Ramírez S, Cevallos MA. 2016. Complete genome sequence of a multidrug-resistant *Acinetobacter baumannii* isolate obtained from a Mexican hospital (sequence type 422). *Genome Announc* 4(3):e00583-16. doi:10.1128/genomeA.00583-16.

Copyright © 2016 Castro-Jaimes et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Miguel A. Cevallos, mac@ccg.unam.mx.

Acinetobacter baumannii is a Gram-negative bacterium that has emerged as a dangerous pathogen worldwide, causing a variety of nosocomial infections, especially in severely ill patients in intensive care units and patients with hematologic malignancies on chemotherapy (1). *A. baumannii* isolates are resistant to most of the currently used antibiotics, including carbapenems, and they are able to survive for extended periods of time under a broad range of environmental conditions, such as dry surfaces. These characteristics promote *A. baumannii* outbreaks in the hospital setting. Most of the isolates causing outbreaks worldwide belong to the international clonal lineages I and II (2), clonal complexes 1 and 2, respectively, as defined by the multilocus sequence typing (MLST) Pasteur scheme (3).

Here, we present the complete genome sequence and annotation of *A. baumannii* 3207, an isolate not related to the international clonal lineages but instead classified as sequence type 422 according to the MLST Pasteur scheme (3). This multidrug-resistant isolate was recovered from bronchoalveolar lavage fluid of a 45-year-old female patient admitted to the Hospital Universitario de Nuevo León, Mexico, a tertiary care center, in August 2008.

The complete genome of *A. baumannii* 3207 was determined on a PacBio RSII platform. To do this, a large insert library (20 to 25 kb) was constructed and sequenced, using one SMRT cell with a P6 polymerase and C4 chemistry combination (P6-C4) with a 180-min movie. The SMRT cells produced 551,632,314 post-filter polymerase reads with a mean read length of 10,718 bp. Subreads were assembled *de novo* using the PacBio RS hierarchical genome assembly process (HGAP) protocol version 3 in SMRT analysis version 2.3 (Pacific Biosciences) (4). To improve regions of low coverage, a library with a 2 × 300-bp paired-end configuration

was sequenced on an Illumina MiSeq platform. The sequencing yielded 3,117,310 reads, resulting in 909,042,928 bp. A hybrid assembly was constructed with the Illumina MiSeq reads and PacBio RII subreads using SPAdes version 3.5.0 (5). The contigs corresponding to the chromosome and the plasmids were circularized using a Perl script available at <https://github.com/jfass/apc>. Functional annotation was performed with the NCBI Prokaryotic Genome Annotation Pipeline.

The *A. baumannii* 3207 genome consists of a circular chromosome (3,998,013 bp) and two plasmids: pAba3207a (13,479 bp) and pAba3207b (80,547 bp). The chromosome has six rRNA operons, 74 tRNAs, and 3,674 coding sequences (CDSs) and embraces most of the antibiotic resistance genes, conferring resistance to aminoglycosides, beta-lactams, carbapenems (blaOXA-65), cephalosporins, fosmidomycin, fusaric acid, phenicols, and tetracyclines. The smallest plasmid has 19 CDSs, including one carbapenem resistance gene (blaOXA-58), whereas pAba3207b has 97 CDSs.

This and a previous report (6) are the first steps toward understanding the genomic diversity of *A. baumannii* in Mexico, not only at some loci but at the genome level. However, to get a more comprehensive picture, comparative analyses of many more isolates are warranted.

Nucleotide sequence accession numbers. The genome sequence of isolate 3207 was deposited in GenBank under the accession numbers CP015364, CP015365, and CP015366.

ACKNOWLEDGMENTS

PacBio sequencing was performed at the Duke GCB Genome Sequencing Shared Resource (Durham, NC, USA). Illumina sequencing was conducted at MacroGen, (Seoul, South Korea).

This work was supported by the Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica PAPIIT, grant number IA200515. A.D.S.-C. and L.G.-M. are doctoral students from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), and received fellowships from CONACYT.

FUNDING INFORMATION

This research was funded by the Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica, PAPIIT (UNAM).

REFERENCES

1. Antunes LC, Visca P, Towner KJ. 2014. *Acinetobacter baumannii*: evolution of a global pathogen. *Pathog Dis* 71:292–301. <http://dx.doi.org/10.1111/2049-632X.12125>.
2. Zarrilli R, Pournaras S, Giannouli M, Tsakris A. 2013. Global evolution of multidrug-resistant *Acinetobacter baumannii* clonal lineages. *Int J Antimicrob Agents* 41:11–19. <http://dx.doi.org/10.1016/j.ijantimicag.2012.09.008>.
3. Diancourt L, Passet V, Nemeč A, Dijkshoorn L, Brisse S. 2010. The population structure of *Acinetobacter baumannii*: expanding multiresistant clones from an ancestral susceptible genetic pool. *PLoS One* 5:e10034. <http://dx.doi.org/10.1371/journal.pone.0010034>.
4. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <http://dx.doi.org/10.1038/nmeth.2474>.
5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
6. Graña-Miraglia L, Lozano L, Castro-Jaimes S, Cevallos MA, Volkow P, Castillo-Ramírez S. 2016. First genome sequence of a Mexican multidrug-resistant *Acinetobacter baumannii* isolate. *Genome Announc* 4(3):e00156-16. <http://dx.doi.org/10.1128/genomeA.00156-16>.