

Aproximación genómica y fenotípica al estudio del papel de la cápsula en la virulencia de *Vibrio vulnificus*

Rubén Salvador Clavell

Máster Universitario de Bioinformática y Bioestadística

Área 3 – Evolución Molecular

Directora externa UV: Carmen Amaro González

Consultora UOC: Dorcas J. Orengo Ferriz

Profesor responsable de la asignatura: Laura Calvet Liñan

junio de 2022



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL DE MÁSTER

Título del trabajo:	<i>Aproximación genómica y fenotípica al estudio del papel de la cápsula en la virulencia de Vibrio vulnificus</i>
Nombre del autor:	<i>Rubén Salvador Clavell</i>
Nombre del consultor/a:	<i>Dorcas J. Orengo Ferriz</i>
Nombre del PRA:	<i>Laura Calvet Liñan</i>
Fecha de entrega:	<i>06/2022</i>
Titulación:	<i>Máster Universitario de Bioinformática y Bioestadística</i>
Área del Trabajo Final:	<i>Área 3 – Evolución Molecular</i>
Idioma del trabajo:	<i>Castellano</i>
Número de créditos:	<i>15</i>
Palabras clave	<i>Vibrio vulnificus, cápsula, genómica comparada</i>

Resumen del Trabajo:

Vibrio vulnificus es un patógeno zoonótico que provoca la muerte del hospedador por septicemia gracias a la producción de una cápsula protectora. En esta especie se ha descrito que puede darse la reversión del fenotipo capsulado (colonias opacas, O) a no capsulado (colonias translúcidas, T) y viceversa, con una frecuencia muy baja. Desentrañar las bases genéticas de la reversión permitiría no solo entender la patogénesis de esta bacteria sino también diseñar estrategias de control. El objetivo de este TFM fue dilucidar este mecanismo mediante una aproximación genómica combinada con una aproximación fenotípica usando una cepa clínica no caracterizada. Para ello, se secuenciaron y compararon los genomas de las variantes O y T de esta cepa y se realizaron ensayos para relacionar la morfología colonial con los fenotipos predichos en la especie. Los resultados obtenidos sugieren que la cepa estudiada pertenece a una nueva serovariedad dentro del Linaje 1 en la que la reversión del fenotipo de O a T y viceversa se debe a fenómenos de regulación aún por determinar y no a cambios genéticos reversibles. En este nuevo serotipo, la producción de la cápsula también confiere capacidad septicémica pero, a diferencia de los descritos, la frecuencia con la que revierte la variante T a O aumenta con los pases en medio de cultivo por lo que se hipotetiza que la reversión se debe a un fenómeno de regulación transcripcional relacionado con la adaptación al medio. Finalmente, el genoma de ambas variantes presenta un inovirus cuya caracterización merece futuros estudios.

Abstract:

Vibrio vulnificus is a zoonotic pathogen that causes the death of its host by septicemia due to the production of a protective capsule. Reversion from capsulated (opaque colonies, O) to non-capsulated (translucent colonies, T) phenotype and vice versa has been described in this species but with a very low frequency. Unravelling the genetic basis of the reversion would allow not only to understand the pathogenesis of this bacterium but also to design control strategies. The aim of this TFM was to elucidate this mechanism by a genomic approach combined with a phenotypic approach using an uncharacterized clinical strain. For this purpose, the genomes of the O and T variants of this strain were sequenced and compared and, at the same time, assays were performed to relate colonial morphology to predicted phenotypes in the species. The results obtained suggest that the strain studied belongs to a new serovar within Lineage 1 in which the reversion of the phenotype from O to T and vice versa is due to regulatory phenomena yet to be determined and not to reversible genetic changes. In this new serotype, the production of the capsule also confers septicemic capacity but, unlike those described, the reversion from T to O increases with the passages in culture medium suggesting that the reversion is due to a transcriptional regulation phenomenon related to adaptation to the environment. Finally, the genome of both variants presents an inovirus whose characterization deserves future studies.

Índice

Listado de figuras	8
Listado de tablas.....	11
1. Resumen.....	12
2. Introducción	14
2.1. Contexto y justificación del Trabajo.....	14
2.2. Objetivos del Trabajo	15
2.2.1. Objetivos generales.....	15
2.2.2. Objetivos específicos	15
2.3. Enfoque y método seguido	16
2.3.1. Metodología <i>in silico</i>	16
2.3.2. Metodologías <i>in vitro</i>	17
2.4. Planificación del Trabajo	18
2.4.1. Tareas	18
2.4.2. Hitos.....	18
2.4.3. Análisis de riesgos.....	19
2.4.4. Calendario.....	20
2.5. Breve resumen de contribuciones y productos obtenidos.....	21
2.6. Breve descripción de los otros capítulos de la memoria	21
3. Estado del arte.....	22
3.1. Ensamblaje.....	22
3.2. Mapeo.....	22
3.3. Detección de profagos.....	23
4. Metodología.....	24
4.1. Obtención de las muestras a analizar	24
4.1.1. Cultivo bacteriano.....	24
4.1.2. Obtención del cepario	24
4.1.3. Extracción, purificación de ADN y secuenciación	25
4.2. Metodología <i>in silico</i>	25
4.2.1. Filtrado de las lecturas, ensamblaje primario y visualización.....	25
4.2.2. Ensamblaje híbrido.....	27
4.2.3. Mapeo.....	28
4.2.4. Anotación de los genomas.....	28
4.2.5. Determinación del linaje filogenético de la cepa CECT9109.....	28
4.2.6. Identificación de los loci <i>cps</i>	28

4.2.7.	Identificación y análisis de la presencia de profagos.....	29
4.3.	Metodología <i>in vitro</i>	30
4.3.1.	Extracción rutinaria de ADN, PCR y electroforesis	30
4.3.2.	Tinción capsular.....	30
4.3.3.	Microscopía electrónica de transmisión (TEM).....	30
4.3.4.	Cálculo de la frecuencia de reversión capsular	31
4.3.5.	Ensayo de resistencia al suero humano con/sin hierro.....	32
5.	Resultados.....	33
5.1.	Resultados <i>in silico</i>	33
5.1.1.	Secuenciación por Illumina <i>MiSeq</i> , filtrado y ensamblaje primario	33
5.1.2.	Secuenciación por PacBio, filtrado y ensamblaje primario	35
5.1.3.	Ensamblaje híbrido.....	37
5.1.4.	Mapeo de los genomas	39
5.1.5.	Anotación de los genomas.....	40
5.1.6.	Detección del linaje filogenético	40
5.1.7.	Identificación de los loci <i>cps</i>	40
5.1.8.	Identificación de la presencia de profagos	44
5.2.	Resultados <i>in vitro</i>	45
5.2.1.	Clasificación de la cepa a nivel de patovariedad.....	45
5.2.2.	Observación de la cápsula.....	45
5.2.3.	Cálculo de la frecuencia de reversión capsular	46
5.2.4.	Análisis de resistencia al suero humano con/sin hierro.....	49
6.	Discusión	50
7.	Conclusiones.....	54
7.1.	Conclusiones.....	54
7.2.	Líneas de futuro.....	55
7.3.	Seguimiento de la planificación.....	56
8.	Glosario.....	57
9.	Bibliografía.....	58
10.	Anexos	62
10.1.	Anexo 1: Comandos (Ubuntu).....	62
10.2.	Anexo 2: Medios de cultivo	65
10.3.	Anexo 3: PCR y electroforesis.....	66
10.4.	Anexo 4: Imágenes de resultados	68

Listado de figuras

- Figura 1. Diagrama de Gantt de las tareas a desarrollar en este TFM..... 20
- Figura 2. Imagen ejemplo del cultivo de la cepa CECT4999 en medio sólido mediante triple estría tras 24 h de cultivo a 28 °C..... 24
- Figura 3. Imágenes del análisis de calidad mediante FastQC del entorno Galaxy para las lecturas de Illumina MiSeq de la cepa CECT9109(O) (A y C: forward; B y D: reverse) previo filtrado con PrinSeq (A y B) y tras filtrado (C y D). Se obtuvieron 1.125.928 lecturas, con longitud entre 35 y 151 pb en el caso de las lecturas crudas (A y B), mientras que tras el filtrado quedaron 1.123.256 lecturas, con longitud entre 50 y 151 pb (C y D)..... 33
- Figura 4. Imágenes del análisis de calidad mediante FastQC del entorno Galaxy para las lecturas de Illumina MiSeq de la cepa CECT9109(T) (A y C: forward; B y D: reverse) previo filtrado con PrinSeq (A y B) y tras filtrado (C y D). Se obtuvieron 703.278 lecturas, con longitud entre 35 y 301 pb en el caso de las lecturas crudas (A y B), mientras que tras el filtrado quedaron 585.226 lecturas, con longitud entre 50 y 301 pb (C y D)..... 34
- Figura 5. Ensamblado de las lecturas obtenidas mediante Illumina MiSeq de las variantes O (A y C) y T (B y D) de la cepa CECT9109. El ensamblaje de las lecturas filtradas se realizó mediante SPAdes (A y B) y Unicycler (C y D), obteniéndose los grafos (representación visual) con la herramienta informática Bandage. 35
- Figura 6. Gráfico de calidad obtenido mediante NanoPlot de las lecturas Hi-Fi de PacBio de la cepa CECT9109(O). Se obtuvieron 93.423 lecturas, con longitud entre 512 y 29.089 pb. 36
- Figura 7. Imagen del ensamblado del genoma de la cepa CECT9109(O), obtenido mediante la herramienta informática Flye, utilizando el fichero FASTQ con las lecturas Hi-Fi de BacBio, y representado con Bandage. Contig 1 en morado, y contig 2 en turquesa..... 37
- Figura 8. Imagen del ensamblado del genoma cerrado de la cepa CECT9109(O), obtenido mediante la herramienta informática Unicycler, empleando solamente los ficheros FASTQ de Illumina y de PacBio (Hi-Fi), y representado con Bandage. Contig 1 en morado, y contig 2 en verde..... 38
- Figura 9. Imagen del ensamblado del genoma cerrado de la cepa CECT9109(T) a partir del mapeo de las lecturas filtradas de Illumina (archivos FASTQ) con el genoma cerrado de la cepa CECT9109(O). La imagen obtenida con Bandage (A) muestra los dos cromosomas circularizados. Con el mapeo se obtuvo el llamado de variantes tras filtrado por calidad de cambio nucleotídico (B), donde se aprecian en total 6 diferencias en el genoma de las dos variantes de la cepa CECT9109..... 39
- Figura 10. Imágenes de los ficheros GenBank de la anotación mediante la herramienta informática Prokka de las variantes O (A) y T (B) de la cepa en estudio. Se muestra el primer gen en ambos genomas (dnaA), no observándose diferencias en este caso, ya que por llamada de variantes no se observó ninguna diferencia de calidad en esta primera región de los genomas. 40
- Figura 11. Operón cps (codificador de la cápsula) descrito en la cepa MO6-24/O (A, obtenido de Chatzidakis-Livanis et al. 7) y representado con Bandage (B, genoma obtenido des NCBI Assembly ASM18658v1 [GenBank assembly accession: GCA_000186585.1]). También se localizó esta región cps en el genoma cerrado de las variantes O (C) y T (D) de la cepa CECT9109, con Bandage. 41
- Figura 12. Región del operón cps con 1.000 nucleótidos agua-arriba y aguas-abajo. Los genes que forman este operón se indican con diferentes colores, correspondiéndose con los genes descritos en la Figura 11. Cepas MO6-24/O (A), CECT9109(O) (B) y CECT9109(T) (C). 42
- Figura 13. Región del operón cps, con los diferentes genes indicados mediante el programa de anotación Prokka, para las cepas MO6-24/O (A), CECT9109(O) (B) y CECT9109(T) (C). 42

Figura 14. Imagen obtenida con el programa informático Easyfig, donde se representan los genes de interés en la región del operón cps, viendo similitudes entre los tres genomas de estudio (A: MO6-24/O; B: CECT9109(O), y C: CECT9109(T)) mediante colorimetría. Se observan en morado los genes descritos en la figura anterior, y en gris aquellos genes de las variantes de la cepa CECT9109 que no son idénticos con los de la cepa de referencia. El color azul de la leyenda indica el grado similitud entre las secuencias analizadas; el color rojo indica el grado de inversión, que en este caso no se presenta.	43
Figura 15. Resultados de la PCR dirigida al gen marcador de la pv. piscis (vep07). Se tomó como control positivo la cepa R99.	45
Figura 16. Imágenes de las variantes O (A) y T (B) de la cepa CECT9109 teñidas con nigrosina.	45
Figura 17. Imágenes tomadas por microscopía electrónica de transmisión de las variantes O (A) y T (B) de la cepa CECT9109. La flecha verde indica las células sin cápsula y la naranja las células con cápsula. Barra de escala = 500 nm.	46
Figura 18. Imagen tomada de una placa cultivada en césped (8º día de experimento) de la cepa CECT9109(T). Las flechas verdes indican la presencia de colonias traslúcidas, mientras que las flechas naranjas indican las colonias opacas.	46
Figura 19. Imagen tomada de una placa cultivada en césped (3º día de experimento) de la cepa CECT9109(O). La flecha naranja indica la fracción opaca de la colonia, mientras que la flecha verde indica la fracción traslúcida.	47
Figura 20. Variante rugosa (R) de la cepa CECT9109 cultivada en placa (A) y en cultivo líquido (B). Se observa el aspecto granuloso (rugoso) del cultivo de la variante R (parte derecha del panel A) en comparación con la variante O (parte izquierda del panel A). La flecha rosa (B) indica el biofilm formado tras 24 h de cultivo de la cepa CECT9109(R) en agitación, característico de este tipo de variantes.	47
Figura 21. Identificación de los revertantes por PCR dirigida al gen marcador de especie vvhA. Los revertantes se nombraron como RSC seguido de un número y “tr” (si revertió de O a T), “rg” (si fue de T a R) y “op” (para el caso de T a O). La muestra denominada “biofilm” consistió en el análisis de este gen marcador en una muestra del biofilm formado tras 24 h de cultivo en agitación del revertante RSC7rg. La presencia de este gen corrobora que todas las muestras analizadas pertenecen a la especie de estudio y que no hubo contaminación en las placas y cepas de reversión.	48
Figura 22. Árbol filogenómico NJ con 81 cepas ya analizadas previamente ⁴³ junto con las variantes O y T de la cepa CECT9109, obtenido mediante PARSNP y visualizado con iTOL. Se tomó como cepa referencia la YJ016. Se observa que ambas variantes de la cepa de estudio pertenecen al linaje 1.	68
Figura 23. Imagen obtenida con Clustal Omega en el que se comparan las secuencias de aminoácidos de la proteína resultante del gen analizado en el que se ha encontrado la diferencia entre ambas las variantes O y T de la cepa CECT9109 con el llamado de variantes. La flecha negra indica el aminoácido resultante diferente en la variante T. Se observa una elevada similitud entre la secuencia de aminoácidos descrita para este gen y las secuencias obtenidas mediante la anotación con Prokka para los genomas estudiados.	69
Figura 24. Ensayo de la resistencia al suero de las variantes O y T de la cepa CECT9109; las imágenes muestran el cultivo de estas bacterias en PBS 1X a modo de control del experimento. Estas imágenes corresponden a placas sembradas por sectores con muestras cultivadas en PBS a tiempo 0 (A) y tras 4 h de incubación (B). En cada sector se sembraron 3 gotas de 10 µL correspondientes a diluciones decimales diferentes (10 ⁰ en el caso de la siembra directa a partir del cultivo en PBS, y 10 ⁻¹ y 10 ⁻² a partir de dicho cultivo).	70
Figura 25. Ensayo de la resistencia al suero de la cepa CECT9109(O). Imágenes a tiempo cero de cultivo (A y B) y tras 4 h de cultivo a 37 °C (C y D), cultivadas en suero humano (HS), suero humano suplementado con hierro (HS + Fe), suero humano inactivado (HS in) y suero humano	

inactivado suplementado con hierro (HS in + Fe). Las imágenes de los paneles A y C se corresponden con las bacterias provenientes de una colonia aislada en medio sólido, mientras que los paneles B y D se corresponden con las bacterias provenientes de medio líquido (10 µL). En cada sector se sembraron 3 gotas de 10 µL correspondientes a diluciones decimales diferentes (10^0 en el caso de la siembra directa a partir del cultivo en los diferentes sueros, y 10^{-1} y 10^{-2} a partir de dichos cultivos)..... 70

Figura 26. Ensayo de la resistencia al suero de la cepa CECT9109(T). Imágenes a tiempo cero de cultivo (A y B) y tras 4 h de cultivo a 37 °C (C y D), cultivadas en suero humano (HS), suero humano suplementado con hierro (HS + Fe), suero humano inactivado (HS in) y suero humano inactivado suplementado con hierro (HS in + Fe). Las imágenes de los paneles A y C se corresponden con las bacterias provenientes de una colonia aislada en medio sólido, mientras que los paneles B y D se corresponden con las bacterias provenientes de medio líquido (10 µL). En cada sector se sembraron 3 gotas de 10 µL correspondientes a diluciones decimales diferentes (10^0 en el caso de la siembra directa a partir del cultivo en los diferentes sueros, y 10^{-1} y 10^{-2} a partir de dichos cultivos)..... 71

Listado de tablas

Tabla 1. Tareas planificadas junto con su duración en días para la realización del TFM.	18
Tabla 2. Resultado de la comparativa entre los diferentes ensambladores (HGAP4 y Flye) y la herramienta de scaffolding (Quickmerge) para el ensamblado de las lecturas largas Hi-Fi de PacBio.	36
Tabla 3. Resultado de la comparativa entre las diferentes técnicas de ensamblaje híbrido, combinando las lecturas cortas filtradas de Illumina y las lecturas largas Hi-Fi de PacBio. En naranja se muestran los ensamblados todos los ficheros FASTQ y el fichero FASTA correspondiente, y en azul, los ensamblados con los ficheros FASTQ de Illumina y el fichero FASTA correspondiente.	38
Tabla 4. Comparativa de los diferentes buscadores de profagos en las secuencias de los genomas de las variantes O y T de la cepa CECT9109.	44
Tabla 5. Frecuencia de reversión de la variante T a la O durante los 10 días de experimento. Las unidades representadas son nº colonias revertidas/volumen de células sembradas en placa (10.000 ufc*/placa). (*ufc: unidades formadoras de colonia)	47
Tabla 6. Ensayo de resistencia al suero de las variantes O y T de la cepa CECT9109, tanto a partir de colonia aislada en medio sólido como con 10 µL de cultivo en medio líquido. Se muestran los valores obtenidos del recuento celular al momento inicial del ensayo (t = 0 h) y tras 4 y 24 h, tanto en PBS 1X (control de no toxicidad celular ni de resistencia), en suero humano (HS), suero humano suplementado con hierro (HS + Fe), suero humano inactivado (HS in) y suero humano inactivado suplementado con hierro (HS in + Fe). Los valores representan el total de células en el vial de cultivo.	49
Tabla 7. Listado de componentes necesarios para la amplificación mediante PCR de los genes vvha y vep07 para la identificación de V. vulnificus y de la pv. piscis, respectivamente.	66
Tabla 8. Programa para la ejecución de la PCR.	67

1. Resumen

V. vulnificus es un patógeno que habita zonas cálidas y que, debido al cambio climático, se está extendiendo a zonas más frías. Esta especie puede causar muerte rápida por septicemia en distintas especies de animales marinos (principalmente peces de piscifactoría) y en el ser humano tanto por ingestión como por contacto. Además, se comporta como un patógeno zoonótico porque la vibriosis animal puede transmitirse directamente del animal enfermo al ser humano por contacto.

En esta especie se ha descrito que la producción de una cápsula por fuera de la membrana externa protege a la bacteria de la inmunidad innata en sangre y que esta producción está sometida a un fenómeno de reversión (producción vs no producción) de base genética desconocida. Hay que destacar que las células productoras de cápsula desarrollan colonias opacas (O) y las no productoras translúcidas (T) y que en el laboratorio se puede dar la reversión de O a T y viceversa con una frecuencia inferior a 10^{-5} . Desentrañar las bases genéticas de la reversión permitiría no solo entender la patogénesis de esta bacteria sino también diseñar estrategias de control.

En consecuencia, el objetivo de este TFM fue dilucidar la base genética de la reversión bajo la hipótesis de que se debe a cambios genéticos reversibles que afectan al operón biosintético de la cápsula (*cps*) usando una aproximación genómica combinada con una aproximación fenotípica. Para ello, se secuenciaron y compararon los genomas de las variantes O y T de una cepa clínica no caracterizada y, al mismo tiempo, se realizaron ensayos para relacionar la morfología colonial con los fenotipos predichos en la especie. En primer lugar, se determinó cuál era la mejor técnica de ensamblado de las lecturas fruto de la secuenciación, a continuación, se mapeó el genoma de la variante T frente al genoma cerrado de la variante O, y finalmente se anotaron ambos genomas. Una vez ensamblados los genomas, se posicionó la cepa objeto de estudio en el árbol filogenómico de la especie constatando que pertenece al Linaje 1, descrito como el de mayor relevancia en salud pública. A continuación, se comparó el operón *cps* y las zonas aguas-arriba y abajo de este en los genomas O y T de la cepa seleccionada usando como referencia el genoma de la una cepa de serotipo CPS-1, el serotipo en el que se ha estudiado el fenómeno de la reversión. No se identificó ningún elemento genético móvil ni tampoco cambios de secuencia significativos entre las variantes O y T y solo se detectó un único cambio nucleotídico no sinónimo en el gen *wlaL* que difícilmente podría explicar la reversión. Las regiones *cps* de la cepa objeto de estudio difirieron de la utilizada como referencia lo que sugiere que la cepa seleccionada no pertenece al serotipo CPS-1. Finalmente, con varias herramientas de búsqueda de virus (o fagos), se encontró la secuencia de un inovirus en los

genomas de ambas variantes y en la misma posición, por lo que tampoco se pudo relacionar con el fenómeno de la reversión.

En paralelo, se analizó *in vitro* la reversión capsular en este nuevo serotipo. Primero, se constató que las células de las colonias O estaban rodeadas de cápsula al microscopio electrónico de transmisión, al observar una capa altamente electrodensa por fuera de la membrana externa. A continuación, se determinó la frecuencia de reversión de las variantes O a T y T a O y se constató que se producía la reversión capsular del fenotipo T a O con una frecuencia que aumentaba a lo largo de los pases en medios de cultivo, llegando a un 50%. Finalmente, se correlacionó la producción de la cápsula con la resistencia al complemento del suero sanguíneo y, por tanto, con el potencial septicémico.

En conclusión, los resultados obtenidos sugieren que la cepa estudiada pertenece a una nueva serovariedad dentro del Linaje 1 en la que la reversión del fenotipo de O a T y viceversa se debe probablemente a fenómenos de regulación aún por determinar y no a cambios genéticos reversibles. En este nuevo serotipo, la producción de la cápsula también confiere capacidad septicémica, pero, a diferencia de los descritos, la frecuencia con la que revierte la variante T a O aumenta con los pases en medio de cultivo por lo que se puede hipotetizar que la reversión se debe a un fenómeno de regulación transcripcional relacionado con la adaptación al medio. Finalmente, el genoma de ambas variantes presenta un inovirus cuya caracterización merece futuros estudios.

2. Introducción

2.1. Contexto y justificación del Trabajo

Vibrio vulnificus es una bacteria Gram negativa, anaerobia facultativa y móvil por un flagelo polar. La especie pertenece al género *Vibrio* y habita en ecosistemas de agua salobre localizados en zonas templadas, tropicales y subtropicales. Debido al calentamiento global, esta bacteria se encuentra en expansión a zonas tradicionalmente más frías, como el mar Báltico ¹. Se considera un patógeno multi-hospedador, capaz de causar enfermedad (vibriosis) en distintas especies de animales acuáticos y en el ser humano. Además, se considera un patógeno zoonótico ya que la vibriosis puede transmitirse de animales enfermos al ser humano ².

La vibriosis de los peces se presenta en dos modalidades relacionadas con la ruta de infección: contacto o ingestión. En ambos casos, puede derivar en septicemia y muerte. Cuando infecta por contacto, la bacteria coloniza las mucosas de la piel y las branquias e ingresa directamente en sangre resistiendo el sistema inmunitario y causando la septicemia ³.

La vibriosis humana también se produce por contacto o ingestión. Cuando la bacteria se ingiere con marisco o pescado crudo o poco cocinado, causa gastroenteritis y cuando infecta por contacto, causa graves lesiones en la epidermis y la dermis que pueden requerir intervención quirúrgica. Independientemente de la ruta de infección, el patógeno puede pasar a la sangre y la enfermedad progresar a septicemia y muerte en menos de 48 horas si el paciente presenta elevados niveles de hierro en sangre debido a patologías subyacentes como la hemocromatosis o la cirrosis ⁴.

En términos generales, la virulencia de *V. vulnificus* es multifactorial dependiendo tanto de la producción de múltiples toxinas y exoenzimas, que dañan los tejidos, como de una envoltura protectora rica en cápsula y en antígeno O de cadena larga ⁵. La cápsula, en concreto, destaca porque protege a la bacteria del sistema inmunitario innato del ser humano y su expresión es óptima a altas concentraciones de hierro ⁶. Así, solo las cepas que producen cápsula son capaces de resistir el sistema inmunitario y causar septicemia en el ser humano.

Desde hace tiempo se sabe que la producción de cápsula es reversible, de manera que variantes capsuladas de la misma cepa pueden revertir a no capsuladas y viceversa ⁷. Ambas variantes pueden reconocerse en los medios de cultivo sólidos porque la capsulada desarrolla colonias opacas a la luz (O) y la no capsulada, translúcidas (T). No obstante, a pesar de la importancia de la cápsula en la virulencia de la especie, el mecanismo molecular de reversión se desconoce. Nuestra hipótesis, basada en los estudios realizados en el serotipo capsular CPS-1, es que esta reversión se debe a cambios genéticos reversibles,

por lo que analizando los genomas de las variantes O y T de una misma cepa, esos cambios podrían identificarse y el mecanismo dilucidarse.

Por todo ello, es relevante realizar un estudio tanto genómico como fenotípico que clarifique el mecanismo de reversión de fenotipo capsulado a no capsulado y viceversa y su papel en la virulencia de esta especie de *Vibrio*. Con ello, podrían diseñarse métodos de control de la actividad virulenta de estas bacterias, impidiendo que la invasión en los seres vivos que infectan pueda llegar a causar vibriosis y, en definitiva, septicemia.

2.2. Objetivos del Trabajo

2.2.1. Objetivos generales

1. Identificar el mecanismo molecular de inducción de la formación/pérdida de la cápsula a partir de la información genómica de la cepa CECT9109.
2. Identificar las diferencias fenotípicas en cuanto a la formación/pérdida de la cápsula en las variantes opaca/translúcida de la cepa CECT9109.

2.2.2. Objetivos específicos

1. Identificar el mecanismo molecular de inducción de la formación/pérdida de la cápsula a partir de la información genómica de la cepa CECT9109.
 - 1.1. Clasificar la cepa a nivel de patovar mediante PCR.
 - 1.2. Secuenciar los genomas de las variantes O y T de la cepa CECT9109; realizar el ensamblaje híbrido *de novo* y el mapeo. Anotar los genomas.
 - 1.3. Determinar el linaje filogenético de la cepa en estudio.
 - 1.4. Identificar los loci *cps* (codificantes de la cápsula) mediante llamado de variantes (detectar posibles InDel o SNP).
 - 1.5. Identificar *in silico* la presencia de profagos.
2. Identificar las diferencias fenotípicas entre las variantes O y T de la cepa CECT9109.
 - 2.1. Observar la cápsula al microscopio óptico mediante tinción de cápsulas y al microscopio electrónico de transmisión (TEM).

- 2.2. Calcular la frecuencia de la reversión O-T y T-O en medio basal.
- 2.3. Relacionar la producción de cápsula con la resistencia al suero humano con/sin hierro.

2.3. Enfoque y método seguido

El objetivo principal del presente Trabajo Fin de Máster (TFM) era averiguar cuál es la base genética del fenómeno de la reversión de fenotipo virulento (capsulado) a no virulento (no capsulado) y viceversa en el patógeno zoonótico *V. vulnificus* usando herramientas genómicas y bioinformáticas. Esta aproximación constituye una novedad puesto que no existe ningún trabajo parecido en esta y en otras especies bacterianas que sufren el fenómeno de la reversión. Este objetivo entraña una dificultad añadida y es que esta especie, a diferencia de la mayoría de las bacterias, posee dos cromosomas con distinta información genética en vez de uno, lo que dificulta el ensamblado de los genomas.

El laboratorio de la Dra. Amaro dispone de cientos de cepas de la especie *V. vulnificus* de entre las cuales se eligió la cepa CECT9109 para el estudio y la cepa CECT4999 (también llamada R99) como control. La cepa CECT9109 fue aislada en Suecia a partir de un caso de septicemia primaria humana tras consumo de ostra y se escogió porque reversionó de capsulada a no capsulada de forma espontánea en el laboratorio. La cepa R99 se aisló en España de anguila enferma, pertenece al Linaje 2, a la patovariedad (pv.) *piscis*, (engloba las cepas virulentas para peces), y al clado zoonótico (agrupa las cepas de la patovariedad que, además, son zoonóticas). Se desconoce tanto el linaje al que pertenece la cepa CECT9109 como si se trata de una cepa de la pv. *piscis*.

2.3.1. Metodología *in silico*

Las secuencias genómicas de la cepa CECT9109 se obtuvieron usando las tecnologías de ultrasecuenciación Illumina *MiSeq*⁸ y *Pacific Bioscience* o PacBio⁹. El uso combinado de ambas tecnologías es muy común ya que la tecnología Illumina permite obtener lecturas de secuencias (*reads*) cortas (200 pb) de alta calidad (pocos errores nucleotídicos) y la PacBio, *reads* muy largos (más de 10 kpb) aunque con un mayor número de errores, por lo que ambas tecnologías se complementan. Normalmente, se realiza una primera aproximación mediante Illumina y una segunda con PacBio y, así, se ensamblan los *contigs* para conseguir un genoma cerrado, en este caso, cada cromosoma en un solo *contig*. En otras palabras, la combinación de estas dos tecnologías ofrece un resultado de alta calidad y precisión.

Una vez obtenidas las secuencias por ambas tecnologías, se analizó la calidad de las lecturas mediante programas informáticos como *FastQC* y se filtró mediante *PrinSeq*¹⁰. A continuación, el ensamblaje *de novo* de estas lecturas y el ensamblaje híbrido (cadenas cortas y largas) se realizó usando diferentes programas informáticos, como son *SPAdes genome assembler*¹¹ y *Unicycler*¹², entre otros, y comparando los resultados obtenidos.

Para la obtención del árbol filogenético se utilizó el modelo *neighbour-joining* (NJ) con el programa *PARSNP*¹³ y para la visualización del árbol la herramienta *iTOL*.

Las comparaciones entre las secuencias de ambas variantes se realizaron usando *BLASTn* y se representaron gráficamente usando la herramienta informática *Bandage*¹⁴. Finalmente, el análisis de regiones *cps* y de profagos se llevó a cabo usando otras herramientas informáticas como son *Clustal Omega* y *Phaster*, entre otras¹⁵.

2.3.2. Metodologías *in vitro*

En primer lugar, se procedió a determinar si la cepa objeto de estudio pertenecía a la pv. *piscis* para lo que se realizó una amplificación de ADN por reacción en cadena de la polimerasa (PCR) dirigida al gen plasmídico *vep07*, gen que codifica una lipoproteína de membrana externa que confiere resistencia a la fagocitosis y al complemento (conjunto de proteínas del suero sanguíneo que es capaz de lisis bacterias) de los peces^{16,17}. La cepa R99 se usó como control positivo.

En segundo lugar se procedió a la visualización de la cápsula por lo que se realizaron preparaciones de cultivos bacterianos que se procesaron para su observación mediante microscopía electrónica de transmisión (TEM)¹⁸ usando el servicio de Microscopía de los servicios generales de la Universitat de València (UV). Esta técnica es muy utilizada para observar estructuras biológicas micro y nanoscópicas. Además, también se realizaron tinciones negativas en portaobjetos que permiten la visualización de ciertas cápsulas con un microscopio óptico convencional¹⁹.

Finalmente, para determinar el potencial septicémico de cada variante de la cepa CECT9109 se hicieron crecer en suero humano comercial, adicionado o no con FeCl_3 10 μM y se calculó el porcentaje de supervivencia tras 4 y 24 h de incubación a 37 °C²⁰. La adición de hierro al medio es esencial puesto que favorece la formación de la cápsula por la bacteria y, en consecuencia, su resistencia al suero.

2.4. Planificación del Trabajo

2.4.1. Tareas

Tabla 1. Tareas planificadas junto con su duración en días para la realización del TFM.

Tarea	Duración (días)
Obtención del cepario (cepas de estudio y controles)	7
Cultivo de las cepas, y extracción y purificación del ADN	7
Análisis de las cepas a nivel de patovar mediante PCR	2
Secuenciación de los genomas de las cepas de estudio	19
Filtraje, ensamblaje híbrido y mapeo	8
Anotación de los genomas	7
Determinación del linaje filogenético de la cepa CECT9109	15
Identificación de los loci <i>cps</i>	21
Identificación de la presencia de profagos	17
Observación de la cápsula al TEM	2
Tinción de la cápsula	3
Cálculo de la frecuencia de reversión capsular	10
Relación de la cápsula con la resistencia al suero humano	3
Redacción de la memoria	13
Elaboración de la presentación	6

2.4.2. Hitos

El primero fue la secuenciación de los genomas de las variantes O y T de la cepa en estudio y su análisis *in silico*, el cual ha comprendido la mayor parte del tiempo del presente TFM. Una vez los genomas estuvieron ensamblados y mapeados, se destinó gran parte del tiempo a la identificación de los loci *cps* y, en menor medida, al análisis de la presencia de profagos en ambos genomas.

Dado que a simple vista ya se distinguían diferencias en cuanto a la morfología colonial de las cepas en estudio (presencia/ausencia de cápsula), el visionado por microscopía, la tinción e inmunotinción de la cápsula y los experimentos *in vitro*, así como el cálculo de la frecuencia de reversión capsular se pudieron realizar de manera paralela al estudio bioinformático.

Por último, se realizó la escritura de la memoria de este TFM con su posterior presentación para la defensa del mismo.

Así pues, los hitos de este trabajo fueron:

- Obtención de las secuencias de los dos cromosomas de las variantes O y T de la cepa en estudio mediante secuenciación, mapeo y ensamblaje.
- Identificación de las posibles causas de la reversión mediante la búsqueda de diferencias en los *loci* relacionados con la biosíntesis de la cápsula y la identificación de profagos.
- Correlacionar la presencia o ausencia de cápsula con fenotipo

- Escritura de la memoria.
- Elaboración de la presentación.

2.4.3. Análisis de riesgos

El riesgo más importante que se identificó para la realización de este trabajo era la posible falta de tiempo disponible para la consecución de todos los objetivos planteados. Al tratarse de un trabajo sobre todo bioinformático, aunque también con varios puntos prácticos en el laboratorio de experimentación, se previó que existían factores que podían llegar a ralentizar la obtención de los resultados, como eran:

- Infraestructura informática: poca capacidad de rendimiento del ordenador, por lo que los análisis bioinformáticos pueden ir lentos.
 - ✓ Solución: disponer de infraestructura de la propia Universitat de València para el ensamblado y mapeado de los genomas, y para la detección de los *cps*.
- Disponibilidad del servicio de secuenciación contratado, y tiempo de espera hasta la obtención de las secuencias de las cepas de estudio.
 - ✓ Solución: adelantar otros ensayos previstos a realizar más adelante.
- No identificación de diferencias a nivel nucleotídico entre las secuencias de ambas variantes de la cepa de estudio, debido a la posibilidad de que estas diferencias se encontraran a nivel de cantidad de ARN.
 - ✓ Solución: análisis de la expresión diferencial entre ambas variantes mediante *microarrays* y PCR cuantitativa.
- Compra y llegada de los fungibles y reactivos necesarios para la realización de la parte experimental en el laboratorio.
 - ✓ Solución: planificación de los experimentos y compra del material con tiempo de antelación y a comerciales o distribuidores de la región.
- Disponibilidad del servicio de microscopía electrónica de la Universitat de València.
 - ✓ Solución: contratación de los mismos servicios en la Universitat Politècnica de València.

Teniendo en cuenta estos factores de riesgo se consideró que los objetivos específicos 1.1 a 2.2 eran los objetivos primarios (de imperativo cumplimiento para el trabajo) mientras el objetivo específico 2.3 era secundario y solo se realizaría siempre que se dieran los factores de tiempo y recursos necesarios.

2.4.4. Calendario

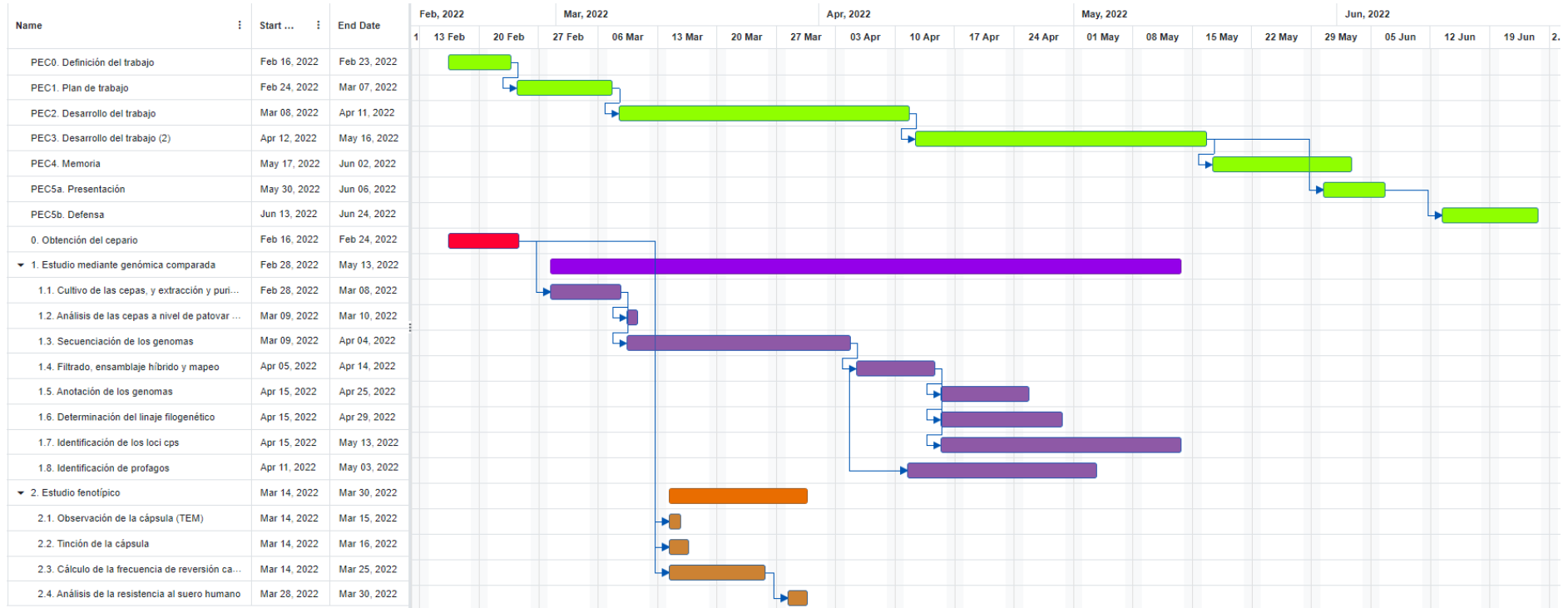


Figura 1. Diagrama de Gantt de las tareas a desarrollar en este TFM.

2.5. Breve resumen de contribuciones y productos obtenidos

Con la realización del TFM se consiguió conocer la secuencia del genoma de las variantes O y T de la cepa CECT9109 de *V. vulnificus* y cerrar los dos cromosomas. A partir de las secuencias, se realizó el análisis bioinformático en búsqueda de los *loci* codificadores de cápsula y de profagos, comparando los resultados obtenidos en ambas variantes. Con ello, se pretendió encontrar diferencias que explicaran el proceso de reversión (i.e., identificación de inversiones, inserciones o fagos).

De manera paralela, se obtuvieron evidencias fenotípicas de la producción de la cápsula observando las cepas al microscopio óptico y electrónico.

Finalmente, se ha realizado una memoria de todo el trabajo realizado, que incluye una introducción del tema tratado, los materiales y la metodología empleada, los resultados obtenidos y las conclusiones de este. A modo de simplificación, se mostrará el trabajo mediante una presentación virtual.

Así pues, hemos obtenido:

A) resultados científicos:

- El genoma cerrado de los dos cromosomas de las dos variantes de la cepa de estudio.
- Evidencias de la causa de la reversión
- Evidencias de que la cápsula es esencial en la cepa CECT9109 para la resistencia al suero y la septicemia

B) resultados entregables:

- El plan de trabajo, documento en el que se definieron los objetivos y calendario a seguir para completar el TFM.
- La memoria de todo el trabajo realizado, introduciendo el tema tratado, los materiales y la metodología empleada, los resultados obtenidos y las conclusiones de este.
- La presentación virtual que a modo de simplificación mostrará el trabajo realizado.

2.6. Breve descripción de los otros capítulos de la memoria

Los siguientes puntos de este TFM constituyen el desarrollo de este. En primer lugar, se presenta *el estado del arte*, es decir, el conocimiento actual sobre las tecnologías de secuenciación y análisis de genomas, así como sobre la importancia de la cápsula en la virulencia de *V. vulnificus* y lo que se conoce sobre el mecanismo de reversión. También se muestra la metodología empleada a lo largo de todo el trabajo, siguiendo con los resultados obtenidos y su análisis crítico. Finalmente, se explican las conclusiones del TFM, junto con la bibliografía y los anexos existentes.

3. Estado del arte

Desde el desarrollo de las nuevas tecnologías de secuenciación (NGS por sus siglas en inglés), se han implementado multitud de herramientas bioinformáticas que ayudan al procesamiento, análisis y comprensión de los datos proporcionados por estos sistemas. El uso de una aplicación informática u otra depende, en gran medida, del tipo de dato a analizar, de la precisión requerida en el estudio y, en definitiva, de la manera de trabajar de la persona.

A continuación, se presenta un listado de parte de estas herramientas informáticas especializadas en las NGS.

3.1. Ensamblaje

En la actualidad, coexisten multitud de herramientas para el ensamblaje de los genomas. *Unicycler*^{12,21} es, por excelencia, una de las más utilizadas, pero está especializada sobre todo en el ensamblaje a partir de lecturas cortas. No obstante, existen otras aplicaciones, como *SPAdes* que también resultan útiles para el ensamblaje con lecturas cortas¹¹.

Para el ensamblaje con lecturas largas, ensambladores como *Canu*, *Flye* o *Shasta*, entre otros, son buenos candidatos^{22,23}; cabe destacar que PacBio posee sus propios ensambladores, como por ejemplo *HGAP*²⁴.

En el caso de *Unicycler*, también puede funcionar como ensamblador híbrido, donde primero genera un grafo de ensamblado a partir de las lecturas cortas de Illumina, y utiliza el alineamiento de las lecturas largas (PacBio) para realizar un *scaffolding* de los *contigs* hasta dar con un genoma completo y cerrado.

Por lo general, estas herramientas informáticas son de fácil uso, pero pueden necesitar un alto tiempo computacional o requisitos concretos de *hardware*.

En definitiva, se puede decir que *Unicycler* ofrece un buen resultado en cuanto a ensamblaje de lecturas de Illumina, pero también para ensamblaje híbrido (combinación de lecturas cortas y largas). Si se dispone de lecturas largas, herramientas como *Flye* o *HGAP* (en caso de disponer del sistema PacBio) pueden ser buenas candidatas.

3.2. Mapeo

Una de las posibilidades para obtener un genoma cerrado sin un coste económico muy elevado consiste en mapear lecturas cortas de Illumina sobre un genoma ya cerrado.

Existen diferentes programas capaces de realizar mapeos de genomas enteros, obteniendo buenos resultados. Estos pueden ser *Bowtie*, *BWA*, *Novoalign*, entre otros ^{25,26}. Aunque *BWA* ofrece resultados buenos incluso con lecturas largas, *Bowtie* presenta un mapeo con mayor precisión y cobertura del alineamiento.

Por otro lado, *Snippy* ²⁷, del gestor de paquetes *Conda*, es otra herramienta de mapeo de código abierto que alinea lecturas cortas sobre genomas cerrados, obteniendo además un fichero de llamado de variantes (VCF), en el que se identifican los cambios producidos en el genoma resultante respecto del genoma de referencia. Este archivo resulta de gran interés cuando se pretende estudiar diferencias entre genomas de diferentes bacterias.

3.3. Detección de profagos

Finalmente, la detección de regiones genómicas pertenecientes a bacteriófagos dentro del genoma estudiado es relevante para este estudio porque la inserción de un profago, o su cambio de posición en el genoma, puede llevar consigo la pérdida de cápsula o su ganancia. La detección *in silico* se puede realizar con muchas herramientas informáticas, más o menos precisas, o que se han desarrollado de manera específica hacia tipos de fagos concretos.

Aplicaciones genéricas pueden ser *Phaster*, *PHISDetector*, *VirSorter* o *Prokaryotic virus Host Predictor (PHP* por sus siglas en inglés) ²⁸⁻³². Estas se basan en bases de datos de fagos descritos en la bibliografía, normalmente detectando regiones parecidas con estas bases de datos.

En cambio, para la identificación de tipos de fagos menos descritos, como por ejemplo los inovirus, existen otras herramientas más especializadas, como es el caso de *Inovirus detector* ³³.

4. Metodología

4.1. Obtención de las muestras a analizar

4.1.1. Cultivo bacteriano

En este TFM se ha empleado la cepa CECT9109 de *V. vulnificus*, cepa de origen clínico que se aisló de un caso de septicemia primaria humana en Suecia tras consumo de ostra. De manera espontánea, esta cepa reversionó de capsulada (O) a no capsulada (T) en un experimento realizado en el laboratorio, por lo que se disponía de cultivos puros de las variantes O y T. Se desconocía al inicio de este TFM a qué linaje filogenético pertenecía esta cepa y si era o no de la pv. *piscis*.

La cepa CECT4999, también llamada R99, es la cepa que se usa como modelo en el laboratorio de la Dra. Amaro por lo que se incluyó como cepa control. Esta cepa pertenece al Linaje 2, a la pv. *piscis* y al clado zoonótico de la especie. Se aisló de anguila enferma en España y es virulenta para ratones (modelo animal de la vibriosis humana).

El Anexo 2 muestra el listado de medios de cultivo empleados y cómo se prepararon.

4.1.2. Obtención del cepario

Una vez preparados los medios de cultivo, las cepas se descongelaron y se cultivaron en 5 mL de medio líquido durante 24 h a 28 °C en agitación (120 rpm). A continuación, se sembraron placas con medio sólido mediante triple estría para la comprobación de la pureza del cultivo y la morfología colonial (Figura 2).



Figura 2. Imagen ejemplo del cultivo de la cepa CECT4999 en medio sólido mediante triple estría tras 24 h de cultivo a 28 °C.

Tras 24 h de cultivo a 28 °C, se resuspendió una colonia de cada cepa/variante en 5 mL de medio líquido y se incubó otras 24 h a 28 °C en agitación. Pasado

este tiempo, las células se criopreservaron con 20 % de glicerol estéril (Panreac, EE.UU.) hasta su uso.

4.1.3. Extracción, purificación de ADN y secuenciación

El siguiente paso consistió en extraer y purificar el ADN a partir de cultivos de 24 h de ambas variantes. La purificación se realizó utilizando micropartículas (*microbeads*) magnéticas (MagAttract HMW DNA; Quiagen, Alemania) según el protocolo descrito por la empresa fabricante y se obtuvo un ADN de gran pureza.

El ADN purificado se secuenció mediante Illumina *MiSeq System* (Illumina, EE.UU.) en los servicios de secuenciación del FISABIO (Generalitat Valenciana) y mediante PacBio (Pacific Biosciences, EE.UU) en los servicios de la Universitat de València. El uso combinado de ambas tecnologías es muy común ya que la tecnología Illumina permite obtener lecturas de secuencias (*reads*) cortas (150-250 pb) de alta calidad (pocos errores nucleotídicos) y la PacBio, lecturas muy largas (más de 10 kpb) aunque con un mayor número de errores, por lo que ambas tecnologías se complementan. Por ello, para conseguir cerrar un genoma, es decir, tener en un único *contig* un cromosoma en concreto, habitualmente se realiza una primera aproximación mediante Illumina y una segunda con PacBio y, así, se ensamblan las lecturas para conseguir el genoma cerrado (un único *contig*). En otras palabras, la combinación de estas dos tecnologías ofrece un resultado de alta calidad y precisión.

4.2. Metodología *in silico*

4.2.1. Filtrado de las lecturas, ensamblaje primario y visualización

Por un lado, se obtuvieron las lecturas de Illumina *MiSeq*⁸ en formato *FASTQ* de las variantes O y T de la cepa CECT9109. Este formato presenta las diferentes lecturas obtenidas, y cada una de ellas consiste en una primera línea que se corresponde con su encabezado iniciado por "@" con la descripción de la lectura, una segunda línea con la secuencia de nucleótidos, una tercera línea con el símbolo "+", y una cuarta línea con una sucesión de caracteres ASCII que se corresponden a la calidad de cada nucleótido en su propia posición.

Con este fichero, se analizó la calidad de las lecturas con *FastQC* del entorno *Galaxy Europe* (<https://usegalaxy.eu/>). *FastQC* tiene como objetivo proporcionar un informe de control de calidad que permite detectar problemas en la secuenciación, debido al secuenciador o al material de la librería realizada.

Tras este primer análisis de la calidad, se decidió filtrar las lecturas mediante *PrinSeq* (<http://prinseq.sourceforge.net>). Esta aplicación permite filtrar,

reformatear y recortar lecturas de secuencias, proporcionando a su vez estadísticos resumidos de los datos de secuencia. Para el filtrado, se decidió seleccionar aquellas lecturas con un mínimo de 50 pb y, además, con una calidad media de 20 puntos. De igual manera, el fichero que proporciona esta herramienta informática también es un *FASTQ* de lecturas con su calidad.

A continuación, se volvió a analizar la calidad de estas lecturas filtradas con *FastQC*, corroborando que el filtrado se había realizado correctamente.

Finalmente, se realizó el ensamblaje primario con los dos archivos (*forward* y *reverse*) de las lecturas filtradas de Illumina. Para ello, se emplearon dos programas informáticos, comparando los resultados: *SPAdes 3.15*¹¹ y *Unicycler 0.5*¹². En ambos casos, se obtuvo un único fichero ahora en formato *FASTA*, el cual presenta los diferentes *contigs* obtenidos tras el ensamblaje. Estos *contigs* presentan un encabezado (iniciado por el símbolo “>”) en la primera línea y la secuencia de nucleótidos del *contig* en la segunda; en este tipo de fichero no se dispone de la calidad de los nucleótidos de la secuencia obtenida.

Para visualizar el resultado del ensamblaje en un gráfico, se utilizaron los ficheros GFA (fichero de grafo) en la herramienta informática *Bandage*¹⁴, con los parámetros por defecto.

Por otro lado, el sistema de secuenciación PacBio⁹ presenta dos tipos de ficheros con las lecturas: un único fichero con las lecturas crudas y otro fichero con las lecturas filtradas (lecturas *Hi-Fi*) por el propio sistema. Tras la visualización de la calidad de las lecturas *Hi-Fi* (CECT9109(O)) mediante el programa *NanoPlot*³⁴ del gestor de paquetes *Conda* (el cual está optimizado para el control de calidad de lecturas largas), se ensamblaron mediante 2 metodologías diferentes: ensamblaje como tal y *scaffolding* (a partir de dos ficheros *FASTA*; esto permite pulir el ensamblado obtenido solo con las lecturas, mejorando la cobertura del ensamblado). Así pues, se utilizaron las siguientes herramientas informáticas:

- *HGAP4*²⁴: ensamblador propio del sistema PacBio. Esta herramienta utiliza el fichero *FASTQ* con las lecturas *Hi-Fi*.
- *Flye*³⁵: utiliza también el fichero *FASTQ* anterior.
- *Quickmerge*³⁶: se trata de una herramienta que realiza *scaffolding* a partir de dos ficheros *FASTA* (en este caso, tras el ensamblaje con *HGAP4* y *Flye*)

Tras estos ensamblajes, se compararon los resultados de calidad obtenidos usando las herramientas *QUAST 5.0.2*³⁷ y *BUSCO 5.3.2*³⁸. Por un lado, la herramienta *QUAST* permite obtener un resumen del ensamblado, en concreto el número de *contigs*, la N50 (longitud de los *contigs* que, usando *contigs* de igual o mayor tamaño, produce la mitad de la longitud del genoma), la L50 (número más pequeño de *contigs* cuya suma de sus longitudes constituye la mitad del

tamaño del genoma), la longitud total del genoma y el número de genes predichos. Por otro lado, el análisis con *BUSCO* muestra cómo de completo está un ensamblado; por ejemplo, determina si las regiones génicas control están bien ensambladas o muestran errores, comparándolas con su propia base de datos (*vibrionales_Odb10* caso del género *Vibrio*).

4.2.2. Ensamblaje híbrido

Otra metodología más precisa para el ensamblaje de genomas es utilizar lecturas tanto cortas como largas; esto se denomina ensamblaje híbrido. Además, existen dos metodologías diferentes para este tipo de ensamblaje: el uso exclusivo de ficheros FASTQ (de lecturas cortas y largas) y el uso de estos ficheros FASTQ de manera combinada con un fichero FASTA (lecturas ya ensambladas) de alta fidelidad. Esta última metodología sirve para refinar el ensamblado final.

El programa informático para el ensamblado híbrido fue *Unicycler*, que permite combinar lecturas tanto cortas como largas, como también un ensamblado ya realizado previamente^{39,40}. Se realizaron diferentes combinaciones de ficheros en *Unicycler* para el ensamblaje híbrido del genoma de la variante O. Así se combinaron los archivos FASTQ:

- con las lecturas filtradas de Illumina y con las lecturas *Hi-Fi*.
- con las lecturas filtradas de Illumina y con las lecturas *Hi-Fi* y archivo FASTA obtenido tras el ensamblaje con *HGAP4*.
- con las lecturas filtradas de Illumina y con las lecturas *Hi-Fi* y archivo FASTA obtenido tras el ensamblaje con *Flye*.
- con las lecturas filtradas de Illumina y con las lecturas *Hi-Fi* y archivo FASTA obtenido tras el scaffolding con *Quickmerge*.
- con solo las lecturas filtradas de Illumina y archivo FASTA obtenido tras el ensamblaje con *HGAP4*.
- con solo las lecturas filtradas de Illumina y archivo FASTA obtenido tras el ensamblaje con *Flye*.
- con solo las lecturas filtradas de Illumina y archivo FASTA obtenido tras el scaffolding con *Quickmerge*.

Como en el punto anterior, se comparó la calidad de estos ensamblados híbridos mediante el uso de *QUAST* y *BUSCO*.

Finalmente, y sumado al estudio de calidad mostrado anteriormente, también se realizó un pulido del genoma ensamblado definitivo usando la aplicación informática *Pilon 1.23*⁴¹. Para ello, se utilizó el fichero FASTA junto con los ficheros filtrados FASTQ de Illumina (de mayor calidad en general).

4.2.3. Mapeo

El mapeo de la variante T (ficheros FASTQ filtrados de Illumina) de la cepa en estudio se realizó utilizando el genoma cerrado de la variante O (ensamblado pulido obtenido anteriormente; fichero FASTA) como referencia. Para ello, se empleó la aplicación informática *Snippy*²⁷ del gestor *Conda*, con los valores por defecto. Esta herramienta ofrece tanto el fichero FASTA con el genoma mapeado como también un fichero informe sobre el llamado de variantes (VCF), el cual muestra las diferencias nucleotídicas entre la secuencia ensamblada (variante T) respecto de la de referencia (en este caso el genoma de la variante O).

4.2.4. Anotación de los genomas

La anotación de ambos genomas, O y T, de la cepa en estudio se realizó mediante *Prokka 1.14.6*⁴², que ofrece diferentes ficheros, siendo el fichero con las extensión GBK (fichero *GenBank*) el de mayor interés.

4.2.5. Determinación del linaje filogenético de la cepa CECT9109

Con las secuencias cerradas de los genomas de las variantes O y T de la cepa en estudio, se procedió a realizar un estudio filogenómico. En la actualidad se distinguen 5 linajes filogenéticos dentro de la especie *V. vulnificus*, de los que los linajes 1 y 3 son los más relevantes desde el punto de vista de la salud pública⁴³.

Para este análisis se usaron los 81 genomas incluidos en el estudio previo de Roig *et al.*⁴³ junto con los genomas de las variantes O y T. Se tomó como referencia la cepa YJ016, perteneciente al linaje 1. Esta cepa se seleccionó por ser la primera cuyo genoma se secuenció y cerró y por ser la más usada como referencia en este tipo de estudios.

Se empleó la herramienta informática *PARSNP 1.2*¹³, la cual da como resultado un árbol del tipo NJ, con formato *Newick*.

Tras este análisis, el árbol se visualizó con la herramienta de la página web *iTOL* (*Interactive Tree of Life*; <https://itol.embl.de/>).

4.2.6. Identificación de los loci *cps*

En primer lugar, se realizó una búsqueda bibliográfica sobre la región codificante de la cápsula en la especie *V. vulnificus*, obteniéndose el operón en cuestión descrito en la cepa MO6-24/O (operón *cps*: *NCBI Nucleotide* DQ360502.1

[*GenBank*: DQ360502.1] ⁷; genoma de la cepa MO6-24/O: *NCBI Assembly* ASM18658v1 [*GenBank assembly accession*: GCA_000186585.1]).

Para determinar la región exacta donde se sitúan los genes del operón en estudio en los genomas de las variantes O y T se utilizó la herramienta *BLASTn* (ejecutado desde Ubuntu, el cual se descargó desde la página oficial del NCBI). Los ficheros CSV de interés indicaron la posición exacta de inicio y fin del operón de estudio. A continuación, la extracción de la región de interés en cada genoma se realizó mediante un script propio del grupo de investigación de la Dra. Amaro, implementado en *Python 2.0* y se corroboró la correcta extracción mediante la visualización con *Bandage*.

Estas regiones se anotaron mediante *Prokka*, y se volvieron a representar junto con la cepa de referencia mediante *Easyfig* de *Python 2.0*.

Para encontrar diferencias a nivel nucleotídico entre las variantes O y T se procedió a un análisis más fino mediante “llamado de variantes” (VCF) obtenido tras el mapeo.

Finalmente, las secuencias afectadas por el cambio nucleotídico encontrado se analizaron mediante *Clustal Omega*, observando similitudes y diferencias con la proteína en cuestión obtenida de *Uniprot* (o *KEGG*).

4.2.7. Identificación y análisis de la presencia de profagos

Los programas informáticos utilizados para la identificación *in silico* de profagos en los genomas ensamblados a partir de las lecturas de Illumina fueron: *Phaster*, *Inovirus detector* y *VirSorter2*. En el caso de los resultados obtenidos con esta última aplicación informática, el resultado se complementó con la herramienta *CheckV* para eliminar falsos positivos y secuencias demasiado cortas para ser consideradas profagos.

En los tres casos, las aplicaciones comparan secuencias detectadas en el fichero de estudio con bases de datos de fagos ya identificados y publicados. Además, mientras que los programas *Phaster* y *VirSorter2* detectan la presencia de fagos (profagos) en general, *Inovirus detector* está especializado en inovirus. Este género de fagos pertenece a los *Inoviridae*, los cuales se caracterizan por ser una familia de fagos de cadena de ADN sencilla (de 9-10 kpb) que requieren de un intermediario de ADN de doble cadena para su replicación y transcripción, y además no poseen envoltura viral.

Así pues, se analizó la presencia de profagos en los genomas de ambas variantes de la cepa en estudio (CECT9109) y también en el de la cepa 95-8-161, ya secuenciada y ensamblada por el grupo de investigación de la Dra. Amaro; el uso de esta última cepa fue debido a que en estudios anteriores ya se detectó la

presencia de profagos con otros programas informáticos, por lo que fue considerada como control positivo de efectividad de la metodología empleada.

4.3. Metodología *in vitro*

4.3.1. Extracción rutinaria de ADN, PCR y electroforesis

La detección de los genes de interés se realizó mediante extracción de ADN, amplificación por PCR y visualización del amplicón tras electroforesis. En concreto, se analizó la presencia de los genes *vvha* (hemolisina de la especie *V. vulnificus*) y *vep07*.

La metodología usada en este apartado se detalla en el Anexo 3. La extracción del ADN bacteriano se realizó mediante ebullición y el ADN obtenido se guardó a -20 °C hasta su uso. Tras la amplificación por PCR de los genes de interés, se realizó una electroforesis en un gel de agarosa donde se corrió simultáneamente un marcador de pesos moleculares. Al término de la electroforesis, la visualización del gel al transiluminador permitió comprobar si la PCR había amplificado o no la secuencia correspondiente al gen estudiado.

4.3.2. Tinción capsular

Para la visualización directa de la cápsula mediante microscopía óptica, se realizó una tinción negativa con nigrosina o tinta china (Sigma-Aldrich, EE.UU.). Para ello, se depositó una gota de 10 µL procedente de una colonia resuspendida en PBS en un portaobjetos y se dejó secar al aire. A continuación, se colocó una gota de nigrosina en un extremo del portaobjetos y se extendió con otro portaobjetos. Se dejó secar al aire nuevamente.

Las muestras teñidas se observaron con un microscopio de campo claro Leica DM4000 B conectado a una cámara Leica DFC420 (Leica Biosystems, Alemania), tomándose imágenes a 100 aumentos.

4.3.3. Microscopía electrónica de transmisión (TEM)

El detalle de la estructura celular se observó mediante microscopía electrónica de transmisión (TEM, por sus siglas en inglés) en el servicio de microscopía de la Universitat de València. Se partió de una colonia aislada en medio sólido de ambas variantes de la cepa de estudio y se realizó un cultivo *overnight* (18 h) en 10 mL de medio líquido. El cultivo crecido se llevó al servicio de microscopía de la UV para su procesado, que incluyó fijación en resina, tinción y obtención de cortes ultrafinos. Estos se observaron con un microscopio electrónico de

transmisión (TEM) JEOL jem 1010 de alto contraste provisto de cámara digital AMT de 8 Mpx (Jeol USA Inc., EE.UU). Se tomaron imágenes a 8.000 aumentos con una potencia de 80.0 kV.

4.3.4. Cálculo de la frecuencia de reversión capsular

V. vulnificus puede revertir su morfología colonial de manera espontánea o inducida bajo presión de selección como la que representa la exposición al suero humano. En términos generales, la bacteria se recupera de las muestras clínicas con la morfología O y puede revertir de a T tras sucesivos pases en el laboratorio. También se ha reportado una tercera morfología, la rugosa (R), aunque con mucha menor frecuencia.

La cepa CECT9109 se recibió en el laboratorio de la Dra. Amaro como un cultivo puro de colonias O pero tras sucesivos pases se observaron colonias T. Ambos tipos se purificaron.

Para el cálculo de la frecuencia de reversión capsular hay que contar el número de colonias de un tipo y de otro y, como esta frecuencia suele ser muy baja, hay que sembrar múltiples placas.

Así pues, se partió de una colonia aislada en medio sólido de ambas variantes que se cultivó por separado en 5 mL de medio líquido y se incubó a 28 °C durante 24 h en agitación (120 rpm) (primer pase). A continuación, se realizaron varios pases sucesivos sembrando cada vez un volumen de 10 µL a partir del tubo crecido durante 24 h en un nuevo tubo conteniendo 5 mL de medio fresco. Todos los tubos de los pases sucesivos se incubaron a 28 °C durante 24 h en agitación.

En todos los casos, tras las 24 h de incubación, se sembraron 10 placas de medio de cultivo sólido por pase, con 100 µL de cultivo crecido de cada variante y el cultivo se extendió con asa de vidrio hasta su completa absorción. Las placas se incubaron a 28 °C durante 24 h y se contaron las colonias O y T. La frecuencia de reversión se calculó como el porcentaje de revertantes con respecto al total.

De las colonias que presentaban reversión, algunas de ellas se pasaron a placas con medio sólido para estudiar si la reversión se mantenía tras varios cultivos o volvían al estado anterior. Y para corroborar que estas colonias con reversión no eran el producto de una contaminación espontánea en las placas de cultivo, se les realizó una identificación mediante PCR dirigida al gen marcador de especie, el gen *vvhA*, siguiendo el procedimiento anteriormente descrito.

4.3.5. Ensayo de resistencia al suero humano con/sin hierro

La producción de cápsula está íntimamente relacionada con la resistencia al suero sanguíneo dado que el hierro activa la transcripción y la formación de la cápsula favoreciendo la resistencia al complemento del suero y la supervivencia del patógeno en sangre (el complemento está formado por una serie de enzimas que se activan unas a otras en cascada atacando a las bacterias y matándolas). En consecuencia, la variante O debería resistir el efecto bactericida del suero humano y la T no.

Para demostrarlo, crecimos las variantes O y T en suero humano comercial (HS, por sus siglas en inglés; Sigma-Aldrich) y en HS con el complemento inactivado por calor (30 min a 57 °C).

A partir de un cultivo de 24 h de ambas variantes tanto en medio líquido como en sólido, se resuspendieron 10^3 bacterias en 400 μ L de HS, HS+Fe (HS + FeCl_3 10 μ M (Sigma-Aldrich)) o HS inactivado. En paralelo, se repitió el experimento, pero sembrando en PBS, como control de viabilidad celular. Las muestras se incubaron a 37 °C y se realizó un recuento bacteriano por goteo a tiempo 0, a las 4 h y a las 24 h de incubación a 37 °C. El recuento por goteo consiste en sembrar volúmenes de 10 μ L por dilución (3X) en vez de los 100 μ L que se siembran normalmente (este procedimiento se usa cuando se trabaja con volúmenes de cultivo menores de 1 mL). Las placas se incubaron a 37 °C y se leyeron a las 24 h.

5. Resultados

5.1. Resultados *in silico*

5.1.1. Secuenciación por Illumina *MiSeq*, filtrado y ensamblaje primario

En primer lugar, se analizó la calidad de las lecturas crudas (tanto *forward* como *reverse*) obtenidas por Illumina para ambas variantes de la cepa CECT9109. Este análisis se realizó mediante *FastQC*, del entorno Galaxy. Las Figuras 3A y 3B, y 4A y 4B muestran este análisis previo para ambas variantes, O y T respectivamente. Se observó una calidad media que bajaba a partir de posiciones avanzadas en las lecturas (líneas azules en las imágenes).

Por ello, se procedió al filtrado de las lecturas mediante *PrinSeq*, siguiendo los parámetros:

- Longitud mínima requerida de las lecturas: 50 pb.
- Calidad media de las lecturas: 20 pb.

Tras el filtrado, se volvió a analizar la calidad de las lecturas, mostrada en las Figuras 3C y 3D, y 4C y 4D y se observaron lecturas de mayor calidad.

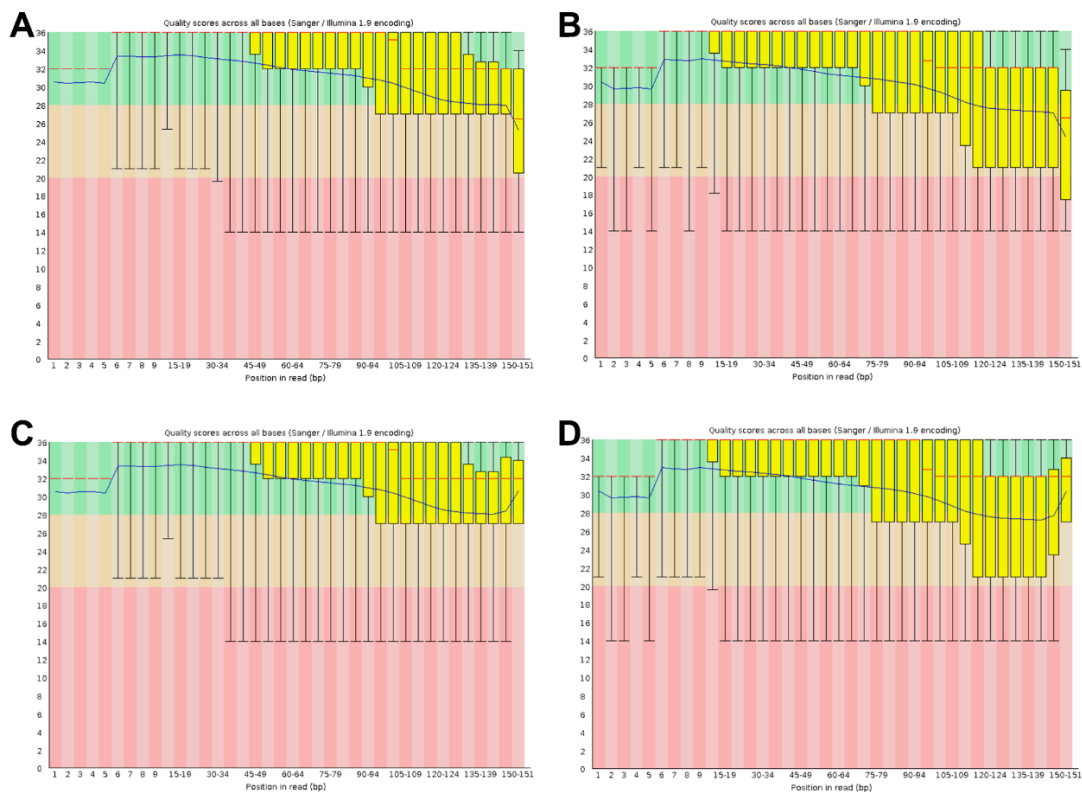


Figura 3. Imágenes del análisis de calidad mediante *FastQC* del entorno Galaxy para las lecturas de Illumina *MiSeq* de la cepa CECT9109(O) (A y C: *forward*; B y D: *reverse*) previo filtrado con *PrinSeq* (A y B) y tras filtrado (C y D). Se obtuvieron 1.125.928 lecturas, con longitud entre 35 y 151 pb en el caso de las lecturas crudas (A y B), mientras que tras el filtrado quedaron 1.123.256 lecturas, con longitud entre 50 y 151 pb (C y D).

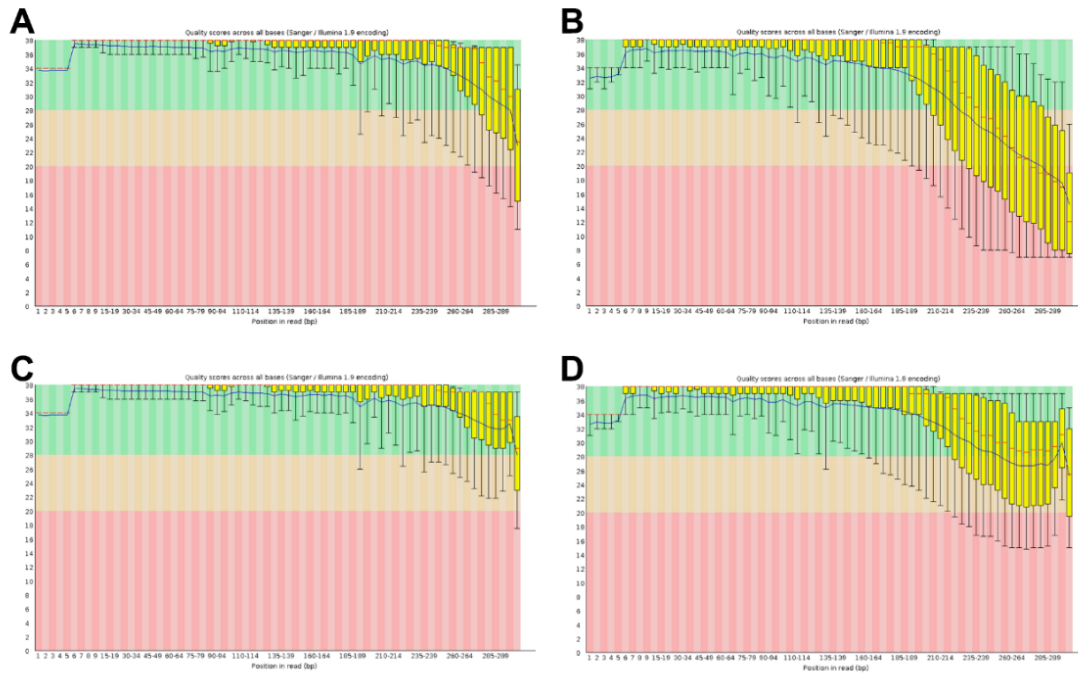


Figura 4. Imágenes del análisis de calidad mediante FastQC del entorno Galaxy para las lecturas de Illumina MiSeq de la cepa CECT9109(T) (A y C: forward; B y D: reverse) previo filtrado con PrinSeq (A y B) y tras filtrado (C y D). Se obtuvieron 703.278 lecturas, con longitud entre 35 y 301 pb en el caso de las lecturas crudas (A y B), mientras que tras el filtrado quedaron 585.226 lecturas, con longitud entre 50 y 301 pb (C y D).

Con las lecturas filtradas, se procedió a su ensamblaje primario usando las herramientas informáticas *SPAdes* y *Unicycler*, con los valores por defecto. La visualización del resultado se realizó mediante *Bandage*, utilizando los ficheros GFA (ficheros grafo). La Figura 5 muestra los ensamblados obtenidos para ambas variantes con las dos aplicaciones utilizadas.

La Figura 5A y 5B muestra el ensamblado con *SPAdes*, donde se obtuvieron 627 nodos para la variante O y 353 nodos para la T; esto indica un gran volumen de *contigs* por lo que el genoma no se consiguió cerrar en ningún caso.

Por otro lado, la Figura 5C y 5D muestra este ensamblado, pero con la herramienta *Unicycler*, la cual produjo 120 nodos para la variante O y 122 para la T; en definitiva, un número muy reducido de nodos comparado con el uso de *SPAdes*. No obstante, tampoco se consiguió cerrar el genoma de las variantes de la cepa en estudio.

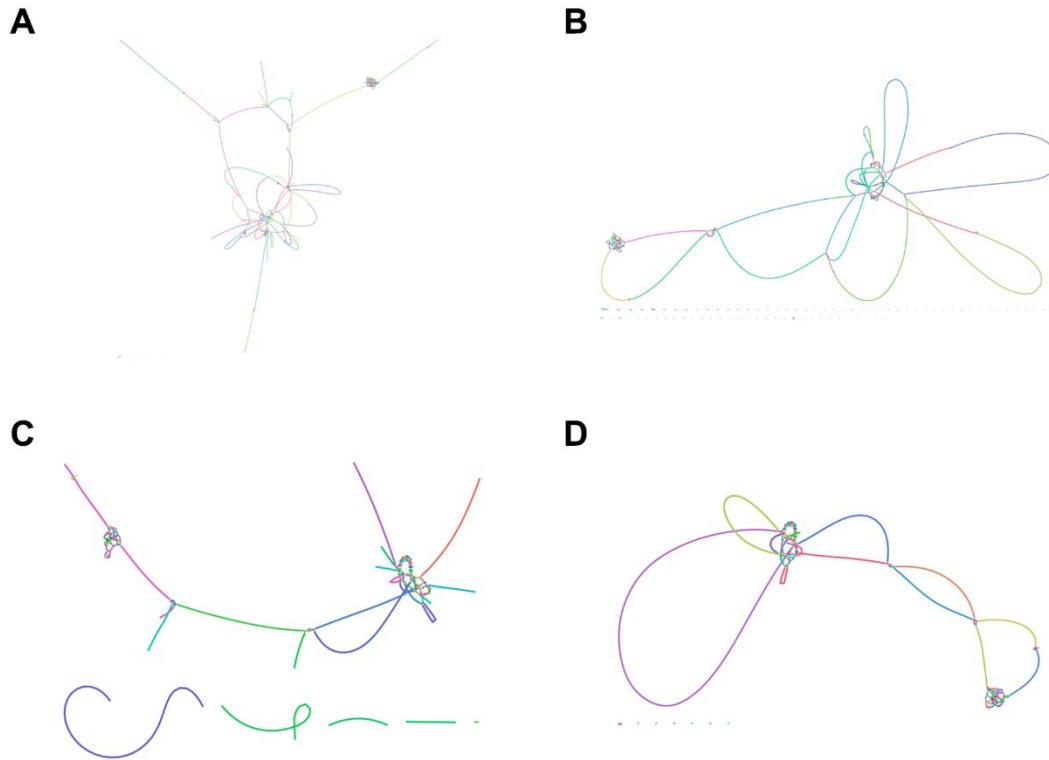


Figura 5. Ensamblado de las lecturas obtenidas mediante Illumina MiSeq de las variantes O (A y C) y T (B y D) de la cepa CECT9109. El ensamblaje de las lecturas filtradas se realizó mediante SPAdes (A y B) y Unicycler (C y D), obteniéndose los grafos (representación visual) con la herramienta informática Bandage.

5.1.2. Secuenciación por PacBio, filtrado y ensamblaje primario

PacBio presenta lecturas largas tanto crudas como filtradas por el propio sistema, las cuales pasan a llamarse lecturas *Hi-Fi*. Tras el análisis de la calidad de estas lecturas filtradas mediante *NanoPlot* (Figura 6), se observó una calidad muy alta (mínima calidad media de 20 pb), llegando a tener lecturas de hasta 25 kpb. Por ello, se decidió mantener y utilizar estas lecturas *Hi-Fi* para los siguientes puntos.



Figura 6. Gráfico de calidad obtenido mediante NanoPlot de las lecturas Hi-Fi de PacBio de la cepa CECT9109(O). Se obtuvieron 93.423 lecturas, con longitud entre 512 y 29.089 pb.

Las lecturas *Hi-Fi* se ensamblaron utilizando dos ensambladores diferentes: *HGAP4* (propio de PacBio) y *Flye*. Y, además, se realizó un *scaffolding* con los ensamblados anteriores mediante *Quickmerge*. Tras los ensamblados, se estudiaron varios parámetros con *QUAST*, mostrándose en la Tabla 2. Se analizó el número de *contigs*, la longitud total del ensamblado, la N50 (longitud de los *contigs* que, usando *contigs* de igual o mayor tamaño, produce la mitad de la longitud del genoma), la L50 (número más pequeño de *contigs* cuya suma de sus longitudes constituye la mitad del tamaño del genoma) y el número de genes predichos. También se analizó la calidad del ensamblado mediante la herramienta *BUSCO*.

Las herramientas *Flye* y *Quickmerge* mostraron ensamblados con 2 *contigs*, mientras que *HGAP4* dio como resultado 3. En cuanto a la longitud total obtenida, las tres herramientas rindieron un resultado parecido, en torno a 5 Mpb. Para los 3 casos, tanto la N50 y la L50 fueron el mismo (el primer *contig* en concreto). Además, el número de genes predicho también fue parecido. Finalmente, la mejor calidad del ensamblado la obtuvo *Flye*, con casi un 100%.

Tabla 2. Resultado de la comparativa entre los diferentes ensambladores (*HGAP4* y *Flye*) y la herramienta de *scaffolding* (*Quickmerge*) para el ensamblado de las lecturas largas Hi-Fi de PacBio.

	nº <i>contigs</i>	longitud total	N50	L50	nº genes predichos	<i>BUSCO</i> (%)
<i>HGAP4</i>	3	5.016.306	3.242.516	1	4.362	99,7 (4)*
<i>Flye</i>	2	4.993.744	3.232.464	1	4.345	99,9 (1)*
<i>Quickmerge</i> (<i>HGAP4</i> + <i>Flye</i>)	2	5.003.796	3.242.516	1	4.352	99,7 (4)*

*número de genes patrón no encontrados de los 1.445 genes en la base de datos empleada por *BUSCO*.

Tras la representación con *Bandage* del grafo obtenido con *Flye* (mejor ensamblado) se observó el segundo *contig* cerrado y circularizado (Figura 7). Por tamaño (1.761.280 pb) se consideró que representaba el cromosoma 2 de la bacteria en estudio. Por el contrario, el *contig* primero (3.232.464 pb, correspondiéndose al cromosoma 1) no se consiguió cerrar.

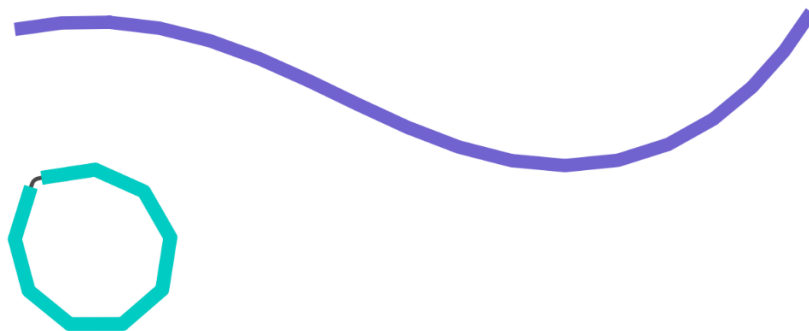


Figura 7. Imagen del ensamblado del genoma de la cepa CECT9109(O), obtenido mediante la herramienta informática *Flye*, utilizando el fichero FASTQ con las lecturas *Hi-Fi* de *BacBio*, y representado con *Bandage*. *Contig* 1 en morado, y *contig* 2 en turquesa.

El resultado anterior demuestra la necesidad de realizar un ensamblaje híbrido con las lecturas cortas de *Illumina* y largas de *PacBio*.

5.1.3. Ensamblaje híbrido

Dado que los puntos anteriores no dieron lugar a un ensamblado perfecto para la variante O, con los genomas cerrados con solo 2 *contigs* y circularizados, se procedió al ensamblaje híbrido empleando las lecturas cortas filtrada de *Illumina* y las lecturas *Hi-Fi* de *PacBio*. Para ello se utilizó *Unicycler*, con las diferentes combinaciones expuestas en el punto de Metodología.

La Tabla 3 muestra los resultados tras el uso de *QUAST* y *BUSCO* para los diferentes parámetros estudiados en los resultados del ensamblaje híbrido.

El mejor resultado se obtuvo con *Unicycler* empleando solamente los archivos FASTQ filtrados de *Illumina* y *PacBio*, dando lugar a 2 *contigs*; el resto de las combinaciones dieron lugar a 4 *contigs*, de igual manera que ocurría en el ensamblado primario de las lecturas *Hi-Fi* con *HGAP4* y *Quickmerge*.

La longitud total del genoma rondó los 5 Mpb, siendo el N50 y el L50 el mismo en todos los casos (primer *contig*). En cuanto al número de genes predichos, solo con los archivos FASTQ se obtuvieron 4.340 genes, mientras que con las combinaciones con ficheros FASTA se obtuvieron 4 genes más.

Finalmente, la herramienta *BUSCO* dio lugar a una calidad de casi el 100% para todos los casos.

Tabla 3. Resultado de la comparativa entre las diferentes técnicas de ensamblaje híbrido, combinando las lecturas cortas filtradas de Illumina y las lecturas largas Hi-Fi de PacBio. En naranja se muestran los ensamblados todos los ficheros FASTQ y el fichero FASTA correspondiente, y en azul, los ensamblados con los ficheros FASTQ de Illumina y el fichero FASTA correspondiente.

	nº contigs	longitud total	N50	L50	nº genes predichos	BUSCO (%)
Unicycler	2	4.991.244	3.229.973	1	4.340	99,9 (1)*
Unicycler + HGAP4	4	4.986.148	3.162.969	1	4.344	99,9 (1)*
Unicycler + Flye	4	4.986.148	3.162.969	1	4.344	99,9 (1)*
Unicycler + Quickmerge	4	4.986.148	3.162.969	1	4.344	99,9 (1)*
Unicycler + HGAP4	4	4.986.148	3.162.969	1	4.344	99,9 (1)*
Unicycler + Flye	4	4.986.148	3.162.969	1	4.344	99,9 (1)*
Unicycler + Quickmerge	4	4.986.148	3.162.969	1	4.344	99,9 (1)*

*número de genes patrón no encontrados de los 1.445 genes en la base de datos empleada por BUSCO.

Así pues, observando los resultados anteriores se decidió que el ensamblado de *Unicycler* empleando solamente los ficheros FASTQ resultó ser el de mayor calidad, por lo que se representó su grafo mediante *Bandage* (Figura 8). Se observaron los 2 contigs circularizados, por lo que sus secuencias estaban cerradas. El primer contig, de 3.229.975 pares de bases, se correspondió al cromosoma 1 de la cepa, mientras que el segundo, de 1.761.272 pb, al cromosoma 2.

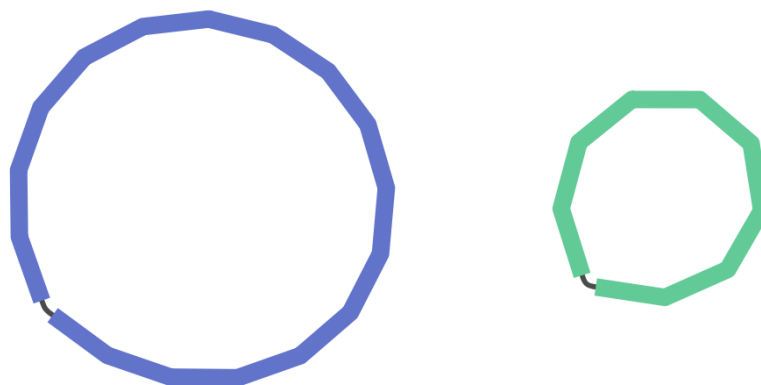


Figura 8. Imagen del ensamblado del genoma cerrado de la cepa CECT9109(O), obtenido mediante la herramienta informática *Unicycler*, empleando solamente los ficheros FASTQ de Illumina y de PacBio (Hi-Fi), y representado con *Bandage*. Contig 1 en morado, y contig 2 en verde.

Finalmente, se realizó el pulido de este genoma cerrado mediante la herramienta *Pilon*, donde en la primera ronda se realizaron 16 cambios y en la segunda y última, un único cambio. El genoma resultante tiene 4.991.247 pb, manteniéndose los 2 contigs.

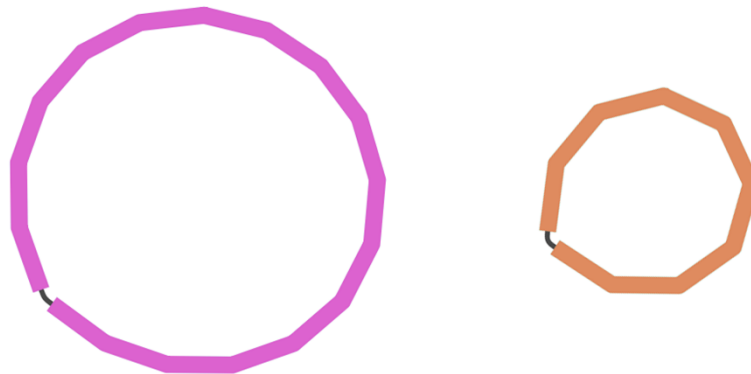
5.1.4. Mapeo de los genomas

Con el genoma cerrado y pulido de la variante O, se realizó el mapeo de la variante T mediante la aplicación informática *Snippy* del gestor *Conda*. Para ello, se utilizaron los archivos FASTQ filtrados de la variante T y el archivo FASTA resultante de la O.

El mapeo dio como resultado el genoma cerrado de la variante translúcida de la cepa de estudio (Figura 9A), con 2 *contigs* con una longitud total de 4.991.247 pb (3.229.975 pb el primer *contig* o cromosoma y 1.761.272 pb el segundo).

Además del fichero FASTA de la variante T obtenido, esta herramienta también ofrece un fichero informe sobre llamado de variantes (VCF), el cual muestra las diferencias nucleotídicas entre la secuencia ensamblada de la variante T respecto de la de referencia (variante O) (Figura 9B). Se distinguieron 6 cambios significativos, 5 de ellos fueron de un único nucleótido y otro de una secuencia de 5 nucleótidos. Los cambios se identificaron en las posiciones 432.913, 1.017.680, 1.174.955, 1.348.474, 2.822.622 y 2.964.050 del cromosoma 1.

A



B

LLAMADO DE VARIANTES (VCF) TRANSLÚCIDA
fileformat=VCFv4.2
FILTER=<ID=PASS;Description="All filters passed">
source=freeBayes v1.3.6
reference=reference/ref.fa
contig=<ID=CECT9109op_chrom1,length=3229975>
contig=<ID=CECT9109op_chrom2,length=1761272>

CHROM	POSITION	ID	REF (op)	ALT (tr)	QUALITY	FILTER	INFO	FORMAT	CECT9109tr_mapeo_definitivo
CECT9109op_chrom1	432913	.	A	T	2719.45	.	AB=0;AO=85;DP=85;QA=3062;QR=0;RO=0;TYPE=snp	GT:DP:RO:QR:AO:QA:GL	1/1:85:0:0:85:3062-275.774-25.5876:0
CECT9109op_chrom1	1017680	.	A	C	1504.29	.	AB=0;AO=49;DP=49;QA=1704;QR=0;RO=0;TYPE=snp	GT:DP:RO:QR:AO:QA:GL	1/1:49:0:0:49:1704-153.628-14.7505:0
CECT9109op_chrom1	1174955	.	T	A	1834.21	.	AB=0;AO=57;DP=57;QA=2075;QR=0;RO=0;TYPE=snp	GT:DP:RO:QR:AO:QA:GL	1/1:57:0:0:57:2075-186.998-17.1587:0
CECT9109op_chrom1	1348474	.	C	A	1506.98	.	AB=0;AO=47;DP=47;QA=1704;QR=0;RO=0;TYPE=snp	GT:DP:RO:QR:AO:QA:GL	1/1:47:0:0:47:1704-153.628-14.1484:0
CECT9109op_chrom1	2822622	.	TAATT	AAATA	2289.52	.	AB=0;AO=71;DP=72;QA=2581;QR=0;RO=0;TYPE=complex	GT:DP:RO:QR:AO:QA:GL	1/1:72:0:0:71:2581-232.348-21.3731:0
CECT9109op_chrom1	2964050	.	A	C	1968.99	.	AB=0;AO=62;DP=62;QA=2224;QR=0;RO=0;TYPE=snp	GT:DP:RO:QR:AO:QA:GL	1/1:62:0:0:62:2224-200.4-18.6639:0

Figura 9. Imagen del ensamblado del genoma cerrado de la cepa CECT9109(T) a partir del mapeo de las lecturas filtradas de Illumina (archivos FASTQ) con el genoma cerrado de la cepa CECT9109(O). La imagen obtenida con Bandage (A) muestra los dos cromosomas circularizados. Con el mapeo se obtuvo el llamado de variantes tras filtrado por calidad de cambio nucleotídico (B), donde se aprecian en total 6 diferencias en el genoma de las dos variantes de la cepa CECT9109.

5.1.5. Anotación de los genomas

Finalmente, ambas secuencias (O y T) se anotaron mediante el programa informático *Prokka* que ofrece diferentes ficheros, siendo el fichero con la extensión GBK (*GenBank*) el de mayor interés (Figura 10A y 10B).

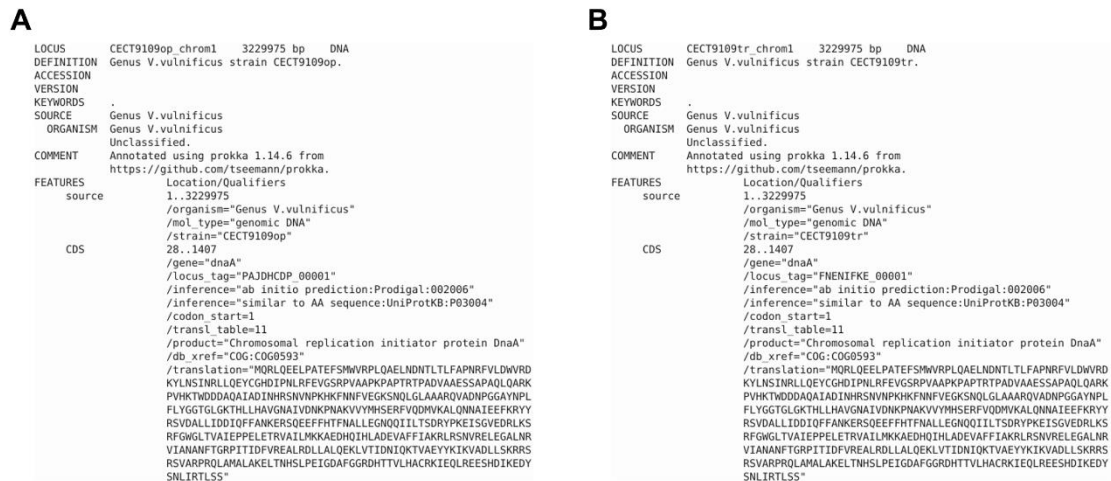


Figura 10. Imágenes de los ficheros GenBank de la anotación mediante la herramienta informática Prokka de las variantes O (A) y T (B) de la cepa en estudio. Se muestra el primer gen en ambos genomas (*dnaA*), no observándose diferencias en este caso, ya que por llamada de variantes no se observó ninguna diferencia de calidad en esta primera región de los genomas.

5.1.6. Detección del linaje filogenético

Para determinar a qué linaje filogenético pertenece la cepa CECT9109 se procedió a un análisis filogenómico usando la herramienta *PARSNP* e incluyendo en el análisis 81 genomas de la misma especie, pertenecientes a los 5 linajes existentes. Uno de estos genomas, el de la cepa YJ016 (linaje 1), se tomó como referencia.

El árbol NJ resultante, con formato *Newick* y visualizado con *iTOL*, mostró que la cepa objeto de estudio pertenece al linaje 1 (Figura 22 del Anexo 4). Como se esperaba, ambas variantes (O y T) resultaron idénticas.

5.1.7. Identificación de los loci *cps*

En primer lugar, se realizó una búsqueda bibliográfica sobre la región codificante de la cápsula en la especie *V. vulnificus*, obteniéndose el operón en cuestión descrito en la cepa MO6-24/O (Figura 11A), como ya se ha descrito en el apartado de Metodología.

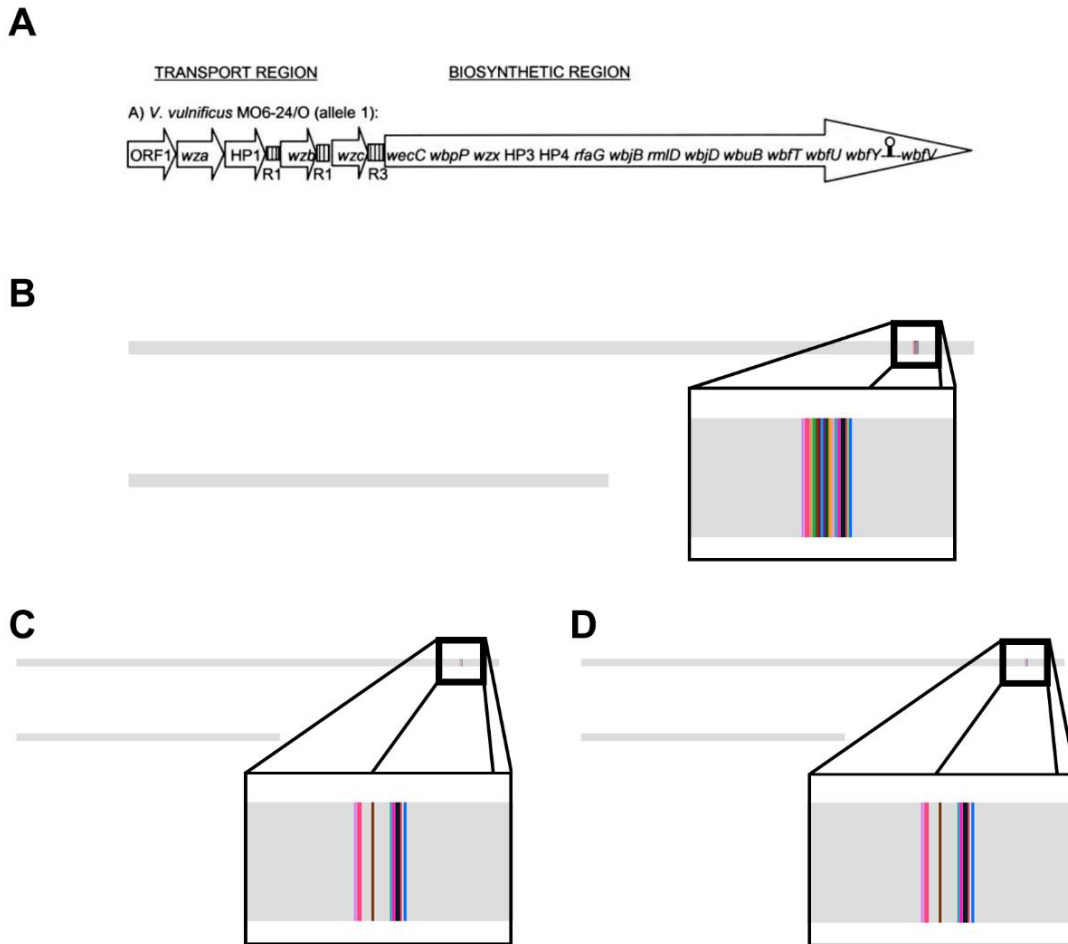


Figura 11. Operón *cps* (codificador de la cápsula) descrito en la cepa MO6-24/O (A, obtenido de Chatzidaki-Livanis et al. ⁷) y representado con Bandage (B, genoma obtenido des NCBI Assembly ASM18658v1 [GenBank assembly accession: GCA_000186585.1]). También se localizó esta región *cps* en el genoma cerrado de las variantes O (C) y T (D) de la cepa CECT9109, con Bandage.

Mediante la herramienta *Bandage*, se observó que los genes *cps* se disponían en el cromosoma 1 de la cepa MO6-24/O, uno tras otro, donde se apreciaron regiones intergénicas solo entre alguno de los genes (Figura 11B). A continuación, se identificaron estos genes también en los genomas de las variantes de la cepa en estudio, observándose que se localizaban en un mismo operón del cromosoma 1 (Figura 11C y 11D). Por lo tanto, se extrajo esta región del operón más 1.000 nucleótidos aguas-arriba y aguas-abajo del operón a partir de los genomas de las dos variantes de la cepa CECT9109. Para determinar la región exacta a extraer, se localizaron los genes en estudio en los genomas mediante *BLASTn*. Los ficheros CSV de interés indicaron la posición exacta de inicio y fin del operón de estudio.

Tras la extracción de la región de interés en cada genoma, se corroboró la correcta extracción de la región del operón usando *Bandage* (Figura 12).

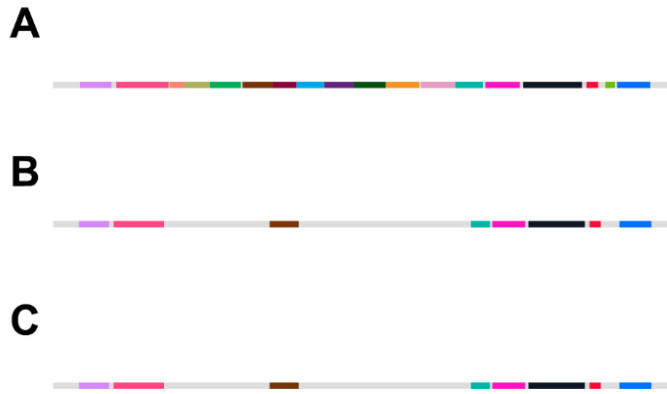


Figura 12. Región del operón *cps* con 1.000 nucleótidos agua-arriba y aguas-abajo. Los genes que forman este operón se indican con diferentes colores, correspondiéndose con los genes descritos en la Figura 11. Cepas MO6-24/O (A), CECT9109(O) (B) y CECT9109(T) (C).

Tras la anotación de estas regiones con *Prokka*, se observó que las variantes de la cepa CECT9109 (Figura 12B y 12C) solo presentaban 8 de los 18 genes descritos en el locus *cps* de la MO6-24/O (Figura 12A), lo que significa que el tipo de cápsula que producen es distinto.

Dado que en los 1.000 nucleótidos aguas-arriba y aguas-abajo del operón de los genomas estudiados no había ningún otro gen relacionado con la cápsula, se repitieron los pasos anteriores, extrayendo esta vez solo la región del operón *cps* (Figura 13).

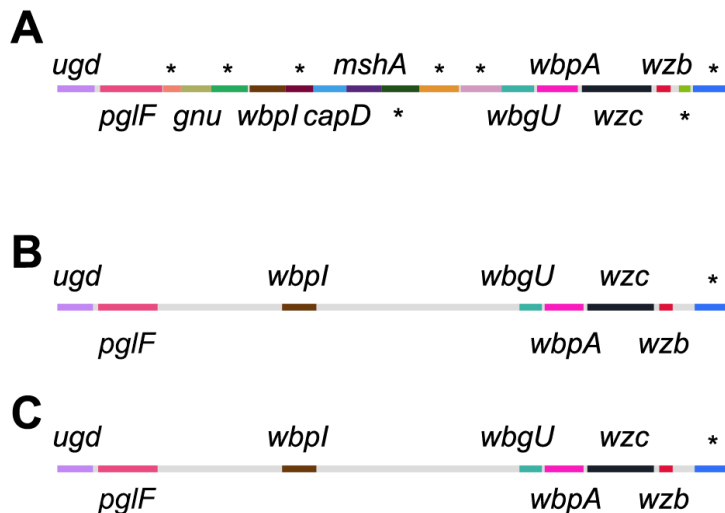


Figura 13. Región del operón *cps*, con los diferentes genes indicados mediante el programa de anotación *Prokka*, para las cepas MO6-24/O (A), CECT9109(O) (B) y CECT9109(T) (C).

Mediante *Easyfig*, se representó esta región *cps* de los tres genomas de estudio. No se observaron grandes diferencias entre la cepa CECT9109 y la cepa de

referencia y, por su parte, las dos variantes O y T de la cepa CECT9109 aparentaban ser idénticas (Figura 14).

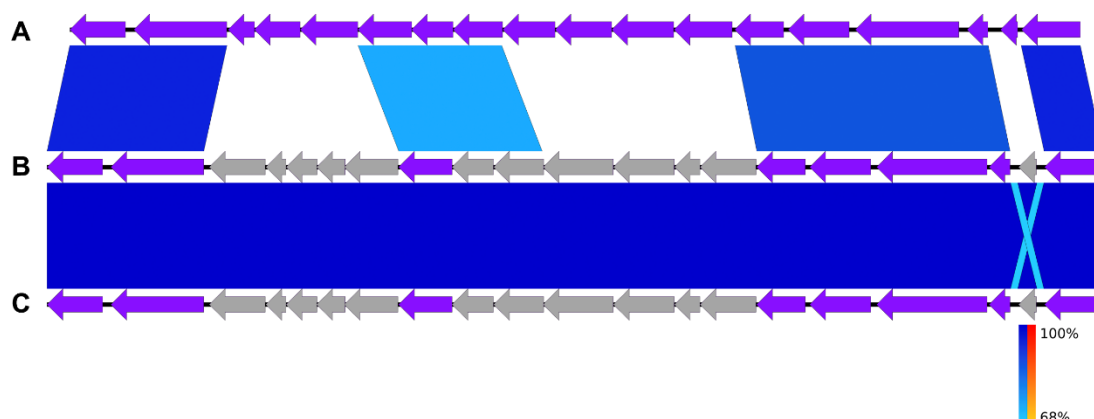


Figura 14. Imagen obtenida con el programa informático Easyfig, donde se representan los genes de interés en la región del operón cps, viendo similitudes entre los tres genomas de estudio (A: MO6-24/O; B: CECT9109(O), y C: CECT9109(T)) mediante colorimetría. Se observan en morado los genes descritos en la figura anterior, y en gris aquellos genes de las variantes de la cepa CECT9109 que no son idénticos con los de la cepa de referencia. El color azul de la leyenda indica el grado similitud entre las secuencias analizadas; el color rojo indica el grado de inversión, que en este caso no se presenta.

Para encontrar diferencias a nivel nucleotídico entre las variantes O y T, se procedió a un análisis más fino mediante "llamado de variantes" (VCF). Solo se observó una posición diferente en el operón en estudio (posición 2.963.187 sobre el genoma entero (cromosoma 1) de la variante T). El cambio observado fue de O: A → T: C, el cual es no sinónimo, lo que supone un cambio aminoacídico.

Tras la traducción a aminoácidos y su visualización mediante *Clustal Omega*, se identificó un cambio en la posición 364 de la proteína: mientras que en la variante O había una "Y" (tirosina; UAU), en la T había una "D" (ácido aspártico; GAU), lo cual se corresponde con el cambio nucleotídico en la secuencia (O: A → T: C). A continuación, se realizó un *BLASTp* con esta proteína resultante y se identificó con una proteína codificada por el gen *wlaL*. Este gen se ha descrito en *Helicobacter pylori* y codifica esta proteína que interviene en la síntesis de polisacáridos de superficie con actividad *UDP-N-acetyl-alpha-D-glucosamine C6 dehidratase* y que también está relacionada con la síntesis del flagelo^{44,45}. Si el cambio nucleotídico es el responsable de que la variante T sea acapsulada quedaría pendiente de posteriores estudios.

Se realizó un nuevo *Clustal Omega* con la secuencia obtenida de *Uniprot* (referencia: O86159) o *KEGG* (referencia: C8J_1061) para la bacteria *Campilobacter jejuni*, y con las secuencias de aminoácidos de las variantes O y T y se encontró que eran muy similares (Figura 23 del Anexo 4).

5.1.8. Identificación de la presencia de profagos

La presencia de profagos en los genomas cerrados de la cepa CECT9109 se analizó usando los programas informáticos *Phaster*, *Inovirus detector* y *VirSorter2* (con comprobación del resultado mediante *CheckV*), estos últimos del gestor *Conda*. Para verificar el funcionamiento de estas herramientas, se utilizó como control positivo el genoma de la cepa de *V. vulnificus* 95-8-161 que presenta un profago, según estudios previos del grupo de la Dra. Amaro.

El análisis reveló la presencia de una región altamente conservada relacionada con genomas de profagos en la misma posición del cromosoma 2 de los genomas de ambas variantes (Tabla 4). Debido al tamaño de dicha región (9-10 kb), se puede afirmar que se trata de un inovirus. Además, el programa *Inovirus detector* lo corroboró, ya que mientras que *Phaster* y *VirSorter2* son genéricos (buscan en bases de datos de fagos en general), *Inovirus detector* está especializado en inovirus (menos descritos en bases de datos). Por lo tanto, aunque *Phaster* mostró que la región detectada codificaba para la cápside del profago, realmente se trataría del recubrimiento proteico de un inovirus, fagos que carecen de cápside *sensu stricto*.

Tabla 4. Comparativa de los diferentes buscadores de profagos en las secuencias de los genomas de las variantes O y T de la cepa CECT9109.

Opaca	nº regiones	longitud (kb)	integridad	estructura	región
<i>Phaster</i>	1	9,7	intacto	cápside	chr2:874998-884708
<i>VirSorter2 + CheckV</i>	1	11,3	parcial-alto	lineal	chr2:874761-886045
<i>Inovirus detector</i>	1	10,1	intacto	recubrimiento	chr2:874998-885132
Translúcida	nº regiones	longitud (kb)	integridad	estructura	región
<i>Phaster</i>	1	9,7	intacto	cápside	chr2:875003-884713
<i>VirSorter2 + CheckV</i>	1	11,3	parcial-alto	lineal	chr2:874766-886050
<i>Inovirus detector</i>	1	10,1	intacto	recubrimiento	chr2:875003-885137

No se apreciaron diferencias nucleotídicas entre las secuencias del inovirus en ambas variantes tras la consulta del fichero VCF ni al corroborarlo mediante el uso de *BLASTn* del NCBI.

5.2. Resultados *in vitro*

5.2.1. Clasificación de la cepa a nivel de patovariedad

Se realizó una PCR dirigida al gen *vep07*, marcador de la pv. *piscis*. Los resultados se muestran en la Figura 15. Solo la cepa control positivo (R99) mostró el amplicón del tamaño esperado. Por lo tanto, la cepa objeto de estudio no pertenece a la pv. *piscis*.

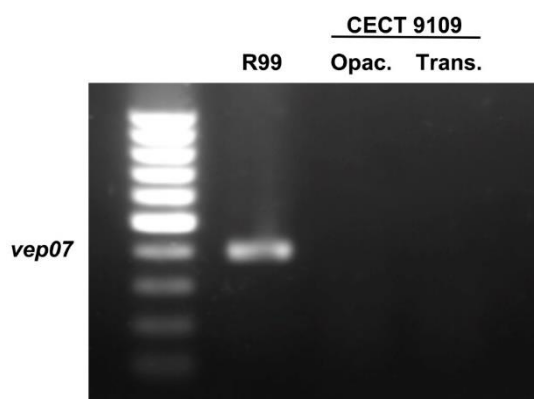


Figura 15. Resultados de la PCR dirigida al gen marcador de la pv. *piscis* (*vep07*). Se tomó como control positivo la cepa R99.

5.2.2. Observación de la cápsula

5.2.2.1. Tinción con nigrosina

Para observar la cápsula al microscopio óptico se realizó una tinción con nigrosina. No se apreciaron diferencias entre ambas variantes. No obstante, este resultado no significa que la bacteria no tenga cápsula, sino que su grosor no es lo suficientemente alto como para ser visualizada por microscopía óptica. En consecuencia, se prepararon muestras para su observación con TEM.

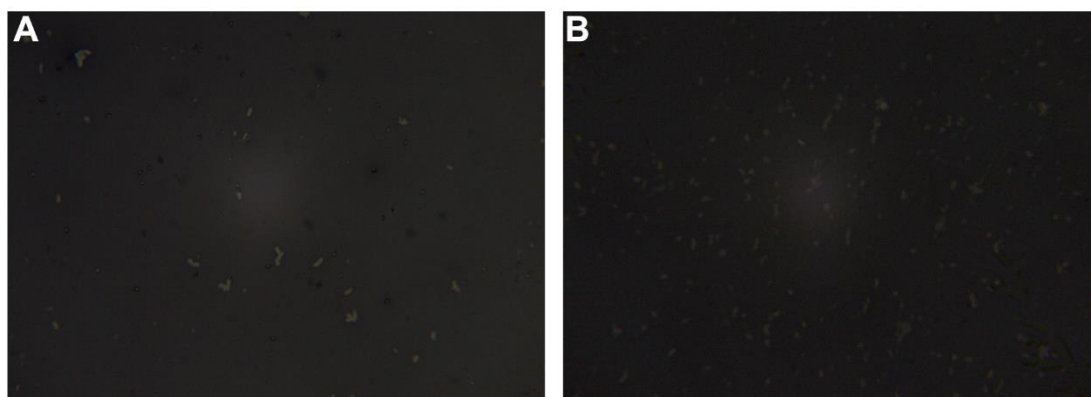


Figura 16. Imágenes de las variantes O (A) y T (B) de la cepa CECT9109 teñidas con nigrosina.

5.2.2.2. TEM

La visualización de las muestras al TEM mostró que las células de la variante O tenían un halo electrodenso entorno a su membrana externa que se correspondería con la cápsula (Figura 17A). Células parecidas junto con otras sin halo electrodenso se observaron en las muestras de la variante T (Figura 17B).

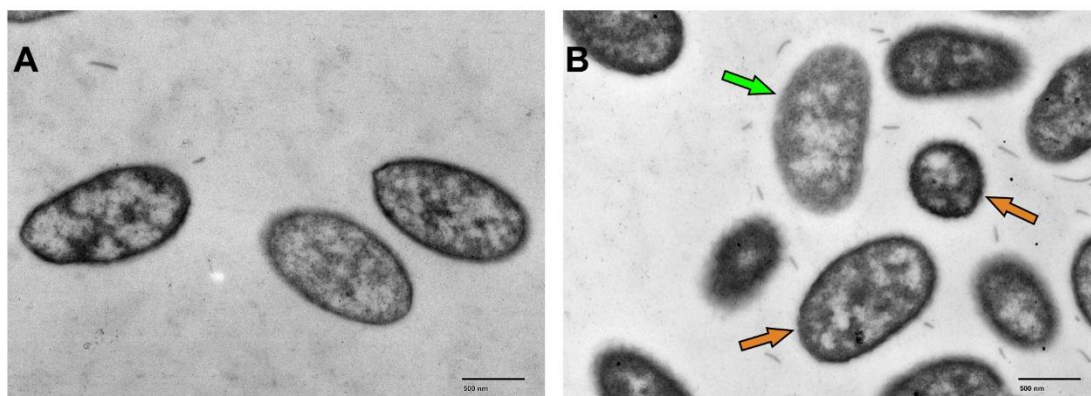


Figura 17. Imágenes tomadas por microscopía electrónica de transmisión de las variantes O (A) y T (B) de la cepa CECT9109. La flecha verde indica las células sin cápsula y la naranja las células con cápsula. Barra de escala = 500 nm.

5.2.3. Cálculo de la frecuencia de reversión capsular

Para averiguar si se producía la reversión de la variante productora de cápsula (O) a la no productora (T) y viceversa, se realizaron pases sucesivos de cultivos en LB-1. La reversión de la variante T a la O se produjo con una frecuencia de $<10^{-7}\%$ en los primeros pases, incrementándose con cada pase hasta alcanzar valores de entre el 50 y el 60% (Figura 18; Tabla 5).

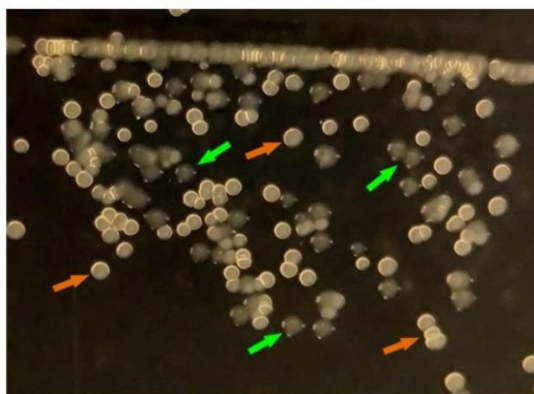


Figura 18. Imagen tomada de una placa cultivada en césped (8º día de experimento) de la cepa CECT9109(T). Las flechas verdes indican la presencia de colonias traslúcidas, mientras que las flechas naranjas indican las colonias opacas.

Tabla 5. Frecuencia de reversión de la variante T a la O durante los 10 días de experimento. Las unidades representadas son nº colonias revertidas/volumen de células sembradas en placa (10.000 ufc*/placa). (*ufc: unidades formadoras de colonia)

Días:	1º	2º	3º	4º	5º	6º	7º	8º	9º	10º
T → O	0	0	1/10 ⁴	1/10 ⁴	2/10 ⁴	3/10 ⁴	1/10 ³	8/10 ²	1/2	2/3

Sin embargo, no se produjo la reversión completa de la variante O a la T y solo se observó un caso de una colonia mixta formada por la variante O y T tras examinar miles de colonias (Figura 19).

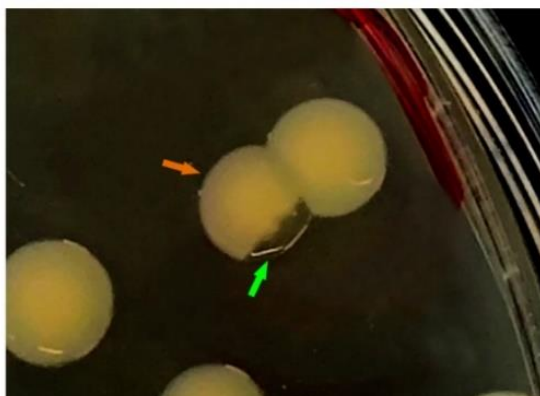


Figura 19. Imagen tomada de una placa cultivada en césped (3º día de experimento) de la cepa CECT9109(O). La flecha naranja indica la fracción opaca de la colonia, mientras que la flecha verde indica la fracción traslúcida.

Sorprendentemente, se detectó una tercera variante de esta misma cepa en placas de la variante T, cuya colonia tenía un aspecto mucho más opaco y una superficie rugosa. La nueva variante podría corresponder a la variante “rugosa” (R) (Figura 20A), la cual es muy infrecuente en *V. vulnificus*.

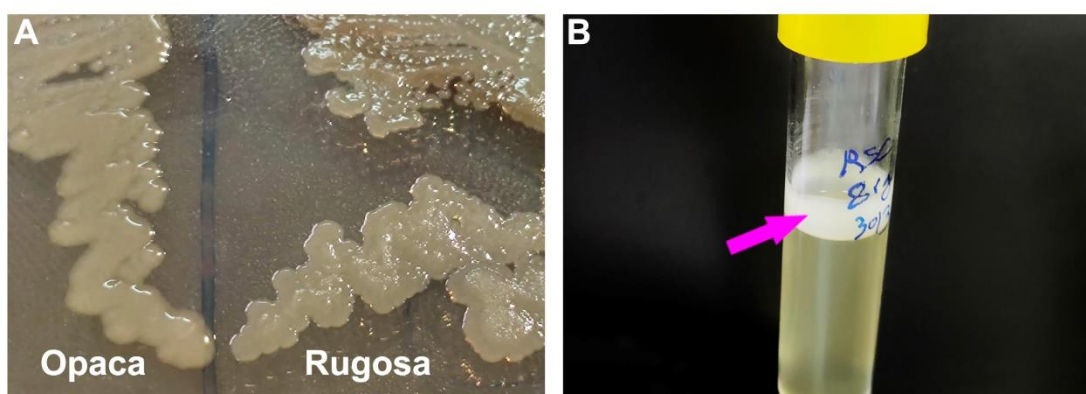


Figura 20. Variante rugosa (R) de la cepa CECT9109 cultivada en placa (A) y en cultivo líquido (B). Se observa el aspecto granulado (rugoso) del cultivo de la variante R (parte derecha del panel A) en comparación con la variante O (parte izquierda del panel A). La flecha rosa (B) indica el biofilm formado tras 24 h de cultivo de la cepa CECT9109(R) en agitación, característico de este tipo de variantes.

Se han relacionado alteraciones en la producción de biofilm con la morfología colonial por lo que se determinó la producción de biofilm de las tres variantes en tubo de ensayo y se comprobó que la variante R era la que más biofilm formaba tras 24 h de cultivo (Figura 20B).

5.2.3.1. Clasificación de las colonias a nivel de la especie *V. vulnificus*

En todos los casos anteriores de reversión, se obtuvieron los revertantes en forma de cultivo puro y se comprobó que eran variantes de *V. vulnificus* y no contaminantes mediante identificación por PCR dirigida al gen marcador de especie *vvhA*, el cual codifica una hemolisina importante en virulencia.

Para ello, se extrajo el ADN de estos revertantes, y se siguió el protocolo de PCR-electroforesis descrito en el apartado de Metodología. En todos los casos se obtuvo el amplicón del tamaño esperado, es decir, todos los revertantes eran *V. vulnificus* (Figura 21).

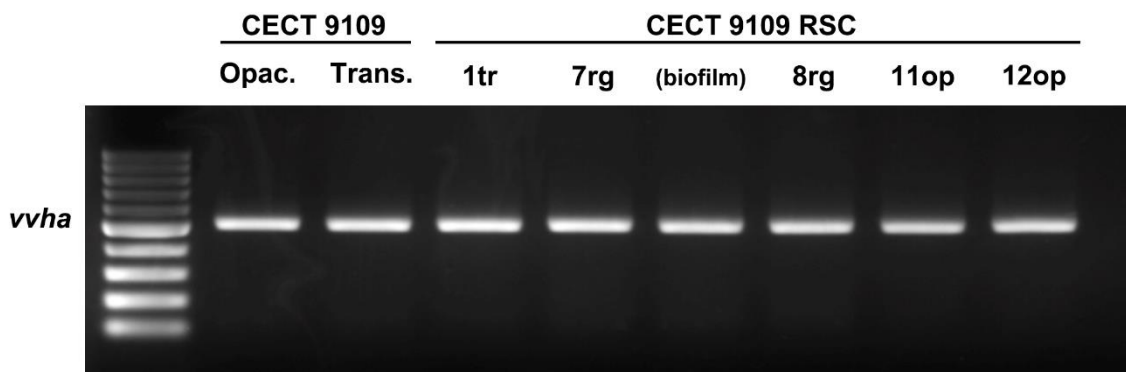


Figura 21. Identificación de los revertantes por PCR dirigida al gen marcador de especie *vvhA*. Los revertantes se nombraron como RSC seguido de un número y "tr" (si reversionó de O a T), "rg" (si fue de T a R) y "op" (para el caso de T a O). La muestra denominada "biofilm" consistió en el análisis de este gen marcador en una muestra del biofilm formado tras 24 h de cultivo en agitación del revertante RSC7rg. La presencia de este gen corrobora que todas las muestras analizadas pertenecen a la especie de estudio y que no hubo contaminación en las placas y cepas de reversión.

Se comprobó que los tres tipos de revertantes mantenían su morfología colonial y se crioconservaron para posteriores estudios fuera del propósito del presente TFM.

5.2.4. Análisis de resistencia al suero humano con/sin hierro

Estudios previos han demostrado que la resistencia al suero humano está relacionada con la producción de cápsula, y esta con la concentración de hierro en el medio. En consecuencia, se determinó la resistencia al suero humano de las variantes O y T.

En primer lugar, se realizó un control de viabilidad de ambos tipos celulares en PBS. Tras 4 h de cultivo, no se observaron diferencias en recuentos bacterianos con respecto al inicial, lo que sugería que ambas variantes son estables en solución salina (Tabla 6, Figura 24 del Anexo 4).

A continuación, ambas variantes se sembraron en HS. La variante O resistió en HS, HS+Fe y HS inactivado con y sin Fe (Tabla 6, Figura 25 del Anexo 4) mientras que la variante T moría tras 4 h de cultivo en HS con y sin hierro y sobrevivía en HS con el complemento inactivado (Tabla 6, Figura 26 del Anexo 4). No se apreciaron diferencias importantes en estos experimentos dependiendo de si el inóculo procedía de una colonia aislada en medio sólido o de medio líquido (10 µL).

Tabla 6. Ensayo de resistencia al suero de las variantes O y T de la cepa CECT9109, tanto a partir de colonia aislada en medio sólido como con 10 µL de cultivo en medio líquido. Se muestran los valores obtenidos del recuento celular al momento inicial del ensayo (t = 0 h) y tras 4 y 24 h, tanto en PBS 1X (control de no toxicidad celular ni de resistencia), en suero humano (HS), suero humano suplementado con hierro (HS + Fe), suero humano inactivado (HS in) y suero humano inactivado suplementado con hierro (HS in + Fe). Los valores representan el total de células en el vial de cultivo.

			PBS 1X	HS	HS + Fe	HS in	HS in + Fe
t = 0 h	O	colonia	3,2×10 ³	6,8×10 ³	8,8×10 ³	6,8×10 ³	5,2×10 ³
		líquido	1,9×10 ³	2,9×10 ³	2,3×10 ³	0,4×10 ³	2,7×10 ³
	T	colonia	6,4×10 ³	6,4×10 ³	4,4×10 ³	7,2×10 ³	6,4×10 ³
		líquido	1,7×10 ³	4,8×10 ³	1,6×10 ³	2,1×10 ³	2,9×10 ³
t = 4 h	O	colonia	7,6×10 ³	10,0×10 ³	6,0×10 ³	8,8×10 ³	5,6×10 ³
		líquido	3,2×10 ³	2,5×10 ³	1,5×10 ³	2,33×10 ³	3,0×10 ³
	T	colonia	6,0×10 ³	0	0	7,2×10 ³	5,6×10 ³
		líquido	3,1×10 ³	0	0	2,0×10 ³	3,3×10 ³
t = 24 h	O	colonia	2,4×10 ³	Incontable*	0	Incontable*	Incontable*
		líquido	1,3×10 ³	0	Incontable*	0,2×10 ³	2,3×10 ³
	T	colonia	1,1×10 ³	0	0	incontable	38,8×10 ³
		líquido	6,8×10 ³	0	0	88,0×10 ³	72,0×10 ³

*Incontable: debido al gran número de colonias crecidas, resultó difícil realizar el recuento.

6. Discusión

V. vulnificus es un patógeno multi-hospedador que puede llegar a ser zoonótico ², y que, debido al cambio climático, se está extendiendo desde zonas acuáticas salobres cálidas, de donde era autóctono, a zonas tradicionalmente más frías, como el mar Báltico ¹. Se considera un patógeno de riesgo porque puede causar la muerte por septicemia tras ingestión de marisco crudo en menos de 24 h ⁴. En consecuencia, resulta de vital importancia conocer cómo este microorganismo causa la septicemia para poder diseñar maneras de combatirlo y reducir su incidencia.

La virulencia de esta especie de *Vibrio* es multifactorial ⁵, destacando la producción de factores lesionales como múltiples toxinas y exoenzimas y la de una cápsula externa que protege al patógeno de la inmunidad innata en sangre. La producción de la cápsula está controlada por la concentración de hierro en el medio siendo máxima a las concentraciones asociadas al alcoholismo o la hemocromatosis, enfermedades que predisponen a la muerte por septicemia ⁶. Curiosamente, se ha descrito en esta especie que la cápsula puede perderse y las colonias O revertir a T y viceversa con una frecuencia entre 10^{-4} y 10^{-5} ⁶. El fenómeno que da lugar a esta reversión fue estudiado en la variante que produce el serotipo capsular CPS-1 pero no se llegó a ninguna conclusión por lo que se sugirió que podría deberse a una regulación transitoria hacia abajo de la expresión del gen *wzb* ⁷. Se desconoce si este fenómeno explica la reversión en el resto de los serotipos.

En el laboratorio de la Dra. Amaro se produjo esta reversión capsular de manera espontánea en una cepa clínica capsulada, la CECT9109, no volviendo la variante T a su estado normal capsulado. El presente TFM se centró en descubrir el mecanismo de reversión producido en esta cepa bacteriana a partir de estudios *in silico* y experimentos *in vitro*.

En primer lugar, ambas variantes O y T de la cepa CECT9109 se secuenciaron con Illumina *MiSeq* ⁸, y la variante O, además, con PacBio ⁹. Mientras que la primera ofrece lecturas cortas de entorno 200 pb y de alta calidad y fiabilidad, la segunda da como resultado lecturas muy largas, de más de 10 kpb. Se ha descrito, pues, que la combinación de ambos tipos de lecturas permite obtener un resultado de alta calidad y precisión ^{21,40}.

Tras testar diferentes herramientas para el ensamblaje, se encontró que *Unicycler* ¹² es la que rindió los mejores resultados, tanto para el ensamblaje de las lecturas cortas como para el ensamblaje híbrido con ambos tipos de lecturas. En este último caso, el ensamblado muestra el genoma cerrado, con 2 *contigs* circularizados, siendo los dos cromosomas de la bacteria. Por lo que respecta al ensamblaje únicamente con las lecturas de PacBio de la variante O, *Flye* ³⁵, por

su lado, consiguió cerrar y circularizar uno de los *contigs*, el correspondiente al cromosoma 2 de la cepa.

En trabajos previos se ha descrito que el uso combinado de tecnologías de secuenciación ofrece muy buenos resultados en cuanto a poder cerrar un genoma ²¹⁻²³. Y es que mientras las lecturas cortas de alta calidad no consiguen ensamblarse completamente, con las largas se facilita este ensamblaje, manteniendo la alta calidad de la secuencia.

A continuación, se procedió al mapeo de las lecturas de Illumina de la variante T utilizando el genoma cerrado de la variante O como referencia. Esta metodología resulta muy útil ya que cuando se esperan pocos cambios nucleotídicos entre secuencias, la técnica del mapeo abarata los costes económicos y permite obtener igualmente un buen resultado ^{25,26,46}.

Se utilizó la herramienta *Snippy* del gestor de programas *Conda*. Esta devolvió el genoma cerrado de la variante T de la cepa en estudio, como también un fichero con el llamado de variantes; este fichero indica los cambios significativos a nivel de nucleótido o de cadena de nucleótidos entre ambas secuencias estudiadas.

Finalmente, las secuencias de los genomas de la cepa CECT9109 se anotaron mediante *Prokka* ⁴². Esta aplicación compara con las bases de datos las secuencias de los genes identificados en los genomas, y los anota en un fichero *GenBank*, facilitando su estudio posterior.

Otro aspecto importante a tener en cuenta cuando se trabaja con patógenos bacterianos es determinar a qué linaje filogenético pertenecen, ya que ello permite comprender la evolución de la especie y predecir la emergencia de nuevas variantes virulentas ^{47,48}.

Tras el uso de la herramienta *PARSNP* ¹³ junto con *iTOL* para la visualización del árbol resultante, se concluyó que la cepa CECT9109 pertenece al Linaje 1, el linaje que engloba las cepas más peligrosas en salud pública que son las que se transmiten oralmente y causan muerte por septicemia primaria ⁴³. Por lo tanto, la cepa estudiada es relevante desde el punto de vista de salud pública.

Asimismo, se determinó si la cepa pertenecía a la pv. *piscis* y se encontró que no, por lo que se trataría de un aislado de Linaje 1 avirulento para peces.

A continuación, se analizó el locus *cps* y se comparó con el mismo locus del serotipo CPS-1 ⁷. Se localizaron 8 genes de los 18 descritos en la cepa MO6-24/O, todos ellos situados en el cromosoma 1, por lo que el tipo de cápsula producido por la cepa CECT9109 es distinto y, en consecuencia, el serotipo capsular también es distinto.

No se observaron diferencias importantes en el locus *cps* de ambas variantes de la cepa CECT9109 tras la visualización mediante *Easyfig*. Por lo tanto, se

procedió a realizar un análisis más fino a partir del llamado de variantes obtenido tras el mapeo de la variante T.

Se observó un solo cambio nucleotídico en la posición 2.964.050 en el locus *cps* del genoma de la variante T. Esta posición se encuentra en el gen *wlaL* y da lugar a un cambio no sinónimo, es decir, que ambas variantes difieren en un solo aminoácido de la proteína WlaL. Si el cambio mencionado en el aminoácido afecta a la funcionalidad de la proteína WlaL, debería demostrarse en posteriores estudios. Por otro lado, en el caso de las regiones aguas-arriba y abajo del gen *wzb* no se observaron diferencias. Estas regiones se han relacionado con un cambio de fenotipo irreversible de O a T y viceversa por un fenómeno de variación de fase ⁶. Este resultado concuerda con lo observado, dado que se produjo reversión, aunque en el caso de la O a la T, esa frecuencia fue muy baja.

El último análisis *in silico* fue la identificación de profagos en los genomas de la cepa en estudio. El análisis con los diferentes programas informáticos no mostró diferencias entre variantes, identificando un inovirus en ambos casos, localizado en la misma posición, por lo que la reversión tampoco puede achacarse al cambio de la posición del provirus en el genoma.

Se continuó con el análisis fenotípico mediante la visualización al TEM que permitió identificar un halo electrondenso entorno a las células de la variante O, que también estaba presente en algunas células de la variante T. Este resultado indica que el cultivo procedente de la variante T contenía una mezcla de células O y T. La mezcla de células O y T puede darse en un mismo cultivo si se ha producido la reversión durante el crecimiento. En consecuencia, la observación al TEM debería repetirse controlando que no se haya producido reversión en el cultivo antes de su procesamiento para microscopía electrónica.

A continuación, se realizó el ensayo de reversión capsular de ambas variantes. Encontramos que la tasa de reversión de O a T y de T a O tras un solo pase era muy baja, inferior a 10^{-7} , pero que se incrementaba considerablemente tras varios pases y solo en el caso de la variante T hasta alcanzar un valor superior al 50%. Este resultado es compatible con un fenómeno transitorio de represión de la transcripción de los genes biosintéticos de la cápsula en la variante T. Este patrón de transcripción iría desapareciendo con los pases sucesivos de modo que la bacteria acabaría revertiendo espontáneamente al fenotipo original, el capsulado. Harían falta estudios de transcriptómica para demostrar esta hipótesis. En cualquier caso, este experimento demostró la tendencia de las células capsuladas que revirtieron a acapsuladas a volver a su estado original ^{49,50} y, en definitiva, a un estado de mayor protección frente al entorno (cabe recordar que la cápsula protege a la bacteria frente al sistema inmunitario y otros estreses ambientales) ¹⁷.

Finalmente, se corroboró que la cápsula del nuevo serotipo protege a las bacterias frente al sistema inmunitario del hospedador dado que la variante O

resistía en suero, pero no la T, que no sobrevivió tras 4 h de cultivo en suero humano activo.

En resumen, se ha conseguido cerrar los genomas de la cepa CECT9109, con un alto grado de precisión. A nivel genómico solo se ha distinguido un cambio no sinónimo en la región del operón *cps* en ambos genomas de la cepa en estudio, que podría estar relacionado o no con la reversión capsular en la variante T. Sin embargo, lo más interesante es que la frecuencia de reversión del fenotipo T al O original se incrementa considerablemente con cada nuevo pase lo que sugiere que el fenómeno de reversión reside en un fenómeno regulatorio como ha sido sugerido para el serotipo CPS-1 de *V. vulnificus* ⁷, y que la variante T tiende a volver a la O original, la protegida frente al estrés ambiental. Finalmente, los experimentos *in vitro* demostraron que la cápsula de este nuevo grupo capsular protege frente a la acción del complemento presente en el suero humano favoreciendo el crecimiento de la bacteria en suero.

7. Conclusiones

7.1. Conclusiones

Como resultado del estudio realizado y en relación con los objetivos planteados, se han alcanzado las siguientes conclusiones:

1. Las variantes O y T de la cepa CECT9109 no pertenecen a la pv. *piscis*, por lo que esta cepa se consideraría avirulenta para peces.
2. Tras la secuenciación y ensamblado o mapeado, se ha conseguido cerrar los genomas de la cepa CECT9109, con un alto grado de precisión.
3. Ambas variantes O y T de la cepa en estudio pertenecen al Linaje 1, linaje que engloba las cepas clínicas más peligrosas ya que se transmiten oralmente y causan muerte por septicemia primaria, siendo pues relevante desde el punto de vista de salud pública.
4. A nivel genómico solo se ha distinguido un cambio no sinónimo en la región del locus *cps*, el cual ha dado lugar a un nuevo aminoácido, pudiendo estar relacionado o no con la reversión capsular en la cepa CECT9109(T).
5. El estudio de la presencia de profagos en los genomas de la cepa en estudio ha revelado la presencia de un inovirus, en la misma región en ambas variantes, por lo que la reversión capsular no puede relacionarse con la presencia y posición de este profago (fago insertado en el genoma).
6. La visualización en microscopía electrónica ha puesto en manifiesto que, mientras la muestra de la variante O se distinguían exclusivamente células capsuladas con alto grado de electrodensidad, en la variante T se observaron bacterias tanto capsuladas como acapsuladas (menos electrodensas).
7. La frecuencia de reversión del fenotipo T al O original se incrementa considerablemente con cada nuevo pase, sugiriendo que el fenómeno de reversión reside en un fenómeno regulatorio, y que la variante T tiende a volver al estado capsulado original, protegiéndose frente a factores ambientales que le provocan estrés.
8. La cápsula de la variante O la protege frente a la acción del complemento presente en el suero humano, favoreciendo su crecimiento en suero.

A la vista de todos los resultados obtenidos, se ha conseguido, en primer lugar, cerrar el genoma de las cepas CECT9109(O) y CECT9109 (T) con un alto grado de precisión. En segundo lugar, se ha localizado la cepa en estudio en el Linaje 1, siendo el más relevante desde el punto de vista clínico. Asimismo, en la región del operón *cps*, el cual es el encargado de la producción de la cápsula, se ha identificado un único cambio no sinónimo, que da lugar a un aminoácido diferente en la variante T; este hallazgo debe estudiarse con más detalle en un futuro. Finalmente, los experimentos *in vitro* han demostrado, por un lado, que la variante T tiende a revertir a su estado capsulado original, hecho que le protege del entorno, tras numerosos pases de cultivo, y por otro, que la cápsula protege a la célula del complemento presente en el suero humano, favoreciendo su crecimiento.

A lo largo de este TFM se han alcanzado todos los objetivos planteados, alterando levemente el inicio de algunas de las tareas a realizar para obtener resultados (identificación preliminar de profagos) mientras se esperaba a la finalización de otras tareas (en concreto, de la obtención de las lecturas de PacBio de la variante O).

7.2. Líneas de futuro

Dado que a nivel genómico solo se ha observado un cambio en la región del operón *cps*, el siguiente paso sería realizar un estudio transcriptómico, identificando qué genes de dicho locus se expresan en determinadas circunstancias, favorables o no a la producción de la cápsula.

Como se ha identificado una secuencia de profago en estos genomas (inovirus), sería interesante su inducción, extracción y purificación *in vitro* para observar su capacidad bacteriófaga sobre la cepa en estudio y si afecta también a otras cepas de la misma especie. Además, también cabe la posibilidad del uso de estos inovirus a modo de vector para introducir otros genes en el genoma.

Sería interesante detectar de la fracción lipopolisacárida (LPS) unida a la cápsula mediante ensayo con antígenos, comparando los resultados obtenidos entre ambas variantes O y T de la cepa CECT9109, y también con la nueva variante descubierta en este trabajo (CECT9109(R)).

Es fundamental repetir en ensayo de visualización de la cápsula al TEM, donde se obtuvieran muestras puras de las tres variantes (O, T y R) de la cepa en estudio.

Finalmente, se debe estudiar la influencia del hierro en el crecimiento de esta cepa.

7.3. Seguimiento de la planificación

A lo largo del desarrollo de este TFM se ha seguido la planificación para alcanzar los objetivos planteados. No obstante, como ya se ha comentado, algunas tareas (como la Identificación de profagos) se adelantaron a la espera de la finalización de otras (en concreto, la obtención de las lecturas largas de PacBio de la cepa CECT9109(O)). Este adelanto de la tarea de Identificación de profagos permitió un estudio preliminar de la presencia de estos profagos en el genoma de ambas variantes O y T a partir del ensamblado primario con las lecturas cortas de Illumina *MiSeq*.

En cuanto a la metodología seguida, ha sido correcta para la obtención de todos los objetivos planteados. Además, ha permitido comparar los resultados de los ensamblajes utilizando los diferentes programas informáticos descritos, consiguiendo, pues, un genoma cerrado de alta calidad y precisión.

8. Glosario

ADN: Ácido desoxirribonucleico

ARN: Ácido ribonucleico

CECT: Colección Española de Cultivos Tipo (Universitat de València)

FeCl₃: Cloruro de hierro

GBK: fichero *GenBank*

HS: *Human Serum* (suero humano)

HS+Fe: suero humano suplementado con hierro

Hi-Fi: High-Fidelity (alta fidelidad)

InDel: inserción-delección

L50: longitud de los *contigs* que, usando *contigs* de igual o mayor tamaño, produce la mitad de la longitud del genoma

Loci *cps*: operón de genes que codifican la producción de cápsula

N50: número más pequeño de *contigs* cuya suma de sus longitudes constituye la mitad del tamaño del genoma

NGS: *Next Generation Sequencing*

NJ: *Neighbour-Joining*

O: variante opaca de la cepa CECT9109

PacBio: *Pacific Bioscience*

pb: pares de bases

PCR: reacción en cadena de la polimerasa

pv.: patovariedad

T: variante translúcida de la cepa CECT9109

R: variante rugosa de la cepa CECT9109

rpm: revoluciones por minuto

SNP: Polimorfismo de nucleótido único

TEM: microscopía electrónica de transmisión

TFM: Trabajo Final de Máster

UV: Universitat de València

V. vulnificus: Vibrio vulnificus

VCF: fichero de llamado de variantes

9. Bibliografía

1. Böer, S. I. *et al.* Temporal and spatial distribution patterns of potentially pathogenic *Vibrio spp.* at recreational beaches of the German north sea. *Microb. Ecol.* **65**, 1052–1067 (2013).
2. Amaro, C. & Biosca, E. G. *Vibrio vulnificus* biotype 2, pathogenic for eels, is also an opportunistic pathogen for humans. *Appl. Environ. Microbiol.* **62**, 1454–1457 (1996).
3. Amaro, C. *et al.* The Fish Pathogen *Vibrio vulnificus* Biotype 2: Epidemiology, Phylogeny, and Virulence Factors Involved in Warm-Water Vibriosis. *Microbiol. Spectr.* **3**, (2015).
4. Oliver, J. D. The Biology of *Vibrio vulnificus*. *Microbiol. Spectr.* **3**, (2015).
5. Li, G. & Wang, M. Y. The role of *Vibrio vulnificus* virulence factors and regulators in its infection-induced sepsis. *Folia Microbiol. (Praha)*. **65**, 265–274 (2020).
6. Pettis, G. S. & Mukerji, A. S. Structure, Function, and Regulation of the Essential Virulence Factor Capsular Polysaccharide of *Vibrio vulnificus*. *Int. J. Mol. Sci.* **21**, (2020).
7. Chatzidaki-Livanis, M., Jones, M. K. & Wright, A. C. Genetic variation in the *Vibrio vulnificus* group 1 capsular polysaccharide operon. *J. Bacteriol.* **188**, 1987–1998 (2006).
8. Ravi, R. K., Walton, K. & Khosroheidari, M. MiSeq: A Next Generation Sequencing Platform for Genomic Analysis. *Methods Mol. Biol.* **1706**, 223–232 (2018).
9. Rhoads, A. & Au, K. F. PacBio Sequencing and Its Applications. *Genomics. Proteomics Bioinformatics* **13**, 278–289 (2015).
10. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
11. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
12. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13**, (2017).
13. Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **15**, 1–15 (2014).
14. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350–3352 (2015).
15. Dong, N. *et al.* Genome analysis of clinical multilocus sequence Type 11

- Klebsiella pneumoniae* from China. *Microb. Genomics* **4**, (2018).
16. Sanjuán, E. & Amaro, C. Multiplex PCR assay for detection of *Vibrio vulnificus* biotype 2 and simultaneous discrimination of serovar E strains. *Appl. Environ. Microbiol.* **73**, 2029–2032 (2007).
 17. Williams, T. C., Ayrapetyan, M., Ryan, H. & Oliver, J. D. Serum Survival of *Vibrio vulnificus*: Role of Genotype, Capsule, Complement, Clinical Origin, and *in Situ* Incubation. *Pathog. (Basel, Switzerland)* **3**, 822–832 (2014).
 18. Harris, J. R. Transmission electron microscopy in molecular structural biology: A historical survey. *Arch. Biochem. Biophys.* **581**, 3–18 (2015).
 19. Moyes, R. B., Reynolds, J. & Breakwell, D. P. Preliminary staining of bacteria: simple stains. *Curr. Protoc. Microbiol.* **Appendix 3**, Appendix 3E (2009).
 20. Hernández-Cabanyero, C. *et al.* Adaptation to host in *Vibrio vulnificus*, a zoonotic pathogen that causes septicemia in fish and humans. *Environ. Microbiol.* **21**, 3118–3139 (2019).
 21. Chen, Z., Erickson, D. L. & Meng, J. Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing. *BMC Genomics* **21**, 631 (2020).
 22. Wick, R. R. & Holt, K. E. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Research* **8**, 2138 (2021).
 23. Gavrielatos, M., Kyriakidis, K., Spandidos, D. & Michalopoulos, I. Benchmarking of next and third generation sequencing technologies and their associated algorithms for de novo genome assembly. *Mol. Med. Rep.* **23**, 251 (2021).
 24. Chin, C.-S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
 25. Hatem, A., Bozdağ, D., Toland, A. E. & Çatalyürek, Ü. V. Benchmarking short sequence mapping tools. *BMC Bioinformatics* **14**, 184 (2013).
 26. Nunn, A., Otto, C., Stadler, P. F. & Langenberger, D. Comprehensive benchmarking of software for mapping whole genome bisulfite data: from read alignment to DNA methylation analysis. *Brief. Bioinform.* **22**, (2021).
 27. Li, Z. P., Pang, B., Lu, X. & Kan, B. [Genomic recombination of the vibrio cholerae serogroup O1 El Tor pandemic strains]. *Zhonghua Yu Fang Yi Xue Za Zhi* **54**, 301–305 (2020).
 28. Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A. & Sun, F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* **5**, 69 (2017).
 29. Lu, C. *et al.* Prokaryotic virus host predictor: a Gaussian model for host prediction of prokaryotic viruses in metagenomics. *BMC Biol.* **19**, 5 (2021).
 30. Hurwitz, B. L., Ponsero, A., Thornton, J. & U'Ren, J. M. Phage hunters:

- Computational strategies for finding phages in large-scale 'omics datasets. *Virus Res.* **244**, 110–115 (2018).
31. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).
 32. Glickman, C., Hendrix, J. & Strong, M. Simulation study and comparative evaluation of viral contiguous sequence identification tools. *BMC Bioinformatics* **22**, 329 (2021).
 33. Roux, S. *et al.* Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nat. Microbiol.* **4**, 1895–1906 (2019).
 34. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
 35. Freire, B., Ladra, S. & Parama, J. R. Memory-Efficient Assembly using Flye. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 1–1 (2021) doi:10.1109/TCBB.2021.3108843.
 36. Chakraborty, M., Baldwin-Brown, J. G., Long, A. D. & Emerson, J. J. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res.* **44**, gkw654 (2016).
 37. Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D. & Gurevich, A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**, i142–i150 (2018).
 38. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
 39. Gan, H. M. & Austin, C. M. Nanopore long reads enable the first complete genome assembly of a Malaysian *Vibrio parahaemolyticus* isolate bearing the pVa plasmid associated with acute hepatopancreatic necrosis disease. *F1000Research* **8**, 2108 (2019).
 40. Austin, C. M. *et al.* De novo genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (*Maccullochella peelii*), from Illumina and Nanopore sequencing read. *Gigascience* **6**, (2017).
 41. Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One* **9**, e112963 (2014).
 42. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
 43. Roig, F. J. *et al.* Phylogeny of *Vibrio vulnificus* from the Analysis of the Core-Genome: Implications for Intra-Species Taxonomy. *Front. Microbiol.* **8**, (2018).
 44. Ishiyama, N. *et al.* Structural Studies of FlaA1 from *Helicobacter pylori* Reveal the Mechanism for Inverting 4,6-Dehydratase Activity. *J. Biol.*

- Chem.* **281**, 24489–24495 (2006).
45. Schirm, M. *et al.* Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol. Microbiol.* **48**, 1579–1592 (2003).
 46. Thankaswamy-Kosalai, S., Sen, P. & Nookaew, I. Evaluation and assessment of read-mapping by multiple next-generation sequencing aligners based on genome-wide characteristics. *Genomics* **109**, 186–191 (2017).
 47. Yang, Z. & Rannala, B. Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* **13**, 303–314 (2012).
 48. Chamberlain, S. A. *et al.* Does phylogeny matter? Assessing the impact of phylogenetic information in ecological meta-analysis. *Ecol. Lett.* **15**, 627–636 (2012).
 49. Phippen, B. L. & Oliver, J. D. Role of Anaerobiosis in Capsule Production and Biofilm Formation in *Vibrio vulnificus*. *Infect. Immun.* **83**, 551–559 (2015).
 50. Biosca, E. G., Llorens, H., Garay, E. & Amaro, C. Presence of a capsule in *Vibrio vulnificus* biotype 2 and its relationship to virulence for eels. *Infect. Immun.* **61**, 1611–1618 (1993).

10. Anexos

10.1. Anexo 1: Comandos (Ubuntu)

Análisis de la calidad

- *NanoPlot (Conda)*

```
NanoPlot -t 10 -o <creación_directorio> -p <prefijo_archivos> --fastq  
<lecturas_largas.fastq>
```

```
NanoPlot -t 10 -o <creación_directorio> -p <prefijo_archivos> --fasta  
<secuencia.fasta>
```

- *QUAST 5.0.2 (Python 2.0)*

```
quast.py <creación_directorio> -t 10 --gene-finding <secuencia.fasta>
```

- *BUSCO 5.3.2 (Python)*

```
BUSCO_v1.1b.py -o <creacion_directorio> -in <secuencia.fasta> -l  
vibrionales_0db10 -m genome
```

Ensambladores

- *SPAdes 3.15 (Python)*

```
spades.py -1 <lecturas_cortas_R1.fastq> -2 <lecturas_cortas_R2.fastq>  
-o <creación_directorio>
```

- *Unicycler 0.5*

```
Unicycler -1 <lecturas_cortas_R1.fastq> -2 <lecturas_cortas_R2.fastq>  
-l <lecturas_largas.fastq> ( -l <secuencia.fasta> ) -t 10 --out <creación_directorio>
```

- *Flye*

```
flye -t 10 --pacbio-fifi <lecturas_largas.fastq> -p <creación_directorio>
```

Scaffolding

- *Quickmerge (Conda)*

```
quickmerge -d <creación_directorio> -q <secuencia1.fasta> -r <secuencia2.fasta>
-p <prefijo_fichero>
```

Pulido del ensamblado

- *Pilon 1.23*

```
bwa index <secuencia.fasta> && bwa mem -t 10 <secuencia.fasta>
<lecturas_cortas_forward.fastq> <lecturas_cortas_reverse.fastq> | samtools
view - -Sb | samtools sort - -@10 -o Pilon_polish/mapped.sorted.bam
```

```
samtools index Pilon_polish/mapped.sorted.bam
```

```
java -Xmx28G -jar pilon-1.23.jar --genome <secuencia.fasta> --fix all --changes
--frags Pilon_polish/mapped.sorted.bam --threads 20 --output
Pilon_polish/roundX | tee Pilon_polish/roundX.pilon
```

Mapeo

- *Snippy (Conda)*

```
snippy --cpus 10 --ram 24 --prefix <prefijo_archivos> --mapqual 60 --mincov 10
--minfrac 0.8 --ref <secuencia_referencia.fasta> --R1 <lecturas_cortas_R2.fastq>
--R2 <lecturas_cortas_R2.fastq> --outdir <creación_directorio>
```

Anotación de los genomas

- *Prokka 1.14.6*

```
prokka --outdir <creación_directorio> --prefix <prefijo_archivos>
--species V.vulnificus --strain <cepa> --kingdom Bacteria <secuencia.fasta>
```

Comparación de genomas

- *BLASTn*

```
blastn -query <secuencia1.fasta> -subject <secuencia2.fasta>
-out <fichero_salida.csv> -outfmt 6
```

Extracción de regiones genómicas de interés

- Programa desarrollado en el grupo de la Dra. Amaro (*Python 2.0*)

```
extract_fasta_from_header.py -f <secuencia.fasta> -header <parte_encabezado>  
-b <posición_inicio> -e <posición_final> -o <fichero_salida.fasta>
```

Filogenia

- *PARSNP 1.2*

```
parsnp -r YJ016.fasta -d <directorio_donde_estan_las_secuencias> -p 10 -c -x
```

Detección de profagos o regiones víricas

- *VirSorter2 (Conda)*

```
virsorter run -w <creación_directorio> -i <secuencia.fasta> --include-groups  
"dsDNAphage,NCLDV,RNA,ssDNA,lavidaviridae" -j 10 all
```

- *CheckV (Conda)*

```
checkv end_to_end "<directorio_anterior>/final-viral-combined.fa"  
"<directorio_anterior>/checkv" -t 2 -d checkv-db-v1.0
```

- *Inovirus detector*

```
while read line
```

```
do
```

```
./Identify_candidate_fragments_from_gbk.pl -g "/media/hector/247f3592-a263-  
4341-99f5-3c250643c87e/TFM_RSC/12-2_Fagos_Inovirus-detector/$line.gbk" -  
p Pfam-A.hmm -sp5 /home/hector/programas/SignalP/signalp-5.0b/bin/ -th  
/home/hector/programas/Tmhmm/tmhmm-2.0c/bin/tmhmm -t 10 &>  
/media/hector/247f3592-a263-4341-99f5-3c250643c87e/TFM_RSC/12-  
2_Fagos_Inovirus-detector/$line.gbk.log
```

```
done < /media/hector/247f3592-a263-4341-99f5-3c250643c87e/TFM_RSC/12-  
2_Fagos_Inovirus-detector/archivos_usados/secuencias
```

```
annot="_annot.gff"
```

```
nucl="_nucl.fna"
```

```
log="_log2"
```

```
while read line
```


do

```
./Get_inovirus_prediction_score_from_gff_fragments.pl -i  
"/media/hector/247f3592-a263-4341-99f5-3c250643c87e/TFM_RSC/12-  
2_Fagos_Inovirus-detector/$line$annot" -f "/media/hector/247f3592-a263-4341-  
99f5-3c250643c87e/TFM_RSC/12-2_Fagos_Inovirus-detector/$line$nucl" &>  
"/media/hector/247f3592-a263-4341-99f5-3c250643c87e/TFM_RSC/12-  
2_Fagos_Inovirus-detector/$line$log"
```

```
done < /media/hector/247f3592-a263-4341-99f5-3c250643c87e/TFM_RSC/12-  
2_Fagos_Inovirus-detector/archivos_usados/putative_regions_annot
```

while read line

do

```
./Identify_candidate_fragments_from_fna.pl -f "/media/hector/247f3592-a263-  
4341-99f5-3c250643c87e/TFM_RSC/12-2_Fagos_Inovirus-detector/$line.fasta"  
-p Pfam-A.hmm -sp /home/hector/programas/SignalP/signalp-4.1/signalp -th  
/home/hector/programas/Tmhmm/tmhmm-2.0c/bin/tmhmm -t 10 &>  
/media/hector/247f3592-a263-4341-99f5-3c250643c87e/TFM_RSC/12-  
2_Fagos_Inovirus-detector/$line.fna.log
```

```
done < /media/hector/247f3592-a263-4341-99f5-3c250643c87e/TFM_RSC/12-  
2_Fagos_Inovirus-detector/archivos_usados/secuencias
```

10.2. Anexo 2: Medios de cultivo

Las bacterias, como todo ser vivo, necesitan nutrirse y un medio en el que vivir. Existen dos posibilidades de medio donde cultivar estas células: medio líquido (también llamado caldo) y medio sólido (sobre placas recubiertas de agar suplementado con diferentes nutrientes).

Por lo que respecta al medio líquido, este consistió en *Lennox Broth medium* (LB; Laboratorios Conda S.A., España) suplementado con 0,5 % de NaCl (Panreac, EE.UU.); este suplemento de sodio hace que el medio líquido tenga una concentración final del 1 %.

Por lo que respecta al medio sólido, se utilizó *Tryptic Soy Agar* (TSA; Liofilchem srl, Italia), también suplementado con 0,5 % de NaCl, teniendo una concentración final del 1 %.

Todos los medios de cultivo se prepararon con agua destilada, y se esterilizaron mediante autoclave.

10.3. Anexo 3: PCR y electroforesis

En primer lugar, se extrajo el ADN de las bacterias mediante la técnica de ebullición. Tras un cultivo en medio sólido de las variantes O y T de la cepa CECT9109 y de la cepa R99, se resuspendió una colonia aislada de cada una de ellas en 100 μ L de agua molecular estéril (HyPure Molecular Biology Grade Water; HyClone Laboratories, EE.UU.), y se llevaron a ebullición durante 10 min, enfriando posteriormente en hielo al menos 5 min. A continuación, se centrifugó el lisado celular a 10.000 rpm durante 10 min a 4 °C. El sobrenadante contiene el ADN, por lo que se guardó a -20 °C hasta su uso.

Para la amplificación por PCR de los genes de interés, se realizó un preparado con los diferentes reactivos necesarios y ADN para esta amplificación.

En la siguiente tabla se listan los componentes necesarios para la realización de la PCR:

Tabla 7. Listado de componentes necesarios para la amplificación mediante PCR de los genes *vwba* y *vep07* para la identificación de *V. vulnificus* y de la *pv. piscis*, respectivamente.

Reactivo (empresa)	Volumen (μ L/muestra)
Agua Molecular (HyClone Laboratories, EE.UU.)	11,0
Speedy Supreme NZYTMTaq 2x Green Master Mix (NZYTech, Portugal)	12,5
Primer del gen diana – Forward (Idtdna, Bélgica)	0,5
Primer del gen diana – Reverse (Idtdna, Bélgica)	0,5
ADN de estudio (propio)	0,5

Y los ciclos junto con su temperatura empleados en estas PCR se muestran en la siguiente tabla. Este programa se estableció a partir de los requerimientos expuestos por parte de la empresa fabricante de la *Master Mix* y de los *primers*.

Tabla 8. Programa para la ejecución de la PCR.

Ciclo(s)	Tiempo (min:seg)	Temperatura (°C)
1	05:00	95,0
	00:02	94,0
30	00:05	55,0
	00:10	72,0
	02:00	72,0
1	∞	4

Finalmente, para la realización de la electroforesis, se preparó un gel de agarosa al 1,8 % (*Agarose D1 Low EEO*; Laboratorios Conda S.A.) suplementado con 1 µL/10 µL de *Red Safe (Nucleic Acid Staining Solution*; iNtRON Biotechnology DR, Corea del Sur), colorante que permitirá la visualización de las bandas en el transiluminador. Una vez el gel de agarosa solidificó, se cargaron 5 µL de muestra y de marcador de pesos moleculares (*GeneRuler* de ADN de 1 kb; ThermoFisher Scientific, EE.UU.) en los pocillos del mismo; el marcador, durante la electroforesis, se distribuirá según pesos moleculares a lo largo del gel, permitiendo la identificación de las bandas aparecidas con su peso molecular (equivalente al número de pares de bases).

La electroforesis se realizó a 100 V durante 30 min, visualizando el resultado de la misma mediante un transiluminador Bio-1000F (Microtek, Taiwan). Gracias al marcador de pesos moleculares, se pudo hacer la correspondencia de las bandas observadas con el número de pares de bases amplificadas, las cuales se corresponden con el número de pares de bases de la sonda génica utilizada (el gen de interés). Así pues, la visualización de banda significa la existencia del gen de estudio.

10.4. Anexo 4: Imágenes de resultados

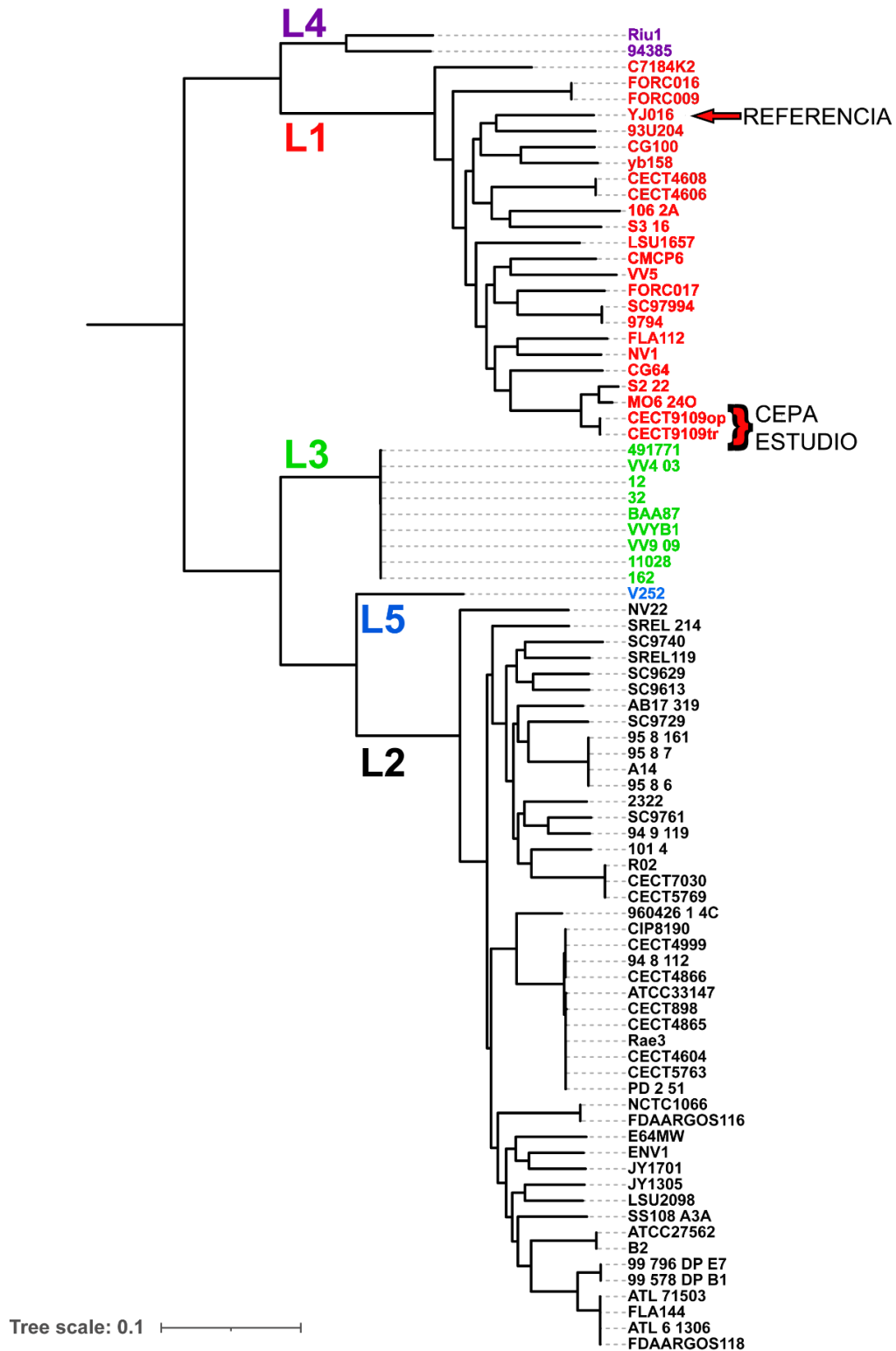


Figura 22. Árbol filogenómico NJ con 81 cepas ya analizadas previamente⁴³ junto con las variantes O y T de la cepa CECT9109, obtenido mediante PARSNP y visualizado con iTOL. Se tomó como cepa referencia la YJ016. Se observa que ambas variantes de la cepa de estudio pertenecen al linaje 1.

CLUSTAL O(1.2.4) multiple sequence alignment

```

uniprot      ----MIFYKSKR---LAFFLTSDIVLILLSVYLAFSLRFSGDIPSIFYHGMVMSAIILLV 53
opaca_manual MDKLAYIWSLPRVHKRLISLAIDTLITFSFFMAIWRHGEVAVSVSSETL-LTLAGTVV 59
translucida_manual MDKLAYIWSLPRVHKRLISLAIDTLITFSFFMAIWRHGEVAVSVSSETL-LTLAGTVV 59
               ::  *      : * : * : * : * : * : * : * : * : * : * : * : *

uniprot      LKLSFLFVFRIVKVAWRFFSLNEARKIFIALLLAEFCFLIFYFFSDFFNPFRSAIVID 113
opaca_manual VTLVTFKLGLYRAVLRYLTFHALTVVVLGALISALSITTFAYFFN---AEPRTVPVIY 116
translucida_manual VTLVTFKLGLYRAVLRYLTFHALTVVVLGALISALSITTFAYFFN---AEPRTVPVIY 116
               : * : : : * : : * : : : : : : : : : : * : : : * : : * : * : *

uniprot      FVLSYMFIGTLRISKRMVDFKPSKMKEEETPCIVVGATSKALHLLKGAKEGSLGLFPVG 173
opaca_manual MTFLLALLCGGARMMVRSIVQAS--RKGCEVLIYAGSTGRQLAIALRN-AETYQVKG 172
translucida_manual MTFLLALLCGGARMMVRSIVQAS--RKGCEVLIYAGSTGRQLAIALRN-AETYQVKG 172
               : : : : * * : * * : : : : : * * * : * : : : : *

uniprot      VVDARKELIGTYCDKFVVEEKEIKISYVEQGV-KTAIAL-RLEQEELKKLFEELVAY-- 229
opaca_manual FIDNDPSLENTIIQGLTVHSSQIISRLVEKQEIEKILLAMPRAISRERKAIIDGLLHLPV 232
translucida_manual FIDNDPSLENTIIQGLTVHSSQIISRLVEKQEIEKILLAMPRAISRERKAIIDGLLHLPV 232
               : * : * : * : : : * : : * : : : : * : : * : * : : * : * : *

uniprot      -----GICDVKIFSFRNEARDISIEDLLARKPKDLDSSAVAFLKDKVVLVSGAGGTI 283
opaca_manual EVLTVPDFKDIVNGNATVDELKDVAIEDLLGRDPVEPNPELMKANIRGKVVMTGAGGSI 292
translucida_manual EVLTVPDFKDIVNGNATVDELKDVAIEDLLGRDPVEPNPELMKANIRGKVVMTGAGGSI 292
               : : * : * : * : * : * : * : * : : : * : : * : * : * : * : *

uniprot      GSELCKQCIKFGAKHLIMVDHSEYNLYKINDDLNLKYE-----KITPILLSILDKQSLD 337
opaca_manual GSELCRQIVRQKPKTLILFELSEYGLYEIDKELSGMVEAMQLEVEIIPLLGSVQRINRLS 352
translucida_manual GSELCRQIVRQKPKTLILFELSEYGLYEIDKELSGMVEAMQLEVEIIPLLGSVQRINRLS 352
               * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

uniprot      EVLKTYKPELILHAAAYKHVPLCEQNPASAVINNIIGTKILCDSAKENKVAKFMISTDK 397
opaca_manual ATMRAFVQVTYHAAAYKHVPLVEYVNVGVRNNVFGTYSSAKAAIEAGVESFVLISTDK 412
translucida_manual ATMRAFVQVTYHAAAYKHVPLVEYVNVGVRNNVFGTYSSAKAAIEAGVESFVLISTDK 412
               : : : : : * : * : * : * : * : * : * : * : * : * : * : * : *
               ↑
uniprot      AVRPTNIMGCTKRVCLEYLTLMSDE-----NFEVACVRFGNVLGSSGSVIPKFKAIAN 451
opaca_manual AVRPTNVMGTSKRMAELALQALAAKENDKVNTRFCMVRFGNVLGSSGSVIPLFKRQIEE 472
translucida_manual AVRPTNVMGTSKRMAELALQALAAKENDKVNTRFCMVRFGNVLGSSGSVIPLFKRQIEE 472
               * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

uniprot      NEPLTLTHPDIVRYFMLVAEAVQLVLQAGAIKGGELFVLDMGKPKVIIDLAKKMLLLSN 511
opaca_manual GQAITVTHPDIIIRYFMTIPEAAQLVIQAGAMGKGGDVFVLDMGEPVKIVDLAKNLIQLSG 532
translucida_manual GQAITVTHPDIIIRYFMTIPEAAQLVIQAGAMGKGGDVFVLDMGEPVKIVDLAKNLIQLSG 532
               : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

uniprot      -----RNDLEIKITGLRKGEKLYEELIDENAKTQYESIFVAKNEKVDLDWLNKE 562
opaca_manual LEVKSSDNPNGDIEIKFTGLRPGEKLYEELIGDNVEGTDHERIMTANEQFLPLEEFNQI 592
translucida_manual LEVKSSDNPNGDIEIKFTGLRPGEKLYEELIGDNVEGTDHERIMTANEQFLPLEEFNQI 592
               : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

uniprot      IENLQ-ICEDISEALLK----IVPEFKHNKEGI----- 590
opaca_manual LDNLDRACHEFDHETIRQILLETPGFNPTDGIQDLVWNAKRKLNASKDKVVEIKVTA 650
translucida_manual LDNLDRACHEFDHETIRQILLETPGFNPTDGIQDLVWNAKRKLNASKDKVVEIKVTA 650
               : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *
    
```

Figura 23. Imagen obtenida con Clustal Omega en el que se comparan las secuencias de aminoácidos de la proteína resultante del gen analizado en el que se ha encontrado la diferencia entre ambas las variantes O y T de la cepa CECT9109 con el llamado de variantes. La flecha negra indica el aminoácido resultante diferente en la variante T. Se observa una elevada similitud entre la secuencia de aminoácidos descrita para este gen y las secuencias obtenidas mediante la anotación con Prokka para los genomas estudiados.

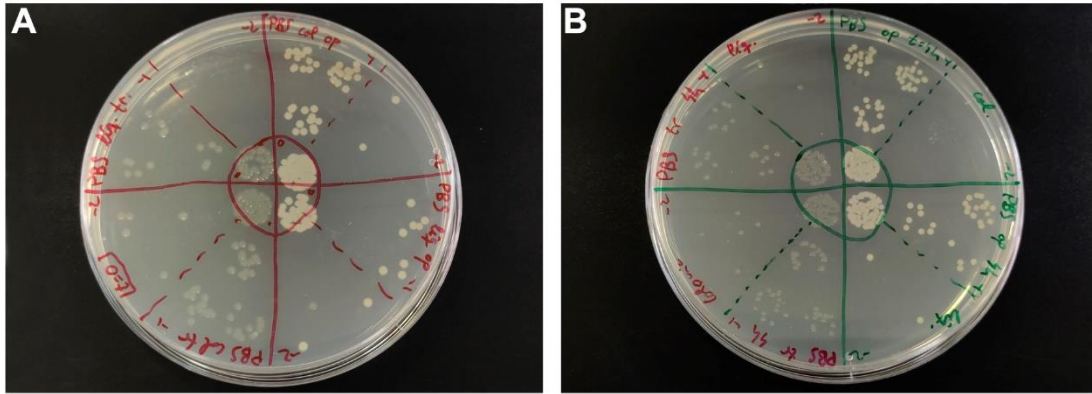


Figura 24. Ensayo de la resistencia al suero de las variantes O y T de la cepa CECT9109; las imágenes muestran el cultivo de estas bacterias en PBS 1X a modo de control del experimento. Estas imágenes corresponden a placas sembradas por sectores con muestras cultivadas en PBS a tiempo 0 (A) y tras 4 h de incubación (B). En cada sector se sembraron 3 gotas de 10 μ L correspondientes a diluciones decimales diferentes (10^0 en el caso de la siembra directa a partir del cultivo en PBS, y 10^{-1} y 10^{-2} a partir de dicho cultivo).

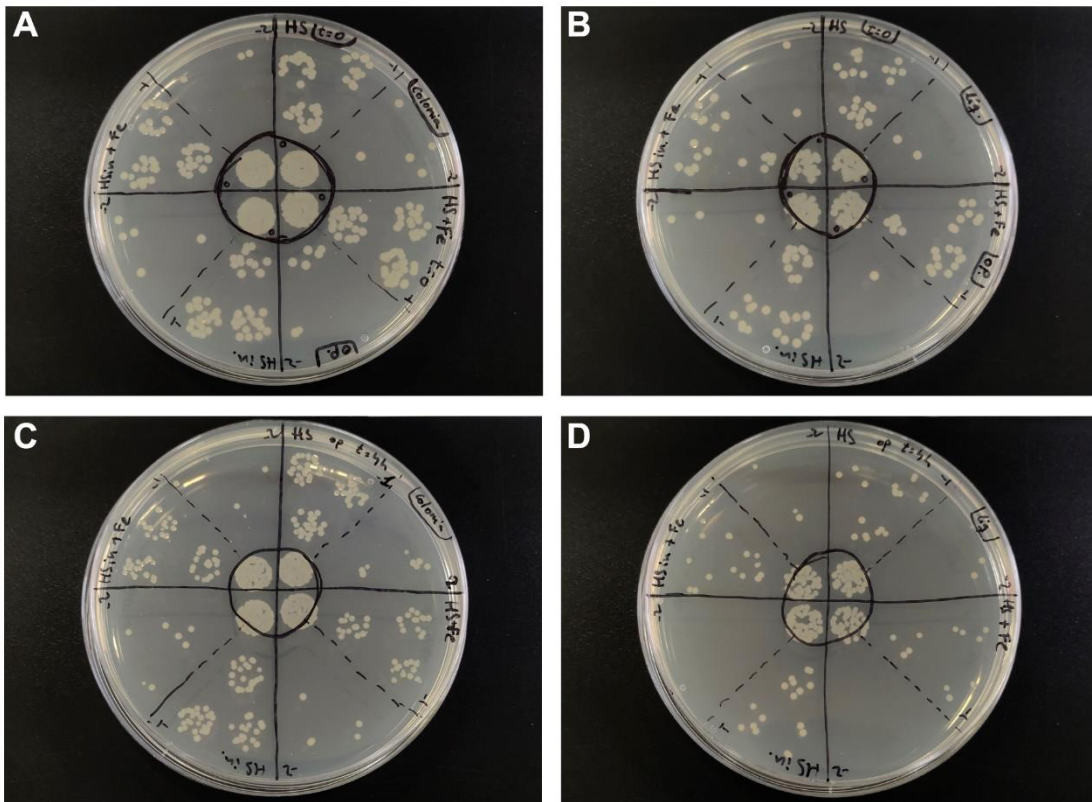


Figura 25. Ensayo de la resistencia al suero de la cepa CECT9109(O). Imágenes a tiempo cero de cultivo (A y B) y tras 4 h de cultivo a 37 °C (C y D), cultivadas en suero humano (HS), suero humano suplementado con hierro (HS + Fe), suero humano inactivado (HS in) y suero humano inactivado suplementado con hierro (HS in + Fe). Las imágenes de los paneles A y C se corresponden con las bacterias provenientes de una colonia aislada en medio sólido, mientras que los paneles B y D se corresponden con las bacterias provenientes de medio líquido (10 μ L). En cada sector se sembraron 3 gotas de 10 μ L correspondientes a diluciones decimales diferentes (10^0 en el caso de la siembra directa a partir del cultivo en los diferentes sueros, y 10^{-1} y 10^{-2} a partir de dichos cultivos).

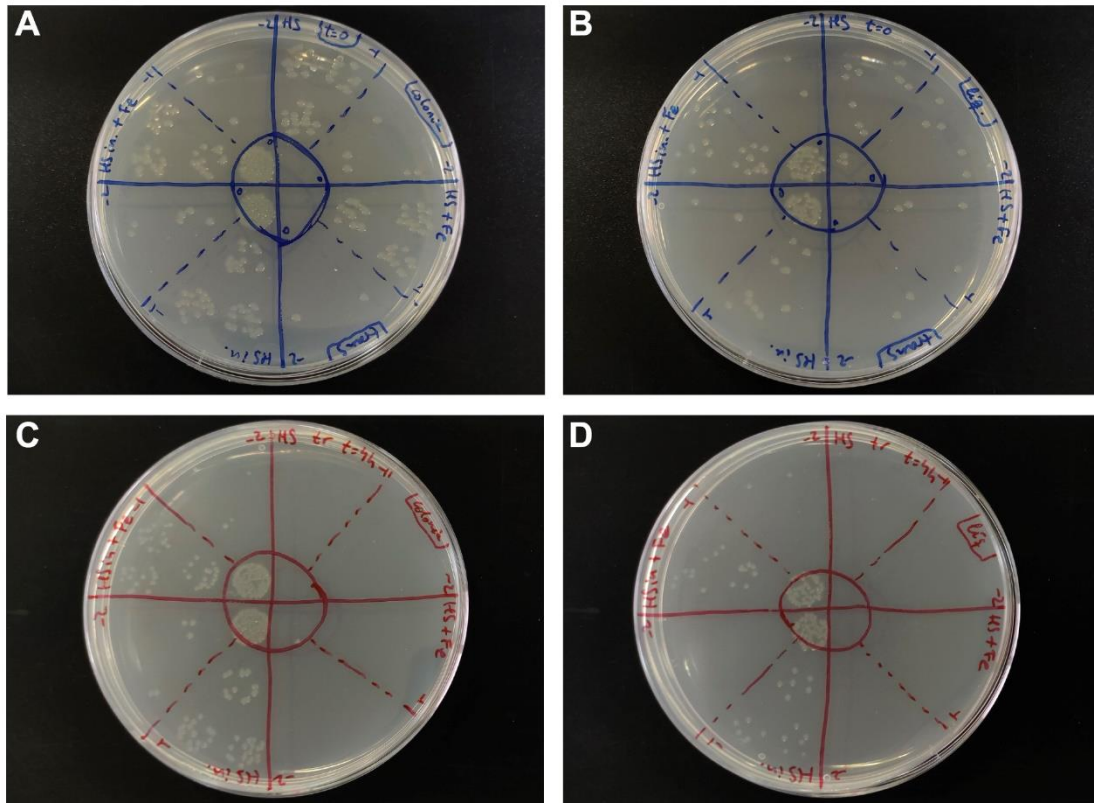


Figura 26. Ensayo de la resistencia al suero de la cepa CECT9109(T). Imágenes a tiempo cero de cultivo (A y B) y tras 4 h de cultivo a 37 °C (C y D), cultivadas en suero humano (HS), suero humano suplementado con hierro (HS + Fe), suero humano inactivado (HS in) y suero humano inactivado suplementado con hierro (HS in + Fe). Las imágenes de los paneles A y C se corresponden con las bacterias provenientes de una colonia aislada en medio sólido, mientras que los paneles B y D se corresponden con las bacterias provenientes de medio líquido (10 μ L). En cada sector se sembraron 3 gotas de 10 μ L correspondientes a diluciones decimales diferentes (10^0 en el caso de la siembra directa a partir del cultivo en los diferentes sueros, y 10^{-1} y 10^{-2} a partir de dichos cultivos).