Epidemiología molecular y genómica de aislados resistentes de *Pseudomonas aeruginosa* de origen hospitalario

Trabajo realizado en la Unidad Mixta Infección y Salud Pública - FISABIO/Univ. Valencia por

Paula Ruiz Hueso

Para optar al grado de Doctora por la Universitat de València

Director:

Fernando González Candelas

Febrero, 2019

Programa de Doctorado en Biomedicina y Biotecnología, código 3102, regulado por el Real Decreto 99/2011. Facultat Ciències Biològiques.

Universitat de València, 2019.

Tesis:

Epidemiología molecular y genómica de aislados resistentes de *Pseudomonas aeruginosa* de origen hospitalario.

Autora:

Paula Ruiz Hueso

Director:

Fernando González Candelas

La investigación desarrollada en este trabajo se realizó en la Unidad Mixta Infección y Salud Pública entre la Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO) y la Universitat de València, en colaboración con el Hospital General Universitario de Valencia, el Hospital General Universitario de Elche y el Hospital Arnau de Vilanova. El proyecto ha sido financiado por el Ministerio de Educación, Cultura y Deporte (MECD), proyectos BFU2014-58656R y BFU2017-89594-R, y la Conselleria de Educación y Cultura de la Generalitat Valenciana, PROMETEO2016-0122.

D. Fernando González Candelas, Doctor en Ciencias Biológicas y Catedrático del Departament de Genètica de la Universitat de València.

CERTIFICA:

Que Dª Paula Ruiz Hueso, Licenciada en Biotecnología por la Universidad Politécnica de Valencia, ha realizado bajo mi dirección el trabajo que lleva por título: "Epidemiología molecular y genómica de aislados resistentes de *Pseudomonas aeruginosa* de origen hospitalario", para optar al Grado de Doctora por la Universitat de València.

Y para que conste, en el cumplimiento de la legislación vigente, firmo el presente certificado en Valencia, a 18 de febrero de 2019.

Fdo.: Dr. Fernando González Candelas.

A mis padres

La ciencia es el fundamento de todo progreso,
que mejora la vida humana y alivia el sufrimiento.

Irène Joliot-Curie

Agradecimientos

En primer lugar, quiero dar las gracias a Fernando por darme la oportunidad de seguir investigando en un área que me gusta tanto y por confiar en mí. He aprendido muchísimo durante todos estos años y me siento muy afortunada de así haya sido.

Gracias a las microbiólogas de los 3 hospitales que han participado: Nuria, Concha, Vicki, Melani y Victoria. Sin vuestra disposición y recopilación de muestras e información esto habría sido imposible de realizar.

Gracias a todos mis compañeros del grupo de Epidemiología y Salud. En especial, a Carlos por ser un apoyo fundamental y porque gracias a él tengo una amiga más a la que adoro y con la que la etapa final se ha hecho mucho más llevadera (Lidia, esto va por ti). También a mis "patitos": Neris, Bea y Martha. Ya hace tiempo que dejaron de serlo, aunque para mí en el fondo siempre lo serán. Fue fantástico acompañaros en vuestros primeros pasos dentro del laboratorio y después teneros como compañeras, estoy muy orgullosa de vosotras. Allá donde vaya sabéis que tenéis un hueco.

Gracias a mis compañeros del área de genómica: doctorandos, alumnos en prácticas, de TFG y máster, técnicos, jefes de grupo, en definitiva, todos aquellos que han formado parte del día a día y con los que he compartido risas y cafés.

Gracias a Leo, Simon, y todos los compañeros del grupo de Julian Parkhill en el Sanger por acogerme durante los 3 meses que estuve en Cambridge y ayudarme a clarificar conceptos en la parte genómica, han sido vitales.

Gracias también a toda la gente que en algún momento ha tenido que sufrir mi mal humor, estrés y aquellos que me han dado mucho ánimo para sobrellevarlo: mi familia y todos mis amigos: los de poblets, los chuferos, los de Verano Azul, mis chicos del tranvía y compañía, los biotecnólogos, los desgranadores y toda la gente que ha ido sumándose en el camino.

Gracias a mis padres. Definitivamente este libro nunca habría sido escrito sin su apoyo y ayuda incondicional, por eso, y sin pretender quitarle mérito a tantas otras personas que han colaborado en este proceso, quisiera dedicarles todo lo que está aquí escrito. Gracias por haberme hecho como soy y por enseñarme a lucharlo todo hasta el

final. Nunca serán suficientes palabras para agradeceros todo cuanto habéis hecho y seguís haciendo. Os quiero.

Por último, y no menos importante, gracias a José. Desde el principio sabías que mi profesión tenía una serie de "inconvenientes" y te agradezco que siempre me hayas apoyado, anteponiendo lo que era bueno para mí. Ahora hay que plantearse nuevos retos, nuevos lugares dónde seguir investigando, pero lo que tengo claro es que será más fácil contigo. Gracias por soportarme, nadie ha sufrido esta tesis más que tú (aparte de mí…) y aun así aquí estás, ya tiene mérito. Te quiero.

Índice

Abreviaturas

ADN - Ácido Desoxirribonucleico

ARN - Ácido Ribonucleico

ASC - *Ascertainment Bias Correction*

BAM - *Binary Alignment Map*

BHI - *Brain Heart Infusion*

BLAST - *Basic Local Alignment Search Tool*

BWA - *Burrows-Wheeler Aligner*

CARD - *Comprehensive Antibiotic Resistance Database*

CMI - Concentración Mínima Inhibitoria

crARN - Ácido Ribonucleico procedente de la estructura CRISPR

CRISPR - *Clustered Regularly Interspaced Short Palindromic Repeats*

FQ – Fibrosis Quística

GC - Guanina-Citosina

GFF – *General Feature Format*

GTR – *General Time Reversible*

HAV - Hospital Arnau de Vilanova

HGUE - Hospital General Universitario de Elche

HGUV - Hospital General Universitario de Valencia

HPD. - *High Probability Density*

Kb - kilobases

LMA - Leucemia Mieloide Aguda

Mb - Megabases

MCMC - *Markov Chain-MonteCarlo*

MM - Mieloma Múltiple

MLST - *Multilocus Sequence Typing*

NCBI - *The National Center for Biotechnology Information*

PAM - *Proto-spacer Adyacent Motiv*

pb – pares de bases

PBS - *Phosphate Buffered Saline*

PCR - *Polymerase Chain Reaction*

PFGE - *Pulsed-field Gel Electrophoresis*

SAM - *Sequence Alignment Map*

SNP – *Single Nuleotide Polymorphism*

s/s/a – sustituciones/sitio/año

ST - *Sequence Type*

TVM – *Transversion Model*

VCF - *Variant Call Format*

Figuras

Tablas

Resumen

Pseudomonas aeruginosa es un patógeno frecuente en entornos hospitalarios que acumula multitud de resistencias a antibióticos. Ante la dificultad de encontrar tratamiento efectivo, los pacientes están mucho tiempo colonizados y aumenta la probabilidad de transmisión.

El principal objetivo del presente trabajo ha sido mostrar cómo el estudio evolutivo a nivel de genomas completos puede ser útil para la detección de transmisiones a diferentes niveles. Se decidió aplicar la misma metodología de reconstrucción de la filogenia a partir de las variantes encontradas en genomas completos que además se dataron por métodos bayesianos. Para ello, hemos contado con muestras de *P. aeruginosa* multirresistentes de 3 hospitales distintos, planteándose situaciones de características muy diferentes: un brote de alcance limitado, un extenso brote con posible dispersión a otras áreas y, por último, estudio retrospectivo de evolución intrapaciente con infecciones recurrentes o de larga duración.

Los resultados muestran cómo la capacidad de resolución mediante este abordaje es muy superior a las de otros métodos, permitiendo confirmar brotes, encontrar transmisiones ajenas a un brote e, incluso, posibles transmisiones en la comunidad a partir del análisis conjunto a mayor escala, definiendo una complejidad de la situación mucho mayor que la visualizada con métodos de rutina.

Adicionalmente, se han incorporado resultados relativos a la resistencia fenotípica frente a la detección genotípica y se ha analizado la estructura CRISPR en este conjunto de muestras, cuyos resultados sugieren que puede ser una alternativa interesante a la utilización de la secuenciación masiva con la combinación CRISPR-MLST.

1. Introducción

1.1 Características de *P. aeruginosa*

1.1.1 Biología básica

Pseudomonas aeruginosa es una bacteria Gram-negativa que pertenece a la clase de las Gamma-Proteobacterias. Este bacilo flagelado es capaz de desplazarse a gran velocidad y sobrevivir en un estado planctónico o formar *biofilms* (Skariyachan *et al.*, 2018). Su control epidemiológico es relevante, ya que se trata de un patógeno oportunista capaz de sobrevivir en todo tipo de medios: agua, suelos, en la superficie de plantas y animales, incluso en medios más difíciles como hidrocarburos, medios con elevada concentración salina, agua destilada o sustancias desinfectantes, soportando, además, temperaturas de hasta 42ºC (Pier. y Ramphal, 2005).

A diferencia de otras Gram-negativas, como las bacterias entéricas crece en aerobiosis. Únicamente puede crecer en un entorno anaeróbico si dispone de alguna fuente de nitrato que actuará como aceptor de electrones. Esta capacidad de adaptación a entornos con menor disponibilidad de oxígeno facilita su infección en determinadas partes del cuerpo de oxigenación más complicada, combinada con su versatilidad metabólica, que le lleva a poder utilizar diferentes compuestos orgánicos para su supervivencia y readaptar su metabolismo cuando así lo requiere (Cornelis y Dingemans, 2013; La Rosa, Johansen y Molin, 2018).

Dada su naturaleza oportunista, se producen infecciones cuando la persona padece una patología de base que conlleva una disminución de su respuesta inmunitaria. Esto puede verse favorecido por la presencia de heridas abiertas o la utilización de dispositivos médicos que faciliten la entrada de la bacteria a modo de vehículo (Hoang *et al.*, 2018). La variedad de infecciones que puede causar es amplia: endocarditis, infección de vías respiratorias, infecciones del sistema nervioso central, infecciones oculares, óseas, articulares, urinarias, gastrointestinales, en piel y tejidos blandos, entre otras. Un aspecto que la convierte en una bacteria especialmente peligrosa es que no solo produce infecciones en la localización en que se ha introducido de forma directa, como puede ocurrir con los cateterismos, sino que es capaz de entrar en el torrente sanguíneo y diseminarse a otras zonas a partir de la infección primaria. Posee mecanismos en cada estadio del proceso infeccioso (colonización, invasión local y diseminación sistémica) que favorecen su inmunidad y su fácil dispersión, como la liberación de enzimas y toxinas

que forman poros en las células eucariotas o que disgregan la estructura del tejido, y la adhesión a las células epiteliales mediante *pili* (Heiniger *et al.*, 2010).

Su persistencia puede verse favorecida por la aparición de fenotipos mucoides que conforman los *biofilms* (Costerton, Stewart y Greenberg, 1999; Høiby *et al.*, 2010). Dicho fenotipo está caracterizado por la producción del exopolisacárido alginato, formando una película protectora alrededor de las colonias que dificulta la actuación del sistema inmunitario y el acceso de los antibióticos, lo que supone un problema añadido para los pacientes con fibrosis quística (FQ). Esta enfermedad rara de herencia recesiva producida por mutaciones en el gen *cftr* (Kreda, Davis y Rose, 2012), lleva asociada una función pulmonar deficiente con acumulación de mucosidad que facilita el crecimiento de bacterias como *P. aeruginosa*. A la larga, las infecciones en pacientes con FQ cronifican, siendo más complicado el tratamiento ante la adaptación de esta bacteria al entorno pulmonar del enfermo (Marvig *et al.*, 2014; Sousa y Pereira, 2014).

1.1.2 Genoma

Esta bacteria fue secuenciada por primera vez en el año 2000 (Olson *et al.*, 2000), mediante la técnica de *shotgun*, y para ello seleccionaron la cepa PAO1, utilizada habitualmente como referencia en el laboratorio. Su genoma circular, con un tamaño de 6,3 Mb (Figura 1.1), fue en ese momento el genoma bacteriano más grande secuenciado. Actualmente se conoce que existen bacterias con genomas de mayor tamaño , llegando a los 14,78 Mb de *Sorangium cellulosum* cepa So0157-2 (Han *et al.*, 2013) o los 16 Mb de *Minicystis rosea* cepa DSM 24000 (Garcia, Gemperlein y Muller, 2014); sin embargo, no es lo habitual entre las bacterias patógenas más estudiadas. A partir de la aparición de nuevas tecnologías de secuenciación que han reducido enormemente los costes en los últimos años, este tipo de estudios es más accesible, por lo que el número de genomas de *P. aeruginosa* en las bases de datos va en aumento (más de 3300 depositados en GenBank a 16/11/2018). Con ello, se ha visto que hay una elevada variabilidad en el tamaño genómico dentro de esta especie, pudiendo variar 1,5-2 Mb entre dos cepas distintas y alcanzando tamaños superiores a 7 Mb, por lo que se ha tratado de determinar el tamaño del pangenoma de esta especie.

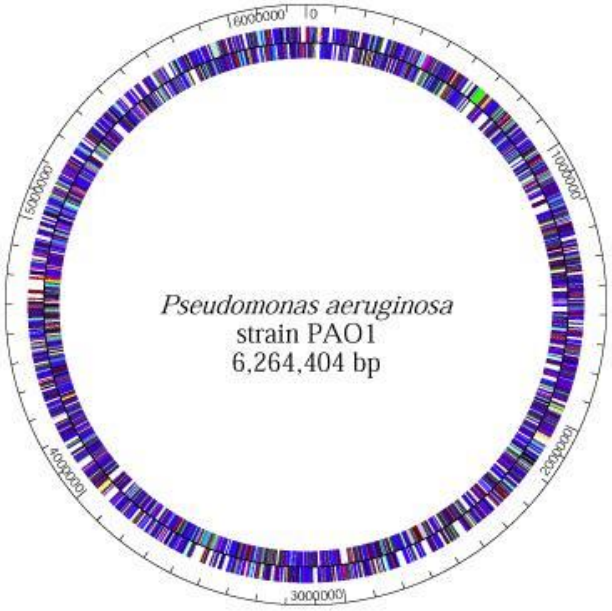


Figura 1.1. Representación del genoma completo de la cepa de referencia PAO1 de la Pseudomonas Genome Database. *Las bandas corresponden a genes, ordenados por dirección de transcripción (exterior: cadena +, interior: cadena -) y coloreados según su función (Winsor et al., 2016).*

El pangenoma es el conjunto de todos los genes de la especie, diferenciándose en genoma "*core*", aquellos genes presentes en todas las cepas, y genoma accesorio, en el que están aquellos que son compartidos por varias cepas. Además, también en el accesorio se definen los específicos de cepa (Rouli *et al.*, 2015). Los genes que forman parte del "*core*" suelen ser genes constitutivos, implicados en funciones vitales para la bacteria; en cambio, en el genoma accesorio se encuentran genes relacionados con transferencia y funciones que les permiten adaptarse al medio, como genes de virulencia, resistencia a antibióticos o relacionados con el metabolismo secundario, además de un alto número de genes con función desconocida (Mosquera-Rendón *et al.*, 2016).

Se ha estimado el tamaño aproximado que tendría el pangenoma de *P. aeruginosa* y, por el momento, no hay un consenso, al encontrarse en constante ampliación. En un estudio de 2015 (Hilker *et al.*, 2015), se estimó, a partir de 20 cepas representativas de los principales complejos clonales, que el genoma "*core*" de la especie constaría de alrededor de 4000 genes, mientras que las estimas del genoma accesorio superarían los 10000 genes. Esto contrasta con un estudio posterior (Mosquera-Rendón *et al.*, 2016) en el que se incluyeron 180 cepas disponibles en la base de datos PATRIC hasta abril de 2014, en el que se determinó que el "*core*" contiene unos 2400 genes y el accesorio 16000 aproximadamente, lo que refleja el alto grado de variabilidad que presenta esta bacteria.

Además de esta amplia variabilidad a nivel de contenido génico, se han estudiado también las tasas de recombinación (Maatallah *et al.*, 2011; Kidd *et al.*, 2012; Dettman, Rodrigue y Kassen, 2014) y de mutación (Oliver, 2010; Oliver y Mena, 2010; López-Causapé *et al.*, 2013; Feliziani *et al.*, 2014) como mecanismos que promueven la rápida adaptación de esta especie, pudiendo verse multiplicada su tasa de mutación por 100 respecto a la tasa de normomutadoras. Aunque no existe un consenso al respecto, un trabajo reciente estima que la tasa de sustitución puede variar entre $4{,}3 \cdot 10^{-6}$ y $1 \cdot 10^{-5}$, encontrando cepas del ST235 con tasas cercanas al rango de $1 \cdot 10^{-3}$ (Miyoshi-Akiyama *et al.*, 2017).

1.2 Brotes e infecciones nosocomiales de patógenos multirresistentes.

El control de las infecciones por bacterias multirresistentes y la búsqueda de nuevos antibióticos se ha convertido en un problema grave de carácter mundial. En 2017, la Organización Mundial de la Salud estableció un listado con las bacterias que constituyen un mayor riesgo para la salud pública (Knols *et al.*, 2016), colocando en la categoría de prioridad crítica a patógenos resistentes a carbapenems, entre ellos *P. aeruginosa*. Los carbapenems son antibióticos de la familia de los betalactámicos, cuyo uso es limitado al ámbito hospitalario como uno de los tratamientos de última línea. Existen genes que codifican enzimas degradadoras o modificadoras de antibióticos; por ejemplo, en el caso de los carbapenems se han reportado enzimas de tipo OXA (Evans y Amyes, 2014), VIM (Segura *et al.*, 2010; Viedma *et al.*, 2013) o NDM (Solé *et al.*, 2011), entre otras, y el mayor problema es que los genes que las codifican pueden ser transferidos horizontalmente a través de transposones, integrones o plásmidos. Esto facilita su dispersión, lo que ha llevado a un aumento en los últimos años de estas y otras resistencias (Meletis, 2016).

La elevada tasa de infecciones causadas por patógenos multirresistentes incrementa la necesidad de llevar a cabo una vigilancia exhaustiva en los hospitales (ECDC, 2014). A este respecto, el personal hospitalario involucrado (microbiólogos y médicos especializados en medicina preventiva) juega un papel fundamental tanto en el registro de casos como en la toma de medidas preventivas ante la sospecha de un brote. Los brotes se definen, según la Organización Mundial de la Salud, como "el aumento de

casos de una enfermedad en exceso respecto a lo que normalmente se esperaría en una comunidad, área geográfica o temporada definida", y estos requieren investigación. Desde los años 80 del siglo pasado, se han ido se han utilizado diferentes metodologías de tipado basadas en marcadores moleculares, dando lugar a una nueva disciplina conocida como Epidemiología Molecular, en la que se combinan los métodos de la Epidemiología clásica con la caracterización genética o molecular del patógeno.

1.2.1 Métodos de tipado

El método con el que se comenzó a distinguir diferentes cepas o agrupaciones de ellas es el tipado serológico. Está basado en la utilización de anticuerpos para discriminar la clase de antígeno somático presente en un aislado bacteriano según el nivel de aglutinación, a partir del cual se establecen los diferentes serotipos. A finales de los años 1980s se configuró un esquema internacional de tipado antigénico específico de *P. aeruginosa* (Liu *et al.*, 1987; Liu y Wang, 1990) para unificar resultados que permite discriminar hasta 20 serogrupos, aunque no siempre es posible ya que existen cepas no aglutinables o que se autoaglutinan.

Con el fin de evitar dichos problemas técnicos y lograr mayor capacidad de discriminación, se desarrollaron otro tipo de métodos de tipado como los basados en perfiles genéticos. Entre los más utilizados se encuentran la electroforesis en campo pulsado (PFGE, siglas en inglés) y la caracterización de secuencias de varios *loci* o "*Multilocus sequence typing*" (MLST). Aunque en ambas técnicas se utiliza el ADN, en la PFGE solamente se determinan las diferencias derivadas del corte de una enzima de restricción a lo largo del genoma. Estas diferencias, por tanto, no se describen a nivel de secuencia, sino que se visualizan en forma de patrón de bandas mediante electroforesis. Un riesgo posible de este método es la determinación de perfiles distintos en aislados con un origen común debido a mutaciones en los sitios de reconocimiento de la enzima que modifica el bandeado. También puede producirse la situación contraria: aislados distantes genealógicamente pueden dar patrones de bandas indistinguibles, tanto por la posibilidad de convergencia en los tamaños generados a partir de distintos puntos de corte, como por la dificultad en discriminar fragmentos grandes de tamaños parecidos.

Por otra parte, con el MLST se realiza la amplificación de fragmentos de varios genes, 7 habitualmente, que se secuencian por Sanger para determinar a qué alelo corresponde mediante comparación con los alelos que se han depositado previamente en

una base de datos. La asignación del "*Sequence type*" dependerá de la combinación de estos 7 alelos, a cada uno de los cuales se le asigna un número al compararlos con los de la base de datos. El esquema de genes es específico de especie, ya que se seleccionan entre los genes que son esenciales para cada una y cumplen los requisitos de variabilidad y capacidad discriminativa. De esta manera, se asegura la tipabilidad de todas las cepas. Gracias a la existencia de repositorios oficiales como PubMLST.org (Jolley, Bray y Maiden, 2018), se registran nuevos alelos y combinaciones de estos que permiten comparar los resultados.

El desarrollo de este tipo de metodologías ha permitido comprobar que *P. aeruginosa* es una bacteria con una estructura poblacional compleja. Los análisis realizados hasta el momento indican que la población sigue una estructura epidémica (Maatallah *et al.*, 2011; Kidd *et al.*, 2012), ya que no hay cepas predominantemente clínicas, ambientales o animales, sino que se encuentran poblaciones muy diversas independientemente del origen (Kidd *et al.*, 2012). La elevada variabilidad queda reflejada en el registro total de 3146 STs depositados en la base de datos específica de *P. aeruginosa* en PubMLST.org.Tras su agrupación en complejos clonales por eBURST y goeBURST (Figura 1.2), podemos observar que existen diferentes subpoblaciones. Estos complejos están definidos por la agrupación de STs que comparten 4 o más alelos de los 7 que conforman el esquema; aquellos conectados de manera directa solamente difieren en 1 alelo. Según el número de conexiones se determina el perfil alélico fundador o central a partir del cual habría surgido el resto. Por ejemplo, el ST175 (Figura 1.3) constituye un complejo clonal de 20 STs.

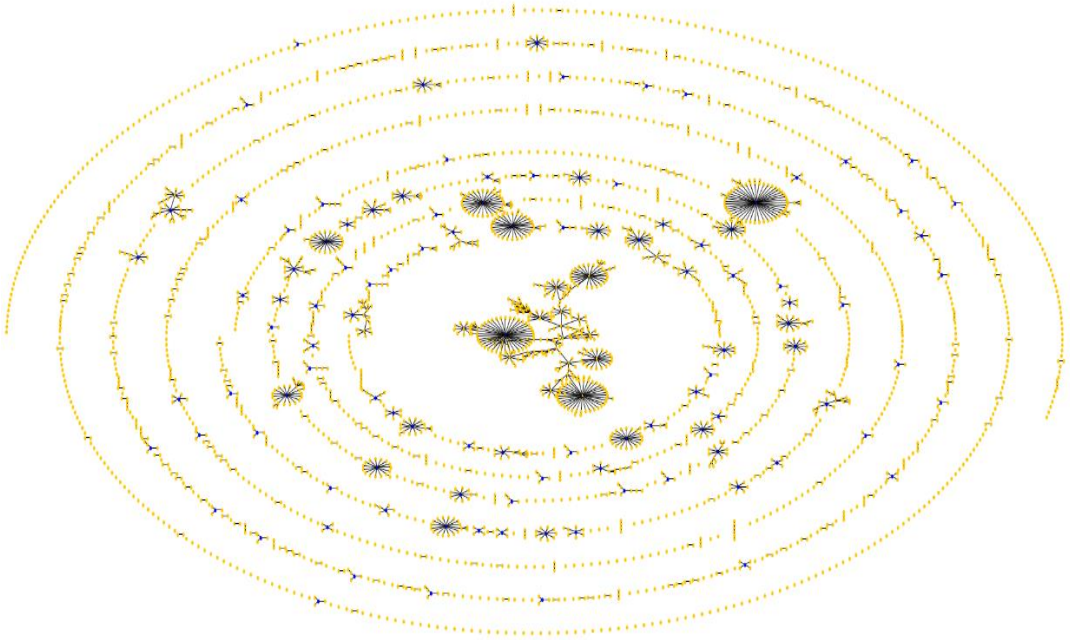


Figura 1.2. Distribución en complejos clonales de los 3146 STs descritos en la base de datos de <u>Pseudomonas aeruginosa</u>. *Gráfico obtenido por eBURST* (Feil *et al.*, 2004) *a partir de los perfiles de MLST depositados en PubMLST* (Jolley, Bray y Maiden, 2018) *a fecha 12/11/2018. Los STs resaltados en azul son los considerados fundadores de los STs con que conectan en la red.*

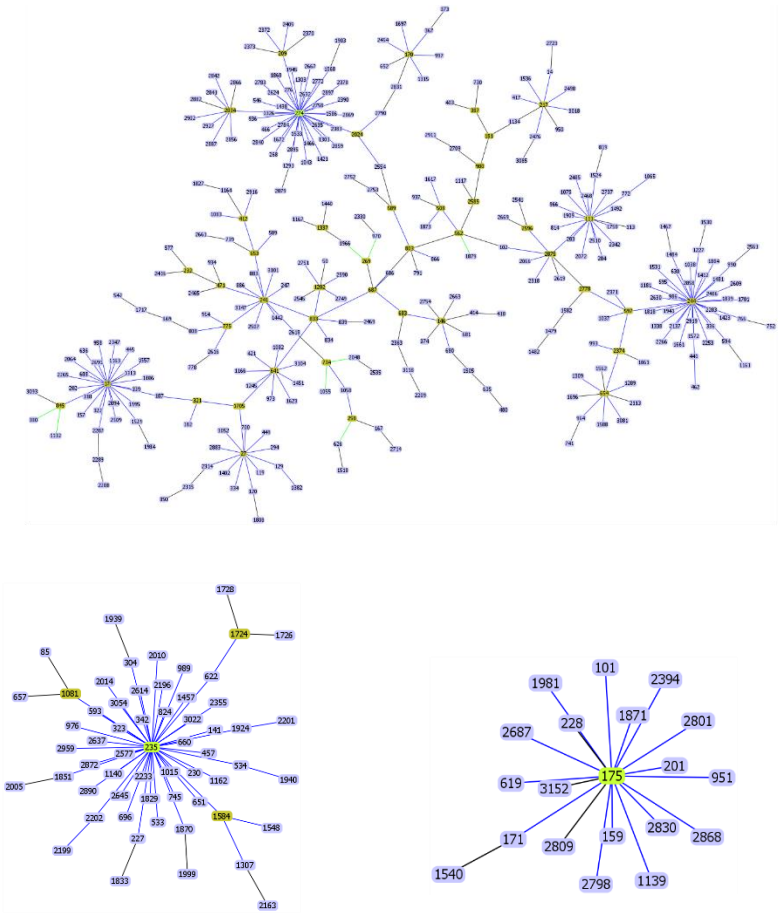


Figura 1.3. Visión detallada de complejos clonales que contienen algunos de los STs más prevalentes en <u>Pseudomonas aeruginosa</u>. *Gráfico obtenido por goeBURST* (Francisco *et al.*, 2009). *Los STs resaltados en amarillo son los considerados fundadores.*

1.2.2 Uso de genomas completos

El desarrollo de nuevas tecnologías de secuenciación y la progresiva disminución de costes ha hecho más accesible su utilización en los últimos años. Una de las plataformas más utilizadas actualmente es Illumina (Steemers y Gunderson, 2005) y sus secuenciadores MiSeq, NextSeq y HiSeq. Esta plataforma utiliza la tecnología de PCR en puente a partir del genoma fragmentado (Figura 1.4). En el proceso, a estos fragmentos se les incorporan adaptadores que contienen una región corta que actúa como índice, para distinguir cada una de las muestras, ya que en el proceso podemos incluir varias muestras simultáneamente; una región complementaria al cebador que se utiliza para la fase de secuenciación y, en el extremo, un fragmento complementario a los oligonucleótidos que revisten la celda de flujo para su anclaje. Esta PCR en puente genera agrupaciones de copias de estos fragmentos, permitiendo en la fase de

secuenciación amplificar la señal de los nucleótidos marcados con distintas fluorescencias. Todo ello se realiza sobre un soporte en el que estarán todos los fragmentos anclados, por lo que el detector irá generando una secuencia/lectura para cada "punto" según la fluorescencia detectada.

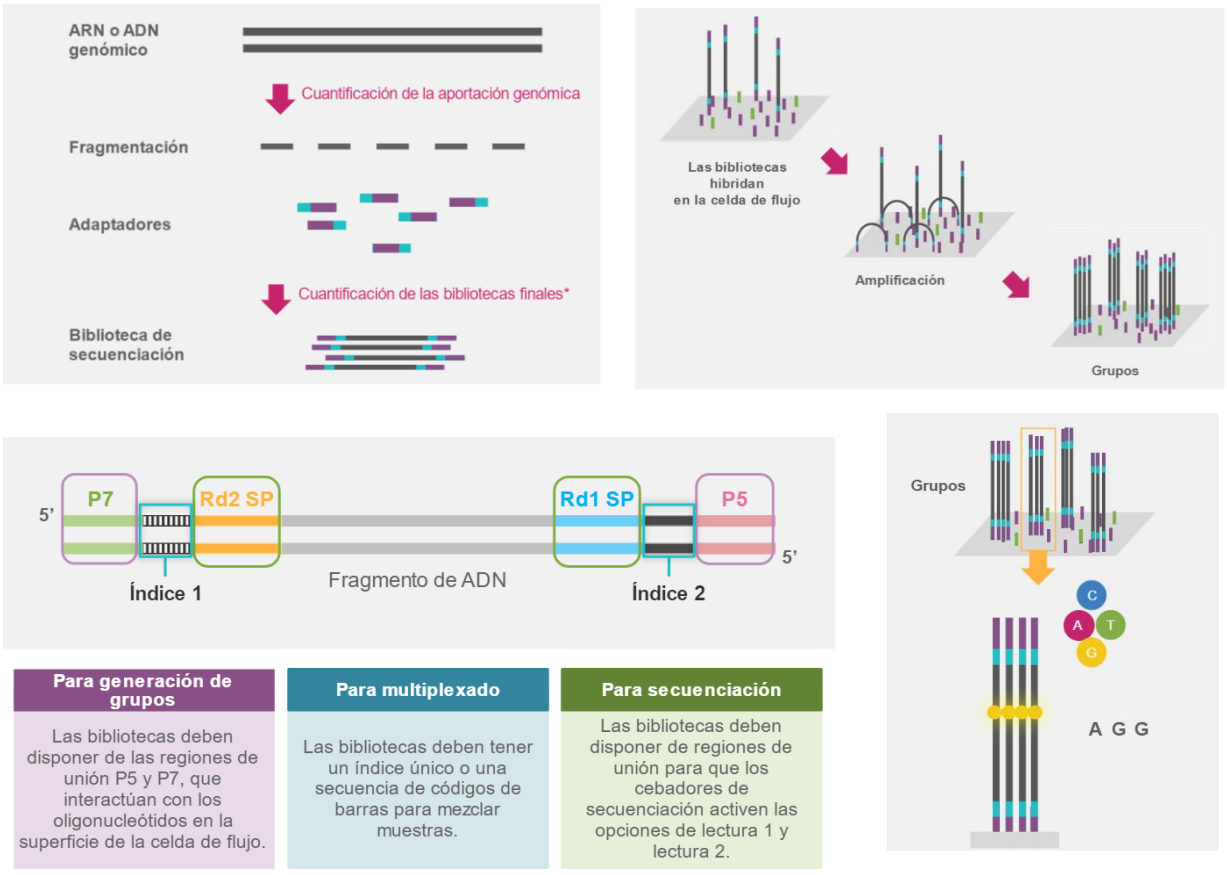


***Figura 1.4. Proceso de secuenciación por Illumina**. (www.Illumina.com)*

En las secuenciaciones de este tipo habitualmente se emplea el sistema "*paired ends*", es decir, para cada fragmento de ADN se generan lecturas "pareadas", cada una de las cuales empieza por un extremo. Esto permite cubrir regiones con repeticiones de difícil secuenciación que queden dentro de esa región, ya que no necesariamente se produce el solapamiento de ambas. La reconstrucción de genomas a partir de las lecturas puede realizarse mediante mapeo frente a una referencia (Boers *et al.*, 2014) o con ensamblado *de novo* (Bianconi *et al.*, 2016). Dependiendo del objetivo y los resultados de la secuenciación en cuanto a calidad y cobertura (número medio de lecturas que cubren una posición) podemos decantarnos por un método u otro. Con coberturas bajas funcionará mejor la reconstrucción a partir de mapeo, pero en genomas tan variables

como el de *P. aeruginosa* existe el riesgo de perder información de regiones pertenecientes al genoma accesorio, por lo que será importante en este caso elegir una referencia lo más cercana posible.

El resultado del mapeo será un genoma consenso que contendrá las variantes encontradas en las lecturas respecto a la referencia utilizada como molde, mientras que con el ensamblaje el resultado final será, idealmente, un genoma completo con toda la información contenida en las lecturas; sin embargo, habitualmente los genomas no consiguen cerrarse sin utilizar otra secuenciación con lecturas más largas, o *long reads*, que permitan unir las secuencias consenso ensambladas, denominadas *contigs* (Madoui *et al.*, 2015). Por tanto, la codificación de posiciones indeterminadas (N) puede deberse a causas distintas: discrepancias en variantes o bajo número de lecturas que impidan determinar la posición en mapeo, o la unión con N con el fin de ordenar los *contigs* en una estructura única denominada *scaffold* (Figura 1.5).

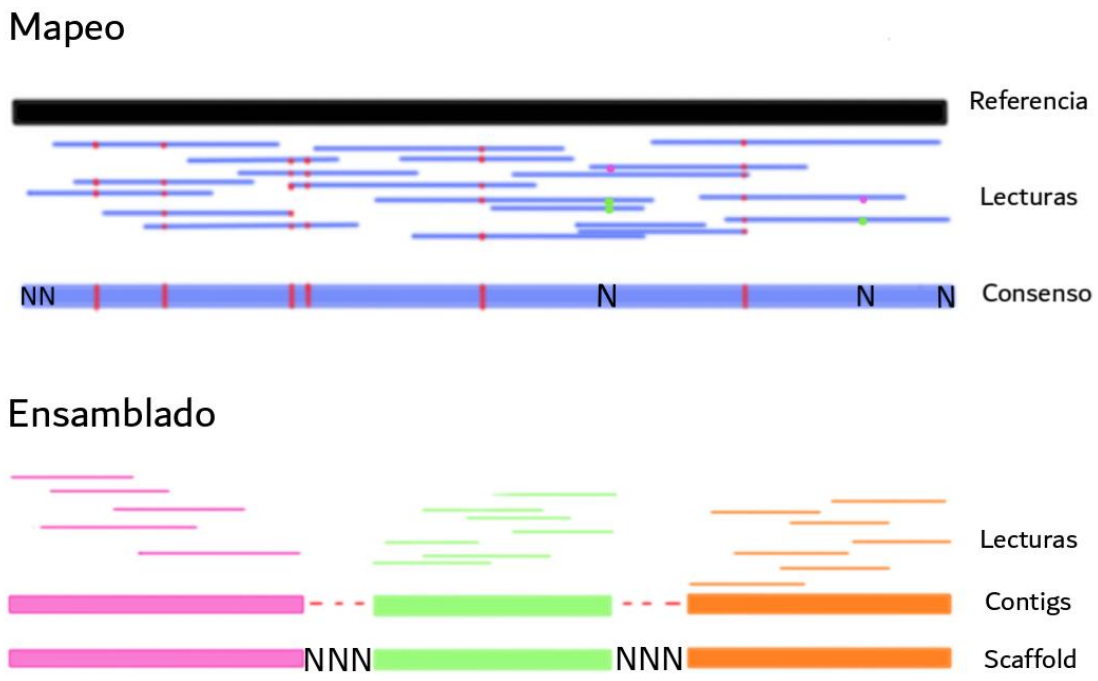


Figura 1.5. Representación gráfica del resultado obtenido con la utilización de técnicas de mapeo o ensamblado. *Mediante mapeo se alinean todas las lecturas respecto a un genoma de referencia, por lo que se detectarán los cambios nucleotídicos e INDELs que pueda haber respecto a esa secuencia. En el ensamblado, por el contrario, se realiza un concatenado de las lecturas solapantes con el fin de reconstruir la secuencia de ADN original, por lo que se podrán obtener regiones como islas genómicas o plásmidos que no serían detectables con mapeo.*

Esta tecnología tiene un poder de discriminación muy superior a la del tipado por MLST (Boers *et al.*, 2014), limitado únicamente a 7 genes frente a los más de 6000 que puede tener un aislado de esta bacteria.

1.2.3 La evolución aplicada al estudio de patógenos en el ámbito clínico

La filogenética es una parte de la biología evolutiva que permite reconstruir las relaciones evolutivas entre diferentes especies. En el caso de los microorganismos, esta herramienta puede utilizarse también dentro de la misma especie, ya que la tasa de evolución, es decir, la velocidad a la que se producen mutaciones que quedan fijadas, es suficientemente rápida para poder diferenciar poblaciones (Pritchard, Stephens y Donnelly, 2000; Gupta y Maiden, 2001; Holt *et al.*, 2013).

La reconstrucción del árbol filogenético con las cepas o especies que se desea estudiar se realiza a partir de un alineamiento de sus secuencias, bien de una región concreta, la concatenación de varios genes o el genoma completo. El resultado será un diagrama con forma de árbol en que se reflejan las distancia entre ellas, habitualmente en una escala de número de sustituciones por sitio respecto al total de bases de dicho alineamiento, y cada cepa o especie se encontrará representada al final de cada una de las ramas formando agrupaciones o clados de aquellas más próximas. El nodo o punto de unión entre ramas corresponde a la estimación del ancestro común más próximo a ambas, con un número que indica el grado de soporte de la rama (soporte *bootstrap* o probabilidad posterior, dependiente del método empleado). Esto es un indicador de la fiabilidad de la agrupación (Figura 1.6).

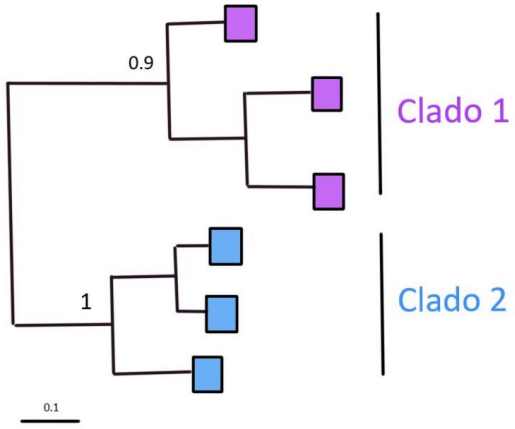


Figura 1.6. Representación de un árbol filogenético. *Se observa que las 6 muestras empleadas para su reconstrucción agrupan en 2 clados diferentes.*

El estudio evolutivo de microorganismos patógenos se ha utilizado con diferentes fines. En clínica se han estudiado transmisiones en hospitales, pudiendo llegar a detectar posibles reintroducciones en el hospital de un cepa persistente procedente de la comunidad como hicieron en el trabajo de Coll *et al.* (2017) con *S. aureus* resistente a meticilina, dejando patente que existen muchos eventos de transmisión no detectados en rutina por métodos habituales o por necesidad del aumento de los controles del entorno. También se ha utilizado como prueba forense en el caso de la transmisión del virus de la hepatitis C por un anestesista que afectó a más de 200 pacientes (González-Candelas *et al.*, 2013) y ha permitido establecer correlaciones entre la evolución humana y un patógeno como *Mycobacterium tuberculosis* (Comas *et al.*, 2013), lo que es muy importante para entender de qué manera el patógeno se adapta al hospedador.

Uno de los aspectos más relevantes y que más preocupan a este respecto es la capacidad de adaptación de *P. aeruginosa*, como se mencionaba anteriormente, en infecciones de pacientes con FQ, dado que la persistencia lleva a muy mal pronóstico. El mayor trabajo realizado al respecto en genomas completos utiliza 474 muestras de 36 pacientes (Marvig *et al.*, 2014). A partir de la determinación de mutaciones observan que en todas ellas se determina la convergencia entre 52 genes propuestos como posibles mutaciones patoadaptativas, facilitando la transición del entorno a pulmón. Sin embargo, en cuanto al aspecto evolutivo, realizan un esquema basado en parsimonia sin llegar a analizar con mayor profundidad mediante métodos más complejos y resolutivos las relaciones filogenéticas entre los aislados, en especial las posibles relaciones entre los aislados de distintos pacientes.

1.3 CRISPR

1.3.1 Función y estructura

Su nombre procede del acrónimo de "*Clustered Regularly Interspaced Short Palindromic Repeats*", aunque este no fue su nombre oficial hasta 2002, tras ser descrito en multitud de bacterias y arqueas diferentes con una estructura común a todas ellas. Este tipo de secuencias repetitivas se identificó por primera vez en 1987 (Ishino *et al.*, 1987), estudiando una proteína enzimática en *E. coli* y sus regiones flanqueantes, que se vio que tenían una estructura peculiar. Posteriormente, fue encontrada en *M. tuberculosis* en 1991, en el estudio de un elemento de inserción (Hermans *et al.*, 1991) y mencionaron que pudiera tratarse de un "punto caliente" en el que estos elemento móviles pudieran insertarse. En 1993, Mojica *et al.* (Mojica, Juez y Rodriguez-Valera, 1993) también la describieron en una arquea, *Haloferax mediterranei*, evidenciando que no era exclusivo de bacterias. No fue hasta el inicio de los años 2000 cuando comenzaron a almacenarse estas secuencias en bases de datos para su identificación y se sugería que podría tratarse de una estructura con funcionalidad propia (Jansen *et al.*, 2002; Mojica *et al.*, 2005).

El sistema CRISPR-Cas está compuesto por operones de genes Cas ("*CRISPR associated sequences*"), que codifican proteínas implicadas en el funcionamiento del sistema, y además uno o varios *arrays* CRISPR (Figura 1.7). Estos últimos se caracterizan por la presencia de repeticiones palindrómicas cortas de tamaño fijo entre las que se intercalan los *spacers*, secuencias de longitud similar a las repeticiones, extraídas a partir del material genético foráneo que ha conseguido acceder al interior de la bacteria (o arquea). El *array* se comporta como un registro de "invasores" para facilitar su eliminación en el caso de que vuelvan a introducirse por medio de interferencia, por lo que se trata de un sistema defensivo (Barrangou *et al.*, 2007).

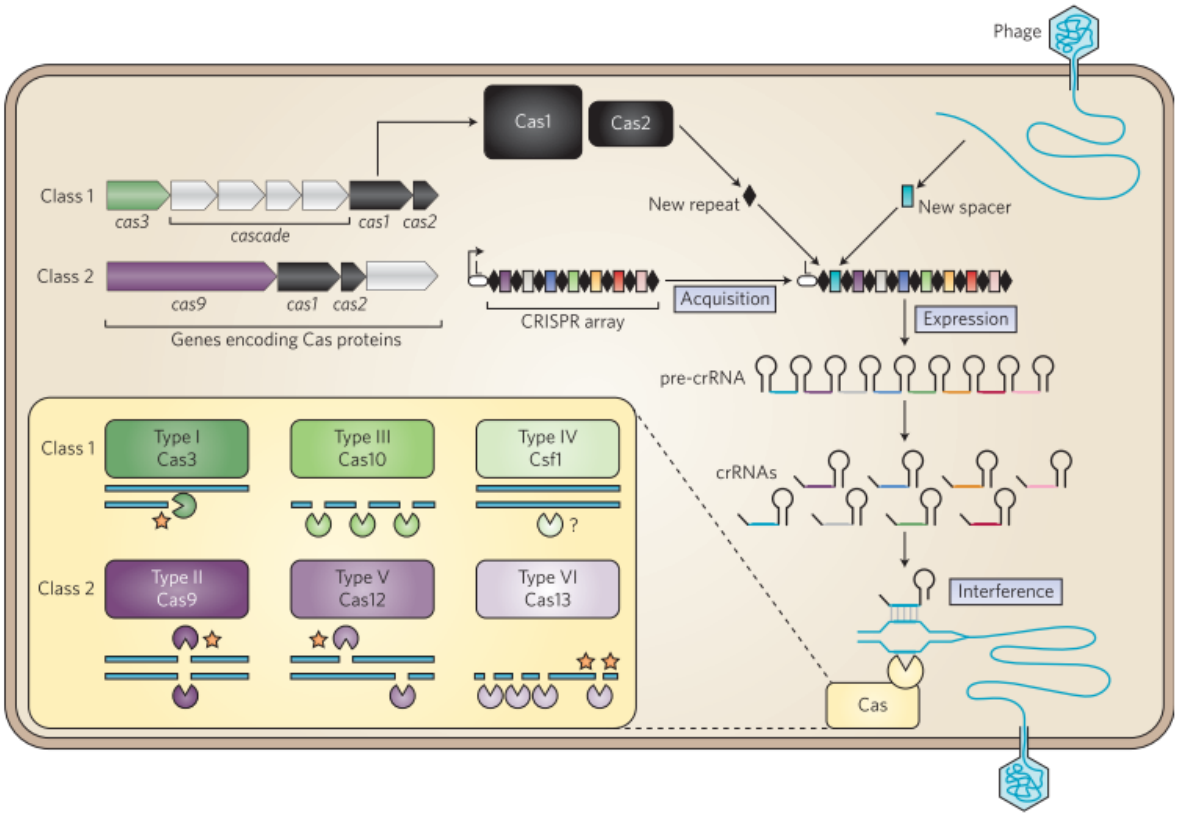


Figura 1.7. Esquema del sistema CRISPR-Cas. (Barrangou y Horvath, 2017)

Cuando se introduce un elemento móvil, como un bacteriófago, se activa la fase de adquisición en la que intervienen Cas1 y Cas2, dos proteínas altamente conservadas en los sistemas CRISPR. El nuevo *spacer* es incorporado al inicio de este *cassette*, en la posición adyacente a la secuencia líder, que es la que marca el lugar de inserción. Al realizarse siempre de esta forma, el *array* permite establecer un registro histórico de las diferentes infecciones o entradas de elementos móviles, ya que la secuencia más distal respecto a la líder corresponderá al evento más alejado en el tiempo.

Además de Cas1 y Cas2, en el sistema CRISPR intervienen otras proteínas Cas que participarán en las etapas ejecutivas, como son la separación en ARNcrs individuales (uno por cada *spacer*) tras la expresión del *array* en una sola molécula ARN (pre-ARNcr), y la interferencia, en la que se produce el corte dirigido del ADN o ARN foráneo. Estas proteínas Cas efectoras son las responsables de que la estructura CRISPR se divida en 2 clases (1 y 2) y estas, a su vez, en tipos. Los sistemas de clase 1 utilizan un complejo multiproteico en cascada junto a una Cas nucleasa; de esta manera, posibilita el reconocimiento de la región complementaria al ARNcr en el material genético diana y el posterior corte. Por el contrario, en los de clase 2 una sola proteína puede llevar a cabo este proceso, como ocurre con Cas9. Los tipos, organizados del I al VI, están distribuidos

en ambas clases y dentro de cada clase se diferencian en el tipo de corte de la proteína efectora que puede ser en simple o doble cadena (en el caso de ser doble, además, romo o cohesivo) o si cortan en una posición concreta respecto a una secuencia PAM ("*Proto-spacer Adyacent Motiv*") o, por el contrario, son con corte aleatorio (Barrangou y Horvath, 2017).

Cada secuencia *spacer* es diferente del resto de las contenidas en el *array*; sin embargo, es posible que para un mismo bacteriófago existan almacenados diferentes *spacers* o que un *spacer* proceda de una secuencia común a bacteriófagos diferentes y permita defenderse de todos ellos. La interferencia será efectiva si existe una similitud del 100% entre el *spacer* y la secuencia foránea y, en el caso de las Cas efectoras que reconocen PAMs, será necesaria su presencia para que la proteína pueda cortar: un cambio nucleotídico en ella mantendría la sensibilidad al bacteriófago a pesar de contar con el *spacer* (Barrangou *et al.*, 2007).

La presencia de este mecanismo de defensa en bacterias y arqueas no está limitado a un taxón concreto: ni a una familia, ni a género, ni siquiera a especie ya que se ha visto que hay variabilidad entre cepas. Por ejemplo, en *K. pneumoniae*, que hasta ahora se pensaba que no tenía CRISPRs (Jansen *et al.*, 2002), en contraste con otras Enterobacterias como *E. coli*, se ha visto que existen cepas en las que sí se encuentra y, además, se relaciona con la susceptibilidad a antibióticos (Li *et al.*, 2018). Dada su presencia en multitud de especies y el grado de variabilidad, en ocasiones se utiliza como método de tipado, como es el caso del "*spoligotyping*" en *Mycobacterium tuberculosis* (Kamerbeek *et al.*, 1997) o el CRISPOL de *Salmonella* (Fabre *et al.*, 2012).

1.3.2 Características propias en *P. aeruginosa*

La estructura CRISPR presenta una elevada variabilidad en las cepas de *P. aeruginosa*. Entre los 6 subtipos que existen dentro del subtipo I (IA-IF) (Makarova *et al.*, 2011), diferenciados en base a los genes que conforman la cascada, en *P. aeruginosa* se han detectado los subtipos I-C, I-E y I-F. El subtipo más frecuente es el I-F, seguido del I-E, siendo I-C descrita por primera vez recientemente (van Belkum *et al.*, 2015). El subtipo I-F ha sido descrito como un mecanismo que, además de eliminar material genético extraño, activa la transcripción de genes de virulencia (Figura 1.8).

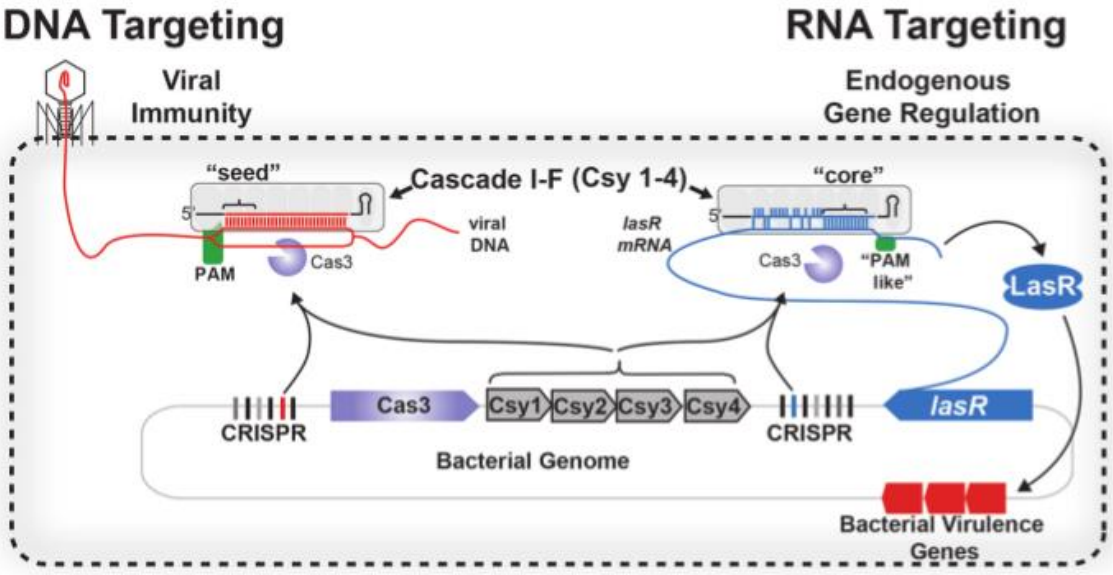


Figura 1.8. Actividad CRISPR I-F como regulador de la expresión. *La unión al ARNm de lasR requiere un par de bases en el extremo opuesto del crARN, una región llamada "*core*", y el reconocimiento de una secuencia de tipo PAM monocatenario. En ambos casos, el enlace de destino recluta a Cas3 para degradación.* (Wiedenheft y Bondy-Denomy, 2017)

Es habitual encontrar cepas que no tienen CRISPR. En estudios a nivel poblacional se ha determinado que menos del 40% de aislados presentan CRISPR (Cady *et al.*, 2011; van Belkum *et al.*, 2015). En un trabajo realizado a gran escala con más de 600 genomas (van Belkum *et al.*, 2015) se han estudiado estas estructuras en relación a su *Sequence Type* y se ha visto una correlación entre el tipo y estructura con pequeñas variaciones dentro de ST, por lo que indican su posible utilidad para estudiar variabilidad dentro de ST.

2. Objetivos

Objetivos

El objetivo central de este trabajo es determinar el grado de utilidad de la tecnología de secuenciación de genomas completos y la aplicación de métodos de estudio evolutivos para el control epidemiológico en *P. aeruginosa* de origen clínico.

Los objetivos específicos son:

- Investigar una situación de sospecha de brote con bajo número de casos.
- Discriminar qué pacientes forman parte de un brote que ha sido identificado previamente e investigar su posible extensión a otras áreas.
- Estudiar la evolución intrapaciente de *P. aeruginosa* en pacientes hospitalizados con el fin de encontrar mecanismos de adaptación al hospedador o rasgos diferenciales del tipo de infección.
- Localizar posibles cambios en la estructura CRISPR entre aislados pertenecientes a un brote.

3. Material y Métodos

Aunque el procedimiento será detallado a continuación, todos los *scripts* utilizados serán incluidos en https://github.com/pauruihu/PhD_scripts para garantizar la reproducibilidad de este trabajo. La Figura 3.1 muestra un esquema general de los diferentes análisis realizados con las muestras estudiadas en esta tesis.

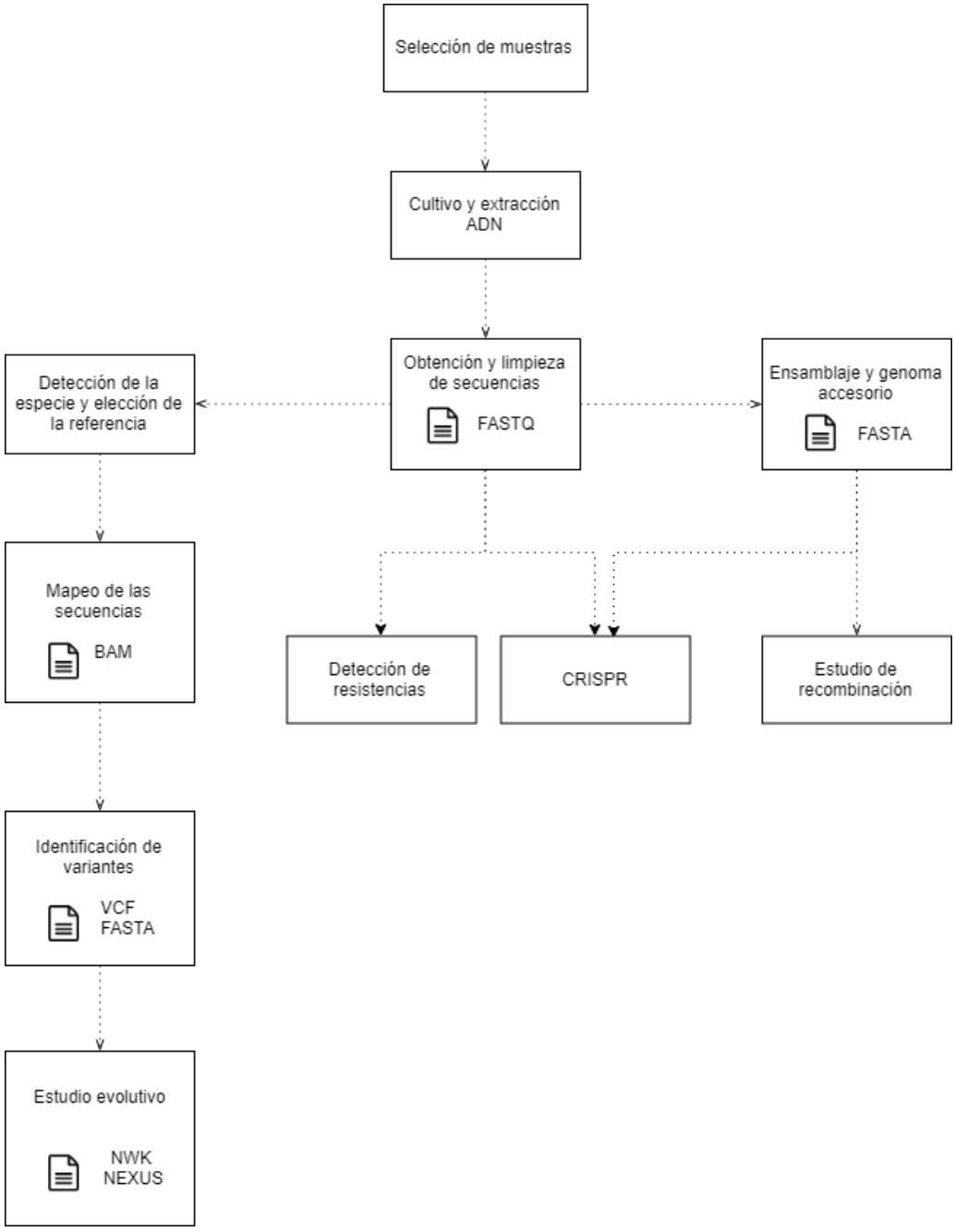


Figura 3.1. Diagrama del proceso de análisis seguido en este trabajo.

3.1 Selección de muestras

Las muestras de *P. aeruginosa* incluidas proceden de 3 hospitales de la Comunidad Valenciana y se seleccionaron con distintos objetivos. El primer proyecto se diseñó en colaboración con el Hospital General Universitario de Valencia (HGUV) con el fin de estudiar la evolución intrapaciente en individuos no afectados de fibrosis quística. Inicialmente se extrajo un registro de los positivos en *P. aeruginosa* por paciente entre enero de 2012 y agosto de 2014. Los criterios principales para la selección de los candidatos fueron: mayor número de aislamientos en localizaciones distintas del organismo, pudiendo proceder de orina, sangre, esputo, exudados, etc., y que entre muestras o grupos de muestras de cada paciente hubiera intervalos amplios de tiempo para facilitar la posibilidad de ver cambios adaptativos, todo ello hasta un tamaño muestral aproximado de 100 para el conjunto de pacientes. Finalmente, se partió de 112 aislados de *P. aeruginosa* procedentes de 18 pacientes. Por otro lado, se realizó el análisis de dos posibles brotes de *P. aeruginosa* en sendos hospitales, el Hospital General Universitario de Elche (HGUE) con 73 muestras, 5 de las cuales fueron clasificadas como ambientales al ser aisladas a partir de aguas recogidas en el interior del hospital, principalmente en grifería; y el Hospital Arnau de Vilanova (HAV) de Valencia, donde se aislaron 12 cepas de 6 pacientes que se encontraban en el Servicio de Hematología (Tablas MS1 y MS2, Material suplementario).

3.2 Cultivo y extracción de ADN

Inicialmente, tras construir una base de datos de los aislados congelados de *P. aeruginosa* agrupados por paciente en el Servicio de Microbiología del HGUV, se seleccionaron aquellos con mayor número de cepas intentando cumplir los criterios anteriormente descritos. En total, se recuperaron 93 cepas procedentes de 16 pacientes. Seguidamente, cada una de las cepas se inoculó en medio de enriquecimiento BHI, aumentando de manera significativa el crecimiento de las colonias seleccionadas y así obtener una mayor cantidad de ADN destinado a la secuenciación masiva. Se utilizaron 700 µL de estos cultivos para la extracción del contenido genómico, que contaban con un volumen aproximado de 1,5 mL. Los sobrantes de los cultivos fueron almacenados a -80ºC en glicerol al 15% como crioprotector.

La extracción partió de un primer paso de centrifugación para precipitar las células y desechar el sobrenadante (el medio BHI) y a continuación resuspender las

células en tampón PBS. Tras la resuspensión, se utilizó el extractor de ácidos nucleicos NUCLISENS® easyMAG® (Biomerieux, Marcy-l'Étoile, Francia), siguiendo su protocolo de lisis interna y purificación con bolas magnéticas. La comprobación del buen estado del ADN genómico se realizó con un gel de agarosa al 0.8% durante una hora a 110 V utilizando un marcador de pesos de 1 kb y GelRed® para el marcaje fluorescente del ADN. Una vez confirmada la calidad del ADN genómico, la cuantificación para la preparación de alícuotas destinadas a la secuenciación masiva se determinó con Qubit utilizando el *dsDNA HS Assay Kit*. Este método parte de un microlitro de muestra al que se añade una mezcla de tampón y el reactivo específico para la medición de dobles cadenas de ADN en un volumen final de 200µL. Una vez determinada la concentración de cada muestra, se realizaron alícuotas de 5 µL a 0.2 ng/µL que se emplearon para la secuenciación.

Los proyectos de detección y análisis de brotes realizados en colaboración con los hospitales de Elche y Arnau de Vilanova partieron del ADN purificado de cada aislado bacteriano que fueron remitidos por los respectivos Servicios de Microbiología, por lo que solamente se procedió a su cuantificación por Qubit y alicuotado a la concentración indicada anteriormente. Únicamente se repitió el sistema de extracción con easyMAG® utilizado en las muestras del HGUV en la segunda remesa de aislados de Elche (muestras de 2016), dado que en este caso fueron enviados los cultivos puros en placa.

3.3 Obtención y limpieza de secuencias

Todas las muestras analizadas en esta tesis fueron secuenciadas por Illumina Miseq 2x300 pb utilizando los kits *Nextera XT DNA library preparation kit* y el *Illumina MiSeq Reagent kit v3*. Los genomas de *P. aeruginosa* del HGUV fueron secuenciados dos veces mediante sendas carreras con 93 muestras. La cobertura media esperada era de 24X en cada carrera, ya que su tamaño genómico promedio es de aproximadamente 6.7 Mb. En cuanto a los brotes, las muestras del HGUE fueron secuenciadas en dos carreras diferentes, dado que se vio que el brote tenía continuidad. En un primer momento se secuenciaron 57, y en el segundo conjunto de muestras, 16. Finalmente, en el caso del brote del Arnau las muestras fueron incluidas en una carrera junto con otro tipo de aislados.

La limpieza de las lecturas crudas se realizó de la misma manera para los tres análisis, aunque en el caso de las secuencias del HGUV previamente fue necesario combinar en un mismo archivo las lecturas de las dos carreras generadas de cada muestra. En primer lugar, se utilizó AUTOADAPT v0.2 (https://github.com/optimuscoprime/autoadapt) para la eliminación de posibles restos de secuencias de los cebadores utilizados en la secuenciación. Se aplicó una calidad mínima de 20 y tamaño mínimo de 50 pb. A continuación, con PRINSEQ V0.20.4 (Schmieder y Edwards, 2011) se eliminaron los fragmentos que pudieran afectar en las etapas posteriores por baja calidad de la secuenciación. Los parámetros se establecieron en una longitud mínima de lectura de 50 pb con eliminación de 10 pb en la región izquierda de la lectura donde se detectaron picos de desequilibrio en los nucleótidos por restos de adaptadores, eliminación en la zona derecha de nucleótidos con calidades media inferiores a 30 y una ventana de 20 pb. Por último, se hizo un cálculo de las estadísticas y comprobación de la mejora de las lecturas con FASTQC v0.11.4 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) y MULTIQC V1.2 (Ewels *et al.*, 2016).

3.4 Detección de la especie y elección de la referencia

Con el fin de evitar posteriores problemas derivados de la secuenciación de otra especie por posibles contaminaciones, se utilizó el programa KRAKEN (Wood y Salzberg, 2014) en diferentes versiones: 0.10.6-a2d113dc8f, con una base de datos en la que se incluían genomas de todas las especies, incluida la humana, y la versión 1.0, con la base de datos preformateada MINIKRAKEN 2014 4GB, que contiene genomas bacterianos, virales y de arqueas. Este programa clasifica taxonómicamente las lecturas en base a *k*-meros. Un *k*-mero es una subsecuencia única de longitud *k*, en este caso, procedente de una lectura. De cada lectura se obtienen todos los *k*-meros posibles y cada uno de ellos se mapea de manera independiente frente al ancestro común más próximo de entre todos los genomas que contienen dicho *k*-mero. Se realiza una clasificación en forma de árbol en que los diferentes nodos corresponden con un taxón, siendo elegida como ruta de clasificación de la lectura aquella con mayor número de *k*-meros mapeados, asignándole

el del nodo alcanzado dentro de la ruta con la clasificación taxonómica más concreta (Figura 3.2).

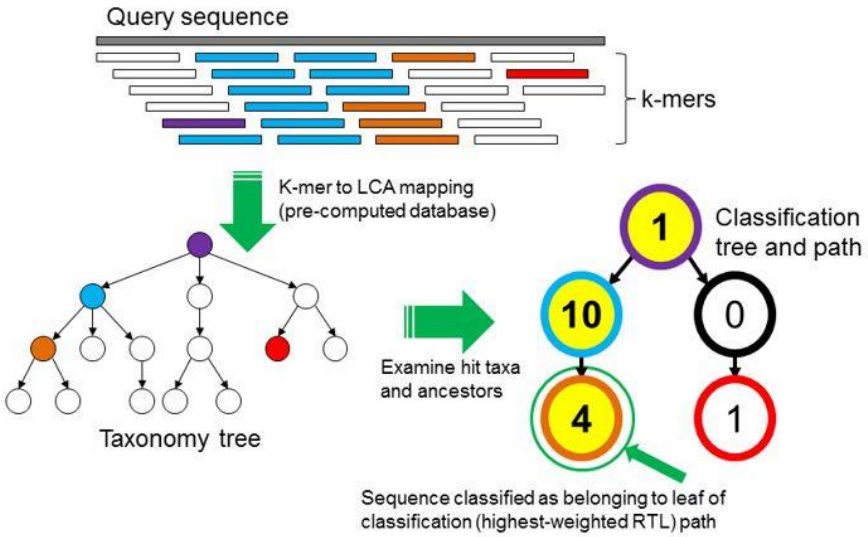


Figura 3.2. Algoritmo de clasificación de KRAKEN. (Wood y Salzberg, 2014)

Aquellas muestras que fueron confirmadas como *P. aeruginosa* se clasificaron siguiendo uno de los métodos más utilizados para el tipado, como es el esquema de *Multilocus Sequence Typing* (MLST) (Curran *et al.*, 2004). Está basado en la determinación, mediante secuenciación, de los alelos presentes en varios genes altamente conservados, habitualmente 7, que en esta especie son *acsA*, *aroE*, *guaA*, *mutL*, *nuoD*, *ppsA* y *trpE*. A los alelos de cada gen se les asigna un número diferente según el orden en que éstos fueron incorporados a la base de datos, que está en continua actualización por el descubrimiento de nuevos alelos. La combinación de los 7 alelos se corresponde con el *sequence type* (ST) de esa cepa, que recibe también un número identificador único, permitiendo la clasificación preliminar de las cepas y, en este caso, también la de seleccionar la referencia más cercana para utilizar en el estudio genómico. El criterio para dicha selección fue utilizar para cada conjunto de muestras aquella referencia con el mismo ST o, en su defecto, la del ST más cercano, es decir, aquel con el que compartiera más alelos de entre todos los genomas completos disponibles en ese momento (enero 2017) en la base de datos la *Pseudomonas Genome DB* (www.pseudomonas.com).

Ya que partimos de lecturas de genomas completos, hemos utilizado el programa SRST2 v0.1.8 (Inouye *et al.*, 2014) para obtener el ST de cada muestra analizada.

Aportando un archivo multifasta, que utiliza como base de datos, realiza un alineamiento de las lecturas frente a cada una de las secuencias contenidas en dicha base de datos, para lo cual recurre a programas de mapeo y procesamiento, como BOWTIE 2 v2.2.6 (Langmead y Salzberg, 2012) y SAMTOOLS V0.1.18 (Li *et al.*, 2009). Para determinar el alelo del gen a que corresponde de entre los contenidos en el multifasta realiza *tests* binomiales y una comparativa de los p-valores obtenidos por alelo, para lo que calcula el *score*: el valor de la pendiente entre el p-valor observado frente al esperado. Se considerará el alelo más cercano aquel con un valor de *score* más bajo y, por tanto, será el que aparezca en la tabla de resultados que devuelve el programa, siempre y cuando cumplan con el 90% de la cobertura y menos del 10% de divergencia. Cuando se utiliza para tipado, además, utiliza una tabla en que se encuentran definidas las combinaciones de alelos que dan lugar al ST, por lo que lo asigna automáticamente en la tabla de resultados. En este caso, hemos obtenido la base de datos de MLST para *P. aeruginosa* utilizando el programa GETMLST.PY, implementado también en el mismo paquete. A continuación, lanzamos el programa utilizando las lecturas limpias por parejas e indicando la base de datos descargada, así como el archivo de texto con los ST definidos, dejando las opciones de mapeo y filtrado por defecto.

De forma accesoria, se han complementado algunos de los resultados de tipado con el uso de programas como ARIBA V2.10.1 (Hunt *et al.*, 2017), a partir de las lecturas, y MLST V2.10-dev (https://github.com/tseemann/mlst), tras el ensamblaje (ver Material suplementario). El algoritmo que lleva a cabo ARIBA es diferente al de SRST2, dado que previo al mapeo agrupa las secuencias de referencia según su similitud con CD-HIT V4.6 (Figura 3.3). Las lecturas que mapean frente al mismo grupo (*cluster*) de secuencias de referencia son ensambladas con FERMI-LITE V0.1 (https://github.com/lh3/fermi-lite). La secuencia resultante, ya sea en uno o varios contigs, es alineada y comparada con las secuencias de dicho *cluster* mediante NUCMER V3.1 (programa del paquete MUMMER), para elegir la referencia más cercana. A continuación se realiza la identificación de variantes (*variant calling*) por partida doble: se alinea el ensamblado respecto a la secuencia referencia y se comparan las variantes, de manera que se confirma que se ha ensamblado correctamente y se determina el grado de integridad de la secuencia resultante; y, por otra parte, se remapean las lecturas con BOWTIE2 V2.2.6 frente a la secuencia ensamblada y se extraen las variantes con SAMTOOLS MPILEUP, de manera que se confirma la presencia de los SNPs y se calcula el grado de cobertura.

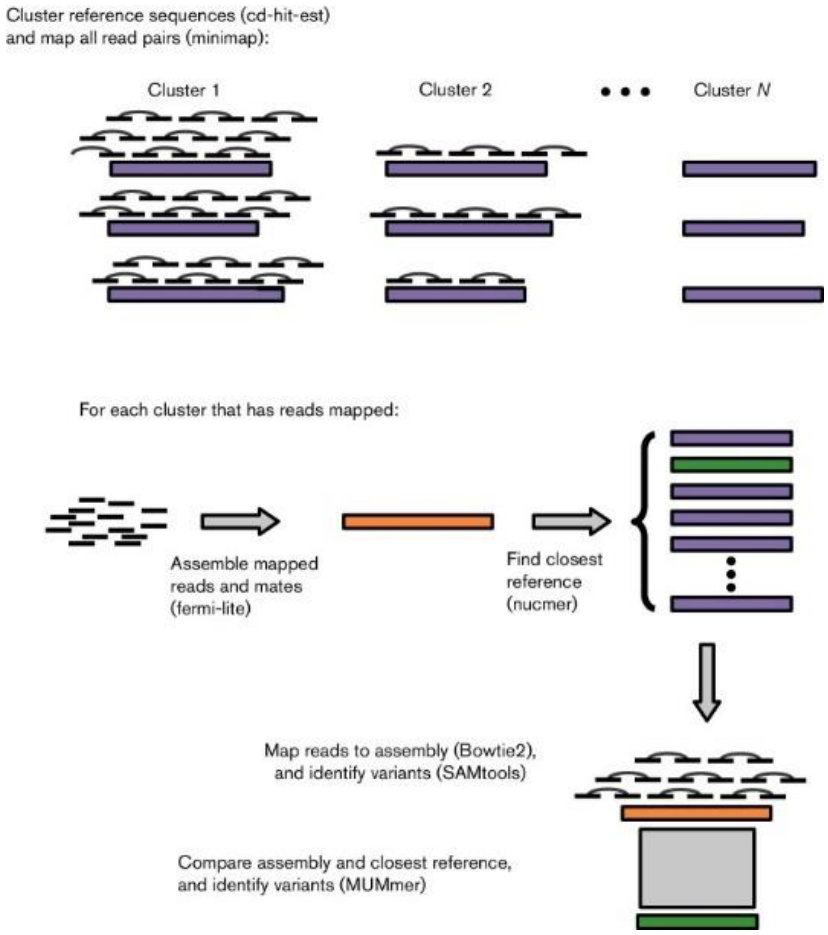


Figura 3.3. Esquema del algoritmo llevado a cabo por ARIBA. (Hunt *et al.*, 2017)

3.5 Mapeo de las secuencias

El primer paso para la obtención de variantes de una cepa es alinear todas las lecturas de una muestra frente a un genoma de referencia lo más cercano posible, conocido como mapeo, un procedimiento alternativo al de la reconstrucción *de novo*. En este caso, utilizando la referencia seleccionada para cada conjunto de muestras, la cepa W16407_3629 para las muestras del HGUV y la cepa H27930_3636 para las *P. aeruginosa* de HGUE y HAV, se mapearon las lecturas utilizando el programa BWA v0.7.12 (Li, 2013) (*Burrows-Wheeler Aligner*) con el algoritmo MEM, el óptimo para lecturas entre 70 pb y 1 Mb procedentes de secuenciación por Illumina. Éste trata de encontrar la coincidencia exacta más larga que cubra cada posición. Para evitar posibles errores o falsos positivos, como podría ocurrir en el caso de genes parálogos, además de

localizar estas regiones, realiza la extensión a la lectura completa, rechazando los alineamientos locales con mejor puntuación frente a los de extensión hasta el extremo si la diferencia entre sus puntuaciones no supera un determinado un umbral. Gracias a este abordaje se pueden alinear lecturas con variaciones reales, reduciendo tanto los sesgos por efecto de la referencia como el número de discrepancias o los *gaps* introducidos. En este paso se han mantenido todos los valores establecidos por defecto, obteniendo como resultado los archivos SAM (*Sequence Alignment Map*), un formato de archivo que contiene el alineamiento de las lecturas frente a la referencia, así como información extra que será útil para el post-procesamiento, como valores de calidad de mapeo de las lecturas o la calidad de las propias lecturas.

Posteriormente se generaron los archivos BAM, el formato binario de los archivos SAM, para conseguir un mejor manejo dado el número de muestras a procesar, ya que estos ocupan menos espacio. Para ello se utilizó SAMTOOLS v1.4 (Li *et al.*, 2009), un programa que permite manipular este tipo de archivos: con el comando *sort* se reorganizaron las lecturas a lo largo del alineamiento, desde aquellas que están más a la izquierda hasta el final del genoma de referencia, y con *index* se genera un índice del contenido del alineamiento para el rápido acceso y procesamiento del archivo binario en pasos sucesivos. A continuación, se realiza con *MarkDuplicates* de PICARD-TOOLS v1.119 (http://broadinstitute.github.io/picard/) (valores por defecto) el marcado y filtrado de lecturas duplicadas, es decir, secuencias producidas por un mismo fragmento de ADN. En ocasiones, estos se producen por algún fallo en la construcción de las librerías o porque el sensor óptico del secuenciador ha detectado dos grupos de amplificación tras la PCR tipo puente, aunque realmente pertenecen al mismo grupo; en este caso se denomina duplicado óptico. Su eliminación es importante, porque pueden llevar a un sesgo en el proceso de llamado de variantes; por ejemplo, si en las lecturas duplicadas hubiera cambios debidos a errores en el proceso de amplificación, al aumentar el número de veces que se ha detectado dicho cambio podrían pasar el filtro de SNP a pesar de ser artefactos.

A pesar del abordaje que utiliza BWA para que el mapeo sea óptimo, es frecuente que se den problemas en las regiones cercanas a las de un *indel* respecto a la referencia, pudiendo considerarse como SNPs aquellos nucleótidos contiguos que han sido erróneamente alineados. Asimismo, es importante tener en cuenta que el algoritmo de mapeo trabaja con cada lectura independientemente, por lo que es probable que haya

lecturas que cubran dicha zona o adyacentes alineadas de forma distinta. Por ello, una vez eliminados los duplicados, se realinean los *indels* utilizando dos programas del paquete de GATK v3.3-0-g37228af (DePristo *et al.*, 2011; Van der Auwera *et al.*, 2013): *RealignerTargetCreator* e *IndelRealigner*, nuevamente con los valores por defecto. Primero, con *RealignerTargetCreator* se construye una lista de aquellas regiones del genoma susceptibles de contener este tipo de fallos. Seguidamente, *IndelRealigner* utiliza la lista para realinear todas las lecturas de dichas zonas si es necesario.

Tras este procedimiento, se realizó el cálculo de cobertura con SAMTOOLS mediante la opción *depth* a partir de los nuevos archivos BAM de alineamiento de lecturas, limpios y realineados. De esta manera, pudimos comprobar si el mapeo se realizó correctamente o si, por el contrario, la referencia pudo no ser adecuada dependiendo del número de lecturas que se quedaron sin mapear, y en ese caso, si dichas lecturas se corresponderían con contaminación.

3.6 Identificación de variantes

La identificación de variantes comienza a partir de los archivos BAM obtenidos tras el mapeo y utilizando SAMTOOLS v1.4 en combinación con BCFTOOLS v1.4a (Li, 2011). Se establece una calidad mínima tanto de mapeo como de llamado de base de 30 (escala phred[1]) en el paso de SAMTOOLS MPILEUP en el que se realiza un resumen de las estadísticas del mapeo y el cálculo de probabilidades del genotipo utilizadas por el siguiente análisis. El resultado es procesado con BCFTOOLS para la asignación de variantes propiamente dicha, considerando ploidía igual a1.

Se marcaron con la etiqueta de baja calidad aquellos SNPs a 3 pb o menos de un *indel* y con una calidad inferior a 20 para su posterior eliminación ante la posibilidad de tratarse de falsos positivos. Además, se fijó un filtro más estricto para determinar si una posición se corresponde con una variante o con la base de la referencia: se estableció que para las variantes deben alinear en dicha posición dos lecturas en *forward* y dos en *reverse* con elevada calidad (lecturas clasificadas por el programa como DP4) y que la proporción total de lecturas que contenga la variante sea, como mínimo, del 75%. En el caso de que

[1] El *phred score* es un parámentro en escala logarítmica que mide la probabilidad de error. Un phred de 30 indica una precisión del 99,9%, o lo que es lo mismo, una probabilidad de error de 1 de cada 1000. (Ewing *et al.*, 1998; Ewing y Green, 1998)

se trate de una posición coincidente con la referencia, solo es necesario una lectura en *forward* y una en *reverse*, manteniendo la proporción del 75%. Si una base no cumple uno de los dos criterios, por ejemplo, en posiciones con una proporción similar de variante y referencia, en la secuencia generada habrá una "N" para esa posición al igual que en sitios delecionados, con ausencia de lectura o que no cumplan los filtros de calidad marcados.

En el siguiente paso, con la opción *view* se filtraron todos aquellos que pasaban el umbral fijado para cada parámetro, quedando excluidas las inserciones. Como último paso se utilizó el programa VCFUTILS.PL del paquete BCFTOOLS para la transformación del VCF (Danecek *et al.*, 2011) a FASTQ y SEQTK v1.0-r31 (https://github.com/lh3/seqtk) para la transformación de FASTQ a FASTA. Así se obtuvo un "pseudogenoma" por aislado, cuyo concatenado en un archivo multifasta se corresponde con el alineamiento de todas las secuencias, puesto que cada posición se corresponde con la posición homóloga del genoma de referencia. Para finalizar dicho alineamiento, tuvo que incorporarse al final del mismo una cantidad de "N"s determinada en cada aislado, ya que es habitual que las posiciones finales del genoma no estén recogidas en el VFC por falta de lecturas; sin embargo, es necesario que todos los genomas posean la misma longitud para que el alineamiento pueda ser reconocido como tal por programas de SNP *calling* o de reconstrucción de árboles filogenéticos; de lo contrario, estos programas no pueden operar.

Por último, se comprobó que el mapeo y llamado de variantes se había realizado de manera adecuada mediante la determinación de la cobertura para cada muestra, entendiendo aquí la cobertura no como el número de lecturas totales mapeadas que vimos anteriormente, sino el número de posiciones que han podido determinarse correctamente (posiciones distintas de "N") respecto del total de bases del genoma de referencia.

Antes de extraer los SNPs del alineamiento, enmascaramos las repeticiones que habían sido previamente localizadas en la referencia con el programa REPEAT-MATCH que forma parte de MUMMER v3.23 (Kurtz *et al.*, 2004). El enmascarado consiste en el cambio de las posiciones que coincidan con las coordenadas de las repeticiones por "N"s a lo largo de todo el alineamiento, para lo cual se utilizó el *script* REMOVE_BLOCKS_FROM_ALN.PY, publicado en el perfil de GitHub del Wellcome Trust Sanger Institute (https://github.com/sanger-pathogens/remove_blocks_from_aln). Aplicando este filtro evitaremos utilizar en el análisis los polimorfismos detectados en

zonas de repeticiones causadas por alineamientos erróneos de lecturas de otras regiones del genoma que contengan dichas repeticiones. Finalmente se utilizaron SNP-SITES V2.1.3 (Keane *et al.*, 2016) o la función seg.sites del paquete APE (Paradis, Claude y Strimmer, 2004) en R para la extracción de las posiciones segregantes.

3.7 Estudio evolutivo

La reconstrucción de árboles filogenéticos y la datación de nodos en los mismos se ha realizado por inferencia bayesiana. Esta metodología, implementada en el programa BEAST (Drummond *et al.*, 2012), permite contrastar una amplia variedad de modelos evolutivos e incorporar, además del alineamiento de ADN, otra información conocida previamente, denominada comúnmente como *priors*, con una velocidad de cálculo asequible y resultados de fácil interpretación. De esta forma, estableceremos el modelo evolutivo más adecuado para cada caso, aunque siempre estaremos condicionados a la disponibilidad de información con que contamos, ya que cuantos más datos le proporcionemos, más se ajustará a la realidad.

Los métodos que se usan tradicionalmente para la reconstrucción de filogenias, como la máxima verosimilitud (Felsenstein, 1981) o el *neighbor-joining* (Saitou y Nei, 1987) evalúan si las agrupaciones obtenidas en el árbol están bien soportadas por los datos usados en su construcción mediante remuestreo con datos pseudoaleatorios (técnica de *bootstrap* no paramétrico (Felsenstein, 1985). Sin embargo, al usar métodos bayesianos obtendremos la probabilidad de que las agrupaciones generadas sean correctas en función de los datos originales aportados y los modelos empleados en su obtención. El estadístico que se emplea para ello es la probabilidad posterior ($P(H|D)$), definida por el Teorema de Bayes:

$$P_r(H|D) = \frac{P_r(D|H)xPr(H)}{Pr(D)}$$

Figura 3.4. Teorema de Bayes. *H: hipótesis (árbol); D: datos. El término* $\boldsymbol{P(H|D)}$ *es la probabilidad del árbol dados los datos (probabilidad posterior) y la* $\boldsymbol{P(D|H)}$ *es la probabilidad de los datos dado el árbol (verosimilitud); ambas son probabilidades condicionales en las que se asume que la segunda condición es cierta. Por otro lado, la* $\boldsymbol{P(H)}$ *es la probabilidad previa del árbol, sin tener en cuenta los datos, y la* $\boldsymbol{P(D)}$ *es la probabilidad de los datos, por lo que es un parámetro que será idéntico para todos los árboles.*

La obtención de este parámetro no se lleva a cabo mediante un cálculo directo, lo que supondría un coste computacional muy alto, sino que se realiza una aproximación a dicho valor utilizando el algoritmo de Cadenas de Markov-Monte Carlo[2] (Markov Chain-MonteCarlo, MCMC). Éste muestrea en el espacio de todos los árboles posibles, pudiendo comenzar con un árbol aleatorio o uno que le aportemos como inicial. En cada paso o generación de la cadena se propone una localización distinta del espacio en la que habrán variado ligeramente uno o varios de los parámetros del árbol y del resto del modelo (evolutivo y demográfico, en nuestro caso). El nuevo estado podrá aceptarse y avanzar en la cadena, o rechazarse y repetir el proceso, mediante la comparación de las probabilidades posteriores; es decir, se aceptará cuando exista una mejora de la probabilidad posterior u ocasionalmente cuando el descenso no sobrepase una determinada proporción. Con este procedimiento tenemos la ventaja que en cada iteración solamente necesitamos calcular la *ratio* de probabilidades posteriores, evitando el denominador de la función que aumentaría la complejidad del cálculo, pues representa la probabilidad de observar los datos integrada sobre todos los valores posibles de los parámetros de los modelos tratados, es decir, la topología del árbol, el modelo evolutivo y el modelo demográfico.

La nueva localización del árbol en el estado siguiente habitualmente será próxima a la anterior, por lo que los árboles serán muy parecidos. Puesto que se dan cambios pequeños hasta alcanzar la región de máxima probabilidad, en la que esperamos que se mantenga la cadena de forma más o menos estable, será importante la elección del número de generaciones, ya que un número muy bajo impedirá que la cadena alcance la zona estacionaria. Cuando este número es suficientemente alto, el número de veces que se muestrea un determinado parámetro con cierto valor será proporcional a su probabilidad posterior. El árbol y el resto de parámetros de los modelos que extraigamos del análisis será el consenso de los estados que se han muestreado en la fase estacionaria, y, para las topologías, en cada nodo se encontrarán las probabilidades posteriores correspondientes, que nos indicarán el grado de confianza de cada agrupación.

Además del propio BEAST, que realiza dichos cálculos, hay una serie de programas auxiliares que facilitan su utilización o que permiten estudiar y procesar los resultados a partir de los archivos obtenidos. Previamente a los tests con BEAST v1.8.4,

[2] En concreto utiliza el algoritmo Metropolis–Hastings, una variante dentro de los algoritmos que emplean MCMC.

se empleó el programa TEMPEST v1.5 (Rambaut *et al.*, 2016). Éste evalúa la señal temporal en los datos moleculares, para lo cual calcula por métodos de regresión si la información sobre las fechas de aislamiento de las muestras es relevante respecto al árbol filogenético. Cuando hay una asociación significativa entre divergencia y fechas de muestreo, se dice que hay señal temporal suficiente, por lo que será útil incorporar dichas fechas en el análisis por BEAST como *prior*. Los datos de partida son los alineamientos de SNPs de las *P. aeruginosa* de cada hospital. A partir de ellos se reconstruyeron los árboles filogenéticos correspondientes por máxima verosimilitud con IQTREE v1.5.5 (Nguyen *et al.*, 2015). Se seleccionaron los modelos de sustitución de ADN más adecuados para el cálculo por MODELFINDER, incluido en IQTREE, incorporando 1000 réplicas para el cálculo del *bootstrap* ultraparamétrico (Minh, Nguyen y von Haeseler, 2013) y 1000 réplicas de *bootstrap* para el *SH-like approximate likelihood ratio test* (Guindon *et al.*, 2010), tal y como se recomienda en la documentación. A continuación, se estudiaron los árboles de cada análisis, determinando con TEMPEST para cada árbol si había suficiente señal temporal en cada conjunto de datos o por clados que permitieran la calibración del árbol. En el caso de los *tests* por clados, se decidió establecer las agrupaciones a distintos niveles mediante HIERBAPS (Cheng *et al.*, 2013), la versión que utiliza BAPS 6.0 y Matlab Runtime Component versión 7.13, permitiendo agrupar secuencias de ADN con el fin de revelar la estructura interna de la población que puede ser muy útil cuando se dan altas de tasas de transferencia horizontal.

La preparación de los archivos XML que requiere BEAST y que aporta al programa todos los datos de que disponemos para efectuar los cálculos se realizó con el programa BEAUTi. En cada caso se utilizó *tip dating* según la señal filogenética del árbol o los subclados del mismo. El modelo de sustitución nucleotídica que se aplicó es el GTR + gamma (4 categorías), con las frecuencias de las bases estimadas y el modelo poblacional de crecimiento exponencial. En cuanto al modelo de reloj molecular aplicado, este fue diferente para cada conjunto de muestras, ya que se consideró el estricto como más adecuado para los brotes de Elche y del Arnau, y el modelo de reloj relajado no correlacionado con distribución lognormal para el de Valencia. Asimismo, se utilizó como *prior* la magnitud de la tasa de mutación más baja de la determinadas en un trabajo reciente (Miyoshi-Akiyama *et al.*, 2017) a partir de las secuencias de la base de datos de NCBI de *P. aeruginosa* a nivel de *scaffold*, aplicando por tanto una media de 1E-6 s/s/a (sustituciones/sitio/año) de ucld.mean, o clock.rate, según el modelo de reloj molecular aplicado, tasa tomada también como valor inicial y siguiendo una distribución

exponencial. Las generaciones que se emplearon para la calibración de las muestras de Valencia y del Arnau fueron 100.000.000 y 60.000.000 en el caso de Elche, guardando los resultados cada 10.000 o 6.000 generaciones, respectivamente. Los resultados de los clados estudiados se utilizaron para la calibración de los alineamientos de SNPs de cada *dataset* y reajuste de los *priors*.

3.8 Ensamblaje y genoma accesorio

La reconstrucción de los genomas *de novo* a partir de las lecturas limpias se realizó con SPAdes v3.9.0 (Bankevich *et al.*, 2012). Este ensamblador, a diferencia de otros, permite utilizar simultáneamente varios tamaños de k-meros, además de incluir la utilización del programa BayesHammer (Nikolenko, Korobeynikov y Alekseyev, 2013) para la corrección de errores de lectura mediante el uso de las frecuencias de los k-meros y la corrección de errores en *contigs* con el alineamiento de las lecturas frente a estos tras el ensamblado usando para ello el mapeador BWA. En este trabajo se especificaron los *k*-meros 31, 55, 77 y 101, estableciendo la cobertura umbral a aquella calculada por el programa. Se empleó el modo *careful* para activar el post-procesamiento con BWA y así tratar de reducir el número de bases no concordantes (*mismatches*) y de *indels* cortos. El archivo tipo fasta con los *contigs* que se generó para cada aislado fue filtrado con un script propio para eliminar aquellos *contigs* inferiores a 500 pb. La evaluación de la calidad de los ensamblados se realizó a partir de los recuentos estadísticos determinados con QUAST v4.3 (Gurevich *et al.*, 2013). Este programa hace un recuento del número de *contigs* en que se encuentra fragmentado el genoma y su longitud. Un parámetro importante en este análisis es el N50. Es un valor de longitud de *contig* e indica que el 50% del genoma está contenido en todos los *contigs* iguales o superiores a dicha longitud. Para un buen ensamblaje esperamos que este valor sea más alto, indicando mayores longitudes de *contigs*. El número total de *contigs* por sí solo no es informativo, puesto que en ocasiones hay un elevado número de fragmentos cortos pero la mayor parte del genoma contenida en *contigs* muy extensos. Para ello, también contamos con el parámetro L50, que indica el número de contigs que contienen el 50% del genoma, por lo que para este parámetro esperamos un número bajo en buenos ensamblados. De la misma manera, los parámetros N75 y L75 corresponden al 75% del genoma.

Para comparar el contenido génico es necesaria la anotación previa, es decir, la identificación de los genes reconstruidos y contenidos en los *contigs*, la cual hemos

realizado con PROKKA 1.12 (Seemann, 2014). En un primer paso, éste utiliza PRODIGAL V2.6 (Hyatt *et al.*, 2010) para predecir las regiones codificantes y establecer sus coordenadas. Posteriormente, localiza a nivel de secuencia proteica aquella con mayor similitud según una base de datos que indiquemos tras una búsqueda BLAST v2.4 (Altschul *et al.*, 1990; Camacho *et al.*, 2009) y se asignará su nombre y función. En el caso de que no se encuentre en dicha base de datos ninguna secuencia al valor *e* indicado como umbral (fijado en este trabajo en 10^{-6}), se realizará una nueva búsqueda utilizando bases de datos más amplias, con menor especificidad, entre las que posee almacenadas el programa por defecto.

En primer lugar, se ha utilizado el script *database_maker.py* (https://github.com/rehrlich/prokka_database_maker) con el que se creó una base de datos de *P. aeruginosa* a partir de los genomas almacenados en el NCBI (taxid 287) en mayo de 2017 para el primer paso de anotación de los contigs por PROKKA. Aquellas regiones codificantes que no pudieron anotarse, el programa las volvió a analizar para anotarlas según la base de datos de proteínas *sprot* incluida en el paquete de PROKKA dentro del reino Bacteria, con el mismo valor *e* umbral.

Uno de los archivos que produce PROKKA para cada aislado es el GFF3 (*general format feature version 3*), que contiene la información necesaria para llevar a cabo la comparación del contenido genómico: nombre, cadena en que se encuentra, posición de inicio y fin del gen, etc., además de la propia secuencia. Este archivo se ha analizado con ROARY V3.12.0 (Page *et al.*, 2015), un programa que calcula el pangenoma de un conjunto de muestras de una misma especie. Convierte las secuencias codificantes en proteínas y construye *clusters* con CD-HIT v4.6 a partir de aquellas proteínas que se encuentren completas (que tengan codón de inicio y de parada), descartando aquellas con más del 5% de nucleótidos indeterminados (Ns) o con secuencia inferior a 120 pb. Comienza la *clusterización* partiendo de secuencias de proteínas que coinciden al 100%, reduciendo este valor en 0,5% en cada iteración hasta el 98% de identidad. En el proceso, las proteínas *core*, presentes en >99% de las muestras, son transferidas a *clusters* que se mantendrán al margen para el paso siguiente; es decir, mantiene aquellas secuencias con 100%-98% de identidad y que no pertenezcan al *core*, utilizando una secuencia representativa de cada *cluster*. Las compara entre sí mediante BLASTP con un mínimo de identidad del 95% por defecto y volverá a hacer agrupaciones con MCL (Enright, Van Dongen y Ouzounis, 2002) con el fin de establecer aquellos genes que según los

parámetros se consideren parálogos. Finalmente, los anota, calcula la presencia/ausencia de genes ortólogos en cada genoma y realiza un recuento y clasificación en función de la cantidad de aislados que presenten dichos genes: si forman parte del *core* genómico estricto (presente en el 99% de las muestras) o relajado (95-99%) o del genoma accesorio (presencia inferior al 95% de las muestras). Adicionalmente obtiene otros resultados, como un árbol en el que se reflejan las agrupaciones o "cercanía" entre las muestras en función de los genes que comparten, lo que denomina un árbol de presencia/ausencia. En este estudio se han utilizado los valores por defecto y solicitando, además, la obtención de un alineamiento nucleotídico múltiple de todos los genes que el programa detecte que forman parte del *core*.

3.9 Estudio de recombinación

A la luz de los resultados obtenidos en el análisis del brote del HGUE, consideramos necesario estudiar con detalle la posibilidad de que se hubiesen producido fenómenos de recombinación que afectasen a las muestras estudiadas. Para ello, se analizaron individualmente los genes pertenecientes al *core* según ROARY. En primer lugar, se generaron los archivos multifasta de cada gen (~4000), extrayendo de cada cepa dicho gen según la tabla de correspondencias de ROARY, dado que los genes poseen una etiqueta propia de cepa y su numeración es diferente dependiendo del orden en que se hayan anotado. Posteriormente, se alinearon mediante CLUSTAL-OMEGA V1.2.1 (Sievers *et al.*, 2011) utilizando los parámetros por defecto.

Se reconstruyó el árbol de cada gen en aquellos casos en los que hubo señal suficiente, que se evaluó mediante el test de *likelihood mapping* utilizando IQTREE (Strimmer y von Haeseler, 1997). Este test determina la verosimilitud de las 3 topologías posibles para cada subgrupo de 4 secuencias, cuartetos, tomados entre la totalidad de las muestras. Las verosimilitudes obtenidas se representan gráficamente en un triángulo con 7 regiones claramente diferenciadas (Figura 3.5) de forma que a cada cuarteto le corresponde un punto en el interior del triángulo cuya distancia a cada uno de los vértices es inversamente proporcional a la verosimilitud que corresponde a cada una de las tres topologías posibles. Así, las regiones de los vértices (A1, A2 y A3) corresponden a los casos en que una de las topologías tiene una verosimilitud muy superior a las otras dos; en las zonas intermedias (A13, A23, A12), aquellos árboles que asignan verosimilitudes semejantes a 2 de las 3 topologías posibles; y, finalmente, en la zona central (A*), se sitúan los casos en que ninguno de los 3 árboles posibles tiene una

verosimilitud claramente mayor que el resto. Esto se realiza para cada cuarteto y la proporción de árboles en A1, A2 y A3 nos indicará si existe señal filogenética para dicho gen. En este caso se decidió hacer el test con 1500 cuartetos para que cada secuencia estuviera representada al menos 100 veces.

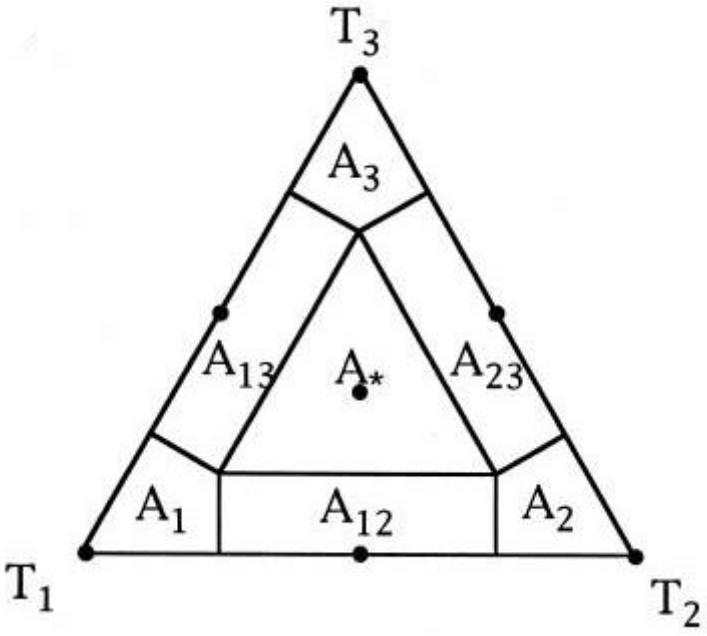


Figura 3.5. Representación gráfica del likelihood mapping. *Modificado de* (Strimmer y von Haeseler, 1997)

El test de topologías se realizó a *posteriori* sobre los árboles obtenidos con suficiente señal filogenética. Mediante este análisis comparamos el alineamiento de cada gen respecto al árbol del *core* genómico y el alineamiento del *core* respecto a los árboles de cada gen, de esta forma comprobaremos si hay congruencia entre ellos de manera que podamos detectar eventos de recombinación. Se hicieron 10000 réplicas para el cálculo de todos los *tests* incluidos en este programa: *bootstrap proportion* (BP), KH test (Kishino y Hasegawa, 1989), SH test (Shimodaira y Hasegawa, 1999), el *expected likelihood weights* (ELW) (Strimmer y Rambaut 2002), *weighted*-KH *and weighted*-SH *tests* y *approximately unbiased* (AU) *test* (Shimodaira 2002). En la tabla de resultados los p-valores nos indicarán si existe congruencia (p-valor > 0,05) o no.

3.10 Detección de resistencias

Los programas empleados anteriormente para realizar la clasificación de las cepas siguiendo el esquema de MLST a partir de lecturas (SRST2 y ARIBA) también se han

utilizado en la detección de genes de resistencias. Para cada programa se ha utilizado una base de datos distinta que contienen distintas colecciones de genes y mutaciones que confieren resistencia frente a una amplia variedad de antibióticos. En SRST2 se ha procedido a la detección de resistencias utilizando la base de datos ARG-Annot (Gupta *et al.*, 2014), que está incluida en el programa. En la versión que se utilizó para la detección en las muestras del HGUV (febrero de 2016) consta de 1683 entradas de genes y mutaciones, mientras que se actualizó a la versión r1 (julio 2016) para el estudio de las *P. aeruginosa* de Elche. Ésta pasó a contener 1654 registros tras la depuración de ciertas secuencias, además de la adición de genes relevantes como el gen *mcr-1* de resistencia a colistina. Por otro lado, se probó el sistema de detección de ARIBA lanzando las lecturas contra la base de datos CARD (*Comprehensive Antibiotic Resistance Database*) (McArthur *et al.*, 2013), descargada en noviembre de 2016 y formateada con ARIBA v2.5. Esta contiene en total 2221 registros, entre los que se encuentran genes de resistencia, variantes y cambios en regiones no codificantes. Se ejecutaron con los parámetros definidos por defecto para cada caso. Adicionalmente, se generó una base de datos propia para la detección de todas las posibles variantes del gen plasmídico de resistencia a colistina *mcr* con SRST2 para las muestras del HGUV, ante su ausencia en el análisis previo.

3.11 CRISPR

La detección de los espaciadores contenidos en loci CRISPR se ha realizado con el programa BLAST, combinado con la aplicación de varios filtros siguiendo la metodología aplicada en England *et al.* (2018) En primer lugar, se ejecuta BLASTn v2.2.31+ ajustando los parámetros para maximizar la cobertura tratándose se secuencias cortas, ya que las repeticiones son de 28 pb. Utiliza las opciones: -word _size 7 -gapopen 3 -gapextend 2 -reward 1 -penalty -1, siendo las lecturas de la muestra las que conforman la base de datos, y las repeticiones específicas de *P. aeruginosa* las secuencias "*query*". El filtrado posterior permite la existencia de secuencias repetitivas degeneradas con un mínimo de identidad del 80% y se conservaron los *hits* con un tamaño mínimo de 24 pb. Dado que las secuencias espaciadoras o *spacers* (Figura 1.7) tienen una longitud aproximada de 32-33 pb, se fijó el máximo en 40 pb para considerarlo como posible secuencia espaciadora, siempre que en ambos extremos flanqueantes de ésta se encontrara la secuencia repetitiva, ambas en el mismo sentido.

Además, contamos con la base de datos de 3152 espaciadores encontrados en esta especie gracias al trabajo de England *et al.* (2018), así como con las combinaciones de espaciadores encontrados por locus, que fueron un total de 878. Estos loci, a su vez, se organizaron en 729 *arrays*, entendiendo *array* como el conjunto de loci presentes en la cepa, ya que hay aislados que pueden tener más de un locus CRISPR aunque no siempre se encuentren los genes *cas* en su estructura. Las secuencias extraídas como posibles espaciadores fueron etiquetadas con su número correspondiente según la base de datos. Para el renombramiento se ejecutó el mismo comando de BLASTn contra la base de datos, pero utilizando solamente el multifasta de las secuencias espaciadoras. Únicamente se asignó nombre si el *best hit* cubría completamente la secuencia para poder detectar nuevos registros.

Se revisó la pérdida de espaciadores tras el filtrado utilizando el comando de BLASTn con las mismas opciones. Se volvió a lanzar el proceso partir de las lecturas de cada aislado, sustituyendo las secuencias de ADN repetitivo del primer método por una base de datos en la que se incorporaron los espaciadores detectados entre todas las muestras de los 3 hospitales.

Por último, se estableció el orden de los espaciadores dentro de cada locus y su correspondencia con los *arrays* descritos en el trabajo de England *et al.* (2018). Ante el elevado número de muestras y la variabilidad de coberturas, se seleccionó una muestra representativa por paciente de las muestras del Hospital General de Valencia y las muestras con variedad en contenido de espaciadores de las muestras de los brotes. Se extrajeron de los genomas, que previamente habíamos ensamblado y anotado, aquellas regiones detectadas por PROKKA como *rpt_family="CRISPR"*, correspondiente a la región con repeticiones de un locus CRISPR. Dichas estructuras fueron analizadas nuevamente con BLASTn, pero utilizando todos los espaciadores que habíamos detectado en el análisis inicial como base de datos, incluso los nuevos encontrados pendientes de confirmar.

3.12 Análisis filogenético conjunto de los aislados de los 3 hospitales

La construcción del árbol filogenético del total de aislados de *P. aeruginosa* procedentes de los 3 centros hospitalarios colaboradores se realizó con los genomas reconstruidos a partir de mapeos. Dado que se utilizó una referencia distinta para el

mapeo de los aislados del Hospital General Universitario de Valencia respecto a los otros dos centros, primero se realizó un alineamiento con PROGRESSIVEMAUVE (Darling, Mau y Perna, 2009) de las dos cepas de referencia (H27930 y W16407). A partir de la reorganización de los bloques que devuelve este programa, se incorporaron al alineamiento los genomas de todos los aislados con el mismo reordenamiento por bloques que su referencia utilizando un script propio (Mat.Supl). El orden se mantuvo, ya que los pseudogenomas generados tras el mapeo tienen la misma estructura y longitud que la secuencia de referencia (ver sección 1.1.5).

El árbol se reconstruyó con IQTREE de la misma manera que con los sets de datos por hospital a partir de los SNPs extraídos con SNP-SITES. Finalmente, se hicieron cálculos de *core* en conjunto y por hospital con el programa BMGE v1.12 (Criscuolo y Gribaldo, 2010), eliminando las regiones con más del 10% de gaps y conservadas al 95%.

4. Resultados

Capítulo 1

4.1 Brote del Hospital Arnau de Vilanova

El primer análisis del presente trabajo corresponde a un posible brote de *P. aeruginosa* detectado en el Hospital Arnau de Vilanova (HAV) de Valencia. Se obtuvieron muestras de 6 pacientes afectados por diferentes enfermedades oncohematológicas, hospitalizados en el Servicio de Hematología en 2016 y 2017 (Tabla 4.1). El aumento de infecciones causadas por este patógeno, aun tratándose de un número bajo de casos y dispersos en el tiempo (más de un año), implicaba un riesgo especialmente alto ante la posibilidad de confirmarse un brote debido a la enfermedad de base de los pacientes afectados, por lo que el principal objetivo de este análisis es establecer si ciertamente se trata de un brote y así puedan tomarse las medidas oportunas. Hay cierta variabilidad en el número de muestras recogidas por paciente y en los tipos de muestra de los que se obtuvieron los cultivos.

Paciente	Muestra	Fecha	Origen	Patología
1	PS21	02/05/2016	Rectal	Linfoma
2	PS03	29/09/2016	Orina	LMA
3	PS06	15/05/2017	Rectal	MM
3	PS09	29/05/2017	Faríngeo	MM
4	PS07	24/05/2017	Faríngeo	LMA
4	PS08	29/05/2017	Rectal	LMA
4	PS10	19/06/2017	Hemo	LMA
4	PS11	06/07/2017	Rectal	LMA
5	PS12	06/07/2017	Catéter	LMA
5	PS13	10/07/2017	Hemo	LMA
5	PS20	03/07/2017	Faríngeo	LMA
6	PS17	24/08/2017	Orina	LMA

Tabla 4.1. Muestras analizadas procedentes del HAV. *LMA: Leucemia Mieloide Aguda; MM: Mieloma Múltiple. Se incluye información de la fecha de aislamiento y tipo de muestra biológica de la cual se obtuvo el cultivo.*

4.1.1 Secuenciación y mapeo

Inicialmente, el total de lecturas extraídas del secuenciador para este conjunto de aislados fue de 26 millones y, tras la limpieza, se redujeron a 22,8 millones de lecturas (Material suplementario, Tabla 7.2). También se vio afectado el tamaño medio de lectura, que disminuyó un promedio de 80 pb sobre las ~280 pb de media de las lecturas brutas. La reducción fue más acusada en las lecturas de uno de los extremos del *paired-end* para todas las secuencias (Figura 4.1). Esto se debió a que las lecturas de dicho extremo cubrían una región más extensa del adaptador empleado en la secuenciación y tuvo que ser eliminado.

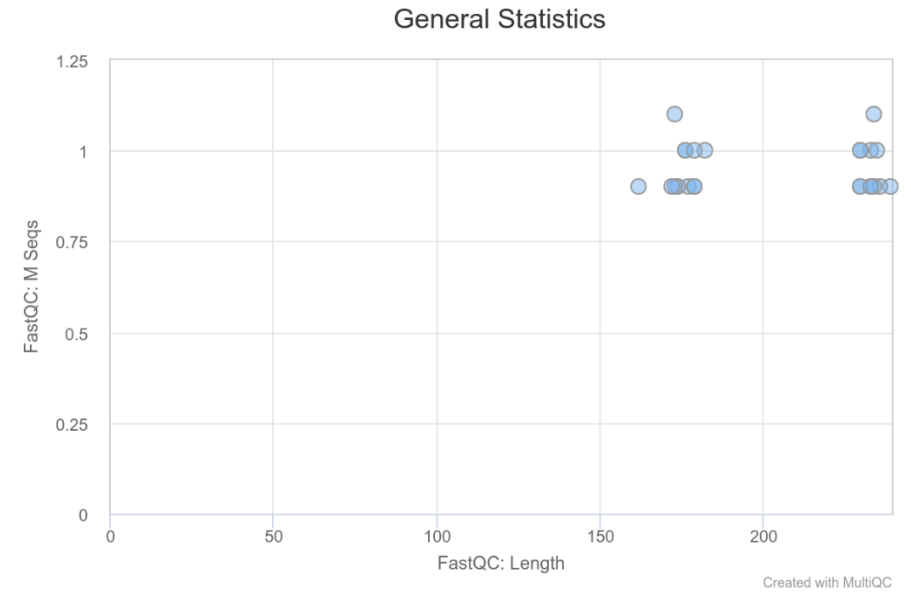


Figura 4.1. Número de secuencias (en millones) frente a la longitud de las lecturas en las 12 muestras del HAV. *Resumen con MultiQC de las estadísticas obtenidas con FastQC tras la limpieza de las muestras.*

Los ~2 millones de lecturas por muestra (~1 millón por cada extremo o *paired-end*), que tienen una calidad superior a un *Phred* de 30 a lo largo de todas las posiciones (Figura 4.2) y sin presencia de adaptadores, pasaron al paso de determinación del *Sequence Type* (ST) y posterior mapeo.

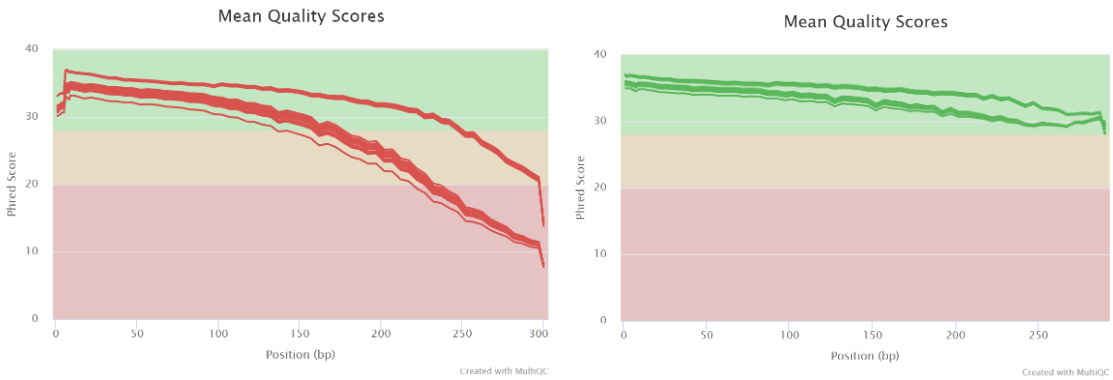


Figura 4.2. Calidad medida según la escala Phred en cada posición de las lecturas antes y después de la limpieza. *Se encuentran representadas los* paired-ends *de todas las muestras.*

Los resultados de tipado con SRST2 indicaron que todas las cepas pertenecían al ST175, por lo que se eligió la misma referencia de mapeo que la utilizada con los genomas procedentes del Hospital General Universitario de Elche, la cepa H27930 (BioSample: SAMN02894354). La cobertura media del conjunto fue de 53 lecturas por sitio, sin grandes variaciones, siendo el máximo de cobertura media de 59x y el mínimo de 48x, con desviaciones típicas del orden de 17x (Tabla 4.2). Además, en todos los casos, aproximadamente el 95% de las lecturas de cada muestra mapeaba frente a la referencia escogida, por lo que se continuó el análisis. A partir del mapeo se determinó la secuencia de cada cepa generando un "pseudogenoma" de la misma longitud que la secuencia de referencia, 6.598.022 pb.

Aplicando los filtros descritos en Material y Métodos (3.6), se identificó para cada posición si esta se correspondía con la misma base que la de la referencia, de una variante o de una posición indeterminada (N). Este protocolo permite evitar realizar el paso de alineamiento de genomas, puesto que todos tienen la misma longitud y solamente se comparan las posiciones encontradas en la referencia. Por ello, el concatenado de todas las muestras en un único fichero con formato multifasta contiene el alineamiento para el post-procesado.

Paciente	Muestra	Cobertura media	Desviación típica	Total lecturas	% lecturas mapeadas	% genoma cubierto
1	PS21	59,78	19,45	2.116.922	90,02	95,33
2	PS03	51,49	16,16	1.824.050	90,37	95,32
3	PS06	51,64	16,04	1.814.255	90,35	95,32
3	PS09	57,6	18,05	2.018.723	90,52	95,34
4	PS07	49,08	15,27	1.762.502	90,37	95,33
4	PS08	53,64	16,26	1.878.157	90,35	95,35
4	PS10	54,3	17,46	1.933.442	89,78	95,32
4	PS11	49,79	16,05	1.744.383	90,49	95,32
5	PS12	48,86	15,46	1.697.683	89,98	95,32
5	PS13	59,25	17,79	2.051.515	89,71	95,35
5	PS20	48,19	16,02	1.726.097	88,73	95,3
6	PS17	54,32	18,5	1.946.061	88,97	95,33

Tabla 4.2. Resultados de cobertura y porcentaje de lecturas mapeadas frente a la cepa H27930. *El porcentaje de genoma cubierto hace referencia a las posiciones del genoma de referencia que han podido ser determinadas con dicho porcentaje de lecturas que mapeaban.*

Tras el enmascarado de regiones repetitivas se determinó que el *core* de las 12 muestras clínicas es de 6.047.994 pb de las 6.598.022 pb totales que tiene el genoma de referencia, considerando como *core* únicamente las posiciones que estén presentes en todas las muestras. Se extrajeron las posiciones variantes con la función *seg.sites* del paquete de R *ape*, utilizando el alineamiento completo y sin tener en cuenta las repeticiones. Obtuvimos un alineamiento de SNPs con una longitud de 22 pb.

PS03											
PS06	3										
PS09	2	1									
PS07	8	9	8								
PS08	8	9	8	0							
PS10	8	9	8	0	0						
PS11	8	9	8	0	0	0					
PS12	2	3	2	8	8	8	8				
PS13	2	3	2	8	8	8	8	0			
PS20	2	3	2	8	8	8	8	0	0		
PS17	8	9	8	10	10	10	10	8	8	8	
PS21	1	2	1	7	7	7	7	1	1	7	1

Tabla 4.3. Estimas de la divergencia entre cepas del HAV. *Se muestran el número de diferencias por pares de secuencias. Todas las posiciones con indeterminaciones (N) fueron eliminadas. En total, se han incluido 16 posiciones del alineamiento de 22 SNPs. Análisis generado con MEGAX. Las casillas de las muestras se han coloreado por paciente para facilitar la interpretación.*

A continuación, se comparó el número de cambios entre aislados para observar el grado de divergencia dentro de paciente y entre pacientes. Como se observa en la matriz realizada con el programa MEGA (Tabla 4.3), en la que se comparan solamente las 16 posiciones compartidas (sin indeterminaciones en ninguna posición) del alineamiento de 22 SNPs, aquellas muestras procedentes del mismo paciente tienen 0 o 1 diferencias al ser comparadas por pares, lo cual era esperable al haberse tomado las muestras en un breve espacio de tiempo y tratándose *a priori* de una infección con una sola cepa.

4.1.2 Análisis filogenético

Estos resultados se ven reflejados en el árbol filogenético construido con IQTREE a partir del alineamiento con las posiciones variantes (Figura 4.3). El modelo de sustitución nucleotídica seleccionado como más adecuado por MODELFINDER durante el proceso es el K2P (ó K80), aunque debemos tener presente la necesidad de una corrección por la ausencia de posiciones constantes que sobreestimaría la longitud de las ramas.

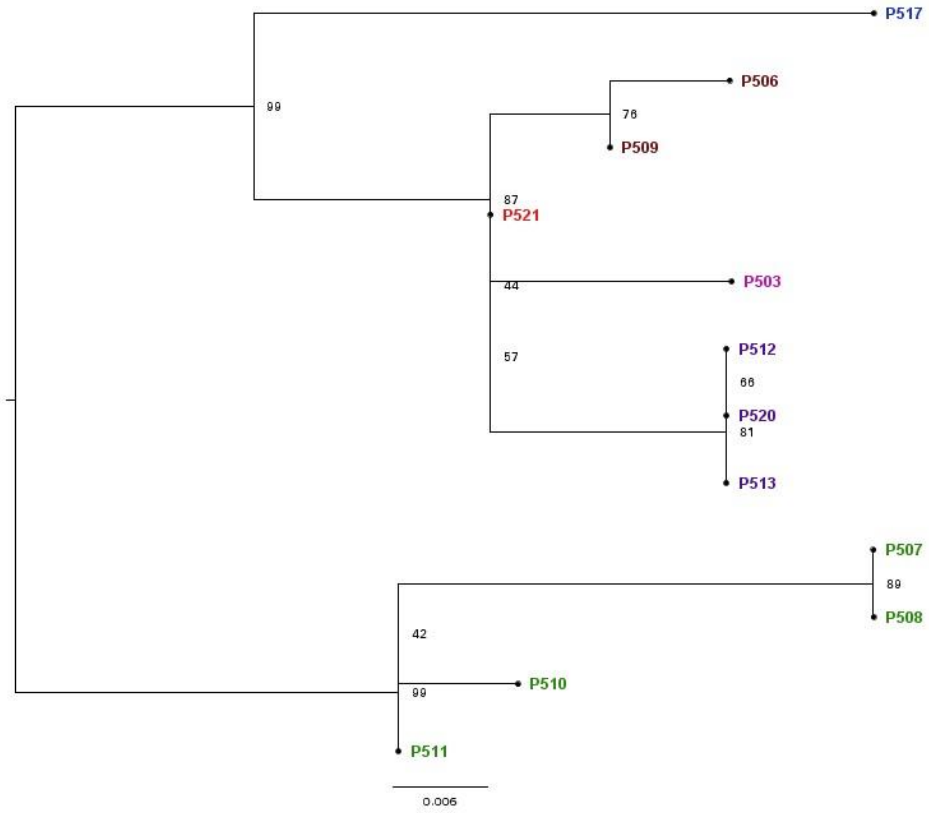


Figura 4.3. Árbol filogenético del brote del HAV construido con IQTREE a partir del alineamiento de 22 SNPs. *Modelo de sustitución empleado: K2P+ASC. El árbol se ha enraizado en el punto medio. Los colores de los extremos corresponden con los diferentes pacientes (orden de arriba a abajo): verde, paciente 4; azul, paciente 6; granate, paciente 3; rojo, paciente 1; rosa, paciente 2; morado, paciente 5.*

Tal y como se vio en la matriz de diferencias, en el árbol queda patente la baja variabilidad entre las 12 cepas estudiadas (<0,0004% del *core*) con una escala fijada en 0,006 sustituciones por sitio. Esto, sumado a la proximidad en el tiempo de las muestras recogidas, lleva a la obtención de una señal filogenética muy baja para el análisis de reloj molecular, como hemos visto en el estudio evolutivo (3.7). Sin embargo, las agrupaciones obtenidas permiten comprobar cómo las diferentes muestras de un mismo paciente conforman clados o grupos monofiléticos: es el caso de los pacientes 4 (PS07, PS08, PS10 y PS11, marcados en verde), 5 (PS12, PS13 y PS20, en morado) y 3 (PS06 y PS09, en granate), reforzando la idea de que han sido infectados por una única cepa. Además, el clado más grande, que contiene las muestras de 4 de los 6 pacientes, es coherente en cuanto a las fechas de aislamiento: la muestra más próxima al ancestro común del clado (PS21) fue la primera en aislarse, en 2016.

4.1.3 Análisis evolutivo

El análisis con TEMPEST de la correlación entre la fecha en que se aislaron las cepas de *P. aeruginosa* de cada paciente y su divergencia respecto al ancestro indica que la señal temporal es limitada. Reestructurando el árbol a la raíz con mejor ajuste, obtenemos una recta de regresión con un coeficiente de determinación bajo (R^2=0,2487) (Figura 4.4), aunque suficiente para probar a realizar un análisis utilizando modelos sencillos mediante BEAST.

Se han realizado los cálculos por triplicado a partir del alineamiento de 22 SNPs, incorporando aparte el número de posiciones constantes (A, 1.014.681; C, 2.014.315; G, 2.009.939; T, 1.015.661) con los parámetros y modelos especificados en el Material y Métodos (3.7).

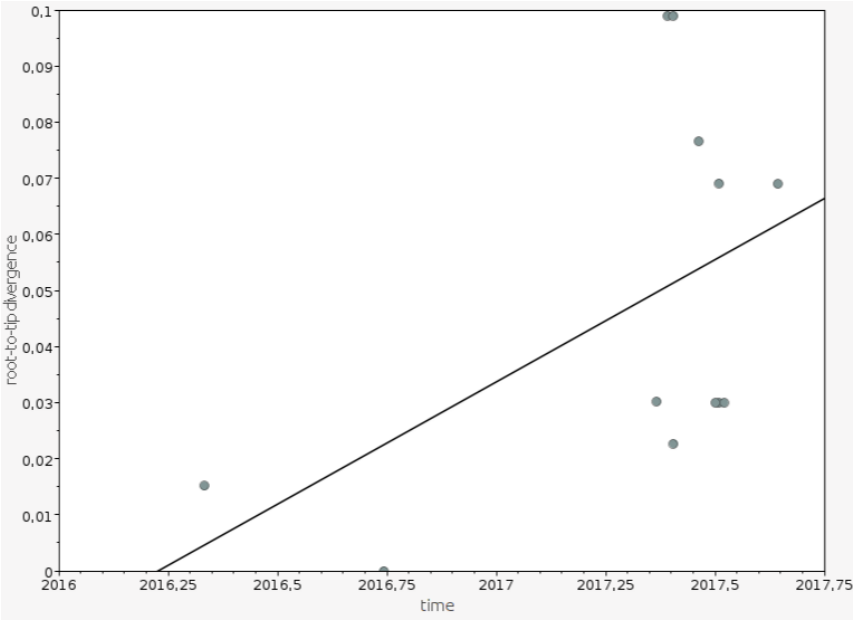


Figura 4.4. Recta de regresión a partir de los datos de toma de muestra y la divergencia de las muestras con respecto al ancestro. *Gráfica obtenida con TempEst a partir del árbol de SNPs y las fechas correspondientes. R^2= 0,2487.*

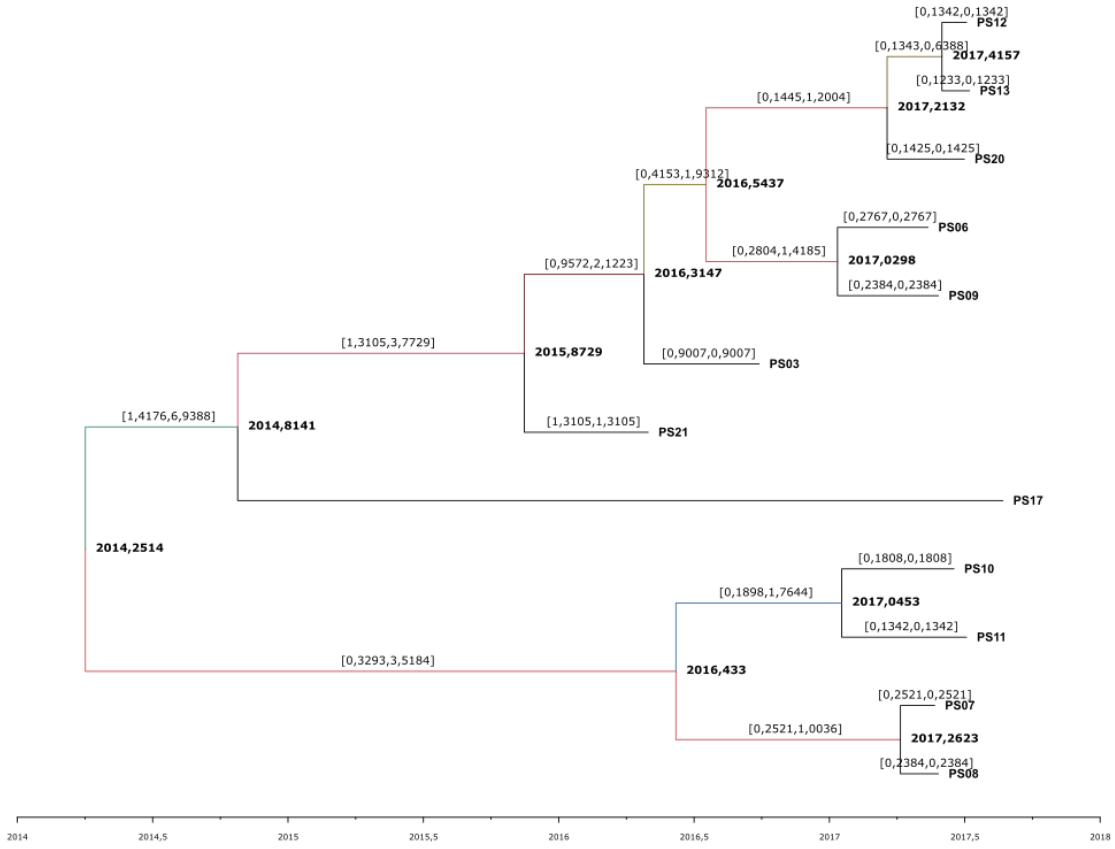


Figura 4.5. Árbol filogenético consenso obtenido por BEAST a partir de las 3 réplicas. *Se representa utilizando una escala temporal con la fecha estimada del ancestro situado en cada nodo y el intervalo de años entre un nodo y el anterior en las ramas. El color de las ramas se corresponde con el valor de la probabilidad posterior.*

A partir de la reconstrucción (Figura 4.5), los parámetros indican que el ancestro común a todas estas muestras sería de hace 3,39 años, es decir, de abril de 2014, teniendo en cuenta que el último aislado recogido tiene fecha de finales de agosto de 2017. Sin embargo, el intervalo de alta densidad posterior al 95% es amplio [1,31, 8,87], por lo que las fechas varían entre mayo de 2016 y octubre de 2008. Dada la reducida señal filogenética de los aislados, no podemos dar por definitivos estos valores, que deben considerarse como las mejores aproximaciones con los datos disponibles. La tasa de evolución obtenida es de 2,98E-7 s/s/a.

4.1.4 Resistencias

La detección de resistencias mediante ARIBA permitió identificar 50 grupos de genes y mutaciones que confieren resistencia a antibióticos, de los cuales 43 son compartidos por las 12 muestras estudiadas. Las agrupaciones las realiza el programa por similitud entre variantes con CDHIT como, por ejemplo, ocurre con el grupo OXA-6 en el que se encuentran 9 alelos de este gen, uno de ellos el OXA-415 detectado en estos aislados de *P. aeruginosa*.

A partir de los 7 cambios restantes, es decir, de los 7 grupos de genes que son detectados en unas cepas y no en otras, se ha generado un dendrograma de presencia-ausencia que agrupa las cepas más similares según grupos compartidos (Figura 4.6).

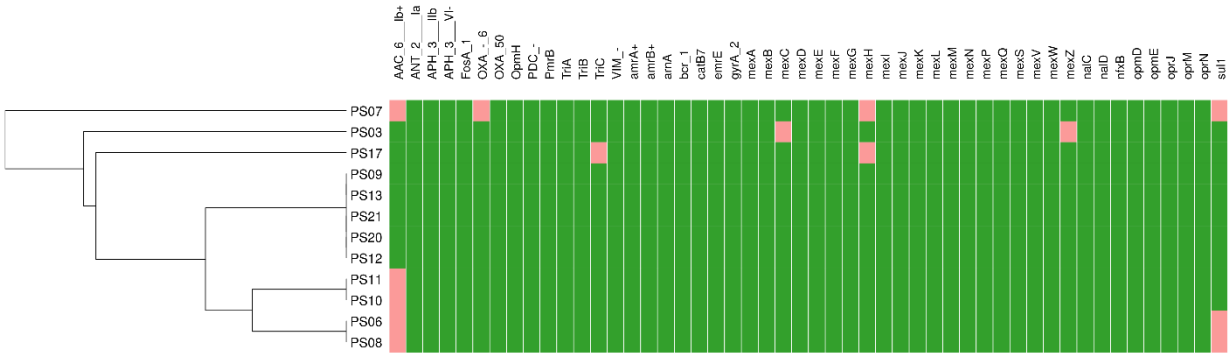


Figura 4.6. Árbol de presencia-ausencia del brote del HAV basado en el contenido en genes o mutaciones causantes de resistencias tras el análisis con ARIBA. *En verde se indican los cambios presentes, los de color rosa son cambios ausentes en esa cepa.*

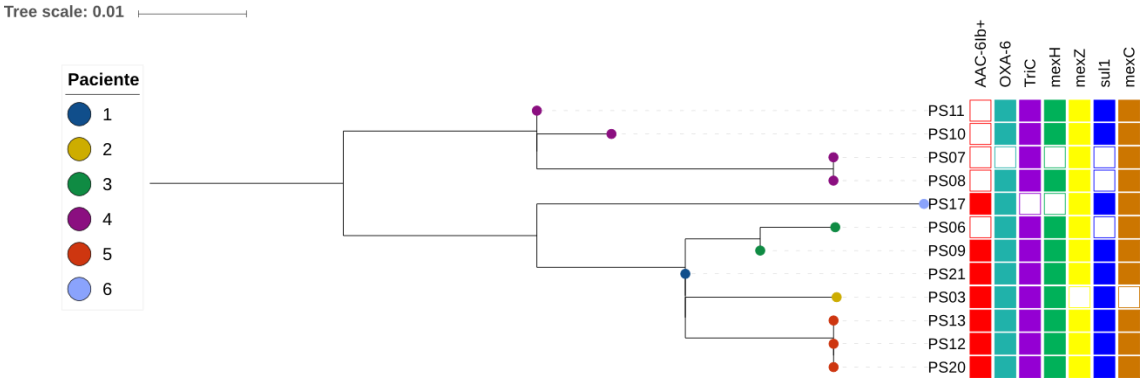


Figura 4.7. Árbol filogenético de SNPs (IQTREE) con metadata. *La información añadida en columnas corresponde con cambios en 7 determinantes antibióticos (presencia del gen o mutación en el mismo), en blanco aquellos que están ausentes. Los círculos de colores indican el paciente al que corresponde cada muestra.*

Entre los 7 grupos, representados también en el árbol filogenético de SNPs (Figura 4.7), hay genes que suelen encontrarse en integrones asociados a otras resistencias (AAC-6Ib y *sulf1*) o plásmidos (OXA) contrariamente al resto (*mexC*, *mexH*, *mexZ* y *TriC*) que, además, salvo *mexZ*, están involucrados en sistemas de resistencia a múltiples antibióticos por bombas de flujo. Observamos que en los pacientes 3 y 4 hay una acumulación de mecanismos de resistencia en muestras sucesivas, lo que sería compatible con la consideración previa de que se trata de una infección con una única cepa que ha ido adaptándose a los tratamientos y/o intercambiando material genético con cepas circulantes, o que exista una coinfección y dicha bacteria sea la que haya transferido los elementos móviles, pudiendo ser incluso de otra especie.

Las cepas contienen genes de resistencia frente a antibióticos de todo tipo: betalactámicos (*OXA*, *PDC*, *VIM*), fosfomicina (*FosA*), fluoroquinolonas (mutación en *gyrA*), aminoglicósidos (*APH(3')-VI*), colistina (cambios en *pmrA* y *pmrB*), entre otros, gracias también a cambios en reguladores y proteínas que intervienen en bombas de flujo que permiten evitar la acción de antibióticos cuyo método de acción implica a un ligando intracelular.

Comparando estos resultados con los perfiles fenotípicos más significativos (Material suplementario, Tabla 7.1) vemos como no hay coincidencia total entre los grupos de determinantes detectados y el fenotipo. Un ejemplo claro es la mutación del gen *pmrB* que confiere resistencia a colistina, presente en todas las muestras y, en cambio, el perfil fenotípico de la muestra PS21 frente a colistina da resultado sensible.

Capítulo 2

4.2 Brote del Hospital General Universitario de Elche

Se sospechó la presencia de un brote de *P. aeruginosa* multirresistente en el Hospital General Universitario de Elche (HGUE) que podría haber comenzado en el Servicio de Nefrología a finales de 2014 o inicios de 2015. El objetivo en este estudio es confirmar si realmente existe un brote y determinar, en ese caso, si hay pacientes bajo sospecha que no pertenezcan al mismo para determinar el alcance. Además, la disponibilidad de algunas muestras ambientales nos permite indagar sobre el posible origen o reservorio de la cepa (o cepas) causantes del brote.

Desde el Servicio de Microbiología nos fueron remitidas 73 muestras de *P. aeruginosa* en dos fases, en su mayoría aisladas a partir de muestras de orina. Inicialmente comenzó el estudio con 57 muestras de 2015, de las cuales 5 eran aislados considerados ambientales ya que procedían de grifería del área, y las 52 muestras restantes correspondían a aislados clínicos, uno por paciente (Material suplementario, Tabla 7.4). Posteriormente, se recibieron 16 muestras clínicas adicionales de 12 pacientes distintos de 2016 dada la posibilidad que el brote se mantuviera activo (Material suplementario, Tabla 7.5). En total, el número de pacientes sospechosos de estar vinculados a un brote por *P. aeruginosa* fue de 64.

4.2.1 Secuenciación masiva y evaluación inicial de las secuencias

Los resultados previos a la limpieza de las lecturas obtenidas por Illumina (2x300 pb) (Material suplementario, Tabla 7.6) indicaron que el tamaño medio de las lecturas era similar entre las diferentes muestras, pero había cambios en el contenido en GC, disminuyendo respecto al 65% esperado. El caso más llamativo es el de los *paired-ends* de la muestra Elche_58, con un 42% de contenido en GC (Figura 4.8). Para comprobar la posible contaminación con otros organismos, se procedió al análisis de las lecturas para su determinación a nivel taxonómico. Según los resultados de KRAKEN, 9 de las muestras clínicas no se corresponden con *P. aeruginosa:* detectamos 6 *Pseudomonas putida*, una *E. coli*, una *Stenotrophomonas maltophilia* y una *Providencia stuartii*, de manera que fueron eliminadas del análisis. Además, tuvo que descartarse la muestra Elche_19 por el

bajo número de lecturas obtenido (513). Las muestras Elche_07, 15 y 23, a pesar del bajo número de lecturas se mantuvieron, aunque con ciertas reservas. Por tanto, el total de pacientes estudiados se redujo a 54.

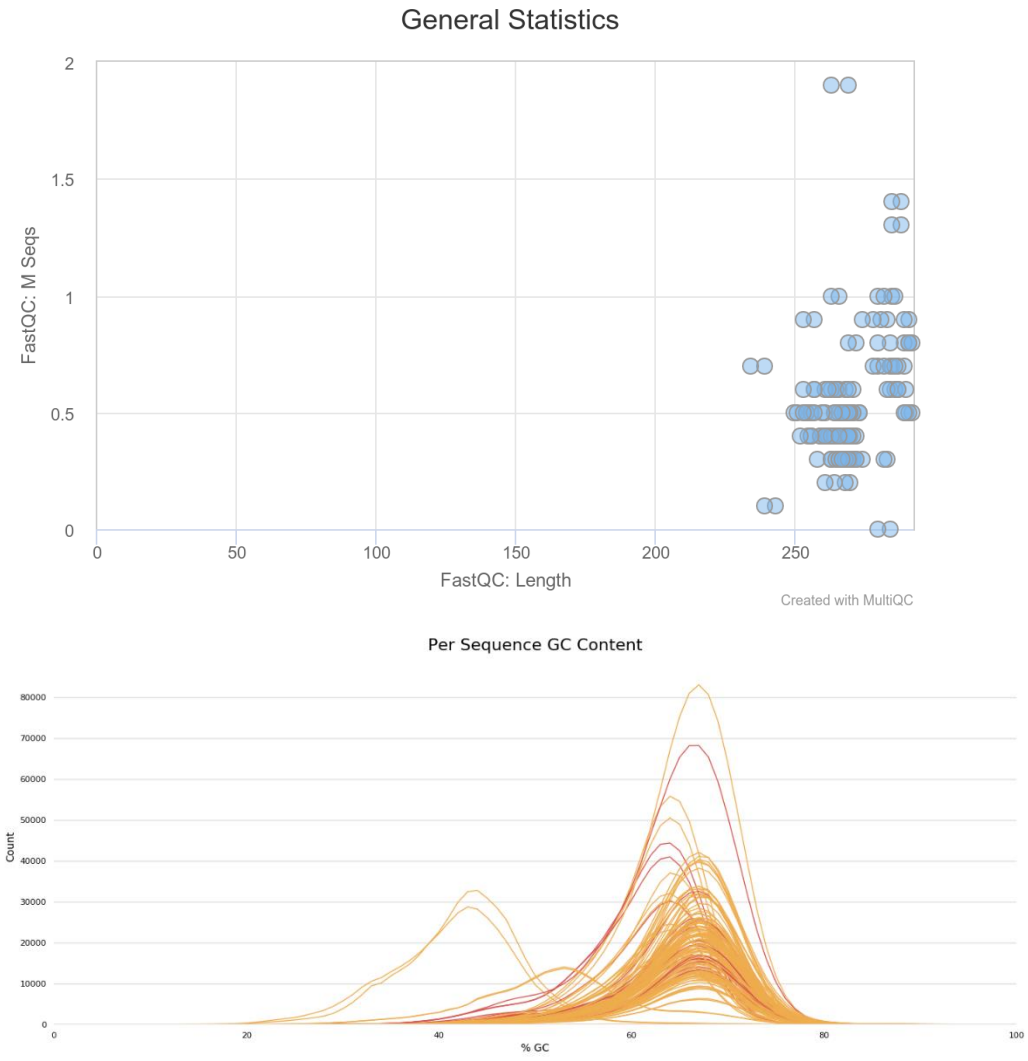


Figura 4.8. Recuento de millones de lecturas por* paired-ends *respecto a su longitud media antes de la limpieza en el total de muestras procedentes del HGUE (arriba); Contenido en %GC frente al número de lecturas para cada* paired-end *(abajo).

4.2.2 Mapeo y obtención de las secuencias de cada muestra

La similitud por *Sequence Type* fue el criterio que se utilizó para seleccionar la cepa más próxima como referencia de mapeo que nos permitiera obtener el máximo de información para todo el conjunto de aislados. El ST mayoritario según el análisis con SRST2 fue el ST175. Sin embargo, entre las 71 cepas de *P. aeruginosa* de las que se dispuso una secuencia genómica completa (Material suplementario, Tabla 7.13) no se encontró ninguna perteneciente a este ST, de manera que se seleccionó aquella con la que compartía el mayor número de alelos de los 7 genes que forman parte de este esquema de MLST. La cepa H27930 con ST389, que comparte 4 alelos con el ST175, fue elegida como cepa de referencia para mapeo. A pesar de ello, la mayoría de muestras cubrían el 95% del genoma; solamente una muestra obtuvo un 88% de posiciones cubiertas (Tabla 4.4)

Muestra	Cobertura Media	Desviación típica	Total lecturas	% lecturas mapeadas	% genoma cubierto
Elche_01	44,24	16,89	1.604.476	90,58	95,92
Elche_02	26,17	10,86	952.200	90,05	95,83
Elche_03	31,3	12,49	1.146.041	89,71	95,87
Elche_04	10,9	5,52	389.062	89,75	93,72
Elche_05	31,52	12,5	1.101.622	90,1	95,91
Elche_06	27,11	11,19	972.109	90,22	95,85
Elche_07	6,58	3,79	240.449	90,12	83,24
Elche_08	30,63	12,19	1.103.728	89,22	95,88
Elche_09	28,13	11,54	984.981	89,8	95,83
Elche_10	21,52	9,15	776.955	90,56	95,69
Elche_11	21,44	9,43	763.577	90,44	95,78
Elche_12	26,71	10,98	946.166	90,04	95,86
Elche_13	19,14	8,12	680.409	89,34	95,72
Elche_14	15,48	6,89	550.734	89,44	95,39
Elche_15	13,18	6,39	471.094	88,95	94,19
Elche_16	23,64	10,83	844.191	89,56	95,76
Elche_17	27,58	11,74	969.239	89,32	95,86
Elche_18	17,67	8,08	636.265	87,24	95,58
Elche_21	19,83	8,84	694.076	89,37	95,68
Elche_22	16,25	7,18	587.533	88,37	95,43
Elche_23	11,64	5,62	416.538	89,39	94,42
Elche_24	17,13	8,42	601.662	90,13	94,83
Elche_25	20,68	8,79	742.756	89,58	95,61
Elche_26	16,2	7,69	578.030	89,35	95,39
Elche_27	23,13	10,58	833.310	89,77	95,75
Elche_28	14,98	7,06	535.480	88,69	95,21

Muestra	Cobertura Media	Desviación típica	Total lecturas	% lecturas mapeadas	% genoma cubierto
Elche_30	15,94	7,08	570.083	89,03	95,46
Elche_32	19,38	8,08	700.241	89,57	95,73
Elche_33	25,03	10,13	919.168	84,88	92,13
Elche_34	20,12	8,52	668.846	93,35	91,99
Elche_35	24,17	9,63	867.733	89,49	95,81
Elche_36	18,56	8,49	685.876	89,05	94,93
Elche_37	28,21	11,62	1.019.043	90,94	94,01
Elche_38	23,33	9,02	868.487	88,46	95,16
Elche_39	22,45	8,91	862.155	88,83	95,76
Elche_40	15,49	7,42	590.722	83,72	91,04
Elche_41	21,17	8,79	795.728	87,33	95,75
Elche_42	19,66	8,29	698.120	89,64	91,41
Elche_43	26,34	10,61	942.948	90,38	95,21
Elche_45	16,63	6,92	611.395	88,22	95,56
Elche_46	31,51	10,59	1.264.767	88,86	95,87
Elche_47	24,69	8,91	946.752	88,98	95,82
Elche_48	26,51	11,08	964.695	89,28	95,85
Elche_49	35,33	15,54	1.242.638	88,9	95,33
Elche_50	32,67	14,43	1.175.841	89,78	95,92
Elche_51	48,63	22,89	1.803.141	87,87	96,02
Elche_52	13,76	6,64	550.874	86,57	94,88
Elche_54	51,06	15,31	1.690.989	90,04	90,57
Elche_57	40,64	11,11	1.428.829	86,77	95,18
Elche_59	37,23	12,54	1.279.374	90,05	95,28
Elche_60	26,77	8,39	966.222	89,48	95,2
Elche_62	22,65	10,45	800.605	83,26	88,57
Elche_63	41,01	12,81	1.368.634	90,31	90,54
Elche_64	28,65	9,85	1.089.463	80,57	91,01
Elche_65	92,01	27,55	3.429.457	88,47	95,39
Elche_66	24,1	8,45	851.589	89,2	95,86
Elche_67	45,2	14,01	1.635.293	89,35	95,96
Elche_68	40,39	14,03	1.380.232	90,43	95,34
Elche_1A	31,49	13,92	1.111.512	89,02	95,91
Elche_2A	15,6	7,23	545.678	89,1	95,45
Elche_3A	34,34	14,72	1.189.811	89,21	95,95
Elche_4A	27,48	10,4	982.179	88,82	95,89
Elche_5A	45,14	20,35	1.573.648	88,49	96,01

Tabla 4.4. Estadísticas de mapeo de las muestras del HGUE. *En la tabla se muestra la cobertura media, es decir, el número de lecturas promedio que cubren una posición del genoma de referencia, además de su desviación típica y el número total de lecturas de la muestra (suma de los* paired ends*). Se han incluido además los porcentajes de lecturas que mapean contra la referencia elegida y de sitios cubiertos por estas lecturas respecto al total del genoma (6.598.022 pb) antes de descartar el repetitivo.*

El alineamiento tras el mapeo tiene 6.598.022 pb, tantas como bases posee la referencia empleada; sin embargo, el número de posiciones conservadas en el 95% de las cepas es de 4.290.736 pb. Podría considerarse un número bajo de posiciones conservadas, pero los criterios con que se han descartado posiciones son elevados para evitar falsos variantes. Entre las posiciones del alineamiento que no se contabilizan (indeterminaciones) se encuentran las deleciones y aquellas posiciones que no pasan el filtrado de calidad para la correcta asignación de base, sumadas a las posiciones eliminadas (no consideradas) por baja cobertura, bien a nivel de todo el genoma o por regiones dependiendo de la muestra, además de las regiones repetitivas que también se eliminaron una vez obtenido el alineamiento.

4.2.3 Análisis filogenético

El árbol se reconstruyó con IQTREE a partir del alineamiento de SNPs extraído del alineamiento procedente del mapeo. Solamente se consideraron posiciones variantes aquellas que tenían 2 o más bases distintas, no teniendo en cuenta como variante las indeterminaciones (N). El alineamiento de SNPs resultante de las 63 muestras clínicas junto con la referencia contiene 106.627 pb. Según MODELFINDER, el modelo más adecuado para el alineamiento aportado es el GTR+ASC+R5, es decir, un modelo general reversible con corrección del sesgo, dado que solo se incluyen posiciones variables, y un modelo de distribución gamma relajado con 5 categorías, lo cual que se recomienda para conjuntos de datos grandes.

La estructura del árbol (Figura 4.9) indica que hay un clado grande que contiene la mayor parte de las muestras (54 en total) y que, *a priori*, conformarían el brote. Las 8 muestras restantes, que provienen de 6 pacientes, quedarían descartadas, si bien se observa un posible origen común entre las muestras Elche_33 y Elche_34 con ST1212, ya que únicamente se diferencian en 3 SNPs. Asimismo, se ha detectado un posible cruce en el etiquetado en las muestras Elche54, Elche62, Elche63 y Elche64, que provienen de 2 pacientes. Se indican en el árbol con etiqueta verde y azul las parejas de muestras de dichos pacientes. Según la agrupación en los distintos clados, cada paciente tendría una muestra del ST699 y del ST348; sin embargo, la amplia distancia filogenética que existe entre estos clados hace que sea muy baja la probabilidad de que ambos pacientes presenten dicha variación entre muestras o que, incluso, presentaran la misma coinfección. En cambio, las etiquetas de las muestras son consecutivas (63 y 64) por lo que esta explicación sería más razonable.

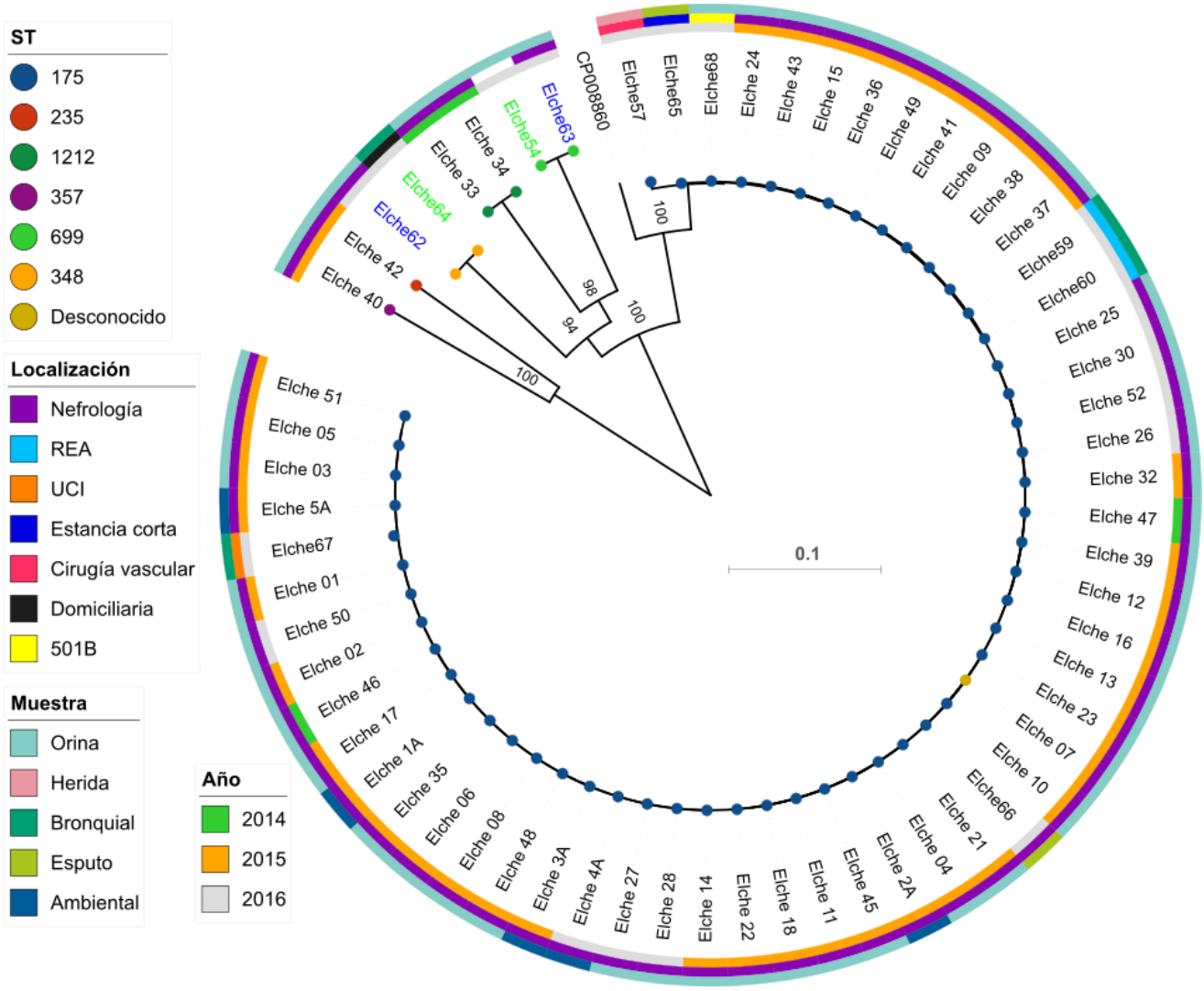


Figura 4.9. Árbol filogenético de los 63 aislados analizados del HGUE junto a la referencia de mapeo construido con IQTREE a partir del alineamiento de SNPs. *Modelo aplicado: GTR+ASC+R5. Alineamiento de SNPs de 106.627 pb. Las muestras marcadas con etiquetas verdes y azules corresponden con 2 pacientes, son muestras de las que se sospecha un error de etiquetado. Los extremos de las ramas se presentan puntos de diferentes colores en función del ST de la muestra; los círculos exteriores que rodean el árbol representan (de fuera hacia adentro) tipo de muestra, localización en el momento de la toma y año, correspondiendo el código de colores con su respectiva leyenda. Los soportes bootstrap de las 6 muestras ajenas al brote no indicados son de 100. Las del brote no se representan para evitar la superposición de los datos, variando estos entre 100 y 14 de bootstrap.*

En el clado grande, el del posible brote, el número de SNPs es de 2.063, <2% de la variabilidad del alineamiento de todas las muestras. Representando únicamente este clado para aumentar la escala (Figura 4.10), vemos que existen diferencias considerables entre Elche57, Elche67 y el resto de muestras, a pesar de pertenecer, como el resto, al

ST175. La matriz de diferencias (Material suplementario, Tabla 7.7) en las posiciones presentes en al menos el 95% de los aislados indica que hay un máximo de 1196 SNPs entre pares de secuencias y un mínimo de 0. Sin embargo, eliminando del análisis las muestras Elche57 y Elche67, el número total de variantes se reduce a 576 con un máximo de 125 SNPs entre dos muestras. La reducción en un orden de magnitud del número de variantes evidencia la amplia distancia que existe entre los dos aislados y el resto del clado, por lo que se descartarán para el estudio del brote.

Hay una elevada variabilidad en el clado, al margen de las dos muestras indicadas, y no existe una relación directa entre su distribución a lo largo del árbol filogenético con el origen, si bien es cierto que los casos externos al Área de Urología son puntuales y mayormente quedan fuera de brote, la fecha de muestreo o el tipo de muestra. Esto podría deberse a que su ancestro común se haya dispersado por diferentes dependencias del hospital mucho tiempo antes de la aparición del brote o la introducción de alguna cepa comunitaria de este ST175.

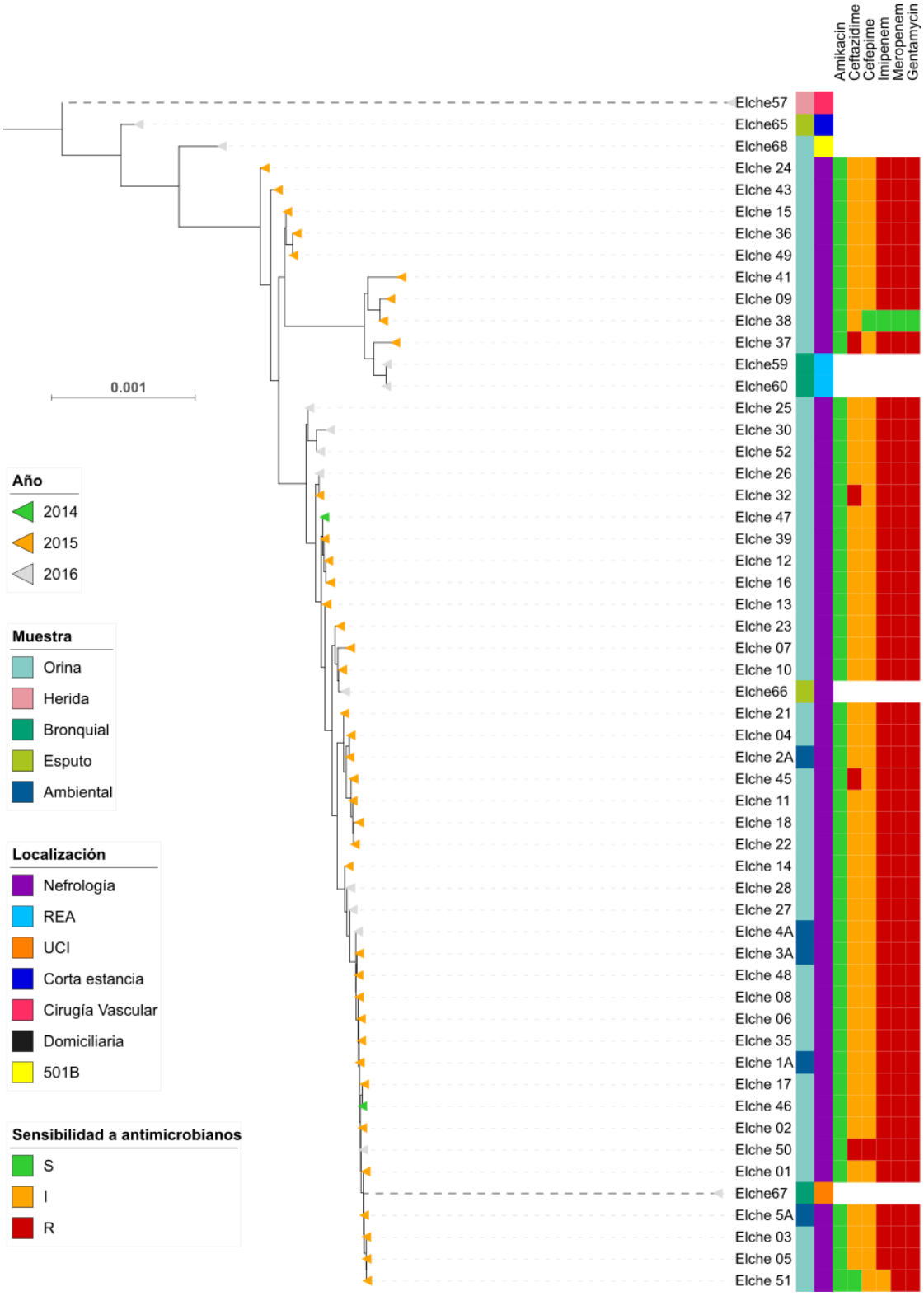
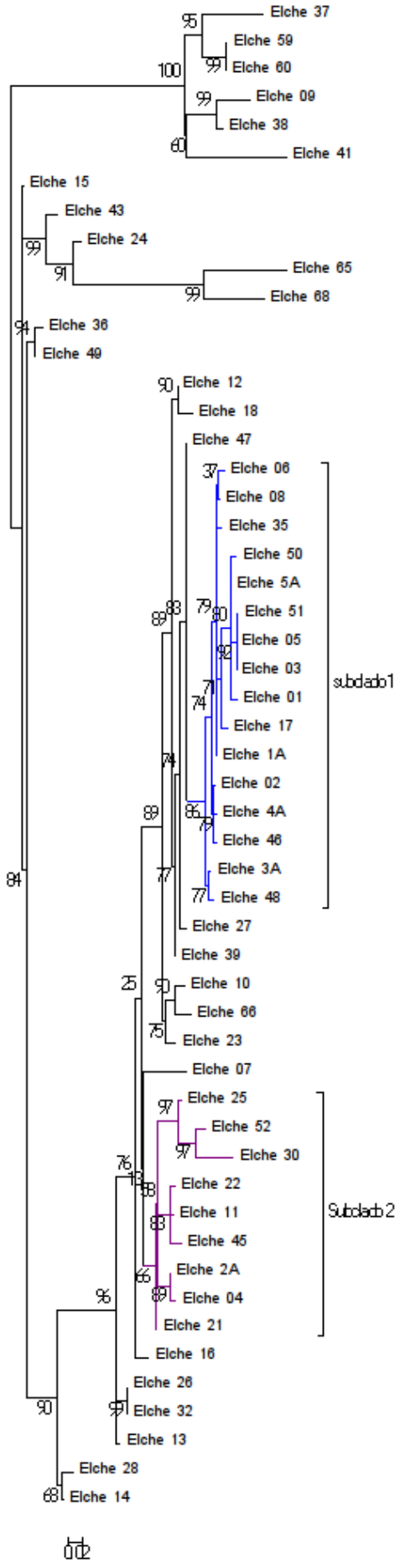


Figura 4.10. Subclado correspondiente al ST175 extraído del árbol filogenético del total de las muestras. *La representación del subclado en que se encuentran las cepas pertenecientes al brote con una escala mayor permite visualizar diferencias significativas. Los triángulos de los extremos de las ramas indica el año de aislamiento. Las columnas, dispuestas a la derecha del árbol, representan por orden el tipo de muestra de origen, la localización en que se encontraba el paciente y, las 6 columnas restantes, el perfil fenotípico de cada uno de los antibióticos indicados en la parte superior.*



4.2.4 Estudio evolutivo

Se ha determinado la fecha del ancestro común de las muestras del clado más grande, identificado como brote, sin las dos muestras más divergentes (Elche57 y Elche67). La señal filogenética que tiene este clado con respecto a las fechas de aislamiento es baja, obteniendo con TempEst un R^2 de 0,117. Por ello, se determinó la señal filogenética de dos subclados con el fin de realizar la calibración en base a un nodo interno con BEAST en caso de que se obtuviera la suficiente señal con alguno de ellos. El subclado 1, comprendido por la muestra Elche_06 a la Elche_48 (Figura 4.11) mantenía una señal baja (R^2=0,16), por lo que se utilizó finalmente el subclado 2 que, a pesar de contener un número pequeño de muestras, tiene una señal de reloj molecular mayor (R^2=0,293).

Se realizaron *tests* con BEAST utilizando el alineamiento de SNPs de las 9 muestras que forman el subclado2, sus fechas de muestreo e indicando como *prior* la tasa de sustitución con distribución exponencial con lambda 1E-6 s/s/a, basado en datos de un estudio previo de *Pseudomonas* (Miyoshi-Akiyama *et al.*, 2017). Se probaron todas las combinaciones de modelo de reloj molecular (estricto, no correlacionado y aleatorio) y modelo poblacional (constante, crecimiento exponencial y *Bayesian skyline* con 3 intervalos) con 60.000.000 de generaciones. Los resultados se compararon en base

Figura 4.11. Árbol sin enraizar del subclado. *Árbol construido con IQTREE con las 53 muestras del brote del HGUE a partir de las cuales realizamos el estudio con BEAST.*

a la verosimilitud marginal calculada por *path sampling* y s*tepping-stone sampling* y se determinó que los modelos que mejor encajaban con los datos eran el de reloj molecular estricto, que consideramos adecuado dada la escasez de datos con que contamos, y el modelo poblacional de crecimiento exponencial, que concuerda con la situación de brote epidémico. Se obtuvo una tasa de evolución de 4,1E-7 s/s/a, con un intervalo de alta densidad de probabilidad (HPD, *High Probability Density*) al 95% de [1.4271E-8, 9.6225E-7] y se confirmaron los resultados obtenidos realizando dos réplicas del análisis. También se comprobó si los datos que aportamos como *prior* están condicionando los resultados. Para ello se lanzó una ejecución en la que no se incluyó el alineamiento sino únicamente los *priors*. En este caso, la tasa de evolución aumentó un orden de magnitud (2,392E-6), por lo que podemos confirmar que el alineamiento está aportando información al análisis.

A continuación, se utilizaron los datos obtenidos con el subclado 2 para datar el ancestro común. Se aplicaron los mismos modelos de sustitución, reloj molecular y población, utilizando como *prior* para la tasa de evolución el valor obtenido, 4,1E-7 s/s/a, siguiendo una distribución normal con desviación típica de 1E-6. Se utilizó como árbol inicial el árbol reconstruido anteriormente, fijando los taxones que conforman el subclado 2 de manera que se mantenga su distribución durante el muestreo. El proceso MCMC se ejecutó con 60.000.000 de generaciones y se realizaron 2 réplicas más con 20.000.000 de generaciones, además de una prueba de los *priors* sin el alineamiento. Como resultado final (Tabla 4.5) obtuvimos una tasa de evolución media de 1,38E-6 s/s/a con un intervalo de HPD al 95% [1.773E-7, 2.699E-6]. El parámetro ESS (*Effective Sample Size*) fue en todos los casos superior a 200, valor a partir del cual se considera que la probabilidad posterior del parámetro se ha calculado de manera óptima y la desviación estándar será reducida. Como vemos, en la prueba sin el alineamiento, el ESS de la verosimilitud del árbol indica que no hay convergencia y, por tanto, únicamente con el valor del *prior* la resolución del árbol no sería buena, encontrando multitud de árboles posibles. Por tanto, el alineamiento está aportando información para la reconstrucción, aunque el *prior* con la distribución y modelos seleccionados son fuertes.

RUN	BURN-IN	CLOCK.RATE (MEAN)	CLOCK.RATE (ESS)	TREE LIKELIHOOD (ESS)
1	6000000	1.3933E-6	4889,7465	9001
1 – 2	2000000	1.3942E-6	1798,8953	7210,2812
1 – 3	2000000	1.3615E-6	1801,848	1820,0153
1 – PRIOR	1469400	9.524E-7	6605,8964	14,6072
COMBINACIÓN	16201800	1.383E-6	5774,8815	13930,8317

Tabla 4.5. Resultados obtenidos por BEAST utilizando la datación por nodo interno. *En todos los casos se lanzó con 20.000.000 de cadenas, salvo el primer test con 60.000.000 y se utilizó como prior la tasa de evolución en distribución normal con media 4,1E-7 y desviación típica 1E-6. Todo son réplicas salvo el PRIOR, en que se testea el grado de información que aporta el prior sin utilizar el alineamiento; además la fila COMBINACIÓN corresponde con los datos finales una vez se obtiene en el consenso de todos estos* runs.

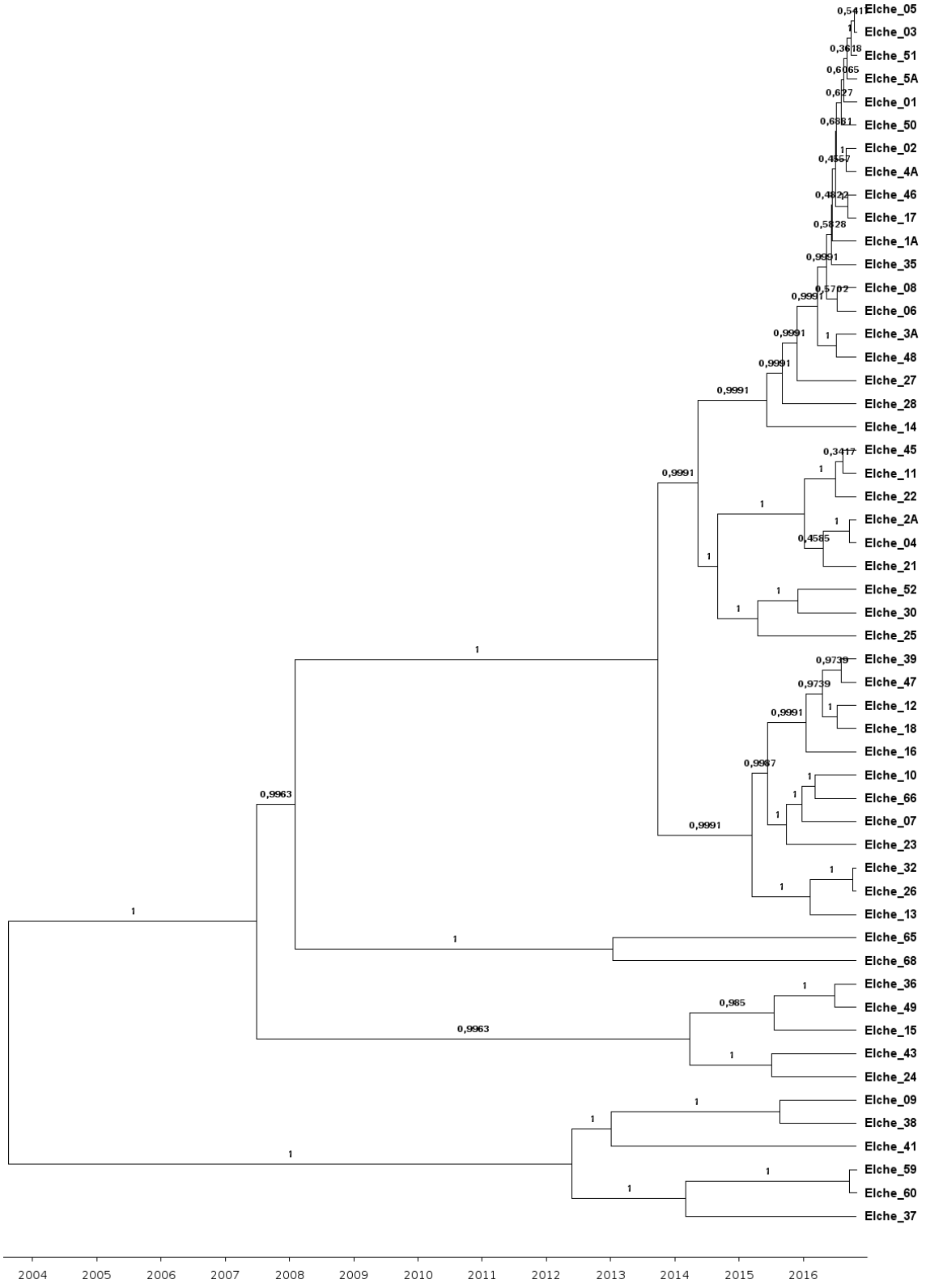


Figura 4.12. Árbol de las muestras procedentes del HGUE del clado del brote obtenido con BEAST. *Este es el árbol consenso a partir de los árboles obtenidos de las 3 ejecuciones independientes de BEAST con reloj molecular estricto, modelo poblacional exponencial y GTR gamma 4 como modelo de sustitución, utilizando los resultados del subclado 2 y fijando su nodo. Se ha tomado la última fecha de muestreo (02/11/2016) como momento "actual" en la escala; se incluyen los valores de posterior en cada rama.*

Según los resultados (Figura 4.12), el ancestro común dataría del año 2003, 13 años antes de la última muestra recogida. Sin embargo, este dato corresponde con la mediana de un intervalo que está comprendido entre los 4,4 y los 47 años, por lo que tenemos un rango excesivamente amplio como para estar seguros de la datación del ancestro común (de 1969 a 2012). En cambio, los nodos internos tienen rangos de tiempo algo más ajustados, como es el caso del subclado 2 cuyo ancestro común más próximo dataría de 2 años atrás, siendo el intervalo entre 0,6 y 7,6 años.

Este resultado podría explicarse por la presencia de varios aislados con un origen común, cuyo ancestro pudo ser introducido en el hospital muchos años antes de la detección de este brote. Con el tiempo es posible que existan diferentes clones circulantes por el hospital, habiendo una colonización persistente. De esta forma se detectarían casos con cierta regularidad, pero con genomas lo suficientemente alejados filogenéticamente como para ser considerados del mismo brote. También podría deberse a una introducción o flujo constante entre la comunidad y el hospital.

4.2.5 *Core* y genoma accesorio

El nivel de variabilidad obtenido a partir del mapeo es algo superior a lo esperado tratándose de un brote de cerca de 2 años de duración, si bien los resultados del apartado anterior apuntan a un origen mucho más antiguo. Se quiso comprobar si pudo producirse algún evento de recombinación respecto a genes del *core* que hubiera introducido cierta variabilidad que permita explicar la agrupación en 4 clados. Con este fin se estudió el contenido del genoma tras el ensamblado con SPAdes. Este abordaje alternativo también permite estudiar aquellos genes que puedan ser compartidos por todos los aislados pero que no estén contenidos en el genoma de referencia, aunque se pueden perder regiones deficientemente cubiertas o genes fragmentados.

Los resultados del ensamblado (Material suplementario, Tabla 7.9) indican que, en líneas generales, el contenido genómico ensamblado está entre 6,4 y 7 Mb, salvo la muestra Elche_07 que contiene 5,4 Mb, previsiblemente por el bajo contenido de lecturas con respecto a la mayoría de muestras, "perdiendo" con ello alrededor de 1 Mb de contenido genómico respecto a lo esperado, lo que va a repercutir en cierta medida en los resultados que obtendremos (Figura 4.13).

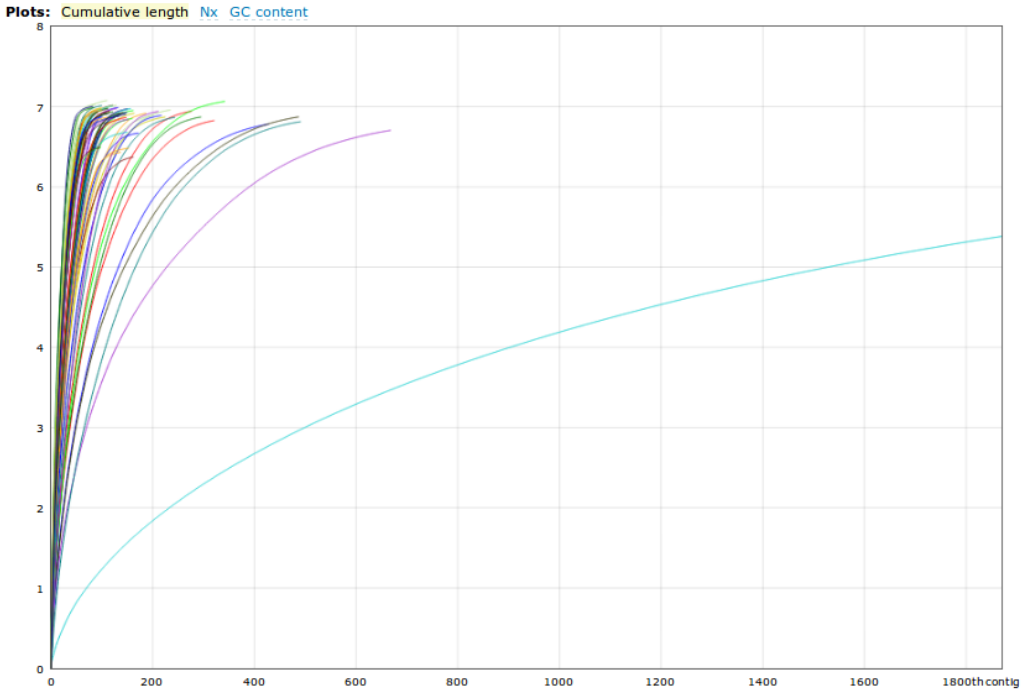


Figura 4.13. Número de contigs vs el contenido total en Mb que contienen acumulativamente. *Las curvas indican en el* plateu *el número de contigs y tamaño de todo el genoma ensamblado. Gráfica obtenida por QUAST.*

El contenido génico total de las 53 muestras que hemos considerado para el estudio evolutivo es de 8.961, siendo el pangenoma que reconstruye el programa ROARY de 8,2 Mb. Por tanto, encontramos alrededor de 2.000 genes más de lo esperado en caso de que fueran totalmente clonales, ya que el genoma es de 6-7 Mb, si bien es cierto que este organismo tiene a presentar una elevada variabilidad por la transferencia de elementos móviles y presencia de plásmidos. En el recuento, 4.030 de los genes son compartidos por mínimo el 99% de las muestras (*core* estricto), 1.884 adicionales están presentes en el 95-99% de las muestras (*core* relajado) y, de los 3.047 genes restantes, 2.150 están presentes en menos del 15% de las cepas (Figura 4.14). Destaca la muestra de Elche_07, tal y como se esperaba por el bajo contenido génico compartido respecto a las demás.

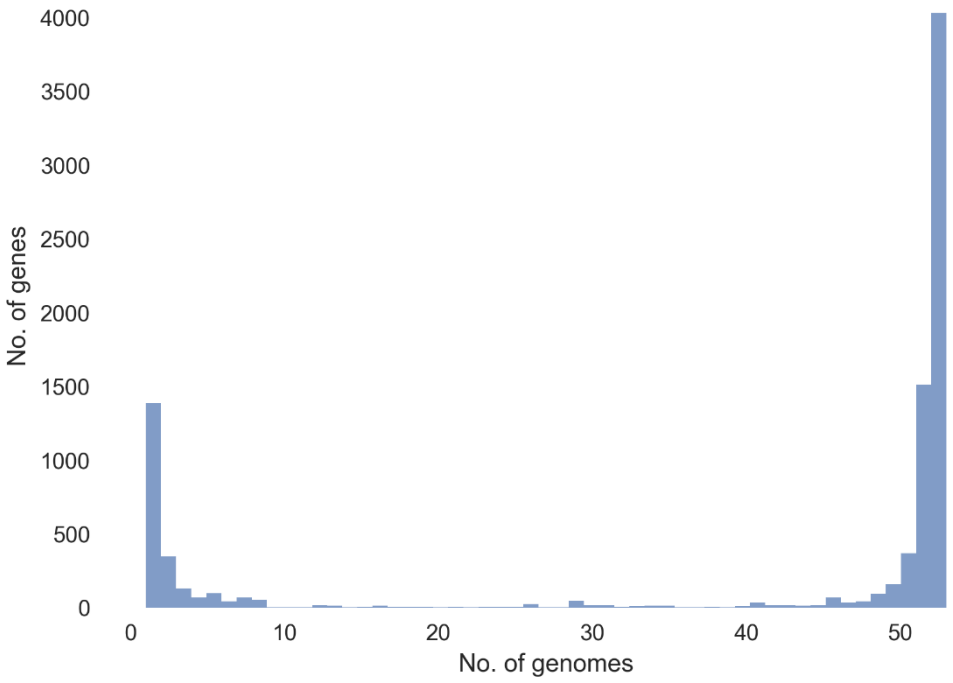


Figura 4.14. Distribución del número del contenido génico en las muestras del brote del HGUE. *Se representan agrupados los genes encontrados en los 53 genomas analizados agrupados en función del número de genomas que los comparten según la clasificación obtenida por* ROARY.

La estructura del árbol generado teniendo en cuenta únicamente la presencia o ausencia de genes (Figura 4.15) indica que hay cierta variabilidad entre los aislados dentro del genoma accesorio, aunque las agrupaciones de estas muestras no se corresponden con el árbol filogenético que obtuvimos a partir del mapeo. Por otro lado, cabe remarcar, como se menciona anteriormente, que, según los resultados de secuenciación y cobertura, es probable que haya aislados de cobertura muy baja con menor número de genes anotados de lo que corresponde, por lo que no toda la variabilidad detectada a nivel de contenido génico sería real. Es el caso de la muestra Elche_07 cuya rama en el árbol es la más distante respecto al resto y menor contenido génico compartido.

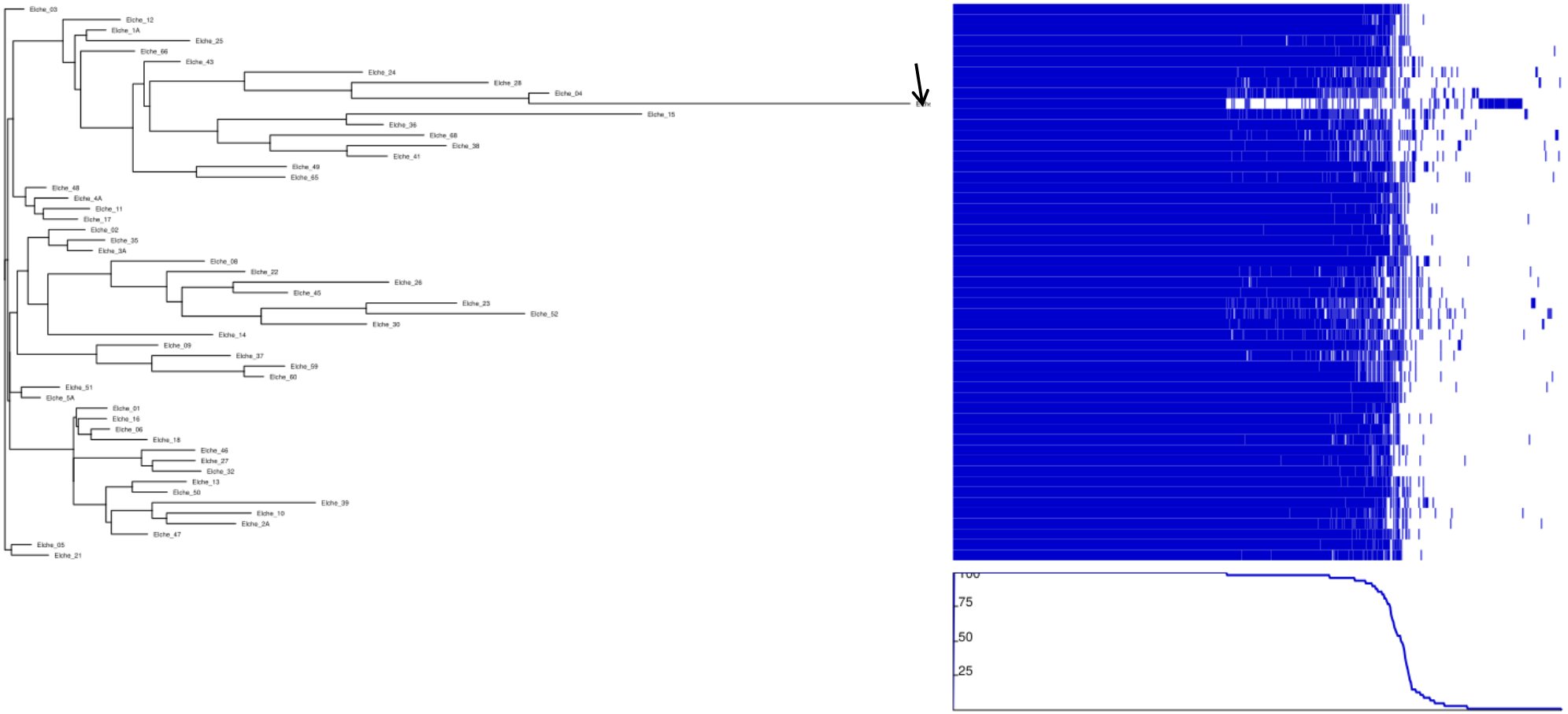


Figura 4.15. Árbol obtenido a partir de la presencia – ausencia de genes de las cepas del brote. *Reconstruido con Phandango a partir del análisis de Roary. En azul se indican los genes contenidos en cada muestra y en la parte inferior se representa el porcentaje de muestras que comparten cada gen. La flecha indica la localización del aislado Elche_07.*

Las funciones biológicas de los genes incluidos en el genoma accesorio (aprox. 3000 genes) fueron clasificadas a partir de los códigos de UniProtUK con BLAST2GO v5.1. Excluimos las regiones correspondientes a ARNs y proteínas hipotéticas, lo que redujo el listado a 804 proteínas, de las que solamente pudo obtenerse el código y su correspondiente clasificación para 203 de ellas.

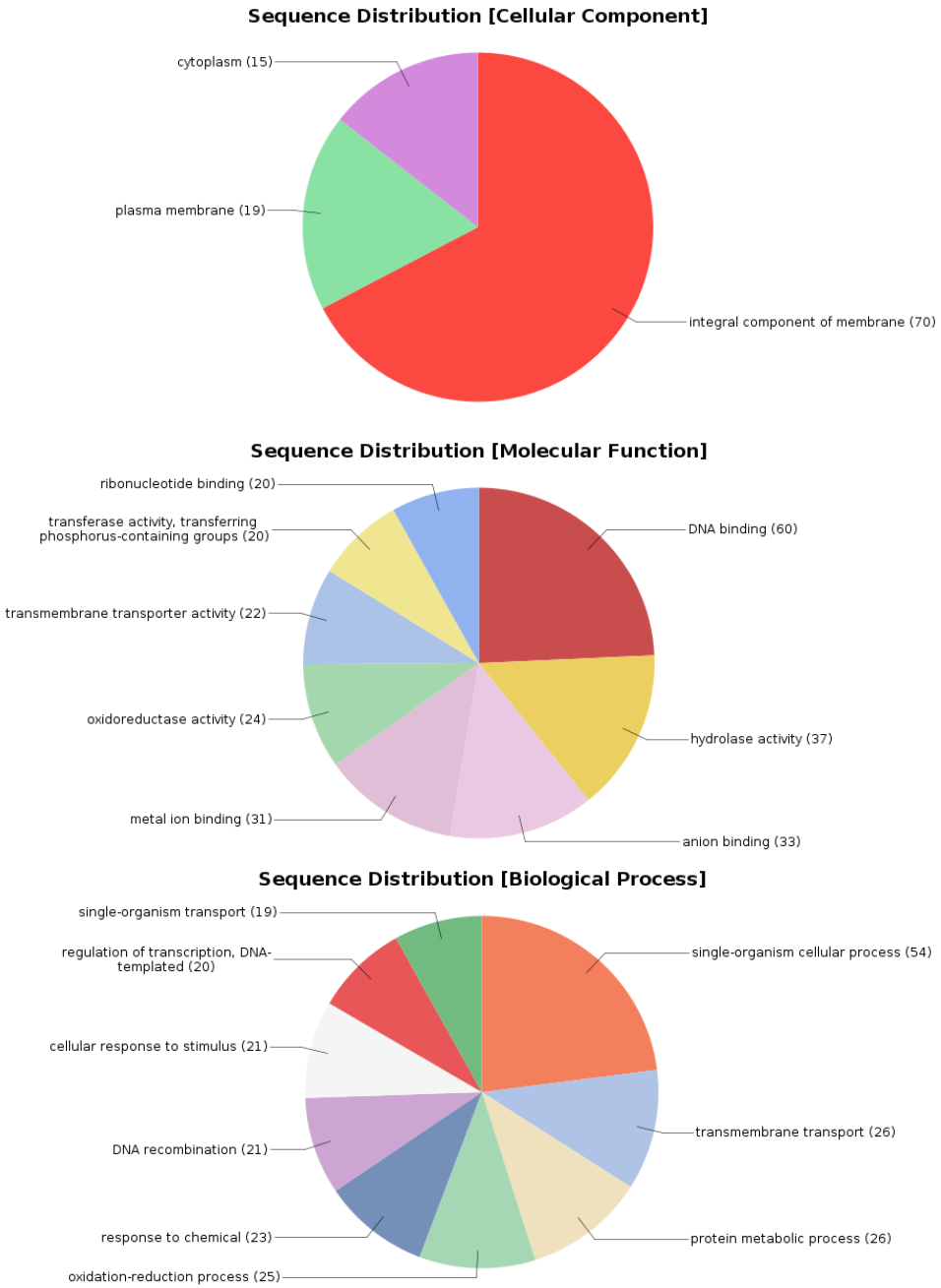


Figura 4.16. Clasificación por localización, función y proceso biológico de las proteínas cuyos genes pertenecen al genoma accesorio en las muestras del brote de HGUE. *Este estudio se ha realizado mediante BLAST2GO a partir de las 203 proteínas que tenían asignado un código UniProt.*

Como se observa en la Figura 4.16, la mayor parte del genoma accesorio analizado con localización definida corresponde a proteínas integradas en membrana; sin embargo, esto está sesgado por la carencia de información, ya que el grupo mayoritario en la clasificación por función molecular son proteínas de unión a ADN. El aspecto más llamativo es su clasificación según el proceso biológico en que participan, teniendo en cuenta su necesidad de adaptación y la presencia de resistencias detectadas fenotípicamente, como son la representación de los grupos de recombinación de ADN, respuesta a químicos o transporte transmembrana.

Por último, se ha visto que de los más de 3.000 genes que conforman el genoma accesorio, 1.388 están presentes en un solo genoma, que pasan a ser tan solo 202 cuando eliminamos proteínas hipotéticas. Aunque están repartidos por las distintas cepas, hay 3 que presentan más genes únicos que las demás: Elche_07, Elche_24 y Elche_68, con alta presencia de genes de transporte, interacción con ADN y motilidad (Material suplementario, Tabla 7.10).

4.2.6 Resistencias (ARIBA, SRST2)

El análisis de resistencias indica que la mayor parte de grupos de determinantes que confieren resistencia son debidos a mutaciones en proteínas de membrana capaces de expulsar múltiples tipos de antibiótico (Figura 4.17), como es el caso de los genes *mex* o los *opr*. Nuevamente, el aislado con más cambios es Elche_07, probablemente de nuevo por la baja cobertura inicial. A pesar de la estrecha relación filogenética entre los aislados del brote a partir de la secuencia de sus genomas completos, observamos una mayor variabilidad en la presencia de genes de resistencia detectados a partir de las lecturas. Las posibles explicaciones para estas discrepancias se discutirán más adelante.

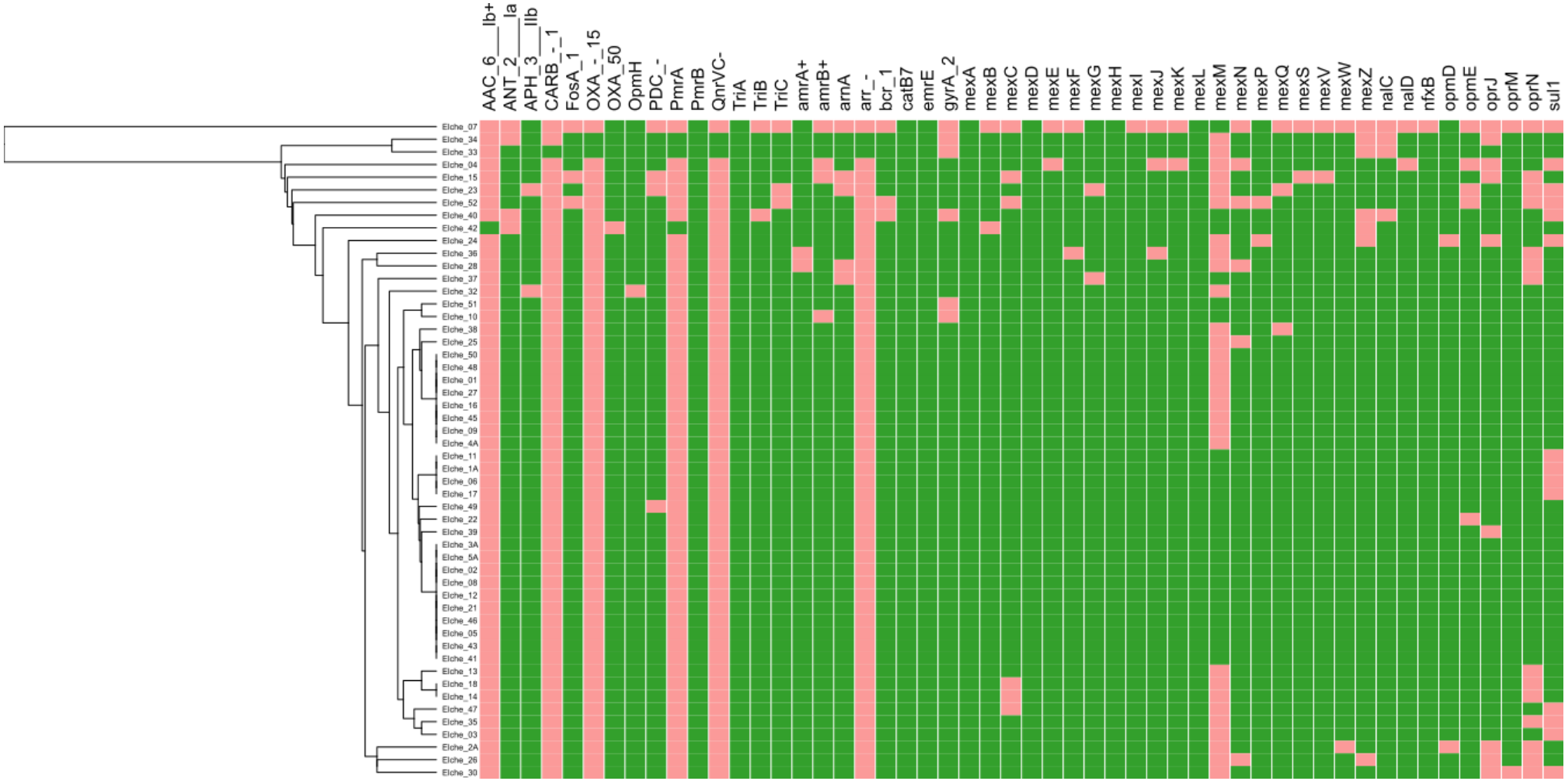


Figura 4.17. Árbol de basado en presencia-ausencia de resistencias por* clusters *en las muestras de HGUE. *Representación obtenida con los análisis de ARIBA utilizando Phandango* (Hadfield *et al.*, 2018). *Las posiciones en verde indican que el gen o la mutación ha sido detectada y las rosas, aquellos no detectados.*

4.2.7 Recombinación

La distribución del árbol que conforman las muestras del brote y la distancia entre ellas indica que es posible que esté tratándose de varios brotes diferentes a partir de un ancestro común a ellos que posiblemente divergiera y diera lugar a transmisiones a partir de diferentes focos. Sin embargo, quisimos comprobar si pudiera existir algún evento de recombinación entre genes que forman parte del *core* en este conjunto de muestras y explicara dichas agrupaciones.

La reconstrucción de los árboles a partir de los genes del *core* de manera individual dio como resultado que únicamente tenían diferencias 10 genes, el resto eran idénticos para todas las cepas, por lo que no se podía reconstruir la filogenia correspondiente. A partir de estos 10 genes se hizo el test de *likelihood mapping* y, como vemos en la Tabla 4.6, solamente se obtuvo señal filogenética suficiente para el gen 15.

Árbol	***1+2+3***	***4+5+6***	***7***
Gen_15	62,07	0	37,93
Gen_213	0	0	100
Gen_453	16,87	0,67	82,47
Gen_1056	20,8	0	79,2
Gen_1068	0	0	100
Gen_1110	30,33	7,27	62,4
Gen_1154	29,6	0	70,4
Gen_2705	0	0	100
Gen_3793	0,93	0,13	98,93
Gen_3965	22,07	0	77,93

Tabla 4.6. Resultados test de* likelihood mapping *de los genes con diferencias entre cepas. *La tabla contiene los resultados del porcentaje de casos en que una topología tiene una verosimilitud superior al resto (1+2+3), dos de las 3 topologías (4+5+6) o ninguna de las 3 tiene una verosimilitud superior al resto (7).*

Se realizó el test de topologías para comprobar el grado de congruencia de la estructura del árbol de este gen con respecto a su conformación utilizando el genoma *core*. Los valores obtenidos (Tabla 4.7) indican que, como esperamos, la topología del alineamiento concuerda con su árbol, pero no con el árbol de genoma ni el alineamiento del genoma con el árbol del gen. Con esta doble confirmación podríamos decir que en este caso tenemos un posible evento de recombinación.

	PB-RELL	P-KH	P-SH	P-WKH	P-WSH	C-ELW	P-AU
ÁRBOL GEN	0.9997	0.9967	1.0000	0.9967	0.9967	0.9997	0.9984
ÁRBOL GENOMA	0.0003	0.0033	0.0033	0.0033	0.0033	0.0003	0.0016
ALINEAM GENOMA – ÁRBOL GEN	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Tabla 4.7. Resultados del test de topologías del gen_15*. P-valor de cada uno de los tests realizados comparando el alineamiento del gen frente a su árbol, al árbol del genoma y el alineamiento del genoma frente al árbol del gen.*

Pb-RELL : bootstrap proportion, RELL (Kishino et al. 1990).
p-KH : p-valor Kishino-Hasegawa test (1989).
p-SH : p-valor Shimodaira-Hasegawa test (2000).
p-WKH : p-valor weighted KH test.
p-WSH : p-valor weighted SH test.
c-ELW : Expected Likelihood Weight (Strimmer y Rambaut 2002).
p-AU : p-valor AU test (Shimodaira, 2002)

El gen codifica para una proteína putativa que controla la longitud de la cola ("*putative tail length tape measure protein*") y en el árbol del gen (Figura 4.18) vemos como hay varias muestras que sí se encuentran en el árbol original aunque con otra conformación.

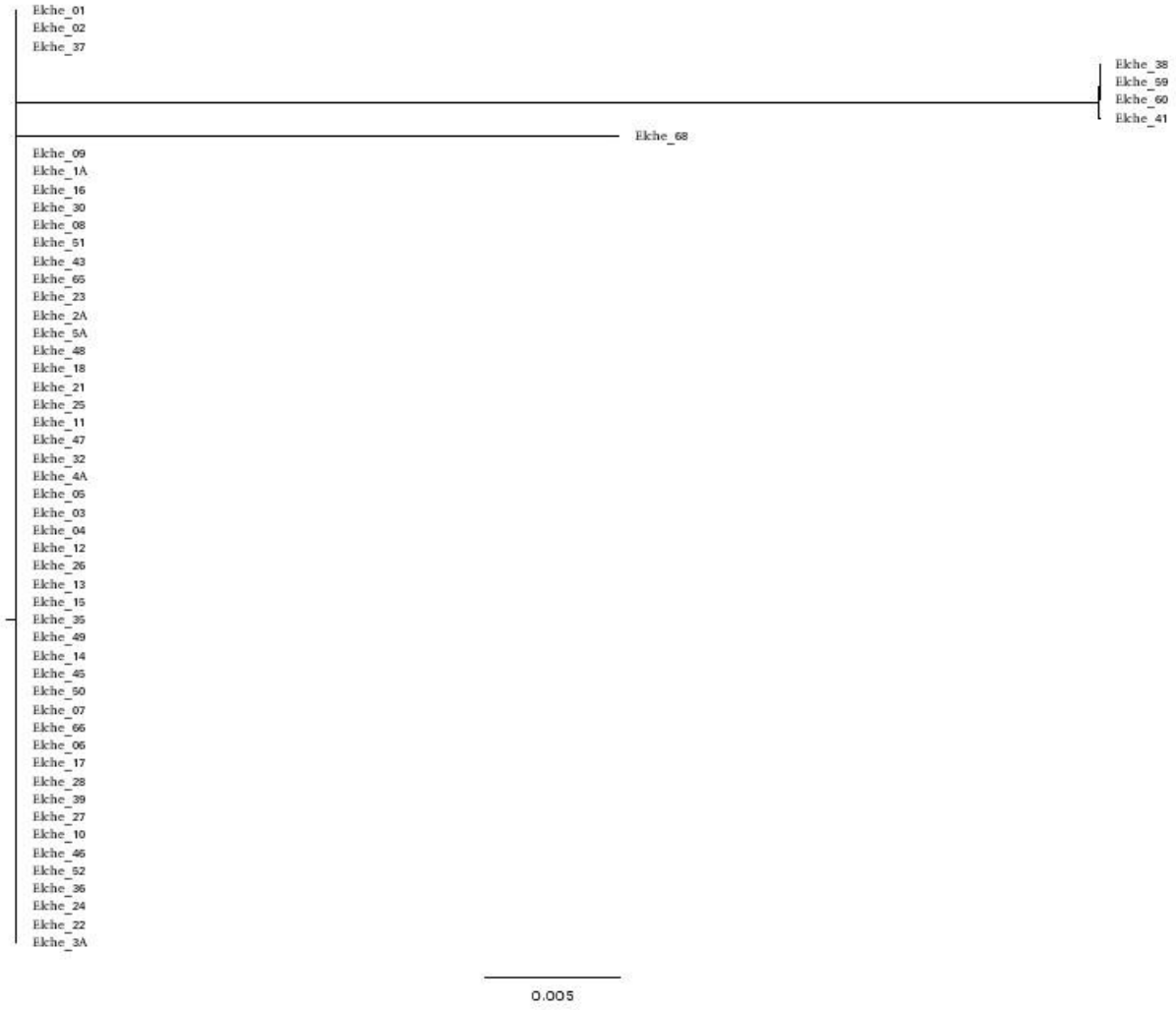


Figura 4.18. Árbol del gen 15. Obtenido tras el análisis con IQTREE.

Capítulo 3

4.3 Hospital General Universitario de Valencia

El proyecto de evolución intrapaciente se diseñó a partir de 112 aislados de *Pseudomonas aeruginosa* de 18 pacientes procedentes del Hospital General Universitario de Valencia (HGUV). Puesto que el objetivo que nos planteamos inicialmente era determinar la variación genómica intrapaciente de las *P. aeruginosa* a lo largo del tiempo y describir cambios adaptativos en diferentes zonas del cuerpo, se seleccionaron aquellos pacientes con mayor número de cepas recogidas en el Servicio de Microbiología entre 2012 y 2014, estando ninguno de ellos afectados de fibrosis quística. Se dio preferencia a aquellas personas con cepas cuyo aislamiento estuviera distribuido en un rango más amplio de tiempo y que se hubieran obtenido a partir de diferentes muestras biológicas (orina, esputo, exudados, etc.). Esto no se cumple en todos los casos, dada la limitación en la disponibilidad de muestras al tratarse de un estudio retrospectivo, por lo que los pacientes pueden dividirse en tres grupos: aquellos que cumplen ambos criterios; aquellos con varias muestras en un rango corto de tiempo provenientes de múltiples zonas del cuerpo, es decir, el criterio de localización; y, en tercer lugar, los que cumplen el criterio de tiempo, contando con muestras tomadas en un intervalo amplio de tiempo pero a partir de una única zona (Material suplementario, Tabla 7.11).

4.3.1 Resultados de secuenciación

Inicialmente, se seleccionaron 112 aislados de *P. aeruginosa*, obtenidos a partir de muestras biológicas de diferentes tipos de 18 pacientes que presentaron infecciones repetidas o de larga duración a lo largo de su estancia hospitalaria. Los aislados, además, fueron clasificados como multirresistentes según los perfiles fenotípicos obtenidos en el Servicio de Microbiología, motivo por el cual se consideró necesario el almacenamiento de la cepa y ha sido posible su recuperación para la realización del presente estudio. Sin embargo, por problemas de localización y de falta de crecimiento en placas, solamente se secuenciaron 93 cepas de 16 pacientes diferentes (Material suplementario, Tabla 7.12).

El contenido en GC en todas las muestras fue el esperado, entre el 64-65%, ya que la cepa de referencia utilizada habitualmente en laboratorio PA01 tiene un contenido aproximado en GC de 66,6%, lo cual parecía indicar que no existieron errores de determinación de la especie en el proceso de aislamiento y cultivo. Tras la limpieza de

las lecturas, se alcanzó una calidad óptima de las bases por posición en la lectura y una mejora en los niveles de contenido en GC que inicialmente poseían un exceso de variabilidad con respecto a la distribución teórica modelizada por el programa FASTQC (Figura 4.19).

A pesar de que el tamaño genómico esperado es de 6-7 Mb en todas las cepas, dado que se trata de la misma especie, se observa una alta variabilidad en el número de lecturas obtenidas por cepa (Figura 4.20). El caso de la muestra 2 del paciente 16 fue crítico, ya que solamente posee 635 lecturas, lo cual obligó a descartarla ante la imposibilidad de cubrir mínimamente el genoma.

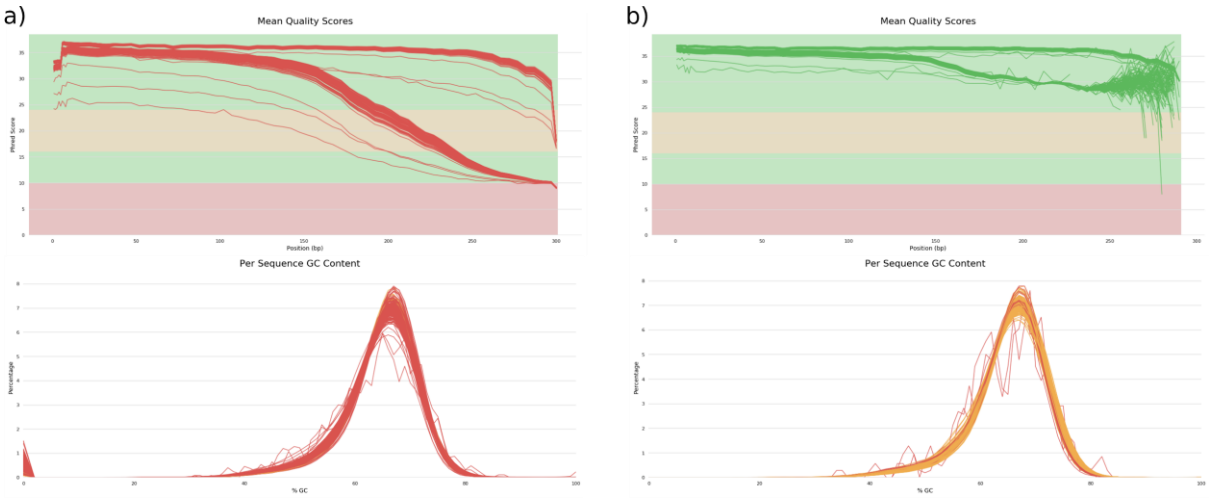


Figura 4.19. Comparativa de la calidad de las lecturas pre- (a) y post-limpieza (b). *En las gráficas superiores se representa la posición de la base en las lecturas frente a la calidad en escala Phred de dicha posición. Las gráficas inferiores corresponden con la distribución en contenido de GC de las lecturas; si la suma de las desviaciones de la distribución normal corresponde con >30% de las lecturas, la distribución será de color rojo y si son de >15% de lecturas, en naranja. Los picos a lo largo de la distribución pueden deberse a contaminaciones posiblemente generados por el proceso de preparación la librería.*

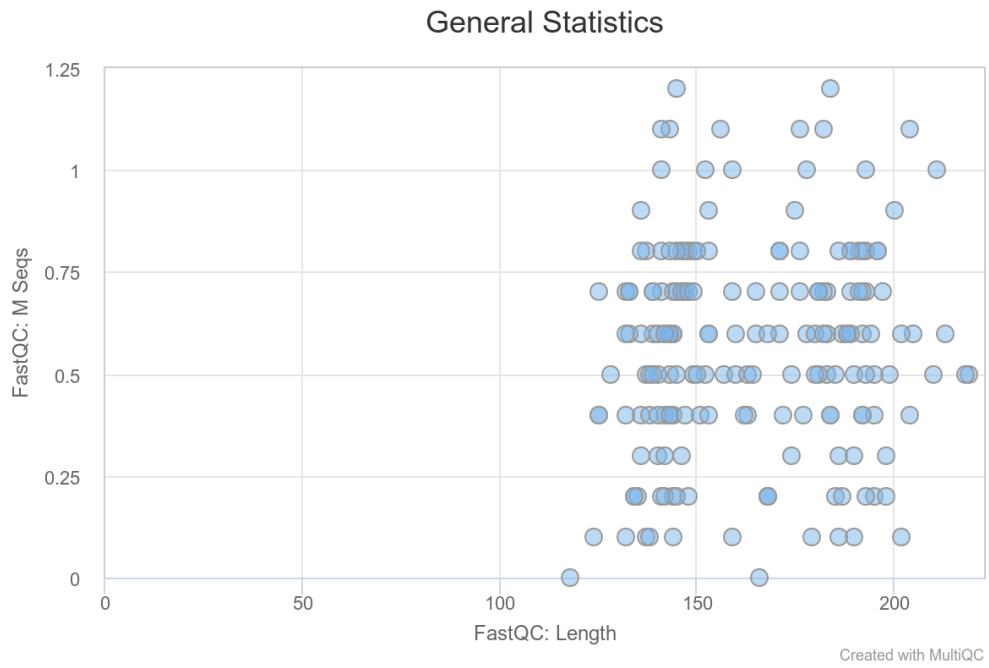


Figura 4.20. Estadística de las lecturas tras la limpieza. *En el eje X se representan la longitud de las lecturas, oscilando aproximadamente entre 125-300 pb, y en el eje Y, los millones de lecturas obtenidos. Cada muestra está representada por 2 puntos con estadísticas similares al tratarse de paired-ends.*

4.3.2 Mapeo y reconstrucción filogenética

La selección del genoma de referencia se basó en los resultados de tipado mediante el procesamiento directo de las lecturas siguiendo el esquema de MLST de *P. aeruginosa*. En su mayoría, las muestras de este estudio pertenecen al ST244 (Material suplementario, Tabla 7.14). Entre los genomas completos disponibles se utilizó la cepa W16407 (NZ_CP008869|gi|976144768) como referencia, la única perteneciente al ST244 (Material suplementario, Tabla 7.13).

En el mapeo de las 92 muestras se obtuvo una cobertura media de 24,78x, aunque con altas diferencias entre las cepas de mayor (P14M9, 51,3x) y menor cobertura (P12M2, 4,9x), tal y como apuntaban los resultados de lecturas por cepa. Estos resultados se detallan en la Tabla 7.14 (Material suplementario). Inicialmente, el protocolo utilizado para mapear y seleccionar las variantes se diseñó con umbrales mucho más restrictivos, como con el aumento del mínimo de proporción de lecturas para la asignación de variantes y/o base de referencia al 90%, o el mínimo de lecturas de alta calidad, inicialmente fijado en 8 lecturas (4 en cada sentido). Esto resultó perjudicial

tanto para las regiones de los genomas con cobertura especialmente baja, ya que la cobertura no es uniforme dentro de una misma muestra, como puede observarse en la Figura 4.21 y Figura 4.22, así como para los genomas que tuvieron una cobertura media baja. Como consecuencia, cada pseudogenoma reconstruido tenía una elevada tasa de posiciones indeterminadas (Ns), siendo indistinguibles en la secuencia final las debidas a deleciones o zonas repetitivas respecto a aquellas que no pasaron los filtros.

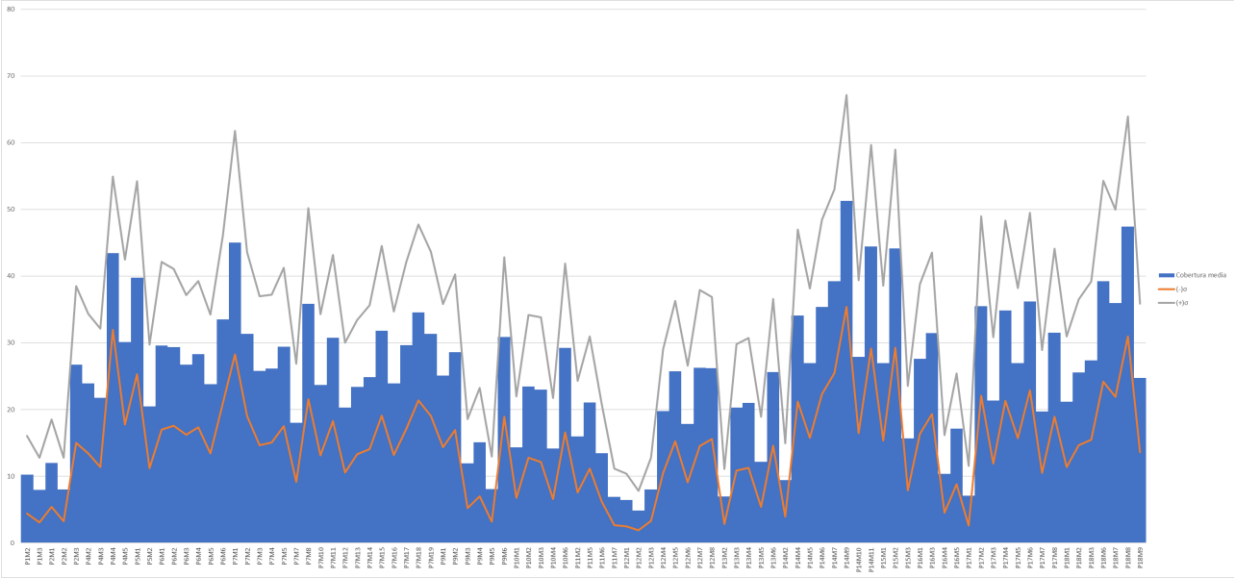


Figura 4.21. Coberturas medias y desviaciones típicas de los aislados del HGUV. *Cada pareja de barras azul y roja corresponde con un aislado diferente, indica el número medio de lecturas por posición del genoma de referencia utilizado.*

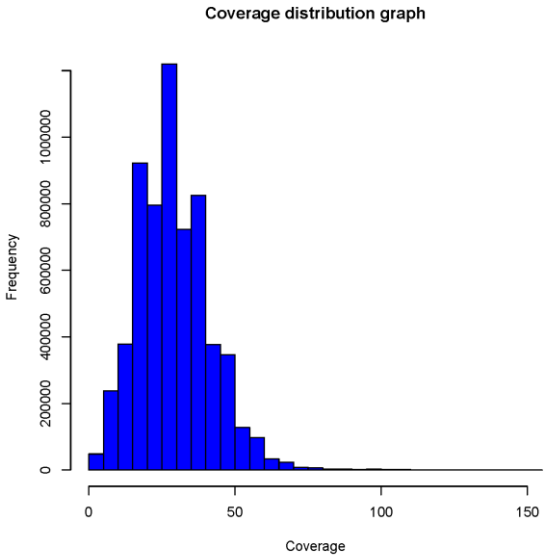


***Figura 4.22*. Distribución de cobertura en la muestra P6M1.** En este gráfico se representa el número de veces que se lee una posición del genoma frente a la frecuencia con que esto se da. En esta muestra la media es 29,5x pero con una deviación típica de 12,6.

Muestra	Cobertura media	% lecturas mapeadas	% posiciones cubiertas en la referencia, restrictivo	% posiciones cubiertas en la referencia, relajado
P12M2	4,8x	92,5	12,6	43,7
P12M1	6,4x	91,9	26,4	59,2
P11M7	6,8x	92	31,6	64,3
P13M2	6,9x	92,3	31,1	63,9
P17M1	7x	90,5	32	63,9
P12M3	8x	92,5	41,6	71,8
P2M2	8x	93,1	42,5	72,8

Tabla 4.8. Muestras con menor cobertura y posiciones que pudieron ser determinadas según la metodología de filtrado. *Comparando las pruebas con el filtrado más restrictivo respecto al finalmente utilizado podemos ver el efecto que tienen las bajas coberturas.*

Finalmente, el alineamiento con las 92 muestras de 6.829.406 pb, la misma longitud que la del genoma de referencia con el *core* estimado al 95% de 4.290.736 pb, tras la aplicación de los filtros tiene un total de 126.670 SNPs. El alineamiento de SNPs se utilizó para la reconstrucción filogenética mediante IQTREE obteniendo un árbol con un clado que contiene la mayor parte de las muestras claramente diferenciado (Figura 4.23).

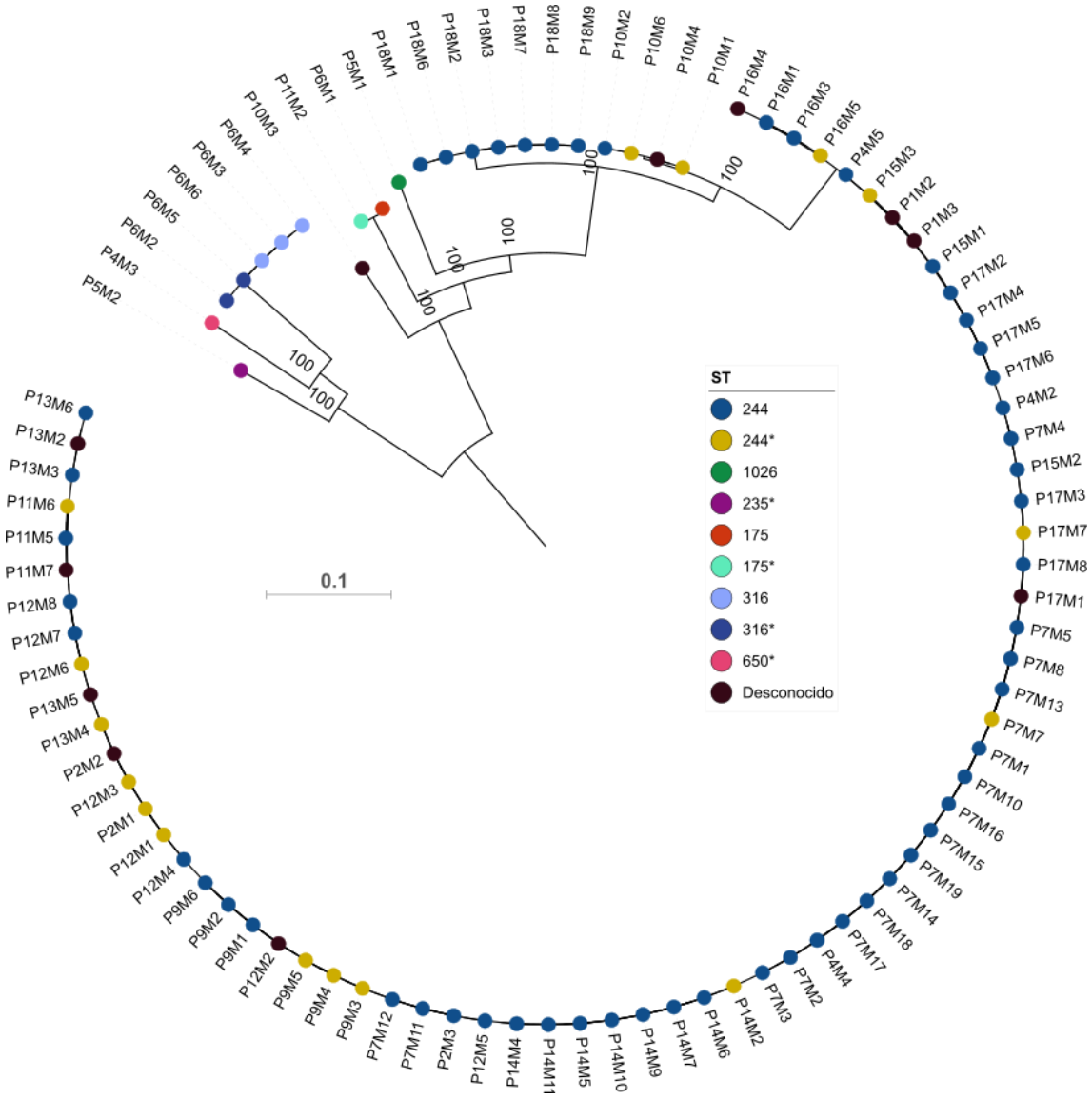


Figura 4.23. Árbol obtenido por IQTREE a partir del alineamiento de SNPs tras el mapeo de las muestras del HGUV*. Modelo utilizado TVM+F+ASC alineamiento de 126.670 SNPs.*

El árbol, obtenido con el modelo de transversión con frecuencias empíricas, modelo FreeRate y corrección de sesgo por ausencia de posiciones constantes, refleja que hay concordancia entre la agrupación por ST y la similitud de los aislados a nivel de genoma completo. De los 16 pacientes estudiados, solamente 3 presentan secuencias que conforman grupos monofiléticos: los pacientes 14, 16 y 18. El 76% de las muestras se encuentran en el clado más grande, siendo sus secuencias muy similares entre sí, ya que presentan únicamente 1493 SNPs de diferencia en conjunto, lo que sugiere un ancestro común muy reciente que podría deberse a la presencia de un brote hospitalario que no fue detectado.

Sin embargo, esta posibilidad se ha propuesto tras visualizar en una misma filogenia, ya que analizando cada paciente por separado tal y como se planteó inicialmente, no encontramos este tipo de agrupaciones, aunque sí cierta variación en el número de variantes que indica la posibilidad de encontrar varios clones en un mismo paciente.

El número de diferencias es elevado, con más de 6,000 SNPs (Tabla 4.9) en aquellos pacientes que presentan aislados de varios ST distribuidos a lo largo del árbol. En el caso del paciente 11, cuyos aislados se encuentran dentro del clado del brote salvo P11M2, vemos cómo su valor de cambios medio se reduce a 16, valor similar al del resto de pacientes del brote.

P10	**6990,20**
P11	6165,33
P12	9,18
P13	3,40
P14	6,68
P15	4,33
P16	9,50
P17	2,68
P18	6,10
P1	18,00
P2	22,67
P4	22093,17
P5	36027,00
P6	14804,27
P7	10,90
P9	5,93

P11	**16,00**
P12	9,18
P13	3,40
P14	6,68
P15	4,33
P16	9,50
P17	2,68
P1	18,00
P2	22,67
P4	95,00
P7	10,89
P9	5,93

Tabla 4.9. Número de diferencias intrapaciente en el total (izquierda) o dentro del ST244 (derecha). *Se ha utilizado para el cálculo el alineamiento de SNPs sin posiciones ambiguas, en total se han mantenido 118.102.*

	P11	P12	P13	P14	P15	P16	P17	P1	P2	P4	P7	P9
P11	16.00											
P12	71,08	9,18										
P13	71,80	8,85	3,40									
P14	79,92	15,545	13,52	6,68								
P15	85,22	16,33	16,80	13,87	4,33							
P16	120,67	42,42	40,45	51,28	23,00	9,50						
P17	86,46	16,75	16,50	12,53	4,67	23,69	2,68					
P1	103,83	35,12	33,40	33,37	13,17	143,25	15,94	18,00				
P2	86,44	16,17	11,80	20,62	23,11	54,25	21,58	44,33	22,67			
P4	99,89	52,62	54,00	58,58	44,11	58,92	46,79	54,33	59,22	95,00		
P7	86,94	20,09	17,85	16,33	11,53	47,79	12,60	34,53	26,88	54,69	10,89	
P9	83,56	9,56	7,57	17,44	20,50	52,83	20,87	37,42	12,72	59,67	22,16	5,93

Tabla 4.10. Número de diferencias entre paciente dentro del clado del brote del HGUV. *En la diagonal principal se recoge también el número promedio de diferencias entre muestras de un mismo paciente (ver tabla 4.9).*

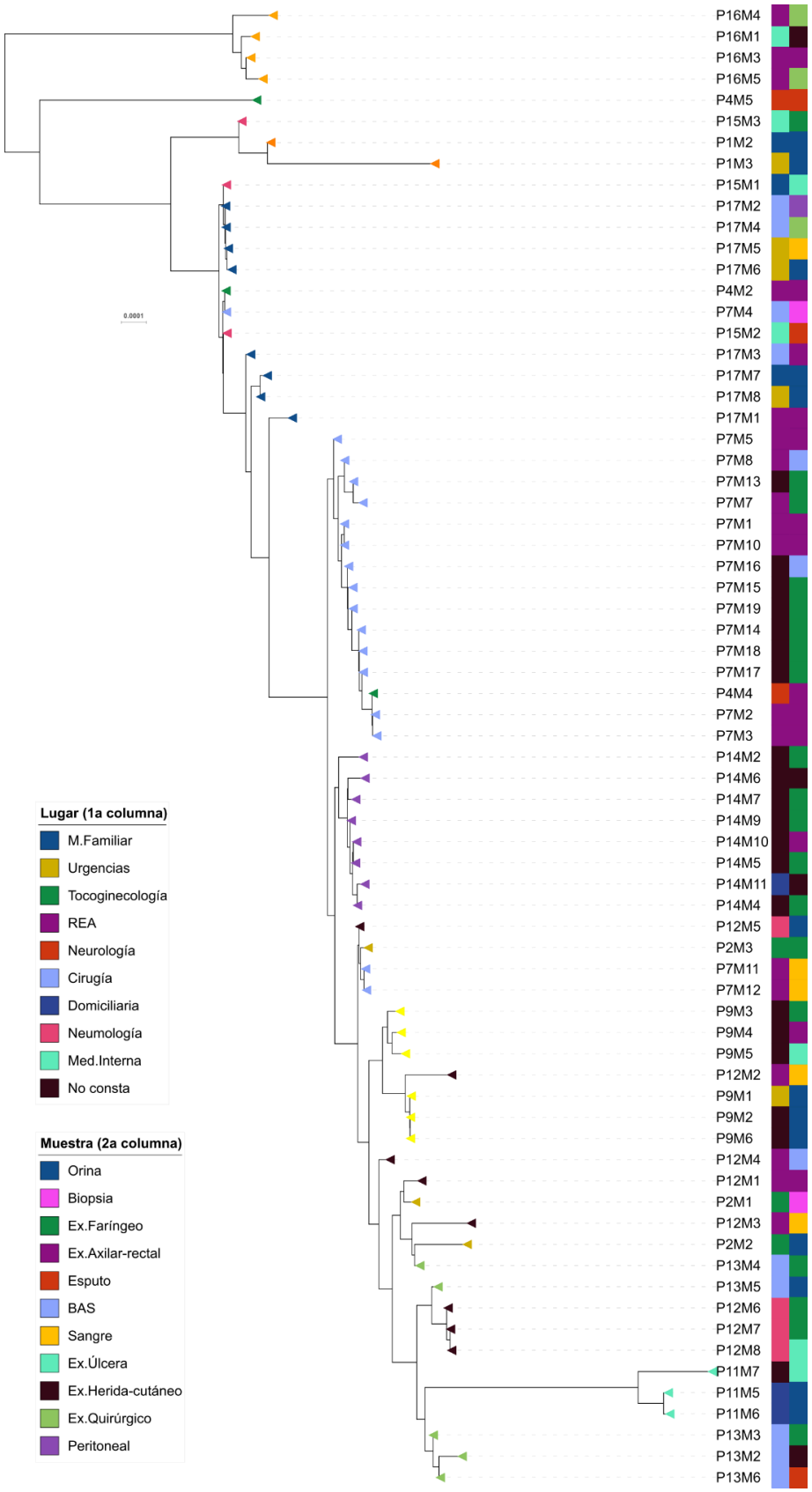


Figura 4.24. Clado del brote putativo del HGUV. *Se representa el subclado extraído del árbol completo a diferente escala. Está incluida la información sobre el área en la que se encontraba el paciente en el momento de toma de muestra (columna 1) y el tipo de muestra biológica a partir de la cual se realizó el aislamiento.*

La estructura del clado (Figura 4.24) no permite establecer ninguna relación entre la localización de los pacientes, el tipo de muestra o las fechas de aislamientos. Aunque los pacientes 14 y 16 son monofiléticos, el resto tienen sus aislados mezclados en diversos clados, sin que exista ningún tipo de relación respecto a los factores indicados. Esto podría deberse, además, a que los pacientes implicados, que han sido en su mayoría ingresados en varias ocasiones durante ese periodo de tiempo (2012-2014), hayan sufrido infecciones repetidas en sucesivos ingresos.

Adicionalmente, se determinó con ARIBA la presencia de grupos de determinantes de resistencia, bien genes o mutaciones, para la combinación de dicha información con el árbol filogenético y el estudio fenotípico realizado en el Servicio de Microbiología. El resultado (Figura 4.25) indica que no hay correlación directa entre ST y brote, pero tampoco respecto al fenotipo de resistencia dada la variación respecto a los antibióticos representados, y tampoco respecto a los determinantes encontrados.

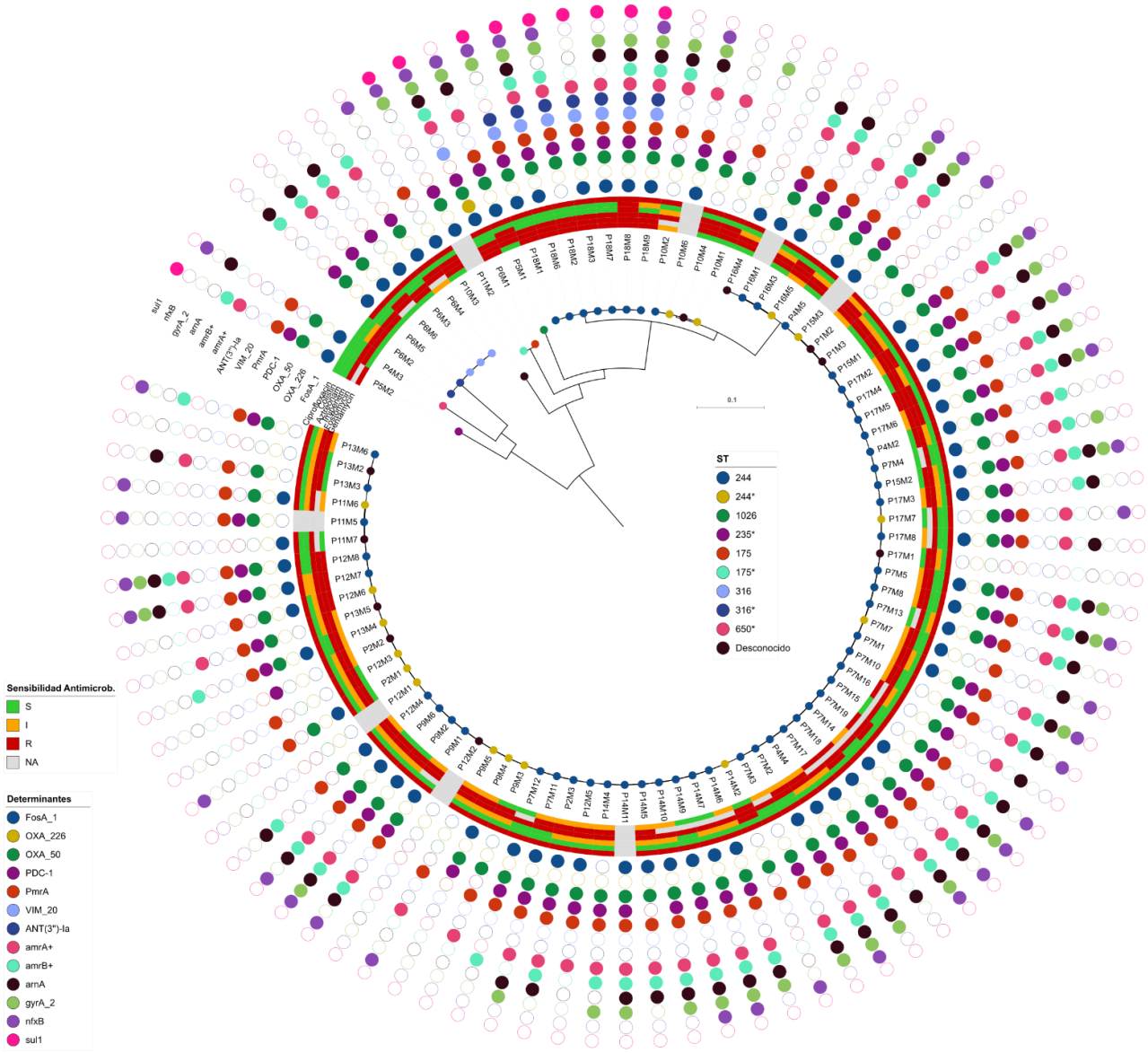


Figura 4.25. Árbol filogenético del conjunto de muestras del HGUV con determinantes y fenotipo de resistencia. *El ST es indicado el extremo de las ramas y en las circunferencias exteriores el fenotipo de resistencia de varios antibióticos representativos y en la región más externa la presencia – ausencia de grupos de determinantes que confieren resistencia a distintos antibióticos.*

4.3.3 Estudio evolutivo

La datación del ancestro común de las muestras del clado más grande (sospecha de brote) se realizó con BEAST de la misma manera que para el brote de Elche (4.2). La señal filogenética del conjunto era bastante baja, R^2=0,018 con la raíz que mejor se adapta a los datos en la función de correlación (Figura 4.26) y, dado el número de muestras, se decidió analizar posibles subclados para la datación a partir de nodo interno mediante BAPS.

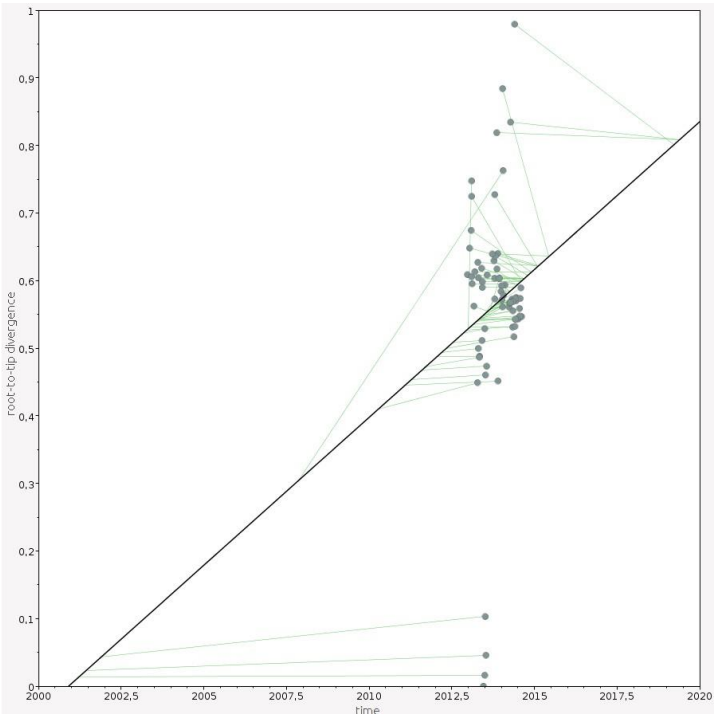


Figura 4.26. Recta de regresión de TempEst a partir del árbol y las fechas del subclado del posible brote en las muestras del HGUV. *En verde se representa la señal de cada muestra respecto al ancestro. Se ha calculado a partir del árbol de SNPs y la* "best-fitting root".

Se eligió el subclado correspondiente al de la mayor parte de las muestras del paciente 7, 15 aislados en el clado en total, con un R^2=0,82. Realizamos con BEAST pruebas con combinaciones de modelos de reloj (estricto, aleatorio y no correlacionado) y de modelos demográficos (constante, de crecimiento exponencial y *Bayesian skyline* de 3 categorías). Las verosimilitudes marginales (Material suplementario, Tabla 7.15) indicaron que los modelos más adecuados para este subclado son los de reloj no correlacionado y crecimiento exponencial, utilizando como prior la tasa de evolución en distribución exponencial con lambda 1E-6 s/s/a. Los resultados de la tasa media eran similares a los obtenidos anteriormente para el subclado 2 de Elche, por lo que utilizamos los mismos parámetros para la datación por nodo. También confirmamos, probando

únicamente los *priors*, que no había convergencia y la tasa de evolución era de 0,16 en ausencia de alineamiento.

Debido a los problemas de convergencia con el conjunto entero, dada la poca señal de las muestras, procedimos a datar el árbol basándonos en un nodo interno. Para ello volvimos a utilizar el prior en distribución normal y cambiamos el modelo de reloj a estricto. El resultado combinado de 3 réplicas (Tabla 4.11) indica una tasa de evolución del "brote" de 1.414E-6 s/s/a (un 95%HPD de [1.5934E-7, 2.7342E-6]) y una verosimilitud del árbol en [-7989078.8752, -7989054.7899] que cambia a [-0, 2.6645E-14] cuando realizamos el mismo proceso sin *priors*. El ancestro común (Figura 4.27) está datado 35,9 años antes, a mediados de 1978 (95% HPD [11.5252, 121.6049]) si consideramos como "presente" en el árbol la fecha del último aislado (12/08/2014). La tasa de evolución es similar a la obtenida anteriormente para el brote de Elche (1.383E-6 s/s/a), lo que resulta coherente al tratarse del mismo tipo de organismo, a pesar de que los datos de partida no permiten obtener resultados más fiables.

	BURN-IN	TASA DE EVOLUCIÓN (S/S/A)	ESS (TASA DE EVOLUCIÓN)	ESS (VEROSIMILITUD DEL ÁRBOL)
2 - 1	30000000	1.4039E-6	2244,27	849,85
2 – 2	6000000	1.4134E-6	4103,97	8542,55
2 – 3	30000000	1.4334E-6	2323,88	1258,91
COMBINADO		1.4148E-6	774,32	552,63

Tabla 4.11 Resultado de las 3 réplicas y su resultado combinado utilizadas para la reconstrucción con BEAST de la filogenia del brote del HGUV.

Existen diferencias en la topología de este árbol respecto al obtenido anteriormente con IQTREE por máxima verosimilitud (Figura 4.28). El cambio más importante se observa entre los clados sombreados en rosa y azul, cuyos ancestros comunes más recientes estarían más alejados de lo que indica el árbol de IQTREE. Esta nueva estructura podría indicarnos que existen varios eventos de transmisión con clones de diferentes orígenes, pudiendo tratarse de varios brotes simultáneos.

.

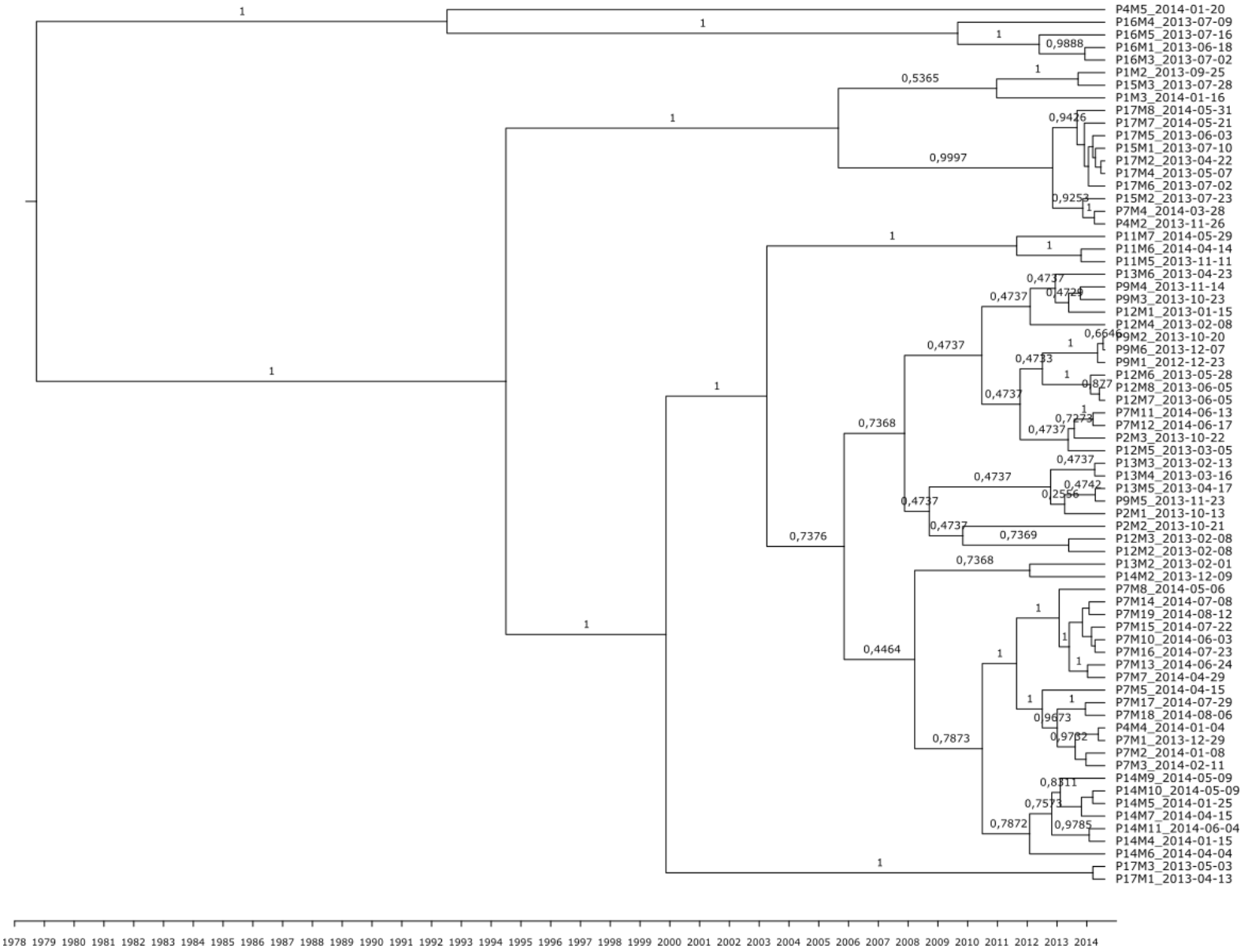


Figura 4.27. Árbol filogenético del clado correspondiente con el posible brote del HGUV datado con BEAST. *El árbol que se muestra es el consenso de árboles de los diferentes* "runs" *independientes en los que se ha empleado datación por nodo interno con los modelos de reloj estricto, población exponencial y una distribución normal del* prior, *en este caso, el de la tasa de evolución con lambda 1E-6.*

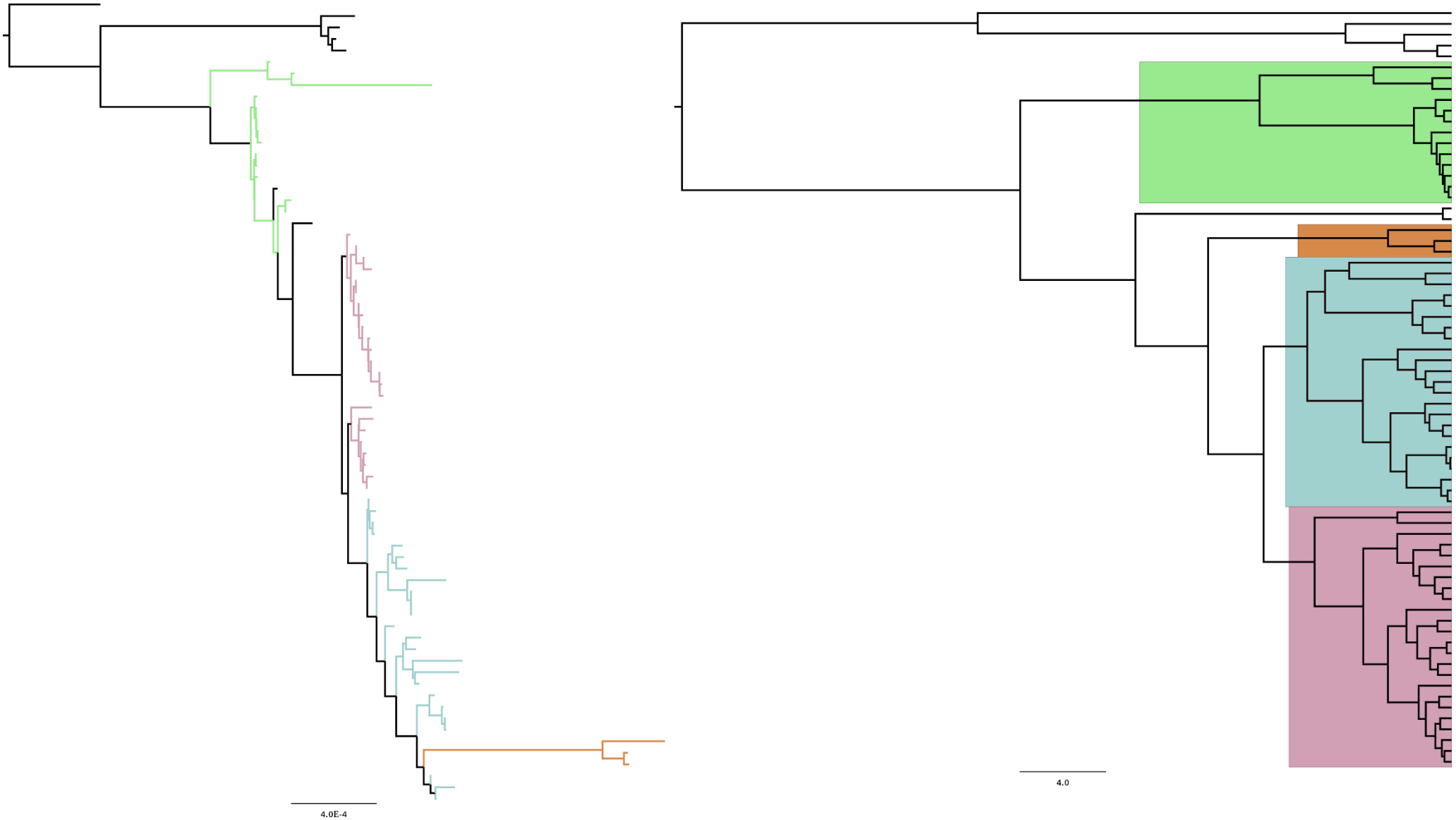


Figura 4.28. Comparación de topologías entre el subclado del brote del HGUV por máxima verosimilitud (IQTREE) frente a la reconstrucción por métodos bayesianos (BEAST).

Se indican en colores diferentes los clados entre los que se observan cambios topológicos y su correspondencia en ambos árboles.

La presencia de un conjunto numeroso de muestras muy próximas filogenéticamente nos hizo sospechar que hubiera un brote no detectado, pero no teníamos más evidencias que permitieran confirmarlo. Solicitamos al Servicio de Microbiología del HGUV información sobre los ingresos de los pacientes implicados en ese periodo para poder revisar las camas y traslados que hubieran podido llevar a la transmisión de este patógeno. Según la información con que contamos (Figura 4.29), el primero de los pacientes en ser ingresado sería el paciente 12 en una cama de Reanimación. La misma cama fue ocupada por los pacientes 9 y 13 sucesivamente durante los dos meses siguientes. Además, vemos cómo en el árbol filogenético las muestras de estos pacientes se encuentran agrupadas en los mismos clados, sin que exista una ordenación clara por paciente. También sabemos que los pacientes 7 y 17 estuvieron en dicha habitación y, aunque del paciente 2 no tenemos constancia de ello, sí estuvo en una habitación de la misma planta. Todos estos datos nos permitirían explicar el subclado rosa coloreado en el árbol de BEAST (Figura 4.28), que sí podrían tener un origen común y su agrupación en este caso tendría más sentido. Además, el paciente 14 estuvo en otra planta en la que permaneció ingresado anteriormente el paciente 9 en fechas posteriores a su paso por Reanimación. Los demás pacientes no parecen haber compartido instalaciones con el resto y su relación no está clara con los datos disponibles.

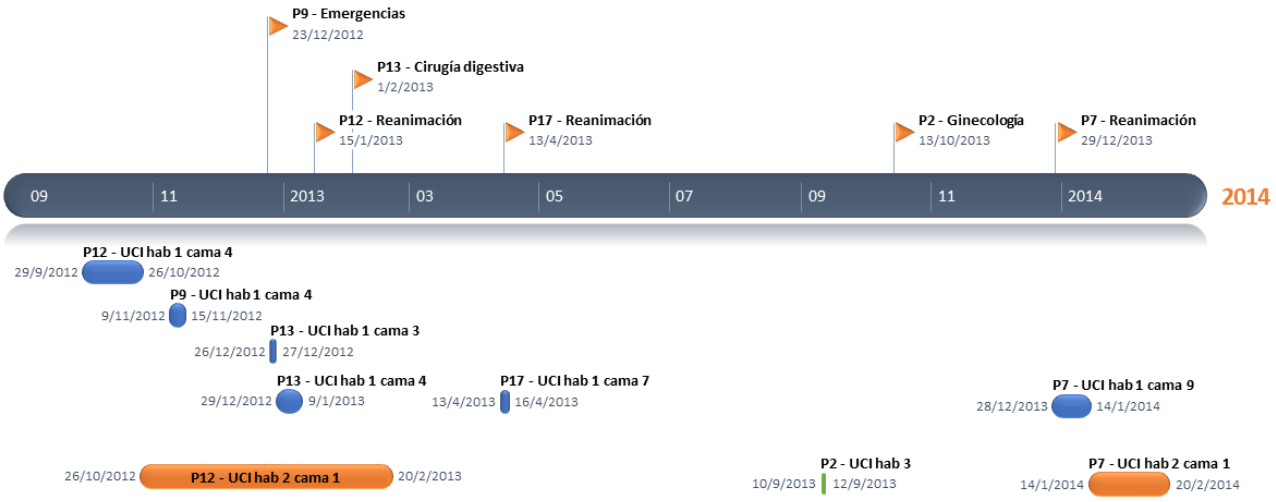


Figura 4.29. Cronología y localización de los pacientes que forman parte del brote del HGUV. *Se han incluido solamente aquellos pacientes de los que se sospecha que forman parte del brote y que han compartido cama o habitación en el área de la UCI, el único lugar donde han coincidido en el intervalo de tiempo aproximado de toma de muestras y, por tanto, posible foco. Las banderas de la parte superior representan la primera muestra de dicho paciente que tenemos, fecha de aislamiento y lugar en que se encontraba el paciente. En la parte inferior, los intervalos de tiempo de ingreso de los pacientes en UCI; del mismo color aquellos que han coincidido en la misma habitación.*

A pesar de no poder establecer una ruta de transmisión, sí podemos confirmar una relación entre pacientes, lo que evidencia la utilidad de las herramientas de análisis filogenético a partir de genomas completos para la detección de brotes. Sería necesario un estudio más amplio de los ingresos anteriores de estos pacientes, puesto que el brote posiblemente comenzó antes, e incorporar también pacientes positivos para *P. aeruginosa* de dicho periodo de tiempo que ayudarían a establecer posibles contactos o zonas que puedan suponer focos de infección.

Capítulo 4

4.4 Análisis conjunto de los genomas de los 3 brotes

Los hospitales participantes en los estudios epidemiológicos de los capítulos 1, 2 y 3 (Hospital Arnau de Vilanova, Hospital General Universitario de Valencia y Hospital General Universitario de Elche) se localizan en áreas geográficas próximas y se ha encontrado el mismo ST mayoritario (ST175) en dos de ellos, HAV y HGUE, por lo que quiso determinar el grado de relación filogenética a nivel genómico entre las cepas de estos hospitales.

La reconstrucción filogenética se realizó con los alineamientos generados a partir del mapeo; sin embargo, el genoma de referencia es diferente en las muestras del HGU de Valencia respecto a los otros dos mapeos, por lo que fue necesario un realineamiento. Para mayor simplicidad, dado que los pseudogenomas de cada muestra tienen la misma estructura y longitud que el genoma de referencia, se alinearon primero los dos genomas de referencia con PROGRESSIVEMAUVE, que permite reordenar los bloques colindantes (Figura 4.30). En este caso se han determinado 5 bloques conservados entre ambas. Posteriormente, se reordenaron los 3 alineamientos en base al orden de su respectiva referencia utilizando un script propio para obtener un alineamiento conjunto.

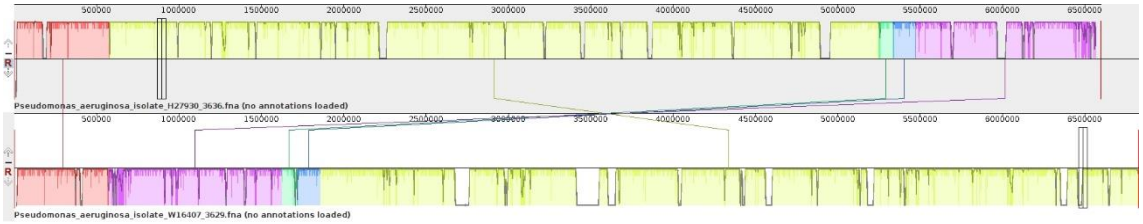


Figura 4.30. Alineamiento por progressiveMauve de los genomas de referencia. *En colores se representan los bloques localmente colineales en cada genoma.*

Se decidió emplear este método frente al uso de genomas ensamblados por el elevado número de muestras (167 en total) con un genoma aproximado de 6,5-7Mb y la alta fragmentación de los genomas, lo que complicaría el cálculo computacional del alineamiento. También se barajó la posibilidad de mapear las muestras del HGU de Valencia, cuyo ST predominante es el ST244, frente a la otra referencia, pero al estar más alejadas filogenéticamente se quería evitar la pérdida de variabilidad dentro de este ST por la ausencia de regiones no conservadas entre las dos referencias.

Análisis conjunto de los genomas de los 3 brotes

El alineamiento completo consta de 167 aislados y las 2 referencias con una longitud de 7.343.753 pb y un *core* al 95% de 5.008.766 pb. Se extrajeron los 183.339 SNPs del alineamiento para su reconstrucción por IQTREE con el modelo GTR+G4. El árbol obtenido por máxima verosimilitud (Figura 4.31) revela la relación entre las muestras del ST175 presentes en los 3 hospitales, compartiendo ancestro entre las procedentes del brote del HAV y el HGUV, además de encontrar próxima una de las muestras del HGUE. Incluso podemos detectar una posible transmisión entre dos muestras prácticamente idénticas entre Elche_42 y P5M2, pertenecientes al ST235.

.

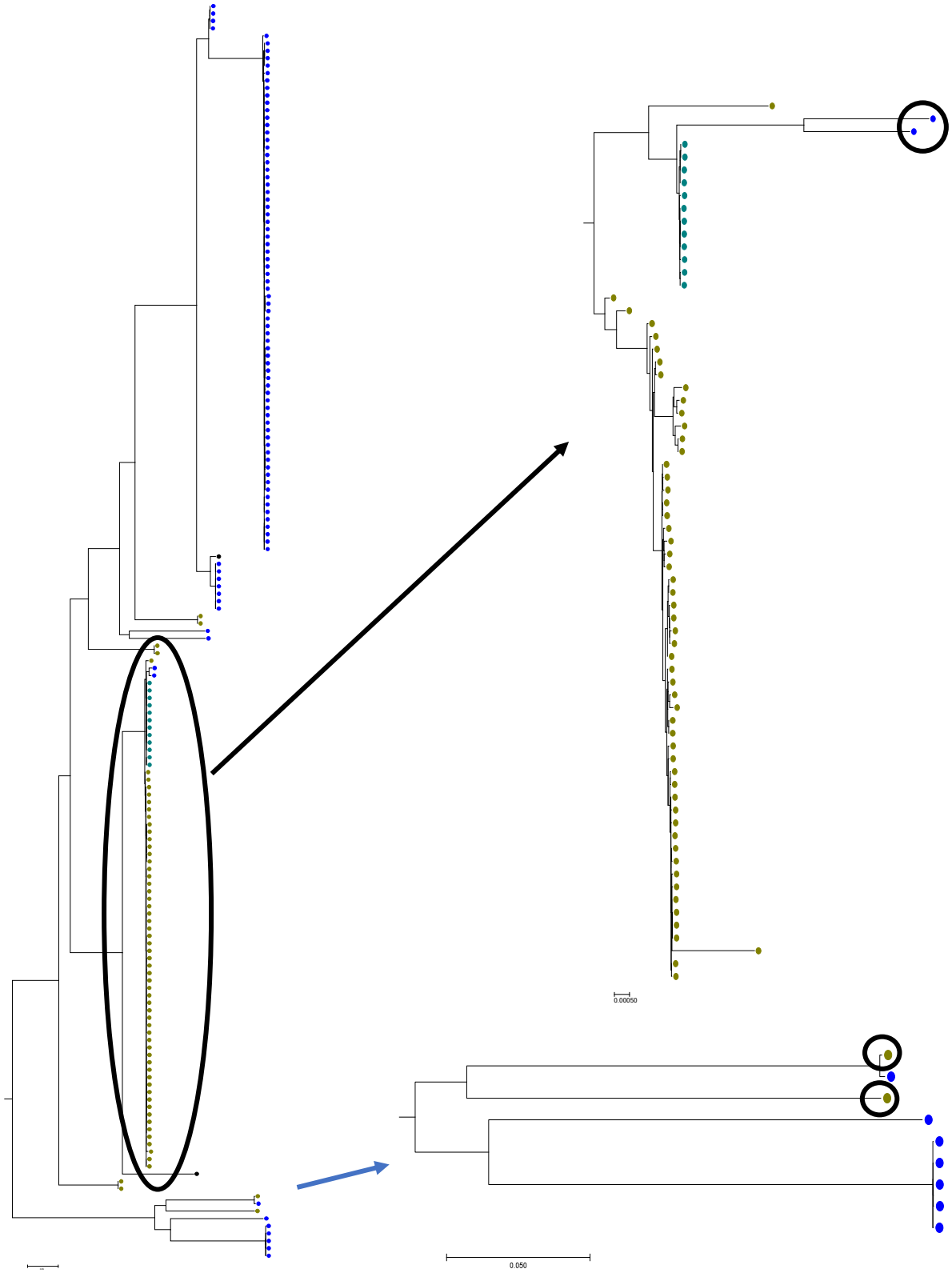


Figura 4.31. Árbol filogenético del alineamiento conjunto de SNPs de los aislados de los 3 hospitales. *El árbol, reconstruido con IQTREE utilizando el modelo GTR+G4 con el alineamiento de 183.339 SNPs. Se incluye además la información de ST en el extremo de cada rama en forma de puntos en función del hospital de procedencia: verde oscuro-HAV, amarillo-verdoso-HGUE y azul-HGUV. Se presentan los subclados del ST175 (arriba) y las de ST235 y ST316, principalmente (abajo)*

Capítulo 5

4.5 Estructura CRISPR en aislados de *P. aeruginosa* de diferentes hospitales

Los aislados hospitalarios descritos en los capítulos anteriores han sido empleados para estudiar en profundidad su estructura CRISPR. En total, se han analizado 167 genomas: 92 del HGUV, 63 del HGUE y 12 del HAV. El trabajo se ha desarrollado en colaboración con la Dra. Rachel Whitaker y Ted Kim, de la Universidad de Illinois Urbana-Champaign.

A partir del procedimiento puesto a punto por England et al. (2018) con aislados de *P. aeruginosa* de pacientes con fibrosis quística, iniciamos la determinación de los *spacers* contenidos en el locus o loci CRISPR. Su detección está basada en la presencia de las repeticiones de CRISPR específicas de *P. aeruginosa* en las lecturas de Illumina. Las repeticiones deben encontrarse en los dos extremos de la región susceptible de ser un *spacer* y en la misma orientación para obtener, en última instancia, la secuencia del *spacer*.

Las regiones repetidas son de alrededor de 28 pb y la longitud del *spacer* de 32 pb, por lo que el tamaño final es de cerca de 90 pb. La longitud de estas lecturas (300 pb) permite aplicar esta metodología, ya que necesitamos que toda la estructura esté contenida en una única lectura; si se utilizara un método de secuenciación de lecturas cortas, esto sería más complicado. Sin embargo, el nivel de cobertura es igualmente relevante, como hemos visto en nuestros resultados.

Además del procedimiento, utilizamos la base de datos de 3152 *spacers* numerados para su identificación que se creó en dicho trabajo, ya que el resultado del programa es la obtención de la secuencia *spacer* y, así, posteriormente tratar de establecer la ordenación de los mismos dentro del locus o loci gracias a los registros de 754 cepas con 878 loci diferentes, teniendo en muchas de ellas combinaciones de loci.

Los primeros resultados de los tres conjuntos de datos indican que hay correlación entre el ST y los loci CRISPR (Figura 4.32). En el brote del Arnau, en que todos los aislados pertenecen al ST175, observamos que el perfil de *spacers* corresponde con el locus 160 (Tabla 4.12). Esta misma situación se produce en el brote de Elche, con

ST175 y L160. Los aislados que quedan fuera del brote en este estudio tienen un ST distinto, de la misma manera que ocurre con el locus CRISPR, e incluso hay aislados sin CRISPRs; sin embargo, seguimos encontrando esta conservación de la estructura CRISPR entre aislados del mismo ST. Un ejemplo de ello son las muestras Elche_62 y Elche_64, ambas del ST348 y L370.

En las muestras de Elche vemos cómo la disminución en la cobertura de los genomas afecta a la eficiencia de este procedimiento. Hay muchos casos de loci en los que alguno de sus *spacers* no ha sido detectado. Aunque pueden producirse pérdidas de *spacers* de manera natural, la ausencia de un patrón hace pensar que la falta de cobertura impide detectarlos, dados los criterios de filtrado. Los resultados de las muestras del HGUV a este respecto son peores, con un alto porcentaje de *spacers* no detectados si consideramos que los loci que hemos definido según los descritos son los correctos. Se determinó que los que mejor coincidían con los descritos son los loci L550 junto al L487, presentes en la mayoría de los aislados que, además, se asocian al ST244.

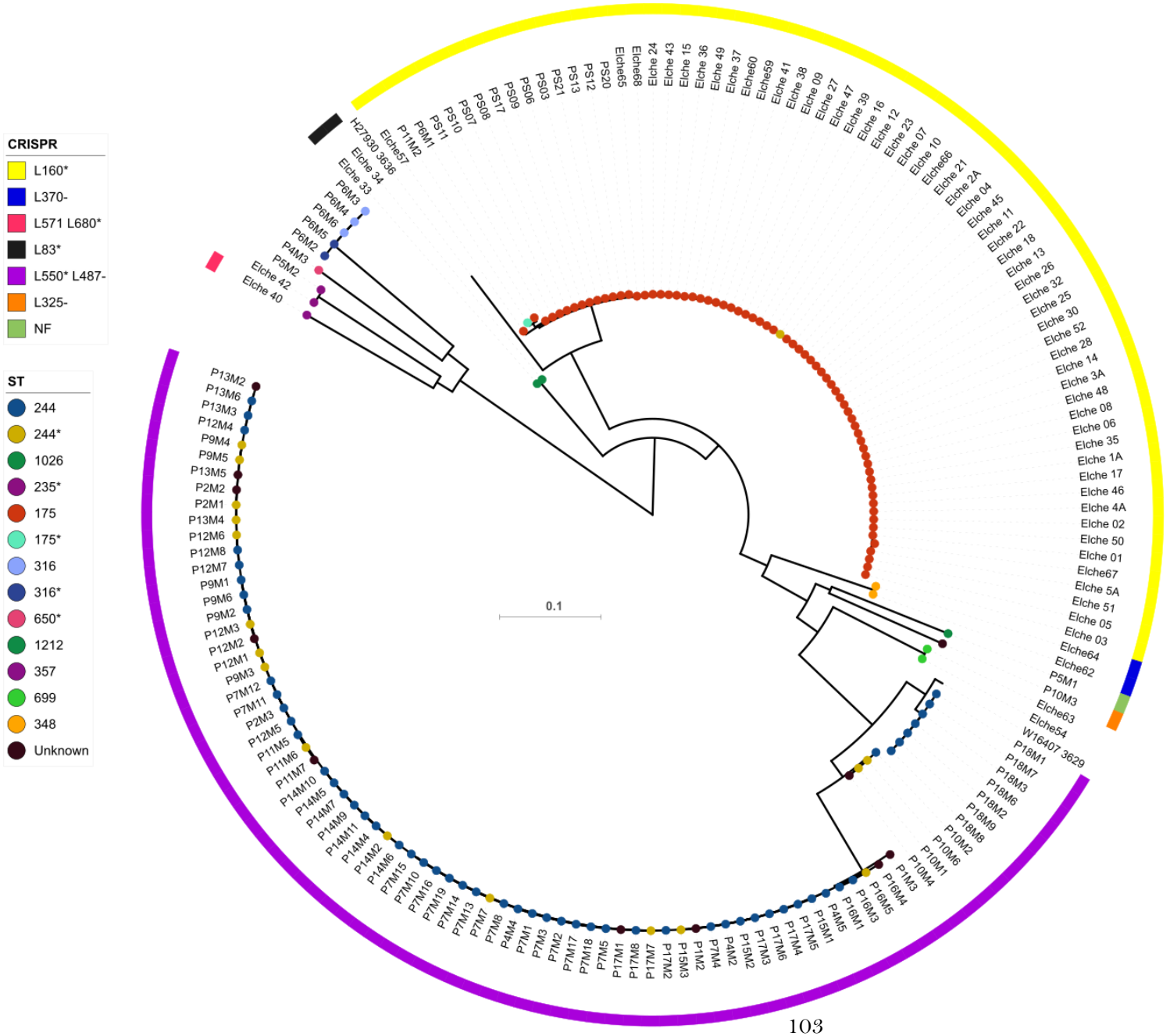


Figura 4.32. Árbol filogenético de los 3 sets de datos conjuntamente con la información de estructura CRISPR y su ST. *El árbol del capítulo anterior construido por IQTREE con el alineamiento de SNPs permite visualizar la correspondencia de estructura CRISPR (con los datos preliminares del primer análisis) con su ST.*

strain	patient	date	locus																
PS21	1	02/05/16	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679	509*	new1	
PS03	2	29/09/16	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679	889*	new2	
PS06	3	15/05/17	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679	509*		
PS07	4	24/05/17	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679			
PS08	4	29/05/17	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679	1105*	1105*	251*
PS09	3	29/05/17	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679	1105*		
PS10	4	19/06/17	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679	1105*		
PS20	5	03/07/17	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679	1105*		
PS11	4	06/07/17	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679			
PS12	5	06/07/17	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679	765*	889*	287*
PS13	5	10/07/17	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679	1033*		
PS17	6	24/08/17	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679	251*		

Tabla 4.12. Resultados CRISPR para las muestras del HAV. *Se marcan con * las secuencias detectadas adicionalmente con algún cambio/mutación.*

Hubo cepas en la que encontramos *spacers* repetidos o con pequeñas variaciones e incluso más de 30 posibles nuevos *spacers* entre todas las muestras de los 3 hospitales, lo que evaluamos posteriormente.

Se construyó una base de datos más pequeña solamente con los *spacers* detectados en todas las muestras, incluyendo las secuencias nuevas encontradas en nuestros aislados. La utilizamos para su detección por BLASTn nuevamente, sin aplicar todos los filtros; de esta manera, evitamos la limitación de detectar las repeticiones en ambos extremos de la lectura de las regiones peor cubiertas. Los resultados, tal y como se esperaba, revelaron un mayor número de secuencias. Quedó reflejado que el cribado del procedimiento anterior estaba dejando de detectar un gran número de secuencias *spacer*. Especialmente en el grupo de muestras del HGUV, con niveles de cobertura más bajos y estructura de CRISPR más compleja, se ha podido ver esta recuperación, marcada por todos los *spacers* de color rojo/amarillo (Tabla 4.14) que anteriormente no se encontraron y cuya casilla correspondiente quedaba marcada en negro (Material suplementario, Tabla 7.17). Por el contrario, la falta de filtros lleva a que se detecten muchas secuencias que se corresponderían con *spacers* que en el locus (o loci) que definimos no se encuentran. Es posible que no pertenezcan a la estructura CRISPR, sino que se trate de secuencias procedentes de profagos que se encuentren en otras zonas del genoma. Ante esta posibilidad, se decidió limitar el estudio completo de *spacers* fuera de locus a algunas muestras por set, puesto que, en su mayoría, además encontramos que tenían variaciones y su identidad no era del 100% (marcados en amarillo, Tabla 4.13).

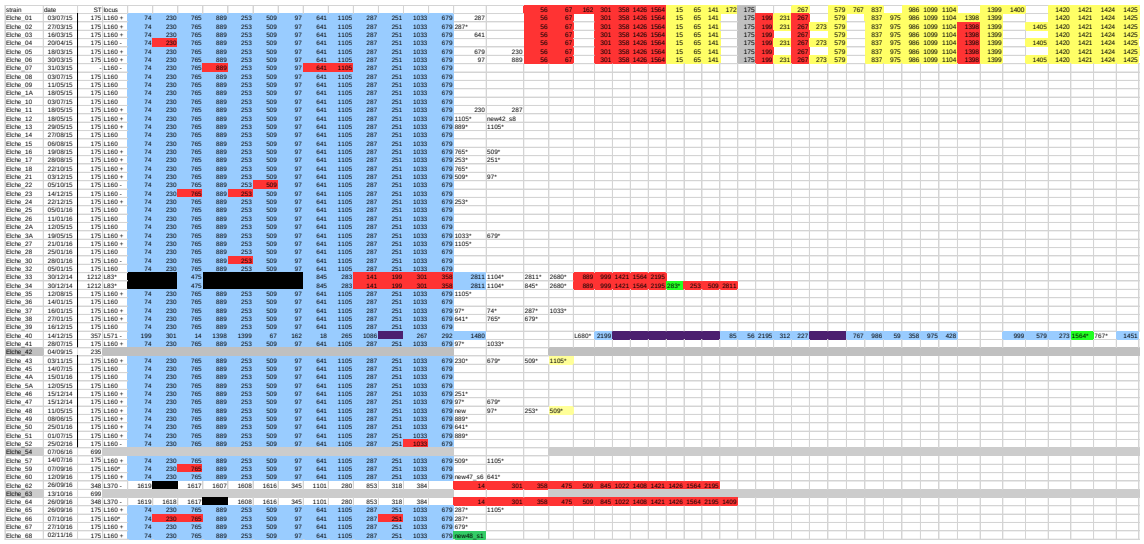


***Tabla 4.13. Resultados muestras del HGUE detectando directamente las secuencias* spacer.** *Las casillas en rojo corresponden con las secuencias que anteriormente no se habían detectado (negro). Las 6 primeras muestras son las únicas que se analizaron por completo en cuanto a espaciadores fuera de locus.*

No se pudieron detectar todos los *spacers* para considerar algunos loci completos, y en la mayoría de casos no se mantenían las mismas "deleciones" entre muestras del mismo paciente, lo que hizo pensar, nuevamente, en el problema de las coberturas. Sin embargo, en las muestras de los pacientes 10 y 18 del HGUV hay una región amplia del L487 (abarcando 12 y 8 *spacers*, respectivamente) que se mantiene en todas las muestras que contienen el locus L487. Estos dos pacientes, cuyos aislados pertenecen al ST244, no forman parte del brote según nuestros resultados mediante filogenias basadas en genomas completos. De confirmarse la diferencia de estructura en CRISPR respecto al resto de muestras del ST244, podría considerarse el uso conjunto de la determinación del ST junto a CRISPR como una posible primera aproximación para la detección de un brote, a la espera de confirmación por medio de estudios filogenéticos, o como soporte a los resultados obtenidos por filogenias.

Estructura CRISPR en aislados de *P. aeruginosa* de diferentes hospitales

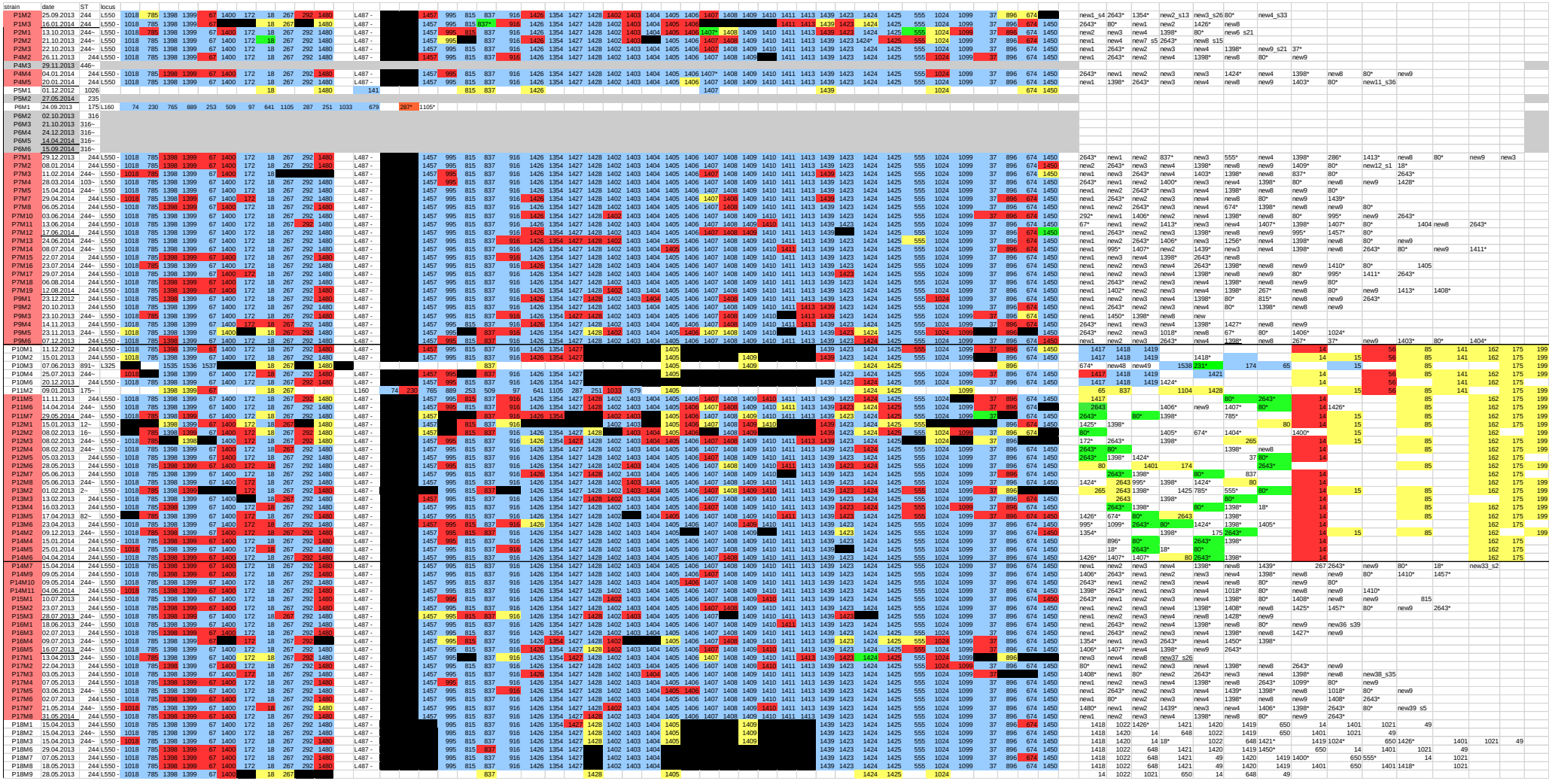


***Tabla 4.14. Resultados CRISPR de las muestras del HGUV detectando directamente las secuencias* spacer.** *Las casillas en rojo corresponden con las secuencias que anteriormente no se habían detectado y las amarillas son las no detectadas anteriormente pero que no tienen el 100% de identidad. Las muestras delimitadas por las líneas son las únicas que se analizaron por completo en cuanto a espaciadores fuera de locus, en verde las que quedan fuera de locus que se detectan tanto en el procedimiento anterior como el de detección directa de secuencia. Muestras en color salmón correspondientes al brote.*

También se trató de definir el orden de los *spacers* correctamente para comprobar si encontramos los loci tal y como hemos estimado, se decidió extraer los loci de CRISPR en los genomas ensamblados de las muestras más representativas en los 3 grupos (Tabla 4.15). Los resultados de anotación por PROKKA de los genomas que habíamos reconstruido previamente con SPADES contenían dicha información. Este programa identificó en estos aislados los *contigs* en que se encontraban y una predicción del número de *spacers* contenidos en cada loci.

Un extracto del informe para el genoma de P10M6:

PROKKA log:

[11:13:17] Searching for CRISPR repeats

[11:13:17] CRISPR1 gnl|HGV|LHBEMMIO_150 86 with 26 spacers

[11:13:17] CRISPR2 gnl|HGV|LHBEMMIO_178 637 with 13 spacers

[11:13:17] Found 2 CRISPRs

Se extrajeron de los *contigs* las regiones anotadas con el nombre de rpt_family="CRISPR" para la asignación del número de *spacers* nuevamente por BLASTn. Al partir de los ensamblados, tenemos los loci completos, por lo que, además de las correspondencias, podemos establecer el orden real de estos *spacers*.

En las muestras del HGUV (Tabla 4.15), detectamos dos loci, tal y como vimos con anterioridad. En algunas muestras, como la P1M2 y la P16M1, parecen observarse 3 loci pero resultan ser 2 por la fragmentación del locus más largo en *contigs* diferentes por la baja cobertura. Confirmamos la existencia de 6 *spacers* nuevos entre los más de 30 detectados como posibles nuevas secuencias con el programa de England, et al (2018), 4 en un locus (L487) y 2 en el otro (L550) que, además, se encuentran en la región más próxima a la secuencia líder, por lo que son incorporaciones recientes. Estas incorporaciones se dan en todas las muestras estudiadas del brote, lo cual respalda un origen común. Uno de los aspectos más relevantes es que la estructura del locus anteriormente denominado L487 es más compleja de lo que se preveía. A partir de la combinación de *spacers* completa, vemos que en las muestras de los pacientes 10 y 18 aparecen secuencias que no forman parte del L487 (del *spacer* 1416 al 1421).

Sample	Locu	num-sp	num-spa_detected		CRISPR LOC	structure																																											
						L281					1412	995	815	837	916	1426	1354	1427	1428	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1413	1414	1415	1416	1417	1418	1419	1420	1421	1422	1423	1424	1425	555	1024	1099	37	896	674	
P10M6	1	26	25	R							1457	995	815	837	916	1426	1354	1427															1416	1417	1418	1419	1420	1421	1439	1423	1424	1425	555	1024	1099	37	896	674	1450
	2	13	12	F	L550+	1018	785	1398	1399	67	1400	172	18	267	292	1480	2576																																
P18M1	1	20	19	R	L280+	1018	785	648	650	1022	1021	49	1401	14	1398	1399	67	1400	172	18	267	292	1480	2576																									
	2	26	25	R	L580-							995	815	837	916	1426	1354	1427		1402	1403	1404													1418	1419	1420	1421	1439	1423	1424	1425	555	1024	1099	37	896	674	1450
P1M2	1	15	15	R	L487-/+	80	new2	new9	new1	new3	1457	995	815	837	916	1426	1354	1427	1428	1402																													
	2	17	16	R	L550+	new37¿?	new8	new4		1018	785	1398	1399	67	1400	172	18	267	292	1480	2576																												
	3	18	18	F		1403	1404	1405	1406	1407	1408	1409	1410	1411	1413	1439	1423	1424	1425	555	1024	37	896																										
P2M3	1	17	15	R	L550+	new37¿?	new8	new4		1018	785	1398	1399	67	1400	172	18	267	292	1480	2576																												
	2	37	35	R	L487-/+		new2	new9	new1	new3	1457	995	815	837	916	1426	1354	1427	1428	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1413									1439	1423	1424	1425	555	1024	1099	37	896	674	1450
P4M5	1	37	35	F		new2	new9	new1	new3		1457	995	815		916	1426	1354	1427	1428	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1413									1439	1423	1424	1425	555	1024	1099	37	896	674	1450
	2	17	15	R	L550+	new37¿?	new8	new4		1018	785	1398	1399	67	1400	172	18	267	292	1480	2576																												
P5M1	1	5	4	F	L396+	141	658	1663	1028																																								
P6M1	1	14	13	R	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679																															
P7M5	1	17	15	F	L550+	new37¿?	new8	new4		1018	785	1398	1399	67	1400	172	18	267	292	1480	2576																												
	2	37	36	R	L487-/+	80	new2	new9	new1	new3	1457	995	815	837	916	1426	1354	1427	1428	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1413									1439	1423	1424	1425	555	1024	1099	37	896	674	1450
P9M2	1	17	15	F	L550+	new37¿?	new8	new4		1018	785	1398	1399	67	1400	172	18	267	292	1480	2576																												
	2	37	36	R	L487-/+	80	new2	new9	new1	new3	1457	995	815	837	916	1426	1354	1427	1428	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1413									1439	1423	1424	1425	555	1024	1099	37	896	674	1450
P11M2	1	12	12	F	L160-	74	230	765	889	253	509	97	641	1105	287	251	1033																																
P11M5	1	17	15	R	L550+	new37¿?	new8	new4		1018	785	1398	1399	67	1400	172	18	267	292	1480	2576																												
	2	37	36	R	L487-/+	80	new2	new9	new1	new3	1457	995	815	837	916	1426	1354	1427	1428	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1413									1439	1423	1424	1425	555	1024	1099	37	896	674	1450
P12M8	1	37	36	F	L487-/+	80	new2	new9	new1	new3	1457	995	815	837	916	1426	1354	1427	1428	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1413									1439	1423	1424	1425	555	1024	1099	37	896	674	1450
	2	17	15	R	L550+	new37¿?	new8	new4		1018	785	1398	1399	67	1400	172	18	267	292	1480	2576																												
P13M4	1	17	15	F	L550+	new37¿?	new8	new4		1018	785	1398	1399	67	1400	172	18	267	292	1480	2576																												
	2	30	29	R	L487-/+	80	new2	new9	new1	new3	1457	995	815	837	916	1426	1354	1427	1428	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1413									1439	1423	1424	1439							
	3	10	9	F	resto L487																																				1424	1425	555	1024	1099	37	896	674	1450
P14M10	1	37	36	F	L487-/+	80	new2	new9	new1	new3	1457	995	815	837	916	1426	1354	1427	1428	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1413									1439	1423	1424	1425	555	1024	1099	37	896	674	1450
	2	17	15	R	L550+	new37¿?	new8	new4		1018	785	1398	1399	67	1400	172	18	267	292	1480	2576																												
P15M1	1	17	15	F	L550+	new37¿?	new8	new4		1018	785	1398	1399	67	1400	172	18	267	292	1480	2576																												
	2	37	36	R	L487-/+	80	new2	new9	new1	new3	1457	995	815	837	916	1426	1354	1427	1428	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1413									1439	1423	1424	1425	555	1024	1099	37	896	674	1450
P16M1	1	24	23	F	L550-/L487-	new37¿?	new8	new4		1018	785	1398	1399	67	1400	172	18	267																						1423	1424	1425	555	1024	1099	37	896	674	1450
	2	28	27	R	L487-/+	80	new2	new9	new1	new3	1457	995	815	837	916	1426	1354	1427	1428	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1413									1439	1423									
	3	4	4	R	resto L550													267	292	1480	2576																												
P17M5	1	17	15	R	L550+	new37¿?	new8	new4		1018	785	1398	1399	67	1400	172	18	267	292	1480	2576																												
	2	37	36	R	L487-/+	80	new2	new9	new1	new3	1457	995	815	837	916	1426	1354	1427	1428	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1413									1439	1423	1424	1425	555	1024	1099	37	896	674	1450
Elche_06	1	14	14	R	L160+	74	230	765	889	253	509	97	641	1104	1105	287	251	1033	679																														
Elche_34	1	8	5	R	¿	283	845	475			1104	2811																																					
Elche_40	1	16	15	F	L571+	199	301	14	1398	1399	67	162	18	265	1086	924	267	292	1480	2576																													
(loci de cepa JTTZ01)	2	6	5	R	L572	273	579	999	1104	953																																							
	3	15	14	R	L570-	428	975	358	59	986	767	1112	1057	227	312	2195	56	85	2199																														
Elche_64	1	13	12	R	L370	384	318	853	280	1101	345	1616	1608	1607	1617	1618	1619																																
PS06	1	14	13	R	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679																															
PS21	1	14	13	R	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679																															
PS03	1	14	13	R	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679																															
Ref brote H27930	1	13		F	L160?	230	765	765	889	253	509	97	641	1105	287	251	1033	679																															
	2	6		F		2809	2808	2807	174	65	2576																																						
	3	30		R		796	857	1686	686	291	274	468	376	276	883			1312				3048?																											
Ref Hgral W16407	1	26		F								995	815	837	916	1426	1354	1427		1402	1403	1404													1418	1419	1420	1421	1439	1423	1424	1425	555	1024	1099	37	896	674	1450
	2	19		R	L280	1018	785	648	650	1022	1021	49	1401	14	1398	1399	67	1400	172	18	267	292	1480																										

Tabla 4.15. Resultados CRISPR de las muestras más representativas de los 3 hospitales (diferenciados por colores) detectando las secuencias spacer y su orden tras el ensamblado. *En morado se marcan las secuencias de spacer nuevas, coincidentes con la base de datos de posibles nuevos spacers; en amarillo, coincidentes con alguna secuencia pero con cambios puntuales; de verde posiciones reales dentro del locus que no coincide con ninguna secuencia de las bases de datos; en azul, las secuencias de spacers que no se encuentran en el esquema del locus que hemos definido.*

Filogenéticamente son muy próximas a las del brote, que no tienen dichos *spacers*, por lo que sería lógico considerar que no se trata de dos loci diferentes, sino que es el mismo con pequeñas variaciones. La deleción de fragmentos dentro del locus puede ocurrir para regular su tamaño, por lo que se ha tratado de encontrar un locus que pudiera dar lugar a ambos perfiles por medio de deleciones, y este sería el L251 (Figura 4.33). Este locus completo podría ser el que contuviera el ancestro común a las cepas del ST244 (Figura 4.34). Sin embargo, para darse la configuración actual deberían haberse producidos 3 deleciones diferentes: una en el ancestro de las cepas del brote y otras dos en las cepas de cada paciente (10 y 18), ya que las cepas del brote no tendrían del *spacer* 1414 al 1421, el paciente 10, del 1425 al 1415, y el paciente 18, del 1405 al 1417. Todo ello considerando que no hayamos perdido esa región en la secuenciación.

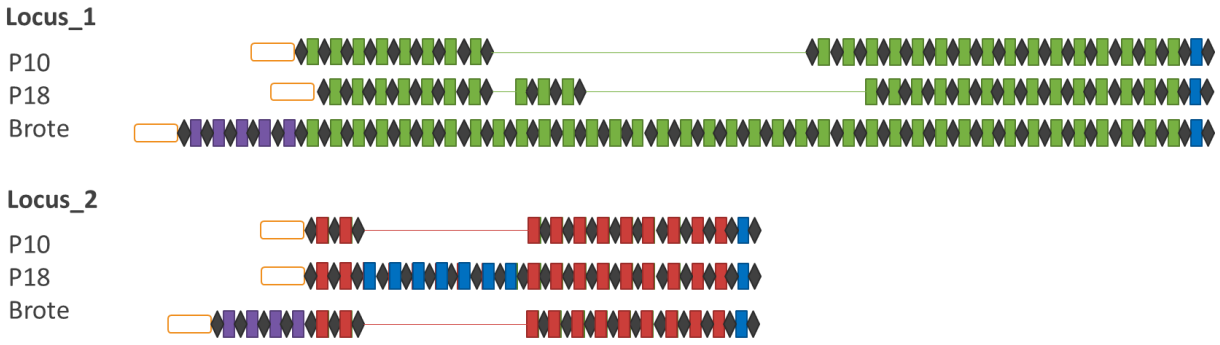


Figura 4.33. Estructura "alineada" de los 2 loci CRISPR encontrados en las muestras del ST175. *Ambos locus se han alineado para que coincidan los espaciadores y se visualice mejor la ausencia de algunos de ellos en la estructura del locus detectada como original en cada caso (L251 y L550, respectivamente). Los rectángulos naranja representan la secuencia líder; los rombos, las repeticiones; y los rectángulos pequeños, los espaciadores. Sus colores dependen de si son nuevos en nuestra base de datos (morado), ya descritos pero no pertenecientes al locus (azul) o los que pertenecen al locus L251 (verde) o L550 (rojo).*

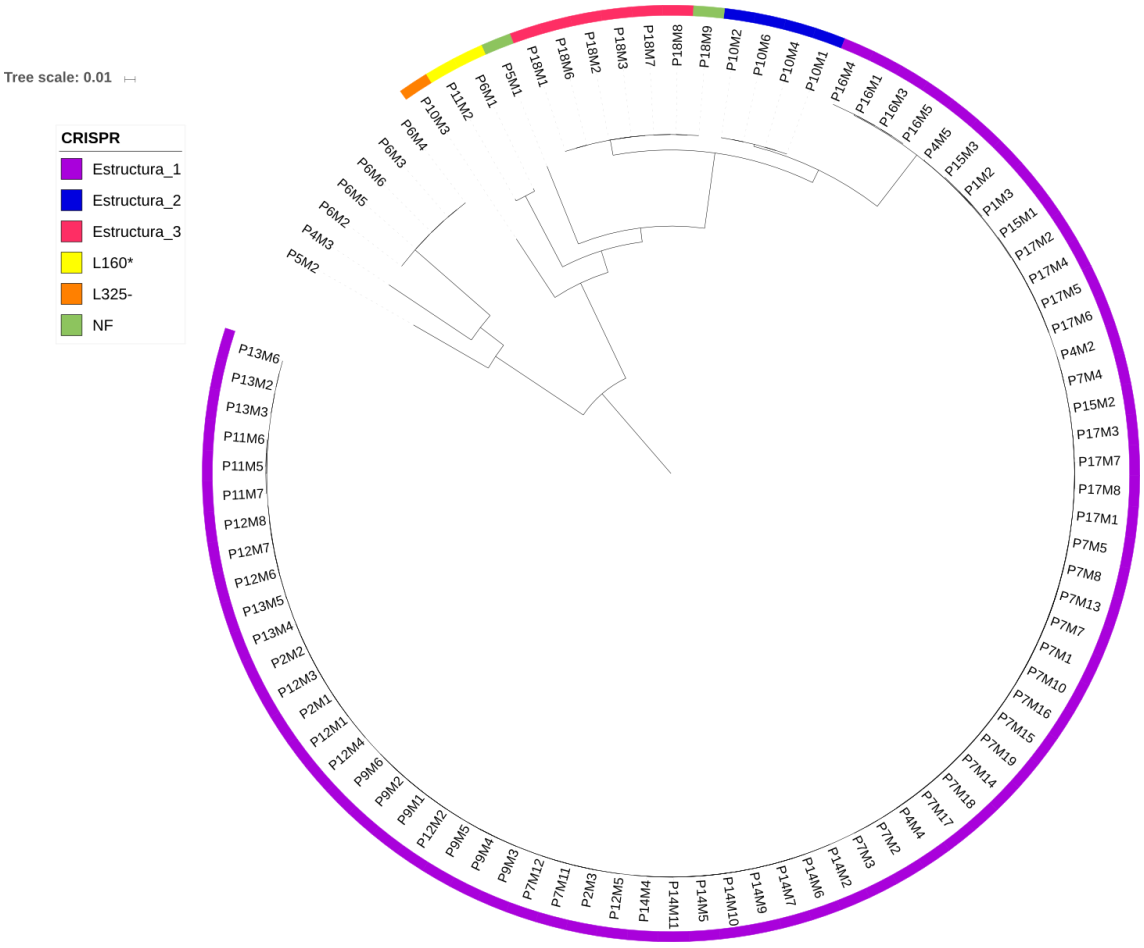


Figura 4.34. Árbol filogenético de las muestras del HGUV a partir de SNPs con la información de las nuevas estructuras de CRISPR detectadas. *La estructura 1, 2 y 3 hace referencia a la combinación de los locus 1 y 2 con estructura diferente descritos en el P10, P18 y resto del brote.*

En las muestras más representativas del brote de Elche no hay grandes cambios, salvo la aparición de una secuencia en la muestra Elche_06 en medio del locus que no corresponde con la estructura del L160, y la muestra Elche_40, que contiene 3 loci casi idénticos, que en la base de datos se corresponden con los de la cepa JTTZ01. Por otro lado, en las muestras del hospital Arnau (Capítulo 1) no existe ninguna variación y todas mantienen el L160.

Para acabar con el estudio del locus L251*, queríamos descartar que pudiera haberse producido una única deleción amplia en todo el conjunto de muestras, siendo esta la explicación más sencilla para el patrón observado. Una posibilidad es que los *spacers* de la región con deleciones (de 1428 a 1421) fueran idénticos a otros de los detectados, ya que sabemos que algunos *spacers* de la base de datos son idénticos, aunque posean nombres diferentes (como ocurre con 1439 y 1422), o con más del 90% de identidad, debido a alguna mutación o errores de lectura. De esta manera, si los *spacers* "extra" en los pacientes 10 y 18 (1416 a 1421), que ocupan la zona delecionada en las muestras del brote, son similares a las 1428 a 1413 que solo encontramos en las del brote, podríamos descartar que hubiera tantas deleciones. Por otro lado, también podrían no haberse detectado porque no hubieran conseguido ensamblarse, por lo que fue necesario volver a buscar dichos *spacers* a partir de las lecturas. Mediante BLASTN nuevamente quedaron descartadas ambas posibilidades, aunque no podemos estar seguros de que no hubiera lecturas porque realmente no esté en el genoma frente a una secuenciación insuficiente.

Se intentó comprobar la procedencia de los *spacers* que han podido ser delecionados en los aislados que constituyen el brote del HGUV. Sin embargo, esto solo ha sido posible para 4 de ellos (Tabla 4.16) y no hemos podido encontrar una explicación a la ventaja de perder la protección frente a estos fagos.

SPACER	**VIRUS**	**CARACTERÍSTICAS**
1419	cB_PaeS_PMG1 - lítico	Se caracteriza por ciertas peculiaridades del ciclo de la infección lítica y forma un halo alrededor de colonias negativas. Los fagos que forman halos tienen la capacidad de destruir polisacáridos en el curso de la evolución, como respuesta a la capacidad de las bacterias para formar biopelículas.
	Bacteriófago D3 - lisogénico	El bacteriófago D3 es capaz de lisogenizar *Pseudomonas aeruginosa* PAO1 (serotipo O5), conversión el antígeno O de O5 a O16 y la O-acetilación de N-acetilfucosamina.
1420	D3112 - lisogénico	Representa uno de los dos grupos distintos de fagos transponibles encontrados en el patógeno oportunista clínicamente relevante, *Pseudomonas aeruginosa.* Infecta por *pili.*
1421	MP22 - lisogénico	Muestra sintenia y similitud de la secuencia de nucleótidos y proteínas respecto a los ORF correspondientes del fago estrechamente relacionado, D3112.
	MP42 - lisógenico	Su genoma es similar a los de los fagos DMS3 y MP22 (lisogénicos) de *P. aeruginosa.*
	JBD30, 88a, 69, 93 - lisogénicos	*unclassified* D3112virus

Tabla 4.16. **Spacers** ***y virus a los que podrían corresponder.*** *Cuatro de los* spacers *que teóricamente se delecionaron en las cepas no pertenecientes al brote, fueron asignados a uno o varios virus de los que pudo proceder por correspondencia de la secuencia por BLAST mediante base de datos de virus propios de* P. aeruginosa.

5. Discusión

Discusión

A lo largo de la presente tesis se han planteado tres situaciones de transmisión nosocomial en las que el estudio de infección por *P. aeruginosa* con los métodos de rutina como el tipado o el estudio fenotípico no son suficientes. Por ello, se ha planteado el estudio de genomas completos del patógeno *P. aeruginosa* como herramienta alternativa, capaz de aportar información relevante para los profesionales sanitarios que no es posible con otros métodos.

En el primer trabajo colaboramos con el HAV y se planteó la necesidad de confirmar una sospecha de brote con pocos casos pero elevada tasa de mortalidad, lo que requería una actuación inmediata. En este tipo de situaciones suelen utilizarse métodos de tipado rápidos, como el MLST (Curran *et al.*, 2004). Sin embargo, el MLST se centra únicamente en una fracción muy pequeña del genoma (7 de los >6000 genes), por lo que no siempre podemos asumir que se trata del mismo clon entre dos aislados con mismo ST. En las 12 muestras analizadas se detecta el perfil ST175 únicamente, un tipo ampliamente distribuido a nivel global (Oliver *et al.*, 2015) y detectado en varios hospitales repartidos por la geografía española (García-Castillo *et al.*, 2011; Gomila *et al.*, 2013), por lo que, al ser tan frecuente, no es suficiente para establecer un origen común de las muestras presuntamente implicadas en un brote.

Mediante la secuenciación de genomas completos es posible observar toda la variabilidad presente en las muestras analizadas y, de esta manera, vemos que en el caso del HAV los aislados analizados difieren en aproximadamente un 0,0003% del genoma con la metodología utilizada, 22 SNPs en total entre las 12 muestras. Además, la reconstrucción filogenética muestra la agrupación de las muestras por paciente con 0 ó 1 diferencia respecto a los SNPs del *core*, lo que respaldaría la idea de que únicamente se ha producido un evento de transmisión en cada uno de ellos de un solo clon, según la información disponible y el número de muestras incluidas en el estudio.

Es cierto que el número de SNPs encontrados es superior al esperado teniendo en cuenta la tasa de mutación de una cepa normomutadora, pudiendo variar de 1-3 SNPs al año (Cramer *et al.*, 2011; Marvig *et al.*, 2013; Snyder *et al.*, 2013). Por otro lado, se han descrito también cepas hipermutadoras en las que la tasa puede superar los 100 SNPs al

año (Oliver *et al.*, 2000; Feliziani *et al.*, 2014), lo que complica definir el número de SNPs máximo para considerar la presencia de un brote.

Un trabajo con planteamiento muy similar publicado recientemente (B J Parcell *et al.*, 2018) realiza la investigación de un brote en una UCI con 6 pacientes afectados en un intervalo de poco más de un año con el fin de establecer el posible foco y comparando, a su vez, con otras metodologías de tipado básico. De la misma manera, observan que solamente con la filogenia, construida por máxima verosimilitud, sin necesidad de *tests* accesorios pueden determinar estas relaciones. Lo significativo en este estudio es que las relaciones entre paciente y foco son confirmadas a pesar del elevado número de SNPs iniciales que son detectados en el alineamiento (varios miles en cada clado). Esto es reducido drásticamente con la aplicación de un filtro de las zonas en las que puede haberse producido recombinación determinadas por un programa específico, como es Gubbins (Croucher *et al.*, 2014), mientas que nosotros contamos con 22 SNPs totales sin aplicar este tipo de filtro. La diferencia tan significativa en el número de variantes de nuestro estudio indica que entre los aislados de nuestro análisis no se ha producido recombinación o, en el caso de que se haya dado, las cepas circulantes no presentan diferencias suficientes para que este evento sea detectado. Esta metodología, como vemos, facilita la detección de eventos de intercambio de material genético, lo que permite realizar filtrados para corregir esta variación no derivada de mutaciones que nos podría enmascarar o dificultar los análisis evolutivos.

En el segundo capítulo se analiza un brote con la misma metodología, pero con ciertas diferencias en los objetivos. En este caso el brote ya está establecido y el número de pacientes es alto y sigue incrementándose, por lo que el interés radica en comprobar el grado de extensión y verificar las posibles fuentes del mismo. La utilización de herramientas de epidemiología genómica ha sido clave para visualizar una situación más compleja de la esperada. La estructura del árbol revela que los casos que forman parte del brote presentan un elevado número de diferencias, formándose subclados que podrían proceder de brotes diferentes a partir de un clon introducido mucho tiempo antes en el hospital y del que pueda haber varios focos, como ocurría en el trabajo de Parcell *et al.* (2018).

Adicionalmente, nos ha permitido discriminar los pacientes infectados de otras dependencias que sí formarían parte de este brote. Un ejemplo de ello son las muestras Elche_59 y Elche_66, que están incluidas en estos subclados, correspondiendo a

pacientes que se encontraban en Reanimación y Neurología, respectivamente, a diferencia del brote inicial, localizado en Urología.

También hemos comprobado cómo aislados que comparten el ST175 son muy distantes en la filogenia respecto al resto como para considerar que tengan un mismo origen próximo en el tiempo, por lo que se pierde la correlación entre ST y brote. Aunque no hayamos podido acotarlo debido a su complejidad, sí podemos descartar los casos más alejados filogenéticamente o detectar eventos de transmisión entre pacientes ajenos al evento investigado, como ocurre con Elche_33 y Elche_34, entre los cuales solamente hay una diferencia de 3 SNPs. Además, recurriendo a los datos epidemiológicos vimos que habían estado en la misma área y la toma de muestra se realizó con pocos días de diferencia. Más aún, hemos podido incluso detectar posibles errores de etiquetado de las muestras ya que, según el árbol, hay dos pacientes con dos muestras cada uno de STs muy diferentes, los mismos en ambos casos. Lo más probable es que esta combinación se deba a un error en el procedimiento y que una doble coinfección con los mismos STs en pacientes diferentes no sea la situación real.

En el tercer capítulo, por el contrario, se pretendía realizar un estudio intrapaciente de la evolución de *P. aeruginosa* con el fin de encontrar cambios adaptativos respecto al hospedador. Se han realizado estudios semejantes en los últimos años con muestras de pulmón en pacientes con fibrosis quística, revelando la existencia de fenotipos hipermutadores gracias a cambios en genes reparadores del ADN como *mutL* y *mutS* (Oliver *et al.*, 2000) que generan un elevado número de cambios entre muestras del mismo paciente (Feliziani *et al.*, 2014), así como comprobar el alto grado de compartimentalización existente entre las poblaciones que colonizan el pulmón (Jorth *et al.*, 2015). Dado que es un campo ampliamente estudiado, se prefirió cambiar el tipo de paciente, por lo que se buscaron aquellos con patologías de base distintas a la fibrosis quística y de diferentes zonas del cuerpo.

Previo al análisis individual de cada paciente, se decidió realizar la reconstrucción filogenética de las muestras en su conjunto. La idea de compartimentalización a gran escala (el cuerpo en este caso) y la evolución independiente nos hizo pensar que obtendríamos un árbol en que las muestras de cada paciente se encontrarían agrupadas monofiléticamente, con mayor o menor distancia entre clados dependiendo del tipo de clon que hubiera producido la infección en primer lugar. Sin embargo, en un único clado están contenidas más del 80% de las muestras con relativamente pocas diferencias entre ellas, además de distribuirse en diferentes subclados y sin poder encontrar una relación

entre esta agrupación con la zona del cuerpo infectada o la localización del paciente en el momento de la toma de muestra. Por lo tanto, con la aplicación de esta metodología sumada a la realización del análisis conjunto, que en ocasiones no se realiza en fibrosis quística (Feliziani *et al.*, 2014), encontramos nuevamente varios casos de muestras estrechamente relacionadas a nivel genómico, compatibles con la existencia de múltiples brotes, sin que los hubiésemos previsto ni buscado de forma activa.

Otro aspecto que nos ha demostrado este análisis es que la capacidad de *P. aeruginosa* de vivir en cualquier tipo de medio y encontrarse en nuestro entorno habitual, impide que esta se encuentre de manera exclusiva en una zona del cuerpo. Es decir, que la posibilidad de que se produzcan infecciones con varios clones simultáneamente en una zona mucho más accesible que el pulmón, como es cualquier parte del cuerpo con un acceso abierto, o de que se produzcan reinfecciones con cepas circulantes distintas a las de la infección previa, incluso de tener una superinfección en la que el paciente pueda ser colonizado posteriormente con una cepa distinta estando ya infectado por otra. La complejidad de la situación ante este patógeno oportunista capaz de producir desde infecciones urinarias, sepsis o neumonías, no permite realizar estudios como el que planteábamos en un primer momento.

En los últimos años se han publicado algunos trabajos de este tipo con brotes amplios y genomas completos obtenidos por secuenciación masiva (Snyder *et al.*, 2013), pero la combinación con la construcción de árboles filogenéticos sigue siendo aún poco frecuente en los estudios de infecciones nosocomiales. Por ejemplo, un trabajo reciente interesante sobre posible infección a través del jabón utilizado en el hospital (Blanc *et al.*, 2016) en el que determinan la poca relevancia de esta fuente como foco de infección ante la obtención de más de 25.000 posiciones variantes (SNPs) entre cepas y tras filtrar regiones posibles recombinantes. Otro trabajo que va más allá es un análisis con *S. aureus* meticilina-resistente en el que, a partir del mismo tipo de abordaje, se detectan transmisiones entre la comunidad y los hospitales (Coll *et al.*, 2017), un ejemplo más de cómo pueden establecerse pequeños grupos de transmisión y posibles flujos entre zonas, en este trabajo comunidad-hospital.

En el cuarto capítulo se ha reconstruido la filogenia de las muestras de todos los hospitales conjuntamente. De forma similar al trabajo de Coll *et al.* (2017), se observa cómo hay pacientes de distintos hospitales que presentan aislados muy próximos filogenéticamente. No hay trabajos a gran escala de este tipo, más allá del mencionado,

aunque algún trabajo anterior reconstruye y compara brotes de dos hospitales sin que exista evidencia de transmisión en el árbol (Witney *et al.*, 2014), si bien es verdad su objetivo es comparar diferencias que justificaran cambios en la tasa de mortalidad, no establecer infecciones cruzadas.

El estudio de los aislados de los 3 hospitales a nivel genómico también ha mostrado cómo este abordaje aplicado por separado o a gran escala revela más información que la aportada por la MLST. También hay que tener en cuenta que las cepas hipermutadoras presentan cambios en el gen *mutL*, una enzima reparadora del ADN, que además forma parte del esquema de tipado de *P. aeruginosa*. Estos cambios ya se han descrito anteriormente, provocando discrepancias entre métodos (García-Castillo *et al.*, 2012; López-Causapé *et al.*, 2013), lo que remarca la necesidad de estudiarlo con detalle. Sin embargo, este método, con costes relativamente bajos y sencilla aplicación es una buena primera aproximación para posteriormente, aplicar la metodología desarrollada en el presente trabajo con el objetivo de investigar posibles transmisiones.

Por último, en el quinto capítulo realizamos un estudio de la estructura CRISPR en los 3 conjuntos de muestras. Podría plantearse como una alternativa al uso de NGS la complementación del ST y el perfil de CRISPR. En el estudio del HGUV hemos encontrado que existía una coincidencia entre la estructura de CRISPR de los aislados del clado del brote frente a las muestras de los pacientes 10 y 18 con el mismo ST, que quedaban alejados filogenéticamente, y por ello descartados. La detección de dicha región por PCR mediante cebadores específicos ha sido ya utilizado como método de tipado, el más conocido el *spoligotyping* en *Mycobacterium tuberculosis* (Groenen *et al.*, 1993), pero también en combinación (Shariat *et al.*, 2013).

La variabilidad encontrada en este caso, además, podría explicarse por deleciones posteriores de determinados *spacers* de una región similar o mediante mecanismos de recombinación, dos mecanismos que parecen estar presentes en varias especies (Lopez-Sanchez *et al.*, 2012; Kupczok, Landan y Dagan, 2015). Sin embargo, serían necesarios más estudios con otros conjuntos de muestras de *P. aeruginosa* para comprobar sí la estructura es suficientemente dinámica para poder realizar dicha distinción y comprobar que el resultado que hemos obtenido no sea meramente circunstancial.

A nivel general, para todos los estudios realizados la metodología empleada es determinante a la hora de calcular las distancias entre aislados ya que, en este caso, elegir mapeo o ensamblaje puede cambiar la conformación o longitud de las ramas. *P.*

aeruginosa es un patógeno altamente variable por lo que, si seleccionamos el mapeo frente a una referencia, es posible que perdamos información (lecturas) de regiones no contenidas en dicha referencia, como puede ocurrir con islas genómicas que, además, contienen elementos importantes en la adaptación (Winstanley *et al.*, 2009; Chowdhury, Scott y Djordjevic, 2017). Por otro lado, el uso de ensamblaje puede no ser una opción óptima en el caso de muestras con baja cobertura, como ocurre en algunas de las muestras de este trabajo o las del HGUV en conjunto, ya que produce ensamblados muy fragmentados, sin posibilidad de distinguir claramente regiones que puedan proceder de plásmido por la corta longitud de la secuencia de algunos *contigs*, introduciendo variabilidad que no debe tenerse en cuenta a la hora de realizar la filogenia.

En este trabajo, se ha optado por la utilización de mapeo contra una referencia próxima con el fin de conseguir cubrir el mayor número de posiciones genómicas posibles en todas las muestras con unos filtros que fueron reduciéndose hasta encontrar un balance entre no perder información para evitar descartar muestras dentro de lo posible y a la vez tener confianza en que los SNPs detectados son reales y no variabilidad introducida en la secuenciación.

La variabilidad encontrada también repercute en el estudio por BEAST, con el que hemos realizado la datación de los ancestros comunes de cada brote o conjunto de brotes. En este trabajo esto ha sido determinado con poca señal filogenética en las situaciones más complejas, como las de HGUE y HGUV, por lo que los resultados no son demasiado fiables. Los valores de las tasas se asemejaban entre los sets del HGUV y HGUE, con una tasa de 1,41E-6 y 1,38E-6 s/s/a, respectivamente, y una tasa de 2,97E-7 s/s/a para las muestras del HAV. Estas tasas se acercan a las estimadas por Miyoshi-Akiyama *et al.* (2017) a partir de los genomas ensamblados depositados en la base de datos, variando entre 4.3E10-6 y 1.0E10-5 s/s/a. Varias razones pueden explicar estas discrepancias, en especial para el HAV, como que la baja señal filogenética y temporal aumenta la influencia de los *priors* en las estimas finales, o que el rango temporal del brote influya en las tasas estimadas, estableciéndose una relación directa entre la antigüedad del ancestro común y las tasas estimadas (González-Candelas *et al.* 2015, Dûchene *et al.* 2017).

Otro aspecto relevante en los resultados de los 3 hospitales es que se muestran claramente cambios en los fenotipos de resistencias, así como variaciones en los tipos de determinantes que confieren resistencia, a pesar de la cercanía evolutiva de los aislados.

Esto puede deberse a un problema con la interpretación de los *tests* y al sesgo que puede producirse por el personal que los realiza. En un estudio de Reino Unido en que realizan control fenotípico amplio (Henwood *et al*., 2001), comprueban cómo, en la repetición de los fenotipos, el 49% de las muestras cambiaba de intermedio o resistente a una lectura de susceptibilidad. Por ello, el control derivado de la comparación de fenotipos no es el mejor marcador de brote. También es cierto que los resultados obtenidos en la secuenciación de las muestras del HGUE y HGUV dan lugar a cobertura muy baja en algunas de las muestras, lo que puede interferir en los resultados de detección de resistencia a pesar de la utilización de mapeos. Hay casos en los que el número de lecturas era inferior al millón entre la suma de lecturas de *forward* y *reverse* y, dado el gran tamaño genómico, cabe la posibilidad de que no se hayan detectado correctamente algunos de los genes entre los que se observan discrepancias tras el análisis por ARIBA.

La aplicación clínica de la secuenciación de genomas completos, tal y como aquí se detallan, a *priori* podría parecer inviable por los costes y la especialización necesaria para realizar el análisis de los datos. Sin embargo, con los años se han reducido enormemente los costes de las tecnologías de secuenciación y son fácilmente asumibles si tenemos en cuenta los costes-beneficios. Según datos del Ministerio de Sanidad, Consumo y Bienestar Social actualizados a 29 de agosto de 2018[3], los costes más frecuentes incluyen neumonía e infecciones de vías urinarias con costes medios por paciente de 4.276,8€ y 3.280,1€, respectivamente. Además, en la lista de los procesos con mayor coste figuran las enfermedades infecciosas y parasitarias (incluyendo la infección por VIH) con procedimientos con una media de 18.434€ por paciente.

También se han realizado estudios al respecto en concreto en pacientes con infecciones por *P. aeruginosa* a lo largo de dos años en el Hospital del Mar de Barcelona (Morales *et al*., 2012), en los que se indica que el tiempo de hospitalización es superior en pacientes infectados por cepas resistentes o multirresistentes y se duplica la tasa de mortalidad. Frente a todo ello, los costes de secuenciación como los afrontados en este trabajo no superan los 50€ y la aplicación de herramientas bioinformáticas solamente implica una primera inversión en infraestructura para computación. Por supuesto, no debe tenerse en cuenta solamente el ahorro que supone aplicar esta metodología en la vigilancia epidemiológica, sino el bienestar de los pacientes y la reducción del número

[3] https://www.mscbs.gob.es/estadEstudios/estadisticas/inforRecopilaciones/anaDesarrolloGDR.htm

de casos en riesgo de infección debida a bacterias que, como en este trabajo, presentan multirresistencias, si se toman medidas con rapidez.

Por todo lo descrito, hemos demostrado cómo la genómica en combinación con datos epidemiológicos son útiles para el control de la dispersión de patógenos tan habituales en el medio como *P. aeruginosa*.

6. Conclusiones

- La utilización de epidemiología genómica puede resolver una sospecha de brote con un número de casos limitado.
- Esta herramienta es muy útil para discriminar en un brote de gran extensión aquellos aislados no pertenecientes al mismo y confirmar su dispersión a otras dependencias del hospital. Además, se han encontrado transmisiones ajenas al brote principal.
- No es posible estudiar la variabilidad intrapaciente en zonas infectadas de fácil acceso y, por tanto, con posibilidad de ser colonizadas a la vez o en tiempos distintos por diferentes cepas. Sin embargo, ha sido clave para la detección de un brote amplio cuya existencia era desconocida.
- La utilización combinada de los datos epidemiológicos con la información obtenida tras la reconstrucción de la filogenia permite establecer causas y vías de transmisión, incluso posibles focos de infección.
- El estudio conjunto de aislados de distintos hospitales de una misma área geográfica revela posibles contactos o transmisiones comunitarias.
- El estudio de la estructura CRISPR podría ser útil para su utilización como complemento con el tipado por MLST, ya que dentro de un mismo ST puede haber grandes cambios dada su naturaleza dinámica.

En resumen, el estudio evolutivo a partir de genomas de *P. aeruginosa* de distintos orígenes hospitalarios es una herramienta eficaz para la investigación y control epidemiológico.

7. Material suplementario

Tabla 7.1. Perfiles de resistencia fenotípicos de las muestras del HAV. *Se incluyen algunos de los antibióticos más relevantes testados en el Servicio de Microbiología. De aquellos en los que existen diferencias en la concentración mínima inhibitoria (CMI), con o sin cambio de fenotipo según los parámetros de la National Committee for Clinical Laboratory Standards (NCCLS), se ha incluido en la tabla.*

Aislado	Marcadores de resistencia	AMP/SUL	CEFOTAXIMA	CMI CEFOTAXIMA	AZTREONAM	CMI AZTREONAM	DORIPENEM	CMI DORIPENEM	IMIPENEM	GENTAMICINA	CMI GENTAMICINA	AMIKACINA	CIPROFLOXACINA	CMI CIPROFLOXACINA	TRIMETO/SULFAMETOX	COLISTINA	AMPICILINA
PS03	-																
PS06	MBL	R	R	>16	R	>4	R	>8	R	R	>32	R	R	>4	S	R	R
PS07	MBL	R	R	>16	R	>4	R	>8	R	I	32	R	R	>4	S	R	R
PS08	-																
PS09	-																
PS10	MBL	R	R	>16	R	>4	R	>8	R	I	32	R	R	>4	S	R	
PS11	MBL	R	R	>16	R	>4	R	>8	R	I	32	R	R	>4	S	R	R
PS12	MBL	R	R	>16	R	>4	R	>8	R	R	>32	R	R	>4	S	R	R
PS13	MBL	R	R	>16	R	>4	R	>8	R	R	>32	R	R	>4	S	R	R
PS17	-	R	R	>16	R	>4	R	>8	R	R	>32	R	R	>4	S	R	R
PS20																	
PS21	MBL	R	R	>32	R	>16	R		R	R	>8	I	R	>2	R	S	

Tabla 7.2. Calidad de las lecturas de las muestras del HAV tras la limpieza. *Las muestras están numeradas por paciente precedidas por PS (*Patient Sample*). El sufijo 1 y 2 corresponde con las lecturas en* paired-end *de una misma muestra obtenidas por Illumina MiSeq. El análisis se generó con FASTQC y MULTIQC después de limpiarlas con AUTOADAPT y PRINSEQ. Tipado por MLST con SRST2.*

Muestra	% duplicados	%GC	Longitud media pre-limpieza	Longitud media post-limpieza	Total secuencias pre-limpieza	Total secuencias post-limpieza	ST
PS03_1	13,82	65	279,84	233,00	1.071.481	919.316	175
PS03_2	10,92	65	285,11	171,51	1.071.481	919.316	
PS06_1	13,98	65	278,93	234,16	1.053.784	916.199	175
PS06_2	11,50	65	284,29	173,49	1.053.784	916.199	
PS07_1	14,22	65	277,70	236,12	1.093.739	889.041	175
PS07_2	9,97	65	284,37	162,39	1.093.739	889.041	
PS08_1	14,60	65	276,69	230,27	1.059.103	948.041	175
PS08_2	12,57	65	281,58	179,19	1.059.103	948.041	
PS09_1	15,71	65	274,81	229,91	1.137.389	1.019.202	175
PS09_2	13,55	65	279,77	178,66	1.137.389	1.019.202	
PS10_1	14,82	65	276,35	229,79	1.097.468	976.284	175
PS10_2	12,14	65	281,85	176,15	1.097.468	976.284	
PS11_1	13,94	65	277,52	230,22	992.610	882.998	175
PS11_2	11,95	65	282,37	178,60	992.610	882.998	
PS12_1	13,90	65	284,46	239,25	1.000.095	863.892	175
PS12_2	11,05	65	288,54	174,30	1.000.095	863.892	
PS13_1	15,92	65	280,16	234,81	1.156.369	1.048.075	175
PS13_2	13,95	65	284,48	181,90	1.156.369	1.048.075	
PS17_1	14,49	65	278,37	233,27	1.120.872	982.525	175
PS17_2	11,84	65	283,62	175,50	1.120.872	982.525	
PS20_1	13,37	65	280,10	232,92	988.733	868.383	175
PS20_2	11,02	65	284,80	177,02	988.733	868.383	
PS21_1	15,90	65	279,47	234,07	1.239.933	1.071.621	175
PS21_2	12,64	65	284,72	172,85	1.239.933	1.071.621	

Tabla 7.3. Resultados de BEAST de los aislados del HAV.

	posterior	prior	likelihood	treeModel, rootHeight	exponential, popSize	exponential, growthRate	alpha	clock,rate	meanRate	treeLikelihood	branchRates	coalescent
mean	-8059338,401	-794	-8059337,607	4,2095	7,6109	0,4408	0,3507	2,98E-07	2,98E-07	-8059337,607	0	-21,8342
stderr of mean	0,8942	0,8883	0,0182	0,0296	0,1079	3,14E-03	2,45E-03	9,92E-10	9,92E-10	0,0182	n/a	0,0402
stdev	18,2725	18,1551	2,9481	4,7074	17,4867	0,5164	0,3991	1,59E-07	1,59E-07	2,9481	n/a	6,4845
variance	333,8825	329,6089	8,6915	22,1594	305,7836	0,2667	0,1593	2,53E-14	2,53E-14	8,6915	n/a	42,0485
median	-8059342,174	-4,7361	-8059337,271	3,3896	4,5078	0,3495	0,2132	2,71E-07	2,71E-07	-8059337,271	n/a	-21,3835
mode	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
geometric mean	n/a	n/a	n/a	3,5823	4,7516	n/a	0,1844	2,59E-07	2,59E-07	n/a	n/a	n/a
95% HPD Interval	[-8059369,0335, -8059298,6305]	[-28,7725, 40,1819]	[-8059343,6323, -8059332,4761]	[1,3108, 8,8666]	[0,2878, 22,4998]	[-0,4188, 1,5917]	[1,0039E-3, 1,1474]	[2,9058E-8, 5,9913E-7]	[2,9058E-8, 5,9913E-7]	[-8059343,6323, -8059332,4761]	n/a	[-35,3349, -8,9983]
auto-correlation time (ACT)	6,47E+05	6,46E+05	10331,3902	10699,4932	10284,3715	10000	10198,0636	10492,1304	10492,1304	10331,3902	n/a	10382,4555
effective sample size (ESS)	417,5773	417,7217	26136,8503	25237644	26256,3444	27003	26478,5563	25736432	25736432	26136,8503	n/a	26008,2982

Tabla 7.4. Muestras de* Pseudomonas aeruginosa *del HGUE; primera parte, Área de Urología. *Cada muestra procede de un paciente distinto, en número del caso corresponde también con el número de muestra. El tipo de muestra está dividido entre pacientes (P) y ambientales (A); las ambientales proceden del agua de los aseos. Se incluyen los datos fenotípicos obtenidos en el Servicio de Microbiología (Sensible, S; Resistente, R; Intermedio, I), así como los perfiles de PFGE.*

Casos	Fecha toma	Tipo muestra	PFGE (perfil)	AMIKA	COLISTINA	CEFTAZIDIMA	CEFEPIME	P/T	IMIPENEM	MEROPENEM	GENTA	TOBRA	CIP	LEVO
1	03/07/15	P	10	S	S	I	I	I	R	R	R	R	R	R
2	27/03/15	P	10	S	S	I	I	I	R	R	R	R	R	R
3	16/03/15	P	10	S	S	I	I	I	R	R	R	R	R	R
4	20/04/15	P	10	S	S	I	I	I	R	R	R	R	R	R
5	18/03/15	P	10	S	S	I	I	I	R	R	R	R	R	R
6	30/03/15	P	10	S	S	I	I	I	R	R	R	R	R	R
7	31/03/15	P	10	S	S	I	I	I	R	R	R	R	R	R
8	03/07/15	P	10	S	S	I	I	I	R	R	R	R	R	R
9	11/05/15	P	10	S	S	I	I	I	R	R	R	R	R	R
10	03/07/15	P	10	S	S	I	I	I	R	R	R	R	R	R
11	18/05/15	P	10	S	S	I	I	I	R	R	R	R	R	R
12	18/05/15	P	10	S	S	I	I	I	R	R	R	R	R	R
13	29/05/15	P	10	S	S	I	I	I	R	R	R	R	R	R
14	27/08/15	P	10	S	S	I	I	I	R	R	R	R	R	R

Casos	Fecha toma	Tipo muestra	PFGE (perfil)	AMIKA	COLISTINA	CEFTAZIDIMA	CEFEPIME	P/T	IMIPENEM	MEROPENEM	GENTA	TOBRA	CIP	LEVO
15	06/08/15	P	10	S	S	I	I	I	R	R	R	R	R	R
16	19/08/15	P	10	S	S	I	I	I	R	R	R	R	R	R
17	28/08/15	P	10	S	S	I	I	I	R	R	R	R	R	R
18	22/10/15	P	10	S	S	I	I	I	R	R	R	R	R	R
19	30/10/15	P	10	S	S	I	I	I	R	R	R	R	R	R
20	09/12/15	P	10	S	S	I	I	I	R	R	R	R	R	R
21	03/12/15	P	10	S	S	I	I	I	R	R	R	R	R	R
22	05/10/15	P	10	S	S	I	I	I	R	R	R	R	R	R
23	14/12/15	P	10	S	S	I	I	I	R	R	R	R	R	R
24	22/12/15	P	10	S	S	I	I	I	R	R	R	R	R	R
25	05/01/16	P	10	S	S	I	I	I	R	R	R	R	R	R
26	11/01/16	P	10	S	S	I	I	I	R	R	R	R	R	R
27	21/01/16	P	10	S	S	I	I	I	R	R	R	R	R	R
28	25/01/16	P	10	S	S	I	I	I	R	R	R	R	R	R
29	28/01/16	P	10	S	S	I	I	I	R	R	R	R	R	R
30	28/01/16	P	10	S	S	I	I	I	R	R	R	R	R	R
31	22/12/14	P	1	S	S	R	R	R	R	R	R	R	R	R
32	05/01/15	P	2	S	S	R	I	S	R	R	R	R	R	R
33	30/12/14	P	3	S	S	S	R	R	R	I	R	R	R	R

Casos	Fecha toma	Tipo muestra	PFGE (perfil)	AMIKA	COLISTINA	CEFTAZIDIMA	CEFEPIME	P/T	IMIPENEM	MEROPENEM	GENTA	TOBRA	CIP	LEVO
34	30/12/14	P	4	S	S	I	R	R	R	R	R	R	R	R
35	12/08/15	P	5	S	S	I	I	S	R	R	R	R	R	R
36	14/01/15	P	6	S	S	I	I	S	R	R	R	R	R	R
37	16/01/15	P	7	S	S	R	I	S	R	R	R	R	R	R
38	27/01/15	P	8	S	S	I	S	S	S	S	S	R	R	R
39	16/12/15	P	9	S	S	I	I	S	R	R	R	R	R	R
40	14/12/15	P	11	S	S	S	I	S	R	R	I	S	R	S
41	28/07/15	P	12	S	S	I	I	I	R	R	R	R	R	R
42	04/09/15	P	13	R	S	S	I	I	R	R	R	R	R	R
43	03/11/15	P	14	S	S	I	I	I	R	R	R	R	R	R
44	20/03/15	P	1	S	S	R	R	R	R	R	R	R	R	R
45	14/07/15	P	7	S	S	R	I	I	R	R	R	R	R	R
46	15/12/14	P	10	S	S	I	I	S	R	R	R	R	R	R
47	15/12/14	P	10	S	S	I	I	S	R	R	R	R	R	R
48	11/05/15	P	10	S	S	I	I	I	R	R	R	R	R	R
49	08/06/15	P	10	S	S	I	I	I	R	R	R	R	R	R
50	25/01/16	P	10	S	S	R	R	R	R	R	R	R	R	R
51	01/07/15	P	10	S	S	S	I	I	IR	R	R	R	R	R
52	25/02/16	P	10	S	S	I	I	I	R	R	R	R	R	R

Casos	Fecha toma	Tipo muestra	PFGE (perfil)	AMIKA	COLISTINA	CEFTAZIDIMA	CEFEPIME	P/T	IMIPENEM	MEROPENEM	GENTA	TOBRA	CIP	LEVO
1A	18/05/15	**A**	10	S	S	I	I	I	R	R	R	R	R	R
2A	12/05/15	**A**	10	S	S	I	I	I	R	R	R	R	R	R
3A	19/05/15	**A**	10	S	S	I	I	I	R	R	R	R	R	R
4A	15/01/16	**A**	10	S	S	I	I	I	R	R	R	R	R	R
5A	12/05/15	**A**	10	S	S	I	I	I	R	R	R	R	R	R

Tabla 7.5. Muestras de* Pseudomonas aeruginosa *del HGUE; segunda parte, posible dispersión a otras áreas.

Pacientes	Nuevas muestras	Servicio /CAMA	Tipo muestra	Fecha	PFGE	CARBAPENEMASAS
1	**53**	MEDICINA INTERNA	ORINA	01/06/2016	1	VIM
2	**54**		ORINA	07/06/2016		N
	64	HOSPITALIZACION A DOMICILIO	BRONCOASPIRADO	26/09/2016		N
3	**55**		ORINA	24/06/2016		N
4	**56**	NEFROLOGIA	ORINA	19/07/2016	1	VIM
5	**57**	CIRUGIA VASCULAR	EXUDADO DE HERIDA	14/07/2016	10	N
6	**58**	ANESTESIA Y REANIMACION	BRONCOASPIRADO	24/08/2016		N
7	**59**	ANESTESIA Y REANIMACION	BRONCOASPIRADO	07/09/2016	10	N
	60		BRONCOASPIRADO	12/09/2016	10	N
8	**61**	NEFROLOGIA	ORINA	08/09/2016	1	VIM
	62		ORINA	26/09/2016	1	VIM
	63		ORINA	13/10/2016	1	VIM
10	**65**	MEDICINA CORTA ESTANCIA	ESPUTO	26/09/2016	10	N
11	**66**	NEFROLOGIA	ESPUTO	07/10/2016	10	N
12	**67**	UCI6	BRONCOASPIRADO	27/10/2016	10	N
13	**68**	501B	ORINA	02/11/2016	10	N

Tabla 7.6. Calidad de las lecturas de las muestras del HGUE antes y después de la limpieza. *Las muestras están numeradas por paciente precedidas por PS (*Patient Sample*). El sufijo 1 y 2 corresponde con las lecturas en* paired-end *de una misma muestra obtenidas por Illumina MiSeq. El análisis se generó con FastQC y MultiQC. Tipado por MLST con ARIBA y confirmados posteriormente con MLST las muestras con discrepancias.*

Muestra	% duplicados	%GC	Longitud media pre	Longitud media post	Total secuencias pre	Total secuencias post	KRAKEN	ST
Elche_01_R1	15,67	65	252,88	224,41	868031	821684	*P. aeruginosa*	175
Elche_01_R2	13,46	65	256,84	173,25	868031	821684		
Elche_02_R1	10,	65	253,26	224,7	507338	479187	*P. aeruginosa*	175
Elche_02_R2	8,28	65	257,03	173,24	507338	479187		
Elche_03_R1	12,46	65	252,62	225,75	618584	583279	*P. aeruginosa*	175
Elche_03_R2	10,33	65	256,82	171,86	618584	583279		
Elche_04_R1	5,1	65	260,81	230,38	207661	195330	*P. aeruginosa*	175
Elche_04_R2	4,14	65	263,99	175,31	207661	195330		
Elche_05_R1	11,09	65	268,55	237,12	591061	553642	*P. aeruginosa*	175
Elche_05_R2	9,	65	271,15	176,87	591061	553642		
Elche_06_R1	10,79	65	259,87	229,29	522092	491145	*P. aeruginosa*	175
Elche_06_R2	8,7	65	263,51	173,	522092	491145		
Elche_07_R1	5,54	65	238,87	213,75	129062	122290	*P. aeruginosa*	NF
Elche_07_R2	4,86	65	242,73	171,02	129062	122290		
Elche_08_R1	11,55	65	261,75	232,3	596713	557205	*P. aeruginosa*	175
Elche_08_R2	9,23	65	265,2	173,39	596713	557205		
Elche_09_R1	11,13	65	266,93	236,75	532728	499823	*P. aeruginosa*	175
Elche_09_R2	9,3	65	269,42	176,93	532728	499823		
Elche_10_R1	8,48	65	252,28	224,35	414671	390386	*P. aeruginosa*	175
Elche_10_R2	6,94	65	256,02	171,59	414671	390386		
Elche_11_R1	7,92	65	265,54	233,32	410739	380807	*P. aeruginosa*	175
Elche_11_R2	6,2	65	268,58	173,02	410739	380807		

Muestra	% duplicados	%GC	Longitud media pre	Longitud media post	Total secuencias pre	Total secuencias post	KRAKEN	ST
Elche_12_R1	9,5	65	266,97	235,11	510234	474813	*P. aeruginosa*	175
Elche_12_R2	7,61	65	269,92	174,36	510234	474813		
Elche_13_R1	7,38	65	268,71	234,74	365289	339909	*P. aeruginosa*	175
Elche_13_R2	5,74	65	270,96	177,44	365289	339909		
Elche_14_R1	5,91	65	267,25	234,26	296396	274655	*P. aeruginosa*	175
Elche_14_R2	4,67	65	269,43	176,91	296396	274655		
Elche_15_R1	5,43	65	265,99	233,41	255435	235133	*P. aeruginosa*	175
Elche_15_R2	4,19	65	268,46	174,32	255435	235133		
Elche_16_R1	8,71	65	261,16	229,96	447537	422754	*P. aeruginosa*	175
Elche_16_R2	7,28	65	263,47	178,6	447537	422754		
Elche_17_R1	10,05	65	267,8	235,89	516413	485726	*P. aeruginosa*	175
Elche_17_R2	8,47	65	269,85	180,24	516413	485726		
Elche_18_R1	6,97	65	272,38	238,08	341738	318454	*P. aeruginosa*	175
Elche_18_R2	5,54	65	274,33	177,46	341738	318454		
Elche_19_R1	0,51	64	279,71	-	1952	-	-	-
Elche_19_R2	0	64	284,2	-	1952	-		
Elche_1A_R1	11,15	65	284,07	244,62	613612	556274	*P. aeruginosa*	175
Elche_1A_R2	8,26	66	285,58	170,48	613612	556274		
Elche_20_R1	15,41	61	269,29	-	751326	-	*P. putida*	-
Elche_20_R2	12,75	62	271,64	-	751326	-		
Elche_21_R1	7,54	65	269,83	237,66	372266	347031	*P. aeruginosa*	175
Elche_21_R2	5,97	65	271,63	180,19	372266	347031		
Elche_22_R1	6,88	65	266,1	230,62	312767	293942	*P. aeruginosa*	175
Elche_22_R2	5,54	65	268,18	177,8	312767	293942		
Elche_23_R1	4,62	65	268,48	230,8	223197	207750	*P. aeruginosa*	175
Elche_23_R2	3,55	65	270,4	176,5	223197	207750		
Elche_24_R1	6,41	65	269,98	234,26	323800	300028	*P. aeruginosa*	175
Elche_24_R2	4,93	65	272,15	176,11	323800	300028		
Elche_25_R1	7,59	65	263,9	229,85	397416	370781	*P. aeruginosa*	175
Elche_25_R2	6,09	65	266,03	176,68	397416	370781		

Muestra	% duplicados	%GC	Longitud media pre	Longitud media post	Total secuencias pre	Total secuencias post	KRAKEN	ST
Elche_26_R1	6,22	65	265,17	232,45	308381	288830	*P. aeruginosa*	175
Elche_26_R2	5,09	65	267,32	178,18	308381	288830		
Elche_27_R1	8,94	65	258,92	228,67	445008	418000	*P. aeruginosa*	175
Elche_27_R2	7,38	65	261,7	175,58	445008	418000		
Elche_28_R1	6,01	65	269,72	234,2	287382	266663	*P. aeruginosa*	175
Elche_28_R2	4,74	65	271,43	178,37	287382	266663		
Elche_29_R1	13,77	52	255,21	-	430590	-	*E. coli*	-
Elche_29_R2	12,73	52	256,45	-	430590	-		
Elche_2A_R1	6,2	65	281,86	245,3	298440	271998	*P. aeruginosa*	175
Elche_2A_R2	4,49	65	283,49	173,77	298440	271998		
Elche_30_R1	6,02	65	268,1	233,41	303848	284369	*P. aeruginosa*	175
Elche_30_R2	4,85	65	270,31	177,76	303848	284369		
Elche_31_R1	8,95	61	261,19	-	411509	-	*P. putida*	-
Elche_31_R2	7,72	62	262,82	-	411509	-		
Elche_32_R1	7,24	65	265,66	232,44	381018	349207	*P. aeruginosa*	175
Elche_32_R2	5,42	65	268,27	171,95	381018	349207		
Elche_33_R1	8,68	64	253,82	225,28	483998	459680	*P. aeruginosa*	1212
Elche_33_R2	7,56	64	255,78	179,06	483998	459680		
Elche_34_R1	7,08	65	259,18	226,86	354914	334239	*P. aeruginosa*	1212
Elche_34_R2	5,99	65	261,56	176,31	354914	334239		
Elche_35_R1	8,24	65	266,5	232,09	462637	433818	*P. aeruginosa*	175
Elche_35_R2	6,76	65	269,13	175,53	462637	433818		
Elche_36_R1	6,7	65	261,4	224,53	365693	342506	*P. aeruginosa*	175
Elche_36_R2	5,39	65	264,55	170,39	365693	342506		
Elche_37_R1	10,25	65	256,94	222,87	554660	510551	*P. aeruginosa*	175
Elche_37_R2	8,03	65	261,02	166,3	554660	510551		
Elche_38_R1	9,43	65	260,78	226,79	475264	435417	*P. aeruginosa*	175
Elche_38_R2	7,12	65	264,64	167,1	475264	435417		
Elche_39_R1	8,49	65	251,22	219,44	470941	431775	*P. aeruginosa*	175
Elche_39_R2	6,44	65	255,92	163,4	470941	431775		

Muestra	% duplicados	%GC	Longitud media pre	Longitud media post	Total secuencias pre	Total secuencias post	KRAKEN	ST
Elche_3A_R1	12,32	65	284,35	246,63	653723	595887	*P. aeruginosa*	175
Elche_3A_R2	9,07	66	285,48	174,8	653723	595887		
Elche_40_R1	6,42	64	263,39	227,76	321254	294737	*P. aeruginosa*	357
Elche_40_R2	4,92	65	266,74	168,59	321254	294737		
Elche_41_R1	8,38	65	265,01	229,89	434291	398043	*P. aeruginosa*	175
Elche_41_R2	6,54	65	268,3	169,22	434291	398043		
Elche_42_R1	7,77	65	264,21	227,82	383336	348846	*P. aeruginosa*	235
Elche_42_R2	5,74	65	267,92	165,97	383336	348846		
Elche_43_R1	9,81	65	268,18	230,92	517360	472513	*P. aeruginosa*	175
Elche_43_R2	7,61	65	271,33	168,58	517360	472513		
Elche_44_R1	12,1	62	262,82	-	555510	-	*P. putida*	-
Elche_44_R2	9,34	62	266,33	-	555510	-		
Elche_45_R1	6,53	65	267,79	232,82	333233	304835	*P. aeruginosa*	175
Elche_45_R2	4,87	65	270,74	171,01	333233	304835		
Elche_46_R1	10,59	65	234,37	206,11	684151	634663	*P. aeruginosa*	175
Elche_46_R2	8,66	65	238,86	160,34	684151	634663		
Elche_47_R1	8,31	65	249,93	217,99	511194	473341	*P. aeruginosa*	175
Elche_47_R2	6,62	65	254,38	165,51	511194	473341		
Elche_48_R1	9,82	65	269,57	233,39	531092	483117	*P. aeruginosa*	175
Elche_48_R2	7,44	65	272,94	168,32	531092	483117		
Elche_49_R1	11,86	65	277,89	242,33	683193	620746	*P. aeruginosa*	175
Elche_49_R2	9,03	66	279,93	173,6	683193	620746		
Elche_4A_R1	10,04	65	270,87	238,39	536650	490994	*P. aeruginosa*	175
Elche_4A_R2	7,47	65	273,47	173,4	536650	490994		
Elche_50_R1	11,13	65	264,96	233,1	635349	586684	*P. aeruginosa*	175
Elche_50_R2	8,9	65	267,91	172,14	635349	586684		
Elche_51_R1	16,96	65	263,08	231,7	979624	903965	*P. aeruginosa*	175
Elche_51_R2	13,37	65	266,1	171,28	979624	903965		
Elche_52_R1	5,88	64	258,43	216,61	296181	274992	*P. aeruginosa*	175
Elche_52_R2	4,76	64	262,76	161,55	296181	274992		

Muestra	% duplicados	%GC	Longitud media pre	Longitud media post	Total secuencias pre	Total secuencias post	KRAKEN	ST
Elche_53_R1	20,71	62	284,55	-	1371720	-	*P. putida*	-
Elche_53_R2	15,31	62	287,77	-	1371720	-		
Elche_54_R1	15,7	65	282,21	235,4	985787	873202	*P. aeruginosa*	699
Elche_54_R2	11,81	65	285,8	177,87	985787	873202		
Elche_55_R1	17,79	66	281,68	-	723107	-	*Stenotrophomonas maltophilia*	-
Elche_55_R2	11,22	65	286,34	-	723107	-		
Elche_56_R1	19,55	62	284,74	-	1283562	-	*P. putida*	-
Elche_56_R2	15,93	61	287,59	-	1283562	-		
Elche_57_R1	13,11	65	290,52	239,7	823604	727019	*P. aeruginosa*	175
Elche_57_R2	9,83	65	292,24	181,76	823604	727019		
Elche_58_R1	30,72	42	274,48	-	939876	-	*Providencia stuartii*	-
Elche_58_R2	27,35	42	277,59	-	939876	-		
Elche_59_R1	13,97	65	286,64	235,94	749302	665136	*P. aeruginosa*	175
Elche_59_R2	10,76	65	289,46	181,86	749302	665136		
Elche_5A_R1	15,49	65	281,17	247,06	862120	790614	*P. aeruginosa*	175
Elche_5A_R2	11,73	65	282,56	176,39	862120	790614		
Elche_60_R1	9,57	65	287,4	224,32	558353	493148	*P. aeruginosa*	175
Elche_60_R2	7,48	65	290,4	175,09	558353	493148		
Elche_61_R1	16,36	62	288,86	-	925253	-	*P. putida*	-
Elche_61_R2	12,96	62	290,86	-	925253	-		
Elche_62_R1	9,35	65	288,6	232,46	457226	403549	*P. aeruginosa*	348
Elche_62_R2	7,2	65	291,15	181,85	457226	403549		
Elche_63_R1	13,31	65	280,39	225,53	774195	705495	*P. aeruginosa*	699
Elche_63_R2	11,23	65	283,64	184,16	774195	705495		
Elche_64_R1	11,59	65	283,27	231,76	632744	556209	*P. aeruginosa*	348
Elche_64_R2	8,75	65	286,91	177,25	632744	556209		
Elche_65_R1	21,5	65	263,42	218,02	1919165	1746609	*P. aeruginosa*	175
Elche_65_R2	17,23	65	269,41	175,71	1919165	1746609		
Elche_66_R1	9,72	66	290,18	238,59	502582	432364	*P. aeruginosa*	175
Elche_66_R2	6,68	66	292,1	174,81	502582	432364		

Muestra	% duplicados	%GC	Longitud media pre	Longitud media post	Total secuencias pre	Total secuencias post	KRAKEN	ST
Elche_67_R1	16,93	65	280,5	235,79	1004984	851948	*P. aeruginosa*	175
Elche_67_R2	10,73	65	284,89	166,5	1004984	851948		
Elche_68_R1	13,67	66	289,01	234,15	784134	706793	*P. aeruginosa*	175
Elche_68_R2	10,95	66	290,89	183,49	784134	706793		

Elche_01																																																						
Elche_02	2																																																					
Elche_03	2	2																																																				
Elche_05	2	2	0																																																			
Elche_06	4	4	4	4																																																		
Elche_07	5	5	5	5	3																																																	
Elche_08	2	2	2	2	4	5																																																
Elche_09	96	96	96	96	98	77	96																																															
Elche_10	4	4	4	4	2	3	4	98																																														
Elche_11	3	3	1	1	5	6	3	97	5																																													
Elche_12	1	1	1	1	3	4	1	95	3	2																																												
Elche_13	2	2	2	2	4	5	2	96	4	3	1																																											
Elche_14	1	1	1	1	3	4	1	95	3	2	0	1																																										
Elche_15	3	3	3	3	5	6	3	94	5	4	2	3	2																																									
Elche_16	3	3	3	3	5	6	3	96	5	4	2	3	2	4																																								
Elche_17	1	1	1	1	3	4	1	95	3	2	0	1	0	2	2																																							
Elche_18	4	4	4	4	6	7	4	98	6	5	3	4	3	5	5	3																																						
Elche_1A	1	1	1	1	3	4	1	95	3	2	0	1	0	2	2	0	3																																					
Elche_21	2	2	2	2	4	5	2	96	4	3	1	2	1	3	3	1	4	1																																				
Elche_22	5	5	3	3	7	8	5	99	7	2	4	5	4	6	6	4	7	4	5																																			
Elche_23	4	4	4	4	6	7	4	98	6	5	3	4	3	5	5	3	6	3	4	7																																		
Elche_24	18	18	18	18	20	17	18	109	20	19	17	18	17	17	19	17	20	17	18	21	19																																	
Elche_25	4	4	2	2	6	6	4	98	6	3	3	4	3	5	5	3	6	3	4	5	6	20																																
Elche_26	1	1	1	1	3	4	1	94	3	2	0	1	0	2	2	0	3	0	1	4	3	17	3																															
Elche_27	2	2	2	2	4	5	2	96	4	3	1	2	1	3	3	1	4	1	2	5	4	18	4	1																														
Elche_28	4	4	4	4	6	7	4	97	6	5	3	4	3	5	1	3	6	3	4	7	6	20	6	3	4																													
Elche_2A	3	3	1	1	5	6	3	96	5	2	2	3	2	4	4	2	5	2	3	4	5	19	1	2	3	5																												
Elche_30	8	8	6	6	10	10	8	101	10	5	7	8	7	9	9	7	10	7	8	3	10	24	8	7	8	10	7																											
Elche_32	1	1	1	1	3	4	1	95	3	2	0	1	0	2	2	0	3	0	1	4	3	17	3	0	1	3	2	7																										
Elche_35	3	3	3	3	5	4	3	97	5	4	2	3	2	4	4	2	5	2	3	6	5	19	5	2	3	5	4	9	2																									
Elche_36	7	7	7	7	9	8	7	101	9	8	6	7	6	8	8	6	9	6	7	10	9	23	9	6	7	9	8	12	6	8																								
Elche_37	86	86	86	86	88	70	86	40	88	87	85	86	85	84	86	85	88	85	86	89	88	99	87	85	86	87	86	91	85	87	91																							
Elche_38	92	92	92	92	94	73	92	12	94	93	91	92	91	90	92	91	94	91	92	95	94	105	94	90	92	93	92	97	91	93	97	36																						
Elche_39	1	1	1	1	3	4	1	95	3	2	0	1	0	2	2	0	3	0	1	4	3	17	3	0	1	3	2	7	0	2	6	85	91																					
Elche_3A	3	3	1	1	5	6	3	97	5	2	2	3	2	4	4	2	5	2	3	4	5	19	3	2	3	5	2	7	2	4	8	87	93	2																				
Elche_41	98	98	98	98	100	74	98	45	100	99	97	98	97	96	98	97	100	97	98	101	99	111	100	96	98	99	98	103	97	99	103	42	41	97	99																			
Elche_43	14	14	14	14	16	14	14	106	16	15	13	14	13	13	15	13	16	13	14	17	16	18	16	13	14	16	15	20	13	15	19	96	102	13	15	108																		
Elche_04	3	3	1	1	5	6	3	92	5	2	2	3	2	4	3	2	5	2	3	4	5	19	1	2	3	4	0	7	2	4	8	84	89	2	2	96	15																	
Elche_45	3	3	1	1	5	6	3	97	5	0	2	3	2	4	4	2	5	2	3	2	5	19	3	2	3	5	2	5	2	4	8	87	93	2	2	99	15	2																
Elche_46	1	1	1	1	3	4	1	95	3	2	0	1	0	2	2	0	3	0	1	4	3	17	3	0	1	3	2	7	0	2	6	85	91	0	2	97	13	2	2															
Elche_47	1	1	1	1	3	4	1	95	3	2	0	1	0	2	2	0	3	0	1	4	3	17	3	0	1	3	2	7	0	2	6	85	91	0	2	97	13	2	2	0														
Elche_48	4	4	2	2	6	7	4	98	6	3	3	4	3	5	5	3	6	3	4	5	6	20	2	3	4	6	1	8	3	5	9	88	94	3	3	100	16	1	3	3	3													
Elche_49	5	5	5	5	7	8	5	99	7	6	4	5	4	6	6	4	7	4	5	8	7	21	7	4	5	7	6	11	4	6	2	89	95	4	6	101	17	6	6	4	4	7												
Elche_4A	2	2	2	2	4	5	2	96	4	3	1	2	1	3	3	1	4	1	2	5	4	18	4	1	2	4	3	8	1	3	7	86	92	1	3	98	14	3	3	1	1	4	5											
Elche_50	3	3	3	3	5	6	3	97	5	4	2	3	2	4	4	2	5	2	3	6	5	19	5	2	3	5	4	9	2	4	8	87	93	2	4	99	15	4	4	2	2	5	6	3										
Elche_51	3	3	1	1	5	6	3	97	5	0	2	3	2	4	4	2	5	2	3	2	5	19	3	2	3	5	2	5	2	4	8	87	93	2	2	99	15	2	0	2	2	3	6	3	4									
Elche_52	8	8	6	6	10	11	8	101	10	5	7	8	7	9	9	7	10	7	8	3	10	24	8	7	8	10	7	0	7	9	13	91	97	7	7	103	20	7	5	7	7	8	11	8	9	5								
Elche_5A	1	1	1	1	3	4	1	95	3	2	0	1	0	2	2	0	3	0	1	4	3	17	3	0	1	3	2	7	0	2	6	85	91	0	2	97	13	2	2	0	0	3	4	1	2	2	7							
Elche57	769	769	769	769	771	749	769	861	769	770	768	769	768	767	770	768	771	768	769	772	771	735	771	768	769	771	770	774	768	770	774	850	857	768	770	841	777	765	770	768	768	771	771	769	770	770	773	768						
Elche59	89	89	89	89	91	68	89	37	91	90	88	89	88	87	89	88	91	88	89	92	91	102	90	88	89	90	89	93	88	90	92	21	33	88	90	40	99	87	90	88	88	91	92	89	90	90	94	88	854					
Elche60	88	88	88	88	90	67	88	36	90	89	87	88	87	86	88	87	90	87	88	91	90	101	89	87	88	89	88	92	87	89	91	20	32	87	89	39	98	86	89	87	87	90	91	88	89	89	93	87	853	1				
Elche65	33	33	33	33	35	29	33	125	34	34	32	33	32	32	34	32	35	32	33	36	35	45	35	32	33	35	34	39	32	34	38	115	121	32	34	123	41	33	34	32	32	35	36	33	34	34	39	32	367	118	117			
Elche66	9	9	9	9	7	7	9	103	7	10	8	9	8	10	10	8	11	8	9	12	11	25	11	8	9	11	10	15	8	10	14	93	99	8	10	105	21	10	10	8	8	11	12	9	10	10	15	8	776	96	95	40		
Elche67	429	428	429	429	431	375	429	523	431	430	428	429	427	428	430	428	431	428	429	432	427	445	431	427	428	430	430	435	428	430	432	505	519	428	430	525	441	424	430	428	428	431	432	428	430	430	428	428	1196	516	515	460	436	
Elche68	30	30	30	30	32	25	30	122	31	31	29	30	29	28	31	29	32	29	30	33	31	16	32	29	30	32	31	35	29	31	34	112	118	29	31	123	30	31	31	29	29	32	33	30	31	31	36	29	445	114	113	39	37	455

Tabla 7.7. Estima de la divergencia entre las secuencias del subclado del ST175 del árbol de muestras de HGUE. *Se basa en el número de diferencias entre pares de secuencias en las 55 muestras que se encuentran en este clado, eliminando para el cálculo aquellas posiciones con menos del 95% de cobertura en el alineamiento. El total de posiciones es de 1437 de las 2063 pb totales.*

Tabla 7.8. Comparativa de los resultados obtenidos por BEAST en los tests realizados en el subclado 2 dentro del clado del brote del HGUE. *Éste está compuesto por 9 muestras, la señal en TempEst con el mejor árbol era de $R^2 = 0{,}293$. En todos los casos se lanzó con 60.000.000 de cadenas cada prueba y se utilizó como* prior *la tasa de evolución en distribución exponencial, tal y como se detalla en el apartado de Métodos.*

RUN	CLOCK	POPULATION MODEL	MARGINAL LOGL (SS)	MARGINAL LOGL (PS)	CLOCK.RATE (MEAN)	CLOCK.RATE (95%HPD)	TREELIKELIHOOD (95%HPD)
1	strict	constant	-8064056.018	-8064056.018	2.009E-7	[3.8226E-12, 5.9662E-7]	[-8064006.3034, -8063994.7336]
1 – PRIOR	strict	constant	-	-	9.496E-7	[2.0064E-9, 3.0292E-6]	[-0, 5.1852E-12]
2	strict	exponential growth	-8056067.380	-8063215.069	4.1E-7	[1.4271E-8, 9.6225E-7]	[-8064007.7837, -8063995.8134]
2 – 2	strict	exponential growth	-	-	4.105E-7	[1.1675E-8, 9.6316E-7]	[-8064007.411, -8063995.4496]
2 – 3	strict	exponential growth	-	-	4.194E-7	[1.3134E-8, 9.6418E-7]	[-8064007.6074, -8063995.6521]
2 – PRIOR	strict	exponential growth	-	-	2.392E-6	[1.4706E-7, 5.3569E-6]	
3	strict	bayesian skyline 3	6083554.148	3061719.709	2.416E-7	[3.2778E-11, 7.3248E-7]	[-8064006.8977, -8063995.0174]
3 – 2	strict	bayesian skyline 3	-8064055.789	-8064057.075	2.38E-7	[1.9077E-11, 7.4232E-7]	[-8064006.6737, -8063994.6313]
3 – PRIOR	strict	bayesian skyline 3	-	-	2.137E-5	[1.2519E-5, 3.3022E-5]	[674.6339, 695.371]
4	uncorrelated	constant	-	-			
5	uncorrelated	exponential growth	-8064055.883	-8064057.959	9.889E-7	[3.9921E-10, 3.769E-6]	[-8063999.9753, -8063986.5931]
6	uncorrelated	bayesian skyline 3	-	-	E-40		
7	random	constant	-8064052.701	-8064054.490	4.963E-7	[1.2656E-10, 1.7597E-6]	[-8064002.8697, -8063989.0031]
8	random	exponential growth	-7539088.785	-7658552.257	6.643E-7	[2.9577E-9, 1.6662E-6]	[-8064003.4194, -8063989.2692]
8 – 2	random	exponential growth	-8026309.362	-8026309.362	6.925E-7	[2.3385E-8, 1.8464E-6]	[-8064003.4049, -8063989.117]
8 – PRIOR	random	exponential growth	-	-	2.266E-6	[5.2394E-9, 5.1408E-6]	[-0, 763.9349]
9	random	bayesian skyline 3	-8064052.830	-8064054.002	5.144E-7	[1.9853E-10, 1.6349E-6]	[-8064003.3383, -8063989.0723]

Tabla 7.9. Estadísticas del ensamblado con SPAdes de las muestras de HGUE.

Assembly	contigs (>= 1000 bp)	contigs (>= 10000 bp)	contigs (>= 50000 bp)	Total length (>= 50000 bp)	contigs	Largest contig	Total length	GC (%)	N50	L50
Elche_01	92	64	45	6223340	92	350296	6919669	66,11	137682	15
Elche_02	132	92	49	5612433	132	350322	6988809	66,08	102665	21
Elche_03	112	74	46	6012250	112	700182	6983039	66,08	128354	16
Elche_04	668	173	15	1312885	668	226544	6704267	66,1	18109	88
Elche_05	100	68	43	6070475	100	469054	6976820	66,09	144265	15
Elche_06	113	75	44	5904240	113	440008	6918952	66,1	125420	16
Elche_07	1870	49	1	70077	1870	70077	5383700	66,01	3488	405
Elche_08	110	72	41	6010887	110	628840	7076526	65,98	133802	14
Elche_09	122	76	45	5938275	122	350270	7017236	66,04	135639	17
Elche_10	164	107	49	5194053	164	496470	6917196	66,04	90229	23
Elche_11	151	98	42	5020487	151	453234	6972394	66,09	95701	21
Elche_12	112	76	43	5752671	112	450629	6911790	66,1	126129	17
Elche_13	162	103	47	5204823	162	321051	6952039	66,05	88498	22
Elche_14	277	164	33	3094234	277	330880	6943261	66,04	46124	41
Elche_15	428	192	25	2065167	428	198968	6781751	66,1	30451	60
Elche_16	140	96	45	5242400	140	290686	6917862	66,1	100041	21
Elche_17	129	89	50	5616256	129	330929	6976459	66,09	106112	22
Elche_18	187	130	43	4452865	187	352124	6915277	66,1	70861	27
Elche_1A	115	79	45	5782295	115	439661	6922009	66,1	122150	18
Elche_21	157	103	48	5195666	157	350234	6975162	66,08	92897	23
Elche_22	235	141	42	3988151	235	270770	6958866	66,05	57429	33
Elche_23	487	195	26	2057475	487	219089	6872198	66,05	29890	63
Elche_24	225	123	37	4159403	225	440015	6871563	66,08	68236	26
Elche_25	147	104	45	5049418	147	449687	6906400	66,09	90886	22
Elche_26	244	142	39	3607983	244	330902	6866402	66,07	52341	36
Elche_27	160	106	40	4704328	160	442379	6852392	66,12	93376	22

Assembly	contigs (>= 1000 bp)	contigs (>= 10000 bp)	contigs (>= 50000 bp)	Total length (>= 50000 bp)	contigs	Largest contig	Total length	GC (%)	N50	L50
Elche_28	321	180	26	2461403	321	250344	6826072	66,11	40485	47
Elche_2A	217	140	44	4168176	217	524853	6890778	66,08	59678	31
Elche_30	295	176	29	2686689	295	208790	6870857	66,1	40899	46
Elche_32	152	106	44	4958284	152	453459	6841490	66,12	92956	23
Elche_33	137	96	48	4981139	137	372406	6431718	66,37	84437	23
Elche_34	162	104	41	4543492	162	325972	6372271	66,38	92113	21
Elche_35	121	80	38	5677084	121	700116	7006068	66,04	132735	13
Elche_36	201	123	43	4571738	201	450585	6890420	66,09	67341	25
Elche_37	113	77	43	5719627	113	533728	6805168	66,05	132791	16
Elche_38	135	90	46	5357693	135	440093	6840911	66,14	112476	19
Elche_39	149	100	48	5299162	149	453099	6928415	66,07	90250	21
Elche_3A	100	65	40	6084758	100	496422	7013376	66,04	162735	14
Elche_40	342	168	39	3271800	342	191823	7065291	65,75	45717	45
Elche_41	148	88	45	5383611	148	365230	6863841	66,11	105653	20
Elche_42	172	106	45	4733781	172	235159	6670088	66,18	81955	24
Elche_43	122	79	46	5783705	122	469175	6935368	66,11	115664	17
Elche_45	211	134	48	4126352	211	183398	6940561	66,05	62849	37
Elche_46	111	74	45	5828002	111	496059	6867641	66,11	126402	16
Elche_47	116	82	46	5786858	116	456949	6946839	66,07	128235	17
Elche_48	111	69	42	6001962	111	453348	6983285	66,08	137682	15
Elche_49	87	59	39	6286740	87	599820	7030989	66,01	177981	12
Elche_4A	112	74	40	5836899	112	484285	6973847	66,09	138417	15
Elche_50	100	66	40	6092028	100	599813	6954517	66,05	155318	12
Elche_51	82	51	36	6463058	82	453690	7002360	66,05	205367	12
Elche_52	491	222	19	1352486	491	123293	6812150	66,08	24214	81
Elche_54	98	66	43	5678171	98	371227	6490355	66,17	148573	15
Elche_57	92	65	44	6177403	92	350250	6949395	66,08	149036	15
Elche_59	101	67	42	5972970	101	540357	6872994	66,1	134023	14
Elche_5A	84	54	36	6301612	84	628984	6986726	66,08	177978	11

Assembly	contigs (>= 1000 bp)	contigs (>= 10000 bp)	contigs (>= 50000 bp)	Total length (>= 50000 bp)	contigs	Largest contig	Total length	GC (%)	N50	L50
Elche_60	122	79	42	5547107	122	454464	6854924	66,13	121628	16
Elche_62	153	93	41	4609337	153	420642	6483044	65,98	93270	22
Elche_63	95	61	42	5780311	95	530384	6488500	66,17	155307	13
Elche_64	150	87	46	5201324	150	399094	6681759	65,98	99642	20
Elche_65	128	58	43	6351780	128	599826	7016503	66,02	138815	13
Elche_66	142	100	47	5307868	142	400667	6924618	66,09	95273	22
Elche_67	99	67	38	5922905	99	599856	6983049	66,08	134059	14
Elche_68	134	66	42	5922444	134	452899	6882527	66,05	147144	15

Tabla 7.10. Genes presentes únicamente en Elche_07, Elche_24 o Elche_68 según los resultados de Roary*. Se eliminaron las que estaban definidas únicamente como "hypothetical protein".*

Gene	Annotation	Elche_07	Elche_24	Elche_68
group_111	DNA helicase			BGNANKAE_06576
group_119	putative ATP-dependent helicase	NHMFGNME_05303		
group_120	putative ATP-dependent helicase	NHMFGNME_03044		
group_1339	type IV B pilus protein			BGNANKAE_06506
oprD	basic amino acid, basic peptide and imipenem outer membrane porin	NHMFGNME_03343		
group_201	conjugal transfer protein			BGNANKAE_02507
group_205	type IV B pilus protein			BGNANKAE_06571
group_21	putative permease of ABC transporter		AFAPEHBO_05938	
group_260	ribose transport protein RbsA	NHMFGNME_04002		
group_3	Ig-like domain repeat protein	NHMFGNME_02901		
group_313	Putative DNA helicase	NHMFGNME_04556		
group_335	Tfp pilus assembly protein ATPase PilU			BGNANKAE_06567
group_344	Putative DNA helicase			BGNANKAE_06310
group_348	chromosome partitioning related protein		AFAPEHBO_06595	
group_367	putative glycosyl hydrolase	NHMFGNME_04982		
group_40	cell division protein FtsY	NHMFGNME_05138		
group_441	Putative 3-oxoacyl-(acyl-carrier-protein) synthase	NHMFGNME_03331		
mexI_1	putative Resistance-Nodulation-Cell Division (RND) efflux transporter	NHMFGNME_00693		
group_47	putative two-component sensor	NHMFGNME_05534		
group_470	putative iron-containing alcohol dehydrogenase	NHMFGNME_04869		
group_4760	DnaK protein	NHMFGNME_01302		
group_4779	putative ABC transporter	NHMFGNME_01562		
group_4786	putative sugar MFS transporter	NHMFGNME_01666		
group_4808	putative protein	NHMFGNME_01869		

Gene	Annotation	Elche_07	Elche_24	Elche_68
group_481	exodeoxyribonuclease V subunit gamma	NHMFGNME_03745		
lis_1	lipoate synthase	NHMFGNME_02668		
recN_2	DNA repair protein	NHMFGNME_02842		
group_4904	NAD-specific glutamate dehydrogenase	NHMFGNME_02844		
group_4938	NAD-specific glutamate dehydrogenase	NHMFGNME_03062		
group_4942	ATP-phosphoribosyltransferase	NHMFGNME_03100		
group_4958	carboxylate--amine ligase	NHMFGNME_03201		
group_4989	dihydroorotase	NHMFGNME_03440		
group_501	helicase	NHMFGNME_03071		
group_5092	twitching motility protein PilI	NHMFGNME_04131		
group_5105	phosphoenolpyruvate synthase	NHMFGNME_04222		
oprI	Outer membrane lipoprotein OprI precursor	NHMFGNME_04593		
group_5244	putative outer membrane protein	NHMFGNME_04934		
group_527	putative oxidoreductase	NHMFGNME_02093		
cupB6_2	fimbrial subunit CupB6	NHMFGNME_05189		
moeA2	molybdenum cofactor biosynthesis protein A2	NHMFGNME_05327		
group_5365	gef-like domain protein	NHMFGNME_05452		
group_5389	hmga2e, putative	NHMFGNME_05565		
group_551	penicillin acylase	NHMFGNME_04739		
qacE_1	QacEDelta1		AFAPEHBO_02137	
group_5924	membrane protein		AFAPEHBO_05790	
group_5927	putative plasmid stabilization protein		AFAPEHBO_05793	
group_5928	tyrosine recombinase XerC		AFAPEHBO_05795	
group_5934	transposase		AFAPEHBO_05954	
group_5935	transposase		AFAPEHBO_05955	
group_5937	osmotically inducible protein C		AFAPEHBO_05957	
group_5938	sugar dehydrogenase		AFAPEHBO_05958	
group_5939	transcriptional regulator protein		AFAPEHBO_05959	
group_5940	cupin		AFAPEHBO_05960	

Gene	Annotation	Elche_07	Elche_24	Elche_68
group_5943	mosc domain protein		AFAPEHBO_06260	
group_5944	transcriptional regulator		AFAPEHBO_06469	
group_5945	Ssb		AFAPEHBO_06470	
group_5946	Hypothetical protein in PFGI-1-like cluster		AFAPEHBO_06471	
group_5947	transcriptional regulator		AFAPEHBO_06472	
group_604	relaxase		AFAPEHBO_05794	
group_618	PilV			BGNANKAE_06584
group_6246	membrane protein			BGNANKAE_02509
group_6248	relaxase			BGNANKAE_02511
group_6250	acetyltransferase			BGNANKAE_03884
group_6251	unknown			BGNANKAE_05221
group_6253	DNA-binding protein			BGNANKAE_05223
group_6254	Bbp36			BGNANKAE_05224
group_6255	transcriptional regulator, LuxR family			BGNANKAE_05228
lexA_2	LexA repressor			BGNANKAE_05229
group_6261	phage-like protein			BGNANKAE_05238
group_6262	helix-turn-helix domain-containing protein			BGNANKAE_05239
group_6263	DNA replication protein DnaC			BGNANKAE_05240
group_6264	helicase DnaB			BGNANKAE_05241
group_6266	Putative metallophosphoesterase			BGNANKAE_05243
group_6268	phage-like protein			BGNANKAE_05245
group_6269	transcriptional regulator			BGNANKAE_05246
group_6270	Putative uncharacterized protein			BGNANKAE_05247
group_6273	nucleoid-associated protein NdpA			BGNANKAE_06301
group_6279	transcriptional regulator			BGNANKAE_06367
group_6281	transposase			BGNANKAE_06369
group_6282	protein containing transposase DDE domain protein			BGNANKAE_06370
group_6284	putative antirepressor			BGNANKAE_06380
group_6289	cp44 protein			BGNANKAE_06499

Gene	Annotation	Elche_07	Elche_24	Elche_68
group_6294	dTDP-D-glucose 4,6-dehydratase			BGNANKAE_06562
amiA_2	N-acetylmuramoyl-L-alanine amidase	NHMFGNME_00489		
group_76	Integrase regulator R			BGNANKAE_06379
group_864	Reverse transcriptase	NHMFGNME_01241		
group_8723	tRNA-Leu(cag)	NHMFGNME_02505		
group_8740	tRNA-Ala(ggc)	NHMFGNME_05292		
group_8928	cold-shock protein		AFAPEHBO_06466	
group_8942	Ssb			BGNANKAE_06381
pltI_1	ATP-binding protein	NHMFGNME_01544		

Tabla 7.11. Muestras de* Pseudomonas aeruginosa *del HGUV consideradas en el estudio. *No se incluyen las muestras descartadas por problemas en el cultivo o tras la secuenciación, por ese motivo en muchos casos los números de las muestras no son consecutivos.*

	Muestra	*Fecha*	*Tipo muestra*	*Procedencia*	*Servicio*
Paciente 1	P1M2	25.09.2013	Orina micción espontánea	Ca - Convento Jerusalén: Medicina Familiar	Medicina Familiar: General
	P1M3	16.01.2014	Orina de sonda permanente	Urgencias: General	Medicina Urgencia: General
Paciente 2	P2M1	13.10.2013	Biopsia	Tocogine: Hg Hosp General	Tocoginecología: Planificación Familiar
	P2M2	21.10.2013	Orina de sonda permanente	Tocogine: Hg Hosp General	Tocoginecología: Planificación Familiar
	P2M3	22.10.2013	Exudado faríngeo	Tocogine: Hg Hosp General	Tocoginecología: Planificación Familiar
Paciente 4	P4M2	26.11.2013	Exudado axilar-rectal	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P4M3	29.11.2013	Bas	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P4M4	04.01.2014	Exudado axilar-rectal	Neurología: Hg Hosp General	Neurología: Neurofisiología
	P4M5	20.01.2014	Esputo	Neurología: Hg Hosp General	Neurología: Neurofisiología
Paciente 5	P5M1	01.12.2012	Esputo	Neumología: Hg Hosp General	Neumología: General
	P5M2	27.05.2014	Esputo	Ap - Paiporta: Medicina Familia	Medicina Familiar: General
Paciente 6	P6M1	24.09.2013	Exudado Herida superficial	Unidad Hospitalización Domicilio: Hospitalización	Unidad Hospitalización Domicilio: General
	P6M2	02.10.2013	Exudado ulcera	Unidad Hospitalización Domicilio: Hospitalización	Unidad Hospitalización Domicilio: General
	P6M3	21.10.2013	Exudado Herida superficial	Unidad Hospitalización Domicilio: Hospitalización	Unidad Hospitalización Domicilio: General
	P6M4	24.12.2013	Exudado Herida superficial	Unidad Hospitalización Domicilio: Hospitalización	Unidad Hospitalización Domicilio: General
	P6M5	14.04.2014	Exudado Herida superficial	Unidad Hospitalización Domicilio: Hospitalización	Unidad Hospitalización Domicilio: General
	P6M6	15.09.2014	Exudado Herida superficial		
Paciente 7	P7M1	29.12.2013	Exudado axilar-rectal	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P7M2	08.01.2014	Exudado axilar-rectal	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica

	Muestra	*Fecha*	*Tipo muestra*	*Procedencia*	*Servicio*
	P7M3	11.02.2014	Exudado axilar-rectal	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P7M4	28.03.2014	Biopsia	Cir.Gral-Digest: Hg Hosp General	Cirugía General Y Digestiva: Vesícula-Biliares
	P7M5	15.04.2014	Exudado axilar-rectal	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P7M7	29.04.2014	Exudado faríngeo	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P7M8	06.05.2014	Bas	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P7M10	03.06.2014	Exudado axilar-rectal	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P7M11	13.06.2014	Sangre por Catéter	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P7M12	17.06.2014	Catéter central	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P7M13	24.06.2014	Exudado faríngeo		
	P7M14	08.07.2014	Exudado faríngeo		
	P7M15	22.07.2014	Exudado faríngeo		
	P7M16	23.07.2014	Bas		
	P7M17	29.07.2014	Exudado faríngeo		
	P7M18	06.08.2014	Exudado faríngeo		
	P7M19	12.08.2014	Exudado faríngeo		
Paciente 9	P9M1	23.12.2012	Orina de sonda permanente	Urgencias: General	Medicina Urgencia: General
	P9M2	20.10.2013	Orina Micción espontanea	Otra Especialidad: Hg Hosp Desconocida	Otra Especialidad: General
	P9M3	23.10.2013	Exudado faríngeo	Otra Especialidad: Hg Hosp Desconocida	Otra Especialidad: General
	P9M4	14.11.2013	Exudado axilar-rectal	Otra Especialidad: Hg Hosp Desconocida	Otra Especialidad: General
	P9M5	23.11.2013	Exudado ulcera	Otra Especialidad: Hg Hosp Desconocida	Otra Especialidad: General
	P9M6	07.12.2013	Orina de sondaje	Otra Especialidad: Hg Hosp Desconocida	Otra Especialidad: General
Paciente 10	P10M1	11.12.2012	Exudado ulcera	Varios: No Consta	Otra Especialidad: General
	P10M2	15.01.2013	Exudado ulcera	Varios: No Consta	Otra Especialidad: General
	P10M3	07.06.2013	Exudado Herida superficial	Uhd: General	Unidad Hospitalización Domicilio: General

	Muestra	*Fecha*	*Tipo muestra*	*Procedencia*	*Servicio*
	P10M4	25.07.2013	Exudado Herida superficial	Unidad Hospitalización Domicilio: Hospitalización	Unidad Hospitalización Domicilio: General
	P10M6	20.12.2013	Exudado Herida superficial	Unidad Hospitalización Domicilio: Hospitalización	Unidad Hospitalización Domicilio: General
Paciente 11	P11M2	09.01.2013	Orina de sondaje	Unidad Hospitalización Domicilio: Hospitalización	Unidad Hospitalización Domicilio: General
	P11M5	11.11.2013	Orina de sonda permanente	Unidad Hospitalización Domicilio: Hospitalización	Unidad Hospitalización Domicilio: General
	P11M6	14.04.2014	Orina de sondaje	Unidad Hospitalización Domicilio: Hospitalización	Unidad Hospitalización Domicilio: General
	P11M7	29.05.2014	Exudado ulcera	Varios: No Consta	Otra Especialidad: General
Paciente 12	P12M1	15.01.2013	Exudado axilar-rectal	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P12M2	08.02.2013	Sangre por Catéter	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P12M3	08.02.2013	Sangre venopunción	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P12M4	08.02.2013	Bas	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P12M5	05.03.2013	Orina de sondaje	Neumología: Hg Hosp General	Neumología: General
	P12M6	28.05.2013	Exudado faríngeo	Neumología: Hg Hosp General	Neumología: General
	P12M7	05.06.2013	Exudado faríngeo	Neumología: Hg Hosp General	Neumología: General
	P12M8	05.06.2013	Exudado ulcera	Neumología: Hg Hosp General	Neumología: General
Paciente 13	P13M2	01.02.2013	Exudado herida profunda	Cir.Gral-Digest: Hg Hosp General	Cirugía General Y Digestiva: Vesícula-Biliares
	P13M3	13.02.2013	Exudado faríngeo	Cir.Gral-Digest: Hg Hosp General	Cirugía General Y Digestiva: Vesícula-Biliares
	P13M4	16.03.2013	Exudado faríngeo	Cir.Gral-Digest: Hg Hosp General	Cirugía General Y Digestiva: Vesícula-Biliares
	P13M5	17.04.2013	Orina Micción espontanea	Cir.Gral-Digest: Hg Hosp General	Cirugía General Y Digestiva: Vesícula-Biliares
	P13M6	23.04.2013	Esputo	Cir.Gral-Digest: Hg Hosp General	Cirugía General Y Digestiva: Vesícula-Biliares
Paciente 14	P14M2	09.12.2013	Exudado faríngeo	Otra Especialidad: Hg Hosp Desconocida	Otra Especialidad: General
	P14M4	15.01.2014	Exudado faríngeo	Otra Especialidad: Hg Hosp Desconocida	Otra Especialidad: General
	P14M5	25.01.2014	Exudado faríngeo	Otra Especialidad: Hg Hosp Desconocida	Otra Especialidad: General
	P14M6	04.04.2014	Exudado cutáneo	Otra Especialidad: Hg Hosp Desconocida	Otra Especialidad: General

	Muestra	*Fecha*	*Tipo muestra*	*Procedencia*	*Servicio*
	P14M7	15.04.2014	Exudado faríngeo	Otra Especialidad: Hg Hosp Desconocida	Otra Especialidad: General
	P14M9	09.05.2014	Exudado faríngeo	Otra Especialidad: Hg Hosp Desconocida	Otra Especialidad: General
	P14M10	09.05.2014	Exudado axilar-rectal	Otra Especialidad: Hg Hosp Desconocida	Otra Especialidad: General
	P14M11	04.06.2014	Exudado Herida superficial	Unidad Hospitalización Domicilio: Hospitalización	Unidad Hospitalización Domicilio: General
Paciente 15	P15M1	10.07.2013	Exudado ulcera	Ap - Torrent 1: Medicina Familia	Medicina Familiar: General
	P15M2	23.07.2013	Esputo	Med.Interna: Hg Hosp General	Medicina Interna: General
	P15M3	28.07.2013	Exudado faríngeo	Med.Interna: Hg Hosp General	Medicina Interna: General
Paciente 16	P16M1	18.06.2013	Exudado herida profunda	Med.Interna: Hg Hosp General	Medicina Interna: General
	P16M3	02.07.2013	Exudado axilar-rectal	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P16M4	09.07.2013	Exudado quirúrgico profundo	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P16M5	16.07.2013	Exudado quirúrgico superficial	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
Paciente 17	P17M1	13.04.2013	Exudado axilar-rectal	Anestesia-Reanimación: Hg Hosp G Med.Intensiva	Anestesia-Reanimación: Reanimación Postquirúrgica
	P17M2	22.04.2013	Líquido peritoneal	Cir.Gral-Digest: Hg Hosp General	Cirugía General Y Digestiva: Vesícula-Biliares
	P17M3	03.05.2013	Exudado axilar-rectal	Cir.Gral-Digest: Hg Hosp General	Cirugía General Y Digestiva: Vesícula-Biliares
	P17M4	07.05.2013	Exudado quirúrgico superficial	Cir.Gral-Digest: Hg Hosp General	Cirugía General Y Digestiva: Vesícula-Biliares
	P17M5	03.06.2013	Sangre venopunción	Urgencias: General	Medicina Urgencia: General
	P17M6	02.07.2013	Orina de sondaje	Urgencias: General	Medicina Urgencia: General
	P17M7	21.05.2014	Orina Micción espontanea	Ap - Millares: Medicina Familia	Medicina Familiar: General
	P17M8	31.05.2014	Orina Micción espontanea	Urgencias: General	Medicina Urgencia: General
Paciente 18	P18M1	15.04.2013	Exudado ótico	Anestesia-Reanimación: Hg Hosp Rean.Cardiaca	Anestesia-Reanimación: Reanimación Postquirúrgica
	P18M2	15.04.2013	Sangre por Catéter	Anestesia-Reanimación: Hg Hosp Rean.Cardiaca	Anestesia-Reanimación: Reanimación Postquirúrgica
	P18M3	15.04.2013	Sangre venopunción	Anestesia-Reanimación: Hg Hosp Rean.Cardiaca	Anestesia-Reanimación: Reanimación Postquirúrgica
	P18M6	29.04.2013	Orina de sondaje	Infecciosos: Hg Hosp General	Unidad Enfermedades Infecciosas: General
	P18M7	07.05.2013	Exudado faríngeo	Infecciosos: Hg Hosp General	Unidad Enfermedades Infecciosas: General

	Muestra	*Fecha*	*Tipo muestra*	*Procedencia*	*Servicio*
	P18M8	18.05.2013	Exudado faríngeo	Infecciosos: Hg Hosp General	Unidad Enfermedades Infecciosas: General
	P18M9	28.05.2013	Exudado faríngeo	Infecciosos: Hg Hosp General	Unidad Enfermedades Infecciosas: General

Tabla 7.12. Calidad de las lecturas de las muestras del HGUV antes y después de la limpieza. *El sufijo 1 y 2 corresponde con las lecturas en* paired-end *de una misma muestra obtenidas por Illumina MiSeq. El análisis se generó con FastQC y MultiQC. Tipado de MLST por SRST2; NF: muestras que no han podido tiparse por ausencia de secuencias suficientes para algún alelo; * aquellas muestras con incertidumbre por presencia de* gaps *o posibles cambios puntuales en algún alelo.*

Muestra	% duplicados	% GC	Longitud media pre-limpieza	Longitud media post-limpieza	Total secuencias pre-limpieza	Total secuencias post-limpieza	ST
P1M2_S1_1	50,75	65	169,44	168,22	270.952	231.560	NF
P1M2_S1_2	50,68	64	176,46	134,48	270.952	231.560	
P1M3_S2_1	50,68	64	191,65	186,84	180.898	156.720	NF
P1M3_S2_2	50,6	64	198,52	141,54	180.898	156.720	
P2M1_S3_1	51,47	65	176,69	174,42	313.246	267.982	244*
P2M1_S3_2	51,26	65	184,36	135,62	313.246	267.982	
P2M2_S4_1	50,66	65	190,06	185,46	180.052	154.506	NF
P2M2_S4_2	50,53	65	196,21	140,71	180.052	154.506	
P2M3_S5_1	52,36	65	157,85	158,9	817.168	691.846	244
P2M3_S5_2	52,09	65	163,63	131,57	817.168	691.846	
P4M2_S6_1	51,96	64	167,83	165,29	699.740	600.444	244
P4M2_S6_2	51,74	64	174,46	132,54	699.740	600.444	
P4M3_S7_1	52,08	64	174,03	170,96	649.068	559.336	650*
P4M3_S7_2	51,82	64	179,84	135,65	649.068	559.336	
P4M4_S93_1	53,98	65	203,28	193,34	1.054.286	984.104	244
P4M4_S93_2	53,57	65	209,57	151,56	1.054.286	984.104	
P4M5_S8_1	52,69	64	185,98	181,19	830.048	719.250	244
P4M5_S8_2	52,29	64	193,12	138,91	830.048	719.250	
P5M1_S9_1	53,26	65	196,98	188,82	917.992	836.166	1026
P5M1_S9_2	52,83	65	204,18	146,08	917.992	836.166	
P5M2_S10_1	52,01	65	199,57	191,86	505.004	448.670	235*
P5M2_S10_2	51,66	65	207,15	144,04	505.004	448.670	
P6M1_S11_1	52,61	65	206,82	196,77	736.706	669.388	175
P6M1_S11_2	52,19	65	214,66	147,55	736.706	669.388	
P6M2_S12_1	52,53	65	201,1	192,18	690.506	617.266	316*
P6M2_S12_2	52,17	65	208,53	144,09	690.506	617.266	
P6M3_S13_1	52,42	65	205,22	194,77	589.572	544.490	316
P6M3_S13_2	52,08	65	212,4	149,43	589.572	544.490	
P6M4_S14_1	52,84	65	213,38	201,51	606.644	564.110	316
P6M4_S14_2	52,46	65	220,87	152,95	606.644	564.110	
P6M5_S15_1	52,09	65	200,88	190,08	542.702	494.148	316*
P6M5_S15_2	51,79	65	207,9	145,26	542.702	494.148	
P6M6_S16_1	53,41	65	199,53	191,28	792.940	716.238	316
P6M6_S16_2	53,01	65	205,39	146,48	792.940	716.238	
P7M1_S91_1	54,09	65	224,08	210,66	1.018.864	957.166	244
P7M1_S91_2	53,66	64	230,9	159,06	1.018.864	957.166	

Muestra	% duplicados	% GC	Longitud media pre-limpieza	Longitud media post-limpieza	Total secuencias pre-limpieza	Total secuencias post-limpieza	ST
P7M10_S22_1	52,5	65	233,1	219,24	494.788	462.650	244
P7M10_S22_2	52,08	65	240,2	160	494.788	462.650	
P7M11_S23_1	52,81	65	198,79	191,93	763.580	689.108	244
P7M11_S23_2	52,57	64	204,63	147,33	763.580	689.108	
P7M12_S24_1	51,8	65	188,	182,82	508.624	454.476	244
P7M12_S24_2	51,58	64	194,05	142,88	508.624	454.476	
P7M13_S25_1	52,39	65	235,47	222,6	476.512	446.576	244
P7M13_S25_2	52,12	65	243,46	162,91	476.512	446.576	
P7M14_S26_1	52,24	65	223,19	210,3	536.760	501.960	244
P7M14_S26_2	52	65	230,35	157,38	536.760	501.960	
P7M15_S27_1	53,11	65	199,51	189,3	781.088	720.340	244
P7M15_S27_2	52,75	64	206,54	147,07	781.088	720.340	
P7M16_S28_1	52,17	64	210,89	198,86	547.402	508.028	244
P7M16_S28_2	51,92	64	216,63	152,08	547.402	508.028	
P7M17_S29_1	52,53	64	189,94	180,9	743.430	686.428	244
P7M17_S29_2	52,34	64	194,88	144,94	743.430	686.428	
P7M18_S30_1	53,66	65	201,83	191,35	854.708	795.912	244
P7M18_S30_2	53,29	65	208,34	148,93	854.708	795.912	
P7M19_S31_1	52,67	65	181,04	175,71	845.042	754.360	244
P7M19_S31_2	52,36	64	185,95	141,15	845.042	754.360	
P7M2_S17_1	52,89	65	189,63	182,36	779.838	715.194	244
P7M2_S17_2	52,59	65	195,77	145,81	779.838	715.194	
P7M3_S18_1	52,49	65	191,86	186,89	656.868	580.106	244
P7M3_S18_2	52,19	65	197,68	143,27	656.868	580.106	
P7M4_S92_1	52,2	64	186,59	179,97	681.652	613.386	244
P7M4_S92_2	51,94	64	192,82	141,95	681.652	613.386	
P7M5_S19_1	52,67	65	216,94	204,84	673.132	618.064	244
P7M5_S19_2	52,34	65	223,47	152,92	673.132	618.064	
P7M7_S20_1	51,57	65	191,23	183,54	442.816	400.012	244*
P7M7_S20_2	51,38	65	198,61	141,74	442.816	400.012	
P7M8_S21_1	53,16	64	199,6	192,93	903.842	813.830	244
P7M8_S21_2	52,87	64	204,81	148,26	903.842	813.830	
P9M1_S32_1	52,34	65	182,71	178,2	674.394	590.968	244
P9M1_S32_2	52,01	64	189,27	138,92	674.394	590.968	
P9M2_S33_1	52,68	65	186,27	183,13	748.258	661.630	244
P9M2_S33_2	52,39	64	191,04	144,39	748.258	661.630	
P9M3_S34_1	51,1	65	205,18	197,83	269.380	240.530	244*
P9M3_S34_2	50,97	64	210,82	148,31	269.380	240.530	
P9M4_S35_1	51,37	64	205,08	197,87	358.184	315.400	244*
P9M4_S35_2	51,16	64	210,84	145,87	358.184	315.400	
P9M5_S36_1	50,7	64	202,86	195,2	172.684	152.444	244*
P9M5_S36_2	50,64	64	209,34	143,51	172.684	152.444	
P9M6_S37_1	53,09	65	200,69	192,85	800.766	697.932	244
P9M6_S37_2	52,6	65	209,21	141,35	800.766	697.932	

Muestra	% duplicados	% GC	Longitud media pre-limpieza	Longitud media post-limpieza	Total secuencias pre-limpieza	Total secuencias post-limpieza	ST
P10M1_S38_1	51,2	65	149,9	152,69	453.232	364.660	244*
P10M1_S38_2	51,02	64	157,15	124,6	453.232	364.660	
P10M2_S39_1	52,17	65	168,58	167,64	696.318	584.928	244
P10M2_S39_2	51,75	65	178,07	131,89	696.318	584.928	
P10M3_S40_1	52,06	65	192,41	185,01	576.686	501.284	NF
P10M3_S40_2	51,75	65	199,97	138,69	576.686	501.284	
P10M4_S41_1	51,5	64		147,29	476.586	391.718	NF
P10M4_S41_2	51,31	64	151,25	125,37	476.586	391.718	
P10M6_S42_1	52,5	65	174,23	170,62	832.248	712.154	244*
P10M6_S42_2	52,15	64	182,16	132,78	832.248	712.154	
P11M2_S43_1	51,73	65	162,66	162,08	478.194	403.568	175*
P11M2_S43_2	51,46	65	170,17	131,64	478.194	403.568	
P11M5_S44_1	51,89	64	164,23	163,85	642.486	526.026	244
P11M5_S44_2	51,57	64	172,15	128,4	642.486	526.026	
P11M6_S45_1	51,41	65	196,3	189,65	327.386	286.700	244*
P11M6_S45_2	51,22	64	203,28	141,69	327.386	286.700	
P11M7_S46_1	50,5	65	190,86	185,77	147.986	125.558	NF
P11M7_S46_2	50,41	64	199,08	137,95	147.986	125.558	
P12M1_S47_1	50,46	64	181,93	178,5	132.654	113.122	244*
P12M1_S47_2	50,4	64	190,2	136,54	132.654	113.122	
P12M2_S48_1	50,33	65	210,15	201,88	75.122	64.558	NF
P12M2_S48_2	50,23	64	218,21	143,79	75.122	64.558	
P12M3_S49_1	50,68	65	168,84	168,44	194.328	164.936	244*
P12M3_S49_2	50,55	65	175,49	134,4	194.328	164.936	
P12M4_S50_1	52,13	65	198,02	191,95	490.412	432.054	244
P12M4_S50_2	51,85	65	205,28	143,39	490.412	432.054	
P12M5_S51_1	52,4	65	145,15	148,85	875.712	714.404	244
P12M5_S51_2	52,14	65	151,74	125,33	875.712	714.404	
P12M6_S52_1	51,92	64	206,72	194,86	441.852	396.770	244*
P12M6_S52_2	51,58	64	216,44	140,17	441.852	396.770	
P12M7_S53_1	52,33	64	195,86	187,64	672.344	600.986	244
P12M7_S53_2	51,96	64	203,16	141,94	672.344	600.986	
P12M8_S54_1	53	65	164,77	164,94	797.702	679.670	244
P12M8_S54_2	52,67	64	171,27	133,47	797.702	679.670	
P13M2_S55_1	50,57	65	158,4	159,4	163.880	135.246	NF
P13M2_S55_2	50,51	64	164,52	132,34	163.880	135.246	
P13M3_S56_1	51,72	65	181,54	180,31	554.464	468.146	244
P13M3_S56_2	51,47	64	188,38	137,92	554.464	468.146	
P13M4_S57_1	51,72	65	176,29	173,99	571.994	494.856	244*
P13M4_S57_2	51,52	64	181,84	137,13	571.994	494.856	
P13M5_S58_1	51,28	64	190,13	185,52	300.530	259.122	NF
P13M5_S58_2	51,04	64	196,66	140,39	300.530	259.122	
P13M6_S59_1	52,31	65	193,09	188,71	657.612	576.322	244
P13M6_S59_2	51,99	64	199,76	143,21	657.612	576.322	

Muestra	% duplicados	% GC	Longitud media pre-limpieza	Longitud media post-limpieza	Total secuencias pre-limpieza	Total secuencias post-limpieza	ST
P14M10_S66_1	52,53	65	183,33	176,04	748.686	673.740	244
P14M10_S66_2	52,23	64	190,18	139,01	748.686	673.740	
P14M11_S67_1	54,09	65	181,49	175,6	1.250.118	1.124.870	244
P14M11_S67_2	53,64	65	187,98	140,7	1.250.118	1.124.870	
P14M2_S60_1	50,83	65	198,21	192,51	207.926	184.220	244*
P14M2_S60_2	50,7	64	205,38	145,22	207.926	184.220	
P14M4_S61_1	53,25	65	206,96	195,89	820.144	758.478	244
P14M4_S61_2	52,9	65	214,84	149,62	820.144	758.478	
P14M5_S62_1	53,43	65	189,46	182,33	709.568	640.502	244
P14M5_S62_2	53,1	64	196,44	142,72	709.568	640.502	
P14M6_S63_1	53,48	65	196,67	186,43	910.870	828.900	244
P14M6_S63_2	52,95	65	204,86	142,68	910.870	828.900	
P14M7_S64_1	53,56	65	185,96	178,5	1.056.220	955.248	244
P14M7_S64_2	53,19	65	192,3	140,88	1.056.220	955.248	
P14M9_S65_1	54,36	65	192,8	183,87	1.357.930	1.249.168	244
P14M9_S65_2	53,92	65	199,45	144,79	1.357.930	1.249.168	
P15M1_S68_1	52,36	65	201,61	193,71	684.012	603.274	244
P15M1_S68_2	52,05	64	208,6	143,97	684.012	603.274	
P15M2_S69_1	53,7	65	186,02	181,8	1.211.060	1.077.524	244
P15M2_S69_2	53,3	65	191,89	142,61	1.211.060	1.077.524	
P15M3_S70_1	51,18	65	179,59	177,18	415.456	357.238	244*
P15M3_S70_2	51,03	65	186,09	137,86	415.456	357.238	
P16M1_S71_1	52,54	65	193,78	188,21	712.208	629.688	244
P16M1_S71_2	52,18	65	200,48	143,32	712.208	629.688	
P16M2_S72_1	50	64	212,26	-	864	-	-
P16M2_S72_2	50	63	224,16	-	864	-	
P16M3_S73_1	53,1	65	172,68	171,27	929.836	803.036	244
P16M3_S73_2	52,75	65	179,11	136,24	929.836	803.036	
P16M4_S74_1	50,82	65	169,32	168,44	271.910	231.732	NF
P16M4_S74_2	50,73	64	176,33	134,98	271.910	231.732	
P16M5_S75_1	51,95	65	171,77	171,8	481.130	408.898	244*
P16M5_S75_2	51,66	65	179,19	135,92	481.130	408.898	
P17M1_S76_1	50,63	64	204,6	189,9	171.194	136.808	NF
P17M1_S76_2	50,41	63	224,88	123,66	171.194	136.808	
P17M2_S77_1	53,11	65	177,23	174,77	1.040.672	891.780	244
P17M2_S77_2	52,6	64	185,13	135,54	1.040.672	891.780	
P17M3_S78_1	51,96	65	183,8	180,78	570.892	496.642	244
P17M3_S78_2	51,63	64	190,38	139,54	570.892	496.642	
P17M4_S79_1	53,38	65	197,44	189,5	901.090	814.290	244
P17M4_S79_2	52,94	64	203,7	145,5	901.090	814.290	
P17M5_S80_1	52,18	65	187,3	182,99	720.886	628.446	244
P17M5_S80_2	51,89	64	194,38	140,09	720.886	628.446	
P17M6_S81_1	53,38	65	200,1	191,86	927.758	847.258	NF
P17M6_S81_2	52,92	65	206,66	147,1	927.758	847.258	

Muestra	% duplicados	% GC	Longitud media pre-limpieza	Longitud media post-limpieza	Total secuencias pre-limpieza	Total secuencias post-limpieza	ST
P17M7_S82_1	52,04	64	192,23	183,51	490.196	446.098	244*
P17M7_S82_2	51,73	64	199,56	142,89	490.196	446.098	
P17M8_S83_1	52,68	65	174,04	170,63	888.858	784.422	244
P17M8_S83_2	52,37	64	180,49	137,02	888.858	784.422	
P18M1_S84_1	52,07	65	214,21	203,56	482.396	442.646	244
P18M1_S84_2	51,81	64	221,72	150,67	482.396	442.646	
P18M2_S85_1	52,18	65	232,26	217,77	529.882	504.894	244
P18M2_S85_2	52,03	65	240,72	163,38	529.882	504.894	
P18M3_S86_1	52,45	65	225,68	212,8	591.160	557.378	244
P18M3_S86_2	52,19	65	233,73	159,97	591.160	557.378	
P18M6_S87_1	54,53	64	212,56	200,46	975.868	910.388	244
P18M6_S87_2	53,91	64	221,07	152,53	975.868	910.388	
P18M7_S88_1	53,52	64	206,67	196,07	870.608	813.294	244
P18M7_S88_2	53,17	64	213,27	153,3	870.608	813.294	
P18M8_S89_1	54,17	65	216,23	204,39	1.120.140	1.051.190	244
P18M8_S89_2	53,8	64	223,03	156,49	1.120.140	1.051.190	
P18M9_S90_1	52,17	64	203,85	192,65	563.736	526.494	244
P18M9_S90_2	52,02	64	210,76	150,4	563.736	526.494	

Tabla 7.13. MLST con SRST2 a partir de los genomas de las cepas de la base de datos para la selección de la referencia. *Hay genomas en los que uno de los genes no ha podido determinarse; en AES-1R hay varios alelos para el gen* acs.

CEPA	ST	*acsA*	*aroE*	*guaA*	*mutL*	*nuoD*	*ppsA*	*trpE*
FRD1	11	17	5	5	4	4	4	3
W36662	17	11	5	1	7	9	4	7
FA-HZ1	27	6	5	6	7	4	6	7
W45909	27	6	5	6	7	4	6	7
CARB01_63	111	17	5	5	4	4	4	3
F30658	111	17	5	5	4	4	4	3
T63266	132	6	20	1	3	4	4	2
LES431	146	6	5	11	3	4	23	1
LESB58	146	6	5	11	3	4	23	1
_12-4-4	152	6	5	19	3	4	6	7
ATCC_27853	155	28	5	36	3	3	13	7
F9670	155	28	5	36	3	3	13	7
S86968	155	28	5	36	3	3	13	7
T38079	155	28	5	36	3	3	13	7
F9676	167	40	5	11	5	4	28	37
YL84	169	40	5	30	5	3	33	14
F63912	198	11	5	11	11	3	27	7
RP73	198	11	5	11	11	3	27	7
NCGM2.S1	235	38	11	3	13	1	2	4
NCGM_1900	235	38	11	3	13	1	2	4
NCGM_1984	235	38	11	3	13	1	2	4
W16407	244	17	5	12	3	14	4	7
M1608	253	4	4	16	12	1	6	3
M37351	253	4	4	16	12	1	6	3
PA140R	253	4	4	16	12	1	6	3
UCBPP-PA14	253	4	4	16	12	1	6	3
X78812	257	35	24	36	11	4	15	14
19BR	277	39	5	9	11	27	5	2
213BR	277	39	5	9	11	27	5	2
SCV20265	299	17	5	36	3	3	7	3
BAMC_07-48	313	47	8	7	6	8	11	40
DN1	316	13	8	9	3	1	6	9
NCGM2	357	2	4	5	3	1	6	11
DK2	386	17	5	11	18	4	10	3
H27930	389	17	22	5	3	1	14	3
PA121617	389	17	22	5	3	1	14	3
DHS01	395	6	5	1	1	1	12	1
F22031	485	11	76	5	3	61	14	3
ATCC_15692	549	7	5	12	3	4	1	7
PAO1	549	7	5	12	3	4	1	7
PAO1OR	549	7	5	12	3	4	1	7
PA1	782	15	3	3	11	1	15	1

CEPA	ST	*acsA*	*aroE*	*guaA*	*mutL*	*nuoD*	*ppsA*	*trpE*
PA1R	782	15	3	3	11	1	15	1
PA1RG	782	15	3	3	11	1	15	1
VRFPA04	823	32	13	24	13	1	6	25
H5708	850	4	5	6	3	4	4	19
USDA-ARS-USMARC-41639	852	11	8	19	115	4	13	18
W60856	959	6	5	11	7	3	70	19
B136-33	1024	2	4	24	3	1	6	25
IOMTU_133	1047	18	8	5	5	1	6	4
H47921	1105	23	5	12	30	1	4	7
PA7	1195	87	34	43	37	53	107	126
M18	1239	16	5	1	3	4	15	7
F23197	1295	11	5	124	67	4	115	3
PACS2	1394	11	5	6	3	74	13	7
VA-134	1767	40	5	36	153	3	7	19
PA_D1	1971	32	190	3	62	8	7	26
PA_D16	1971	32	190	3	62	8	7	26
PA_D21	1971	32	190	3	62	8	7	26
PA_D22	1971	32	190	3	62	8	7	26
PA_D2	1971	32	190	3	62	8	7	26
PA_D25	1971	32	190	3	62	8	7	26
PA_D5	1971	32	190	3	62	8	7	26
PA_D9	1971	32	190	3	62	8	7	26
N17-1	2362	6	5	1	29	92	4	68
AES-1R	649 + alelo	11 + 151	84	11	3	4	4	7
8380	2619	17	5	1	3	4	4	3
MTB-1	2689	5	8	3	5	1	11	3
DSM	-	-	5	1	11	3	6	7
NCTC10332	-	-	5	1	11	3	6	7
T52373	-	16	10	5	3	-	42	7

Tabla 7.14. Estadísticas de mapeo de las muestras del HGUV. *En la tabla se muestra la cobertura media, es decir, el número de lecturas promedio que cubren una posición del genoma de referencia, además de su desviación típica y el número total de lecturas de la muestra (suma de los* paired ends*). Se han incluido además los porcentajes de lecturas que mapean contra la referencia elegida y de sitios cubiertos por estas lecturas respecto al total del genoma (6598022 pb) antes de descartar el repetitivo.*

Cepa	Cobertura media	Desviación típica	Total lecturas	% lecturas mapeadas	% genoma cubierto
P1M2	10,24	5,84	450.633	91,14	80,66
P1M3	7,93	4,84	307.786	90,49	71,80
P2M1	11,96	6,57	510.119	93,28	84,78
P2M2	8,02	4,76	301.845	93,93	72,76
P2M3	26,74	11,75	1.247.879	92,81	92,00
P4M2	23,89	10,45	1.104.564	92,11	91,87
P4M3	21,74	10,38	1.044.117	84,82	88,60
P4M4	43,46	11,48	1.686.433	93,63	92,58
P4M5	30,11	12,39	1.297.519	91,82	92,41
P5M1	39,77	14,49	1.470.643	94,04	85,32
P5M2	20,46	9,27	835.948	88,39	86,67
P6M1	29,56	12,58	1.220.959	87,22	89,38
P6M2	29,34	11,75	1.120.585	93,92	87,50
P6M3	26,7	10,46	989.380	94,4	87,45
P6M4	28,28	10,97	1.015.225	94,58	87,53
P6M5	23,82	10,41	911.489	94,26	87,13
P6M6	33,51	12,57	1.266.323	93,93	87,67
P7M1	45,02	16,75	1.652.758	93,26	91,90
P7M2	31,32	12,29	1.280.906	94	92,43
P7M3	25,81	11,19	1.047.906	94,03	91,03
P7M4	26,15	11,07	1.129.041	91,5	92,06
P7M5	29,39	11,86	1.120.479	93,13	92,10
P7M7	18,01	8,86	750.605	93,26	92,42
P7M8	35,85	14,33	1.451.225	91,86	91,92
P7M10	23,7	10,57	853.295	93,29	92,30
P7M11	30,74	12,44	1.234.492	92,71	92,52
P7M12	20,28	9,72	847.904	93,1	92,41
P7M13	23,36	10,03	831.687	93,34	92,65
P7M14	24,86	10,78	927.356	93,16	92,35
P7M15	31,8	12,72	1.289.525	92,81	92,04
P7M16	23,94	10,75	941.375	92,58	92,19
P7M17	29,62	12,53	1.249.543	92,64	92,43
P7M18	34,56	13,17	1.385.344	92,07	90,18
P7M19	31,34	12,33	1.355.968	92,15	92,53
P9M1	25,09	10,73	1.076.549	92,94	92,02
P9M2	28,59	11,68	1.189.749	92,57	92,31
P9M3	11,9	6,64	462.645	92,47	84,98
P9M4	15,11	8,15	598.859	92,7	88,40
P9M5	8,09	4,88	297.765	92,61	72,36

Cepa	Cobertura media	Desviación típica	Total lecturas	% lecturas mapeadas	% genoma cubierto
P9M6	30,87	11,95	1.249.051	93,27	92,45
P10M1	14,35	7,6	693.303	92,76	87,06
P10M2	23,47	10,7	1.065.701	92,15	91,32
P10M3	22,97	10,83	925.782	94,32	88,48
P10M4	14,16	7,58	742.357	86,22	86,71
P10M6	29,25	12,67	1.277.149	92,26	89,61
P11M2	15,95	8,35	758.123	87,66	86,04
P11M5	21,05	9,9	981.182	92,97	90,92
P11M6	13,44	7,27	548.010	92,61	86,74
P11M7	6,9	4,25	247.625	92,84	64,27
P12M1	6,45	3,95	222.869	92,74	59,19
P12M2	4,87	2,95	128.134	93,05	43,66
P12M3	8,04	4,75	321.700	93,51	71,83
P12M4	19,79	9,27	803.254	93,07	91,21
P12M5	25,75	10,53	1.270.047	92,36	92,09
P12M6	17,85	8,74	740.896	90,86	90,43
P12M7	26,24	11,7	1.108.545	91,71	92,07
P12M8	26,21	10,64	1.221.480	90,26	92,19
P13M2	6,94	4,13	264.885	93,15	63,94
P13M3	20,31	9,47	875.646	92,55	91,23
P13M4	20,97	9,73	921.485	92,86	91,38
P13M5	12,16	6,76	500.692	93,11	85,03
P13M6	25,6	11	1.055.362	92,71	92,14
P14M2	9,44	5,51	358.356	93,59	92,26
P14M4	34,11	12,9	1.363.773	92,05	92,62
P14M5	26,94	11,19	1.153.483	90,53	78,33
P14M6	35,38	13,07	1.444.219	93,39	92,49
P14M7	39,23	13,78	1.654.402	93,15	92,21
P14M9	51,28	15,9	2.114.237	92,46	92,52
P14M10	27,92	11,46	1.219.978	91,88	92,56
P14M11	44,41	15,27	1.904.350	91,87	92,67
P15M1	26,95	11,6	1.104.051	91,96	92,26
P15M2	44,14	14,85	1.862.179	91,81	92,68
P15M3	15,69	7,83	682.915	92,8	89,31
P16M1	27,57	11,25	1.142.127	92,69	92,56
P16M3	31,43	12,13	1.412.242	91,25	92,74
P16M4	10,35	5,83	448.492	92,29	81,01
P16M5	17,11	8,32	761.325	92,34	90,31
P17M1	7,05	4,47	268.126	91,38	63,90
P17M2	35,52	13,44	1.573.687	91,75	92,55
P17M3	21,37	9,49	919.593	92,11	91,61
P17M4	34,84	13,51	1.427.670	91,77	92,55
P17M5	26,95	11,27	1.152.215	92,04	92,21
P17M6	36,21	13,31	1.493.097	90,62	92,61
P17M7	19,71	9,21	838.755	91,88	91,08
P17M8	31,54	12,6	1.407.223	92,2	92,40
P18M1	21,18	9,81	833.326	95,05	94,90
P18M2	25,58	10,91	929.761	95,36	95,60

Cepa	Cobertura media	Desviación típica	Total lecturas	% lecturas mapeadas	% genoma cubierto
P18M3	27,34	11,87	1.019.396	94,97	95,68
P18M6	39,24	15,08	1.545.466	93,28	96,02
P18M7	35,94	14,03	1.425.493	94,5	95,96
P18M8	47,43	16,49	1.820.243	94,6	96,06
P18M9	24,73	11,16	967.928	94,2	91,56

Tabla 7.15. Comparativa de resultados de BEAST con diferentes modelos. p7 HGUV, R² = 0,8232 . *Las condiciones del* run *fueron replicadas (5-2 y 5-3) y testado el* prior *sin alineamiento.*

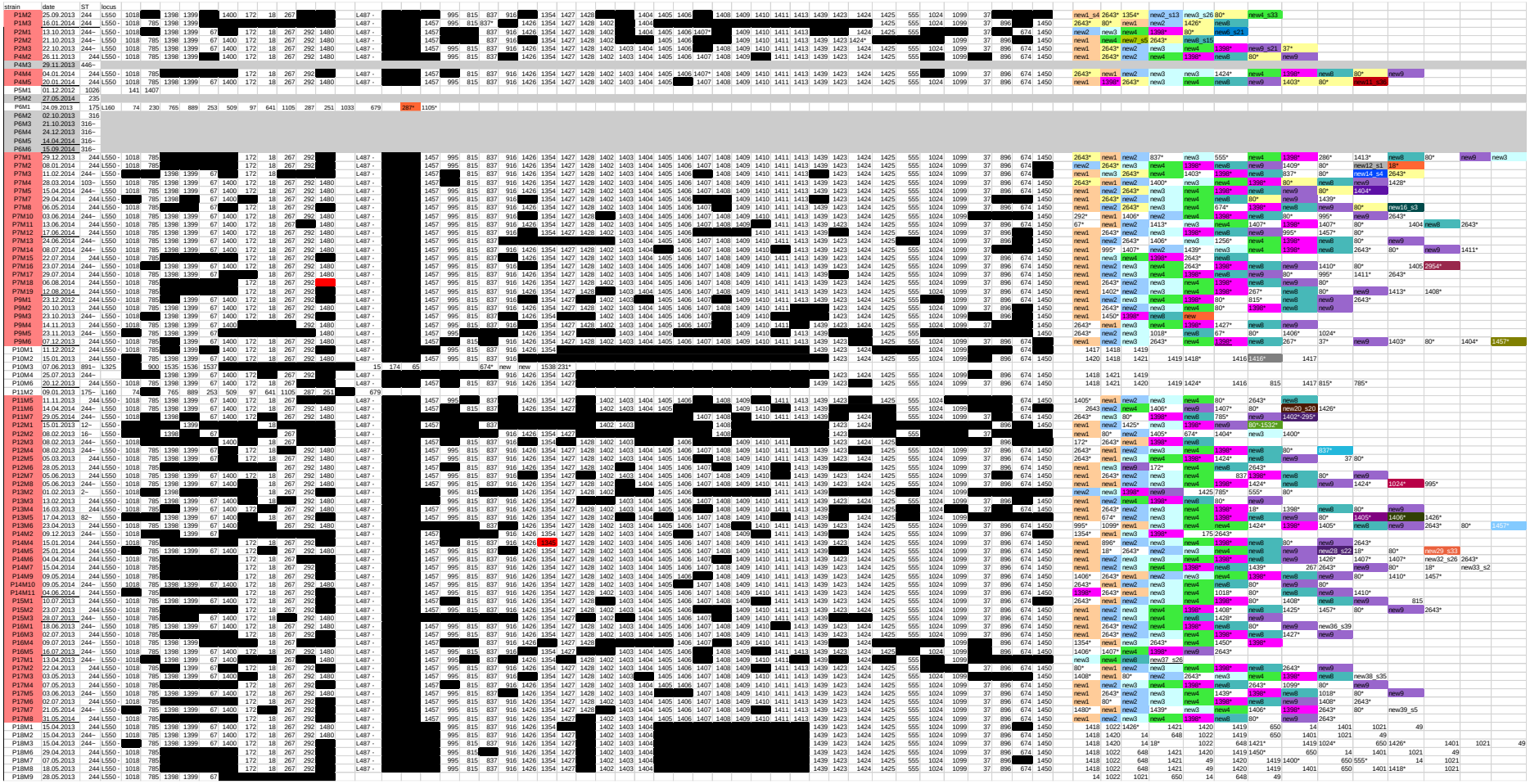
RUN EXPON DIST	CLOCK	POPULATION MODEL	MARGINAL LOGL (SS)	MARGINAL LOGL (PS)	BURN-IN	CLOCK.RATE (MEAN)	CLOCK.RATE (95%HPD)	TREELIKELIHOOD (95%HPD)
1	strict	constant	-7962010.72	-7962010.63	82000000	2.523E-6	[4.9184E-7, 4.4971E-6]	[-7963709.06, -7963695.63]
2	strict	exponential growth	-7963775.50	-7963776.77	90000000	4.818E-7	[2.7653E-8, 1.0745E-6]	[-7963707.48, -7963692.51]
3	strict	bayesian skyline 3	-7963775.26	-7963776.72	85000000	1.595E-7	[3.7061E-10, 4.6818E-7]	[-7963703.15, -7963689.44]
4	uncorrelated	constant			No alcanza la convergencia..	7.31E-10	[1.3036E-313, 2.3682E-16]	
5	uncorrelated	exponential growth	-7961986.39	-7961986.09	80000000 (más cadenas?)	4.407E-7	[2.9921E-9, 1.1832E-6]	[-7963670.39, -7963653.64]
5 – PRIOR	uncorrelated	exponential growth				0,16		
5 – 2	uncorrelated	exponential growth						
5 – 3	uncorrelated	exponential growth			55000000	4.321E-7		
6	uncorrelated	bayesian skyline 3			No alcanza la convergencia...	5.707E-11	[6.725E-106, 1.8638E-21]	
7	random	constant	-7963740.98	-7963741.72	15000000	1.914E-7	[4.5519E-11, 5.7013E-7]	[-7963678.51, -7963660.24]
8	random	exponential growth	-7963763.15	-7963764.63	10000000	1.556E-6	[1.4009E-7, 3.1258E-6]	[-7963692.43, -7963675.61]
9	random	bayesian skyline 3	-7963763.90	-7963765.25	40000000	1.968E-7	[3.2545E-11, 5.8206E-7]	[-7963680.31, -7963661.93]

Tabla 7.16 Resultados CRISPR del HGUE, primer abordaje. *En amarillo las que quedan fuera de locus, en negro las ausencias, verde y naranja las secuencias no clasificadas en locus concreto.*

strain	date	locus																	
Elche_01	03/07/15	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679	287			
Elche_02	27/03/15	L160 +	74	230	765	889	253	509	97	641	1105	287	251	1033	679	new			
Elche_03	16/03/15	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679	641			
Elche_04	20/04/15	L160 -	74		765	889	253	509	97	641	1105	287	251	1033	679				
Elche_05	18/03/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	679	230		
Elche_06	30/03/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	97	889		
Elche_07	31/03/15		74	230	765		253	509	97			287	251	1033	679				
Elche_08	03/07/15	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_09	11/05/15	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_1A	18/05/15	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_10	03/07/15	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_11	18/05/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	230	287		
Elche_12	18/05/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	new	new		
Elche_13	29/05/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	889*	1105*		
Elche_14	27/08/15	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_15	06/08/15	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_16	19/08/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	765*	509*		
Elche_17	28/08/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	253*	251*		
Elche_18	22/10/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	765*			
Elche_21	03/12/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	509*	new		
Elche_22	05/10/15		74	230	765	889	253		97	641	1105	287	251	1033	679				
Elche_23	14/12/15		74	230		889		509	97	641	1105	287	251	1033	679				
Elche_24	22/12/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	253*			
Elche_25	05/01/16	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_26	11/01/16	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_2A	12/05/15	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_3A	19/05/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	1033*	679*		
Elche_27	21/01/16		74	230	765	889	253	509	97	641	1105	287	251	1033	679	1105*			
Elche_28	25/01/16	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_30	28/01/16		74	230	765	889		509	97	641	1105	287	251	1033	679				
Elche_32	05/01/15	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_33	30/12/14		2811	1104*	283	2811*	845	2680*	475										
Elche_34	30/12/14		2811	1104*	283	845*	845	2680*	475	283*									
Elche_35	12/08/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	1105*			
Elche_36	14/01/15	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_37	16/01/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	new	74*	287*	new
Elche_38	27/01/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	641*	new	679*	
Elche_39	16/12/15	L160	74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_40	14/12/15	L571	199	301	14	1398	1399	67	162	18	265	1086		267	292	1480			
Elche_41	28/07/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	97*	1033*		
Elche_42	04/09/15																		
Elche_43	03/11/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	230*	679*	509*	1105*
Elche_45	14/07/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_4A	15/01/16		74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_5A	12/05/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679				
Elche_46	15/12/14		74	230	765	889	253	509	97	641	1105	287	251	1033	679	251*			
Elche_47	15/12/14		74	230	765	889	253	509	97	641	1105	287	251	1033	679	97*	679*		
Elche_48	11/05/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	new	97*	253*	509*
Elche_49	08/06/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	889*			
Elche_50	25/01/16		74	230	765	889	253	509	97	641	1105	287	251	1033	679	641*			
Elche_51	01/07/15		74	230	765	889	253	509	97	641	1105	287	251	1033	679	889*			
Elche_52	25/02/16		74	230	765	889	253	509	97	641	1105	287	251		679				
Elche_54	07/06/16																		
Elche_57	14/07/16		74	230	765	889	253	509	97	641	1105	287	251	1033	679	509*	1105*		
Elche_59	07/09/16		74	230		889	253	509	97	641	1105	287	251	1033	679	97			
Elche_60	12/09/16		74	230	765	889	253	509	97	641	1105	287	251	1033	679	new	641*		
Elche_62	26/09/16	L370 -	1619		1617	1607	1608	1616	345	1101	280	853	318	384					
Elche_63	13/10/16																		
Elche_64	26/09/16	L370 -	1619	1618	1617		1608	1616	345	1101	280	853	318	384					
Elche_65	26/09/16		74	230	765	889	253	509	97	641	1105	287	251	1033	679	287*	1105*		
Elche_66	07/10/16		74			889	253	509	97	641	1105	287		1033	679	287*			
Elche_67	27/10/16		74	230	765	889	253	509	97	641	1105	287	251	1033	679	679*			
Elche_68	02/11/16		74	230	765	889	253	509	97	641	1105	287	251	1033	679	new			

strain																													
Elche_40	L680*	2199							85	56	2195	312	227			767	986	59	358	975	428			999	579	273	1564*	767*	1451

Tabla 7.17. Resultados CRISPR del HGUV, primer abordaje. *En amarillo las que quedan fuera de locus, en negro las ausencias, rojo las que no son 100% idénticas. Resto de colores secuencias no clasificadas en locus concreto que además son nuevas y compartidas por muchas de las cepas. Muestras en color salmón correspondientes al brote.*



7. Material suplementario

8. Bibliografía

Altschul, S. F. *et al.* (1990) «Basic local alignment search tool», *Journal of Molecular Biology*, 215(3), pp. 403–410. doi: 10.1016/S0022-2836(05)80360-2.

Van der Auwera, G. A. *et al.* (2013) «From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline», en *Current Protocols in Bioinformatics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., p. 11.10.1-11.10.33. doi: 10.1002/0471250953.bi1110s43.

Bankevich, A. *et al.* (2012) «SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing», *Journal of Computational Biology*, 19(5), pp. 455–477. doi: 10.1089/cmb.2012.0021.

Barrangou, R. *et al.* (2007) «CRISPR provides acquired resistance against viruses in prokaryotes.», *Science (New York, N.Y.)*. American Association for the Advancement of Science, 315(5819), pp. 1709–12. doi: 10.1126/science.1138140.

Barrangou, R. y Horvath, P. (2017) «A decade of discovery: CRISPR functions and applications», *Nature Microbiology*. Macmillan Publishers Limited, 2(June), pp. 1–9. doi: 10.1038/nmicrobiol.2017.92.

van Belkum, A. *et al.* (2015) «Phylogenetic distribution of CRISPR-Cas systems in antibiotic- resistant pseudomonas aeruginosa», *mBio*, 6(6), pp. 1–13. doi: 10.1128/mBio.01796-15.

Bianconi, I. *et al.* (2016) «Draft Genome Sequences of 40 Pseudomonas aeruginosa Clinical Strains Isolated from the Sputum of a Single Cystic Fibrosis Patient Over an 8-Year Period.», *Genome announcements*. American Society for Microbiology (ASM), 4(6). doi: 10.1128/genomeA.01205-16.

Blanc, D. S. *et al.* (2016) «Hand soap contamination by Pseudomonas aeruginosa in a tertiary care hospital: no evidence of impact on patients», *Journal of Hospital Infection*. W.B. Saunders, 93(1), pp. 63–67. doi: 10.1016/J.JHIN.2016.02.010.

Boers, S. A. *et al.* (2014) «Whole-genome mapping for high-resolution genotyping of Pseudomonas aeruginosa.», *Journal of microbiological methods*. Elsevier B.V., 106, pp. 19–22. doi: 10.1016/j.mimet.2014.07.020.

Cady, K. C. *et al.* (2011) «Prevalence, conservation and functional analysis of Yersinia and Escherichia CRISPR regions in clinical Pseudomonas aeruginosa isolates.», *Microbiology (Reading, England)*. Microbiology Society, 157(Pt 2), pp. 430–7. doi: 10.1099/mic.0.045732-0.

Camacho, C. *et al.* (2009) «BLAST+: architecture and applications», *BMC Bioinformatics*, 10(1), p. 421. doi: 10.1186/1471-2105-10-421.

Cheng, L. *et al.* (2013) «Hierarchical and Spatially Explicit Clustering of DNA Sequences with BAPS Software», *Molecular Biology and Evolution*. Oxford University Press, 30(5), pp. 1224–1228. doi: 10.1093/molbev/mst028.

Coll, F. *et al.* (2017) «Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community.», *Science translational medicine*. American Association for the Advancement of Science, 9(413), p. eaak9745. doi: 10.1126/scitranslmed.aak9745.

Comas, I. *et al.* (2013) «Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans.», *Nature genetics*. NIH Public Access, 45(10), pp. 1176–82. doi: 10.1038/ng.2744.

Cornelis, P. y Dingemans, J. (2013) «Pseudomonas aeruginosa adapts its iron uptake strategies in function of the type of infections», *Frontiers in Cellular and Infection Microbiology*. Frontiers, 3, p. 75. doi: 10.3389/fcimb.2013.00075.

Costerton, J. W., Stewart, P. S. y Greenberg, E. P. (1999) «Bacterial biofilms: A common cause of persistent infections», *Science*, 284(5418), pp. 1318-1322. doi: 10.1126/science.284.5418.1318.

Cramer, N. *et al.* (2011) «Microevolution of the major common Pseudomonas aeruginosa clones C and PA14 in cystic fibrosis lungs», *Environmental Microbiology*, 13(7), pp. 1690-1704. doi: 10.1111/j.1462-2920.2011.02483.x.

Criscuolo, A. y Gribaldo, S. (2010) «BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments», *BMC Evolutionary Biology*. BioMed Central, 10(1), p. 210. doi: 10.1186/1471-2148-10-210.

Croucher, N. J. *et al.* (2014) «Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins», *Nucleic Acids Research*, 44(0), pp. 1-13. doi: 10.1093/nar/gku1196.

Curran, B. *et al.* (2004) «Development of a Multilocus Sequence Typing Scheme for the Opportunistic Pathogen Pseudomonas aeruginosa», *Journal of Clinical Microbiology*, 42(12), pp. 5644-5649. doi: 10.1128/JCM.42.12.5644-5649.2004.

Danecek, P. *et al.* (2011) «The variant call format and VCFtools», *Bioinformatics*, 27(15), pp. 2156-2158. doi: 10.1093/bioinformatics/btr330.

Darling, A. E., Mau, B. y Perna, N. T. (2009) «Progressive Mauve: Multiple alignment of genomes with gene flux and rearrangement». Disponible en: http://arxiv.org/abs/0910.5780.

DePristo, M. A. *et al.* (2011) «A framework for variation discovery and genotyping using next-generation DNA sequencing data», *Nature Genetics*. Nature Publishing Group, 43(5), pp. 491-498. doi: 10.1038/ng.806.

Dettman, J. R., Rodrigue, N. y Kassen, R. (2014) «Genome-wide patterns of recombination in the opportunistic human pathogen pseudomonas aeruginosa», *Genome Biology and Evolution*, 7(1), pp. 18-34. doi: 10.1093/gbe/evu260.

Drummond, A. J. *et al.* (2012) «Bayesian Phylogenetics with BEAUti and the BEAST 1.7», *Molecular Biology and Evolution*, 29(8), pp. 1969-1973. doi: 10.1093/molbev/mss075.

ECDC (2014) *Antimicrobial Resistance surveillance in Europe 2014. Annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net)*. doi: 10.2900/93403.

England, W. E., Kim, T. y Whitaker, R. J. (2018) «Metapopulation Structure of CRISPR-Cas Immunity in Pseudomonas aeruginosa and Its Viruses.», *mSystems*. American Society for Microbiology (ASM), 3(5). doi: 10.1128/mSystems.00075-18.

Enright, A. J., Van Dongen, S. y Ouzounis, C. A. (2002) «An efficient algorithm for large-scale detection of protein families.», *Nucleic acids research*, 30(7), pp. 1575-84. Disponible en: http://www.ncbi.nlm.nih.gov/pubmed/11917018 (Accedido: 16 de noviembre de 2018).

Evans, B. A. y Amyes, S. G. B. (2014) «OXA β -Lactamases», 27(2). doi:

10.1128/CMR.00117-13.

Ewels, P. *et al.* (2016) «MultiQC: summarize analysis results for multiple tools and samples in a single report», *Bioinformatics*. Oxford University Press, 32(19), pp. 3047-3048. doi: 10.1093/bioinformatics/btw354.

Ewing, B. *et al.* (1998) «Base-calling of automated sequencer traces using phred. I. Accuracy assessment.», *Genome research*. Cold Spring Harbor Laboratory Press, 8(3), pp. 175-85. doi: 10.1101/GR.8.3.175.

Fabre, L. *et al.* (2012) «CRISPR typing and subtyping for improved laboratory surveillance of Salmonella infections.», *PloS one*. Public Library of Science, 7(5), p. e36995. doi: 10.1371/journal.pone.0036995.

Feil, E. J. *et al.* (2004) «eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data.», *Journal of bacteriology*, 186(5), pp. 1518-1530. doi: 10.1128/JB.186.5.1518-1530.2004.

Feliziani, S., Marvig, R. L., *et al.* (2014) «Coexistence and Within-Host Evolution of Diversified Lineages of Hypermutable Pseudomonas aeruginosa in Long-term Cystic Fibrosis Infections», *PLoS Genetics*. Editado por I. Matic, 10(10), p. e1004651. doi: 10.1371/journal.pgen.1004651.

Felsenstein, J. (1985) «CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP», *Evolution*, 39(4), pp. 783-791. doi: 10.1111/j.1558-5646.1985.tb00420.x.

Francisco, A. P. *et al.* (2009) «Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach», *BMC Bioinformatics*. BioMed Central, 10(1), p. 152. doi: 10.1186/1471-2105-10-152.

García-Castillo, M. *et al.* (2011) «Wide dispersion of ST175 clone despite high genetic diversity of carbapenem-nonsusceptible Pseudomonas aeruginosa clinical strains in 16 Spanish hospitals», *Journal of Clinical Microbiology*, 49(8), pp. 2905-2910. doi: 10.1128/JCM.00753-11.

García-Castillo, M. *et al.* (2012) «Emergence of a mutL mutation causing multilocus sequence typing-pulsed-field gel electrophoresis discrepancy among Pseudomonas aeruginosa isolates from a cystic fibrosis patient.», *Journal of clinical microbiology*. American Society for Microbiology (ASM), 50(5), pp. 1777-8. doi: 10.1128/JCM.05478-11.

Garcia, R., Gemperlein, K. y Muller, R. (2014) «Minicystis rosea gen. nov., sp. nov., a polyunsaturated fatty acid-rich and steroid-producing soil myxobacterium», *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*. Microbiology Society, 64(Pt 11), pp. 3733-3742. doi: 10.1099/ijs.0.068270-0.

Gomila, M. *et al.* (2013) «Genetic diversity of clinical Pseudomonas aeruginosa isolates in a public hospital in Spain.», *BMC microbiology*, 13, p. 138. doi: 10.1186/1471-2180-13-138.

González-Candelas, F. *et al.* (2013) «Molecular evolution in court: analysis of a large hepatitis C virus outbreak from an evolving source.», *BMC biology*. BioMed Central, 11, p. 76. doi: 10.1186/1741-7007-11-76.

Groenen, P. M. A. *et al.* (1993) «Nature of DNA polymorphism in the direct repeat

cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method», *Molecular Microbiology*. Wiley/Blackwell (10.1111), 10(5), pp. 1057-1065. doi: 10.1111/j.1365-2958.1993.tb00976.x.

Guindon, S. *et al.* (2010) «New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0», *Systematic Biology*. Oxford University Press, 59(3), pp. 307-321. doi: 10.1093/sysbio/syq010.

Gupta, S. K. *et al.* (2014) «ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes.», *Antimicrobial agents and chemotherapy*. American Society for Microbiology, 58(1), pp. 212-20. doi: 10.1128/AAC.01310-13.

Gupta, S. y Maiden, M. C. J. (2001) «Exploring the evolution of diversity in pathogen populations», *Trends in Microbiology*. Elsevier Current Trends, 9(4), pp. 181-185. doi: 10.1016/S0966-842X(01)01986-2.

Gurevich, A. *et al.* (2013) «QUAST: quality assessment tool for genome assemblies», *Bioinformatics*. Oxford University Press, 29(8), pp. 1072-1075. doi: 10.1093/bioinformatics/btt086.

Hadfield, J. *et al.* (2018) «Phandango: an interactive viewer for bacterial population genomics», *Bioinformatics*. Editado por J. Kelso. Oxford University Press, 34(2), pp. 292-293. doi: 10.1093/bioinformatics/btx610.

Han, K. *et al.* (2013) «Extraordinary expansion of a Sorangium cellulosum genome from an alkaline milieu», *Scientific Reports*. Nature Publishing Group, 3(1), p. 2101. doi: 10.1038/srep02101.

Heiniger, R. W. *et al.* (2010) «Infection of human mucosal tissue by Pseudomonas aeruginosa requires sequential and mutually dependent virulence factors and a novel pilus-associated adhesinc mi_1461 1158..1173». doi: 10.1111/j.1462-5822.2010.01461.x.

Henwood, C. J. *et al.* (2001) «Antimicrobial susceptibility of Pseudomonas aeruginosa: results of a UK survey and evaluation of the British Society for Antimicrobial Chemotherapy disc susceptibility test.», *The Journal of antimicrobial chemotherapy*, 47(6), pp. 789-99. Disponible en: http://www.ncbi.nlm.nih.gov/pubmed/11389111 (Accedido: 25 de noviembre de 2018).

Hermans, P. W. *et al.* (1991) «Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains.», *Infection and immunity*. American Society for Microbiology (ASM), 59(8), pp. 2695-705. Disponible en: http://www.ncbi.nlm.nih.gov/pubmed/1649798 (Accedido: 5 de septiembre de 2018).

Hilker, R. *et al.* (2015) «Interclonal gradient of virulence in the *P seudomonas aeruginosa* pangenome from disease and environment», *Environmental Microbiology*. Wiley/Blackwell (10.1111), 17(1), pp. 29-46. doi: 10.1111/1462-2920.12606.

Hoang, S. *et al.* (2018) «Risk factors for colonization and infection by Pseudomonas aeruginosa in patients hospitalized in intensive care units in France», *PLOS ONE*. Editado por Y. R. Kou. Public Library of Science, 13(3), p. e0193300. doi: 10.1371/journal.pone.0193300.

Høiby, N. *et al.* (2010) «Antibiotic resistance of bacterial biofilms», *International Journal of Antimicrobial Agents*, 35(4), pp. 322-332. doi: 10.1016/j.ijantimicag.2009.12.011.

Holt, K. E. *et al.* (2013) «Tracking the establishment of local endemic populations of an emergent enteric pathogen», *PNAS*, p. doi:10.1073/pnas.1308632110. doi: 10.1073/pnas.1308632110.

Hunt, M. *et al.* (2017) «ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads.», *Microbial genomics*. Microbiology Society, 3(10), p. e000131. doi: 10.1099/mgen.0.000131.

Hyatt, D. *et al.* (2010) «Prodigal: prokaryotic gene recognition and translation initiation site identification», *BMC Bioinformatics*, 11(1), p. 119. doi: 10.1186/1471-2105-11-119.

Inouye, M. *et al.* (2014) «SRST2: Rapid genomic surveillance for public health and hospital microbiology labs.», *Genome medicine*. BioMed Central, 6(11), p. 90. doi: 10.1186/s13073-014-0090-6.

Ishino, Y. *et al.* (1987) «Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.», *Journal of bacteriology*. American Society for Microbiology Journals, 169(12), pp. 5429-33. doi: 10.1128/JB.169.12.5429-5433.1987.

Jansen, R. *et al.* (2002) «Identification of genes that are associated with DNA repeats in prokaryotes», *Molecular Microbiology*. Wiley/Blackwell (10.1111), 43(6), pp. 1565-1575. doi: 10.1046/j.1365-2958.2002.02839.x.

Jolley, K. A., Bray, J. E. y Maiden, M. C. J. (2018) «Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications», *Wellcome Open Research*, 3, p. 124. doi: 10.12688/wellcomeopenres.14826.1.

Jorth, P. *et al.* (2015) «Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs», *Cell Host & Microbe*. Elsevier Inc., 18(3), pp. 307-319. doi: 10.1016/j.chom.2015.07.006.

Kamerbeek, J. *et al.* (1997) «Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology.», *Journal of clinical microbiology*. American Society for Microbiology (ASM), 35(4), pp. 907-14. Disponible en: http://www.ncbi.nlm.nih.gov/pubmed/9157152 (Accedido: 29 de noviembre de 2018).

Keane, J. A. *et al.* (2016) «SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments», *Microbial Genomics*, 2(4), p. e000056. doi: 10.1099/mgen.0.000056.

Kidd, T. J. *et al.* (2012) «Pseudomonas aeruginosa Exhibits Frequent Recombination, but Only a Limited Association between Genotype and Ecological Setting», *PLoS ONE*. Editado por S. P. Brown. Public Library of Science, 7(9), p. e44199. doi: 10.1371/journal.pone.0044199.

Kishino, H. y Hasegawa, M. (1989) «Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea», *Journal of Molecular Evolution*. Springer-Verlag, 29(2), pp. 170-179. doi: 10.1007/BF02100115.

Knols, B. G. *et al.* (2016) «Global Priority List Of Antibiotic-Resistant Bacteria To Guide Research, Discovery, And Development Of New Antibiotics», *The Lancet Infectious Diseases*, 9(9), pp. 535-536. doi: 10.1016/S1473-3099(09)70222-1.

Kreda, S. M., Davis, C. W. y Rose, M. C. (2012) «CFTR, mucins, and mucus obstruction in cystic fibrosis.», *Cold Spring Harbor perspectives in medicine*. Cold Spring Harbor

Laboratory Press, 2(9), p. a009589. doi: 10.1101/cshperspect.a009589.

Kupczok, A., Landan, G. y Dagan, T. (2015) «The Contribution of Genetic Recombination to CRISPR Array Evolution», *Genome Biology and Evolution*, 7(7), pp. 1925-1939. doi: 10.1093/gbe/evv113.

Kurtz, S. *et al.* (2004) «Versatile and open software for comparing large genomes.», *Genome Biology*, 5(2), p. R12. doi: 10.1186/gb-2004-5-2-r12.

Langmead, B. y Salzberg, S. L. (2012) «Fast gapped-read alignment with Bowtie 2», *Nature Methods*. Nature Publishing Group, 9(4), pp. 357-359. doi: 10.1038/nmeth.1923.

Li, H.-Y. *et al.* (2018) «Characterization of CRISPR-Cas Systems in Clinical Klebsiella pneumoniae Isolates Uncovers Its Potential Association With Antibiotic Susceptibility», *Frontiers in Microbiology*, 9(July), pp. 1-9. doi: 10.3389/fmicb.2018.01595.

Li, H. *et al.* (2009) «The Sequence Alignment/Map format and SAMtools», *Bioinformatics*, 25(16), pp. 2078-2079. doi: 10.1093/bioinformatics/btp352.

Li, H. (2011) «A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data», *Bioinformatics*, 27(21), pp. 2987-2993. doi: 10.1093/bioinformatics/btr509.

Li, H. (2013) «Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM». Disponible en: http://arxiv.org/abs/1303.3997 (Accedido: 11 de abril de 2018).

Liu, P. V *et al.* (1987) *Comparison of the Chinese Schema and the International Antigenic Typing System for Serotyping Pseudomonas aeruginosa*, *JOURNAL OF CLINICAL MICROBIOLOGY*. Disponible en: http://jcm.asm.org/ (Accedido: 4 de octubre de 2018).

Liu, P. V y Wang, S. (1990) «Three new major somatic antigens of Pseudomonas aeruginosa.», *Journal of clinical microbiology*. American Society for Microbiology (ASM), 28(5), pp. 922-5. Disponible en: http://www.ncbi.nlm.nih.gov/pubmed/2112563 (Accedido: 3 de octubre de 2018).

López-Causapé, C. *et al.* (2013) «Clonal Dissemination, Emergence of Mutator Lineages and Antibiotic Resistance Evolution in Pseudomonas aeruginosa Cystic Fibrosis Chronic Lung Infection.», *PloS one*, 8(8), p. e71001. doi: 10.1371/journal.pone.0071001.

Lopez-Sanchez, M.-J. *et al.* (2012) «The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome», *Molecular Microbiology*, 85(6), pp. 1057-1071. doi: 10.1111/j.1365-2958.2012.08172.x.

Maatallah, M. *et al.* (2011) «Population Structure of Pseudomonas aeruginosa from Five Mediterranean Countries: Evidence for Frequent Recombination and Epidemic Occurrence of CC235», *PLoS ONE*. Editado por R. J. Redfield. Public Library of Science, 6(10), p. e25617. doi: 10.1371/journal.pone.0025617.

Madoui, M.-A. *et al.* (2015) «Genome assembly using Nanopore-guided long and error-free DNA reads.», *BMC genomics*. BioMed Central, 16(1), p. 327. doi: 10.1186/s12864-015-1519-z.

Makarova, K. S. *et al.* (2011) «Evolution and classification of the CRISPR-Cas systems», *Nature Reviews Microbiology*. Nature Publishing Group, 9(6), pp. 467-477. doi:

10.1038/nrmicro2577.

Marvig, R. L. *et al.* (2013) «Genome Analysis of a Transmissible Lineage of Pseudomonas aeruginosa Reveals Pathoadaptive Mutations and Distinct Evolutionary Paths of Hypermutators», 9(9). doi: 10.1371/journal.pgen.1003741.

Marvig, R. L. *et al.* (2014) «Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis.», *Nature genetics*, 47(1), pp. 57-64. doi: 10.1038/ng.3148.

McArthur, A. G. *et al.* (2013) «The comprehensive antibiotic resistance database.», *Antimicrobial agents and chemotherapy*. American Society for Microbiology (ASM), 57(7), pp. 3348-57. doi: 10.1128/AAC.00419-13.

Meletis, G. (2016) «Carbapenem resistance: overview of the problem and future perspectives.», *Therapeutic advances in infectious disease*. SAGE Publications, 3(1), pp. 15-21. doi: 10.1177/2049936115621709.

Minh, B. Q., Nguyen, M. A. T. y von Haeseler, A. (2013) «Ultrafast Approximation for Phylogenetic Bootstrap», *Molecular Biology and Evolution*. Oxford University Press, 30(5), pp. 1188-1195. doi: 10.1093/molbev/mst024.

Miyoshi-Akiyama, T. *et al.* (2017) «Emergence and Spread of Epidemic Multidrug-Resistant Pseudomonas aeruginosa», *Genome Biology and Evolution*. Oxford University Press, 9(12), pp. 3238-3245. doi: 10.1093/gbe/evx243.

Mojica, F. J. M. *et al.* (2005) «Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements», *Journal of Molecular Evolution*. Springer-Verlag, 60(2), pp. 174-182. doi: 10.1007/s00239-004-0046-3.

Mojica, F. J. M., Juez, G. y Rodriguez-Valera, F. (1993) «Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites», *Molecular Microbiology*. Wiley/Blackwell (10.1111), 9(3), pp. 613-621. doi: 10.1111/j.1365-2958.1993.tb01721.x.

Morales, E. *et al.* (2012) *Hospital costs of nosocomial multi-drug resistant Pseudomonas aeruginosa acquisition*. doi: 10.1186/1472-6963-12-122.

Mosquera-Rendón, J. *et al.* (2016) «Pangenome-wide and molecular evolution analyses of the Pseudomonas aeruginosa species.», *BMC genomics*. BioMed Central, 17, p. 45. doi: 10.1186/s12864-016-2364-4.

Nguyen, L.-T. *et al.* (2015) «IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies», *Molecular Biology and Evolution*, 32(1), pp. 268-274. doi: 10.1093/molbev/msu300.

Nikolenko, S. I., Korobeynikov, A. I. y Alekseyev, M. A. (2013) «BayesHammer: Bayesian clustering for error correction in single-cell sequencing», *BMC Genomics*. BioMed Central, 14(Suppl 1), p. S7. doi: 10.1186/1471-2164-14-S1-S7.

Oliver, A. *et al.* (2000) «High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection.», *Science*, 288(2000), pp. 1251-1253. doi: 10.1126/science.288.5469.1251.

Oliver, A. (2010) «Mutators in cystic fibrosis chronic lung infection : Prevalence , mechanisms , and consequences for antimicrobial therapy», *International Journal of Medical Microbiology*. Elsevier GmbH., 300(8), pp. 563-572. doi:

10.1016/j.ijmm.2010.08.009.

Oliver, A. *et al.* (2015) «The increasing threat of Pseudomonas aeruginosa high-risk clones», *Drug Resistance Updates*. Elsevier Ltd, 21-22, pp. 41-59. doi: 10.1016/j.drup.2015.08.002.

Oliver, A. y Mena, A. (2010) «Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance», *Clinical Microbiology and Infection*, 16(7), pp. 798-808. doi: 10.1111/j.1469-0691.2010.03250.x.

Olson, M. V. *et al.* (2000) «Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen.», *Nature*, 406(6799), pp. 959-964. doi: 10.1038/35023079.

Page, A. J. *et al.* (2015) «Roary: Rapid large-scale prokaryote pan genome analysis», *Bioinformatics*, 31(22), pp. 3691-3693. doi: 10.1093/bioinformatics/btv421.

Paradis, E., Claude, J. y Strimmer, K. (2004) «APE: Analyses of Phylogenetics and Evolution in R language.», *Bioinformatics (Oxford, England)*, 20(2), pp. 289-90. Disponible en: http://www.ncbi.nlm.nih.gov/pubmed/14734327 (Accedido: 24 de mayo de 2018).

Parcell, B. J. *et al.* (2018) «Pseudomonas aeruginosa intensive care unit outbreak: winnowing of transmissions with molecular and genomic typing», *Journal of Hospital Infection*, 98, pp. 282-288. doi: 10.1016/j.jhin.2017.12.005.

Pier, G.B. and Ramphal, R. (2005) «Pseudomonas aeruginosa.», en Mandell, G.L., Bennett, J.E. and Dolin, R., E. (ed.) *Mandell, Douglas and Bennett's Principles and Practice of Infectious Disease, 6th Edition*. New York: Churchill Livingstone, pp. 2587-2615.

Pritchard, J. K., Stephens, M. y Donnelly, P. (2000) «Inference of population structure using multilocus genotype data.», *Genetics*. Genetics Soc America, 155(2), pp. 945-959. Disponible en: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461096&tool=pmcentrez&rendertype=abstract.

Rambaut, A. *et al.* (2016) «Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen)», *Virus Evolution*. Oxford University Press, 2(1), p. vew007. doi: 10.1093/ve/vew007.

La Rosa, R., Johansen, H. K. y Molin, S. (2018) «Convergent Metabolic Specialization through Distinct Evolutionary Paths in Pseudomonas aeruginosa.», *mBio*. American Society for Microbiology, 9(2), pp. e00269-18. doi: 10.1128/mBio.00269-18.

Rouli, L. *et al.* (2015) «The bacterial pangenome as a new tool for analysing pathogenic bacteria.», *New microbes and new infections*. Elsevier, 7, pp. 72-85. doi: 10.1016/j.nmni.2015.06.005.

Roy Chowdhury, P., Scott, M. J. y Djordjevic, S. P. (2017) «Genomic islands 1 and 2 carry multiple antibiotic resistance genes in *Pseudomonas aeruginosa* ST235, ST253, ST111 and ST175 and are globally dispersed», *Journal of Antimicrobial Chemotherapy*. Oxford University Press, 72(2), pp. 620-622. doi: 10.1093/jac/dkw471.

Schmieder, R. y Edwards, R. (2011) «Quality control and preprocessing of metagenomic datasets.», *Bioinformatics*. Oxford University Press, 27(6), pp. 863-4. doi: 10.1093/bioinformatics/btr026.

Seemann, T. (2014) «Prokka: rapid prokaryotic genome annotation», *Bioinformatics*, 30(14), pp. 2068-2069. doi: 10.1093/bioinformatics/btu153.

Seemann, T. (sin fecha) *mlst*. Disponible en: https://github.com/tseemann/mlst.

Segura, C. *et al.* (2010) «Spread of plasmids containing the bla VIM-1 and bla CTX-M genes and the qnr determinant in Enterobacter cloacae , Klebsiella pneumoniae and Klebsiella oxytoca isolates», *Journal of Antimicrobial Chemotherapy*, 65(January), pp. 661-665. doi: 10.1093/jac/dkp504.

Shariat, N. *et al.* (2013) «Subtyping of Salmonella enterica serovar Newport outbreak isolates by CRISPR-MVLST and determination of the relationship between CRISPR-MVLST and PFGE results.», *Journal of clinical microbiology*. American Society for Microbiology (ASM), 51(7), pp. 2328-36. doi: 10.1128/JCM.00608-13.

Shimodaira, H. (2002) «An Approximately Unbiased Test of Phylogenetic Tree Selection», *Systematic Biology*. Editado por N. Goldman. Oxford University Press, 51(3), pp. 492-508. doi: 10.1080/10635150290069913.

Shimodaira, H. y Hasegawa, M. (1999) «Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference», *Molecular Biology and Evolution*. Oxford University Press, 16(8), pp. 1114-1116. doi: 10.1093/oxfordjournals.molbev.a026201.

Sievers, F. *et al.* (2011) «Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.», *Molecular systems biology*. EMBO Press, 7(1), p. 539. doi: 10.1038/msb.2011.75.

Skariyachan, S. *et al.* (2018) «Recent perspectives on the molecular basis of biofilm formation by Pseudomonas aeruginosa and approaches for treatment and biofilm dispersal», *Folia Microbiologica*. Folia Microbiologica, 63(4), pp. 413-432. doi: 10.1007/s12223-018-0585-4.

Snyder, L. A. *et al.* (2013) «Epidemiological investigation of Pseudomonas aeruginosa isolates from a six-year-long hospital outbreak using high-throughput whole genome sequencing», *Euro Surveill*. European Centre for Disease Prevention and Control (ECDC) - Health Comunication Unit, 18(17), p. pii=20611. doi: 10.2807/1560-7917.ES2013.18.42.20611.

Solé, M. *et al.* (2011) «First description of an Escherichia coli strain producing NDM-1 carbapenemase in Spain.», *Antimicrobial agents and chemotherapy*, 55(9), pp. 4402-4. doi: 10.1128/AAC.00642-11.

Sousa, A. y Pereira, M. (2014) «Pseudomonas aeruginosa Diversification during Infection Development in Cystic Fibrosis Lungs—A Review», *Pathogens*. Multidisciplinary Digital Publishing Institute, 3(3), pp. 680-703. doi: 10.3390/pathogens3030680.

Steemers, F. J. y Gunderson, K. L. (2005) «Illumina, Inc.», *Pharmacogenomics*, 6(7), pp. 777-782. doi: 10.2217/14622416.6.7.777.

Strimmer, K. y von Haeseler, A. (1997) «Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment.», *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 94(13), pp. 6815-9. Disponible en: http://www.ncbi.nlm.nih.gov/pubmed/9192648 (Accedido: 10 de octubre de 2018).

Strimmer, K. y Rambaut, A. (2002) «Inferring confidence sets of possibly misspecified

gene trees.», *Proceedings. Biological sciences*. The Royal Society, 269(1487), pp. 137-42. doi: 10.1098/rspb.2001.1862.

Viedma, E. *et al.* (2013) «Draft Genome Sequence of VIM-2-Producing Multidrug-Resistant Pseudomonas aeruginosa ST175 , an Epidemic High-Risk Clone», 1(2), pp. 12-13. doi: 10.1128/genomeA.00112-13.Copyright.

Wiedenheft, B. y Bondy-Denomy, J. (2017) «CRISPR control of virulence in Pseudomonas aeruginosa», *Nature Publishing Group*, 27. doi: 10.1038/cr.2017.6.

Winstanley, C. *et al.* (2009) «Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the liverpool epidemic strain of pseudomonas aeruginosa», *Genome Research*, 19(1), pp. 12-23. doi: 10.1101/gr.086082.108.

Witney, A. A. *et al.* (2014) «Genome sequencing and characterization of an extensively drug-resistant sequence type 111 serotype O12 hospital outbreak strain of Pseudomonas aeruginosa», *Clinical Microbiology and Infection*. Elsevier, 20(10), pp. O609-O618. doi: 10.1111/1469-0691.12528.

Wood, D. E. y Salzberg, S. L. (2014) «Kraken: ultrafast metagenomic sequence classification using exact alignments», *Genome Biology*. BioMed Central, 15(3), p. R46. doi: 10.1186/gb-2014-15-3-r46.