



UNIVERSITAT DE
BARCELONA

Duplicaciones génicas y virulencia en *Escherichia coli*

Manuel Bernabeu Lorenzo

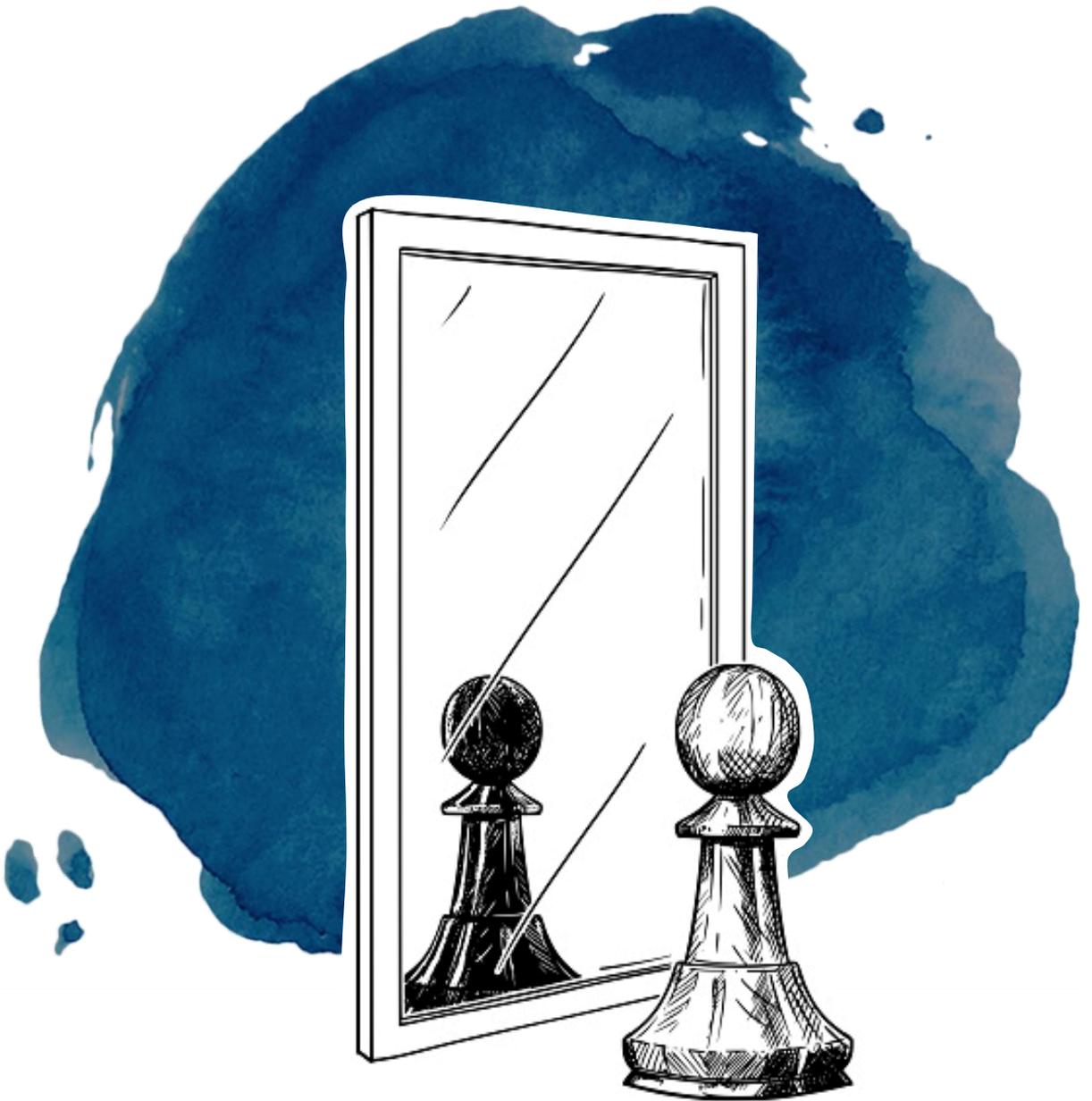
ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Duplicaciones génicas y virulencia en *Escherichia coli*

Manuel Bernabeu Lorenzo



2021



UNIVERSITAT DE BARCELONA

Universidad de Barcelona, Facultad de Biología
Departamento de Genética, Microbiología y Estadística
Sección de Microbiología, Virología y Biotecnología

Duplicaciones génicas y virulencia en *Escherichia coli*

Memoria presentada por Manuel Bernabeu Lorenzo para optar al
grado de Doctor por la Universidad de Barcelona
Programa de Doctorado EEES (Espacio Europeo de Educación Superior)
de Biotecnología

Visto bueno del director y tutor

Prof. Antonio Juárez Giménez

Visto bueno del codirector

Dr. Mário Hüttener Queiroz

Memoria presentada por

Manuel Bernabeu Lorenzo

Barcelona, 2021

La presente tesis doctoral ha sido financiada mediante la Ayuda para la contratación de personal investigador novel (FI-2017) de la Generalitat de Catalunya.

Agradecimientos

El tiempo pasa volando, me han dicho una y otra vez. No les faltaba razón, hace dos días estaba empezando la tesis y, ahora, ya la tengo aquí. Sin embargo, antes de empezar con la tesis propiamente dicha, me reservo este espacio para expresar mis sentimientos. Los que me conocen saben que soy el primero en opinar, que me cuesta muy poco decir lo que pienso. No obstante, los que me conocen todavía mejor saben que conmigo mismo soy como una cebolla, que me pongo capas y que, aunque siempre soy sincero, me cuesta expresar lo que siento. Por eso, escribo estas líneas, para que, además de conocer todo el trabajo realizado estos años, me conozcáis mejor a mí.

Han sido unos años increíbles, una etapa enriquecedora, por supuesto que a nivel laboral pero, lo que más valoro, a nivel personal. Han sido unos años muy movidos, con mucho trabajo y, para colmo, hasta una pandemia. No ha sido fácil, ha sido un camino largo, menudo camino (¡y encima duplicado!). Un camino lleno de baches, de caerme e, incluso, de desesperarme. Pero, por otra parte, y mucho más importante, un camino donde he aprendido a levantarme y seguir y, sobre todo, un camino donde muchos me han dado ese empujón que necesitaba. Les dedico este rincón a todos ellos, porque son parte de este trabajo, de esta etapa y, por supuesto, de mi vida.

En primer lugar, dar las gracias al director de esta tesis, Antonio Juárez, por darme la oportunidad de entrar en su grupo de investigación. Por su compromiso con este trabajo, desde responderme a correos en verano sin conocerme y acompañarme a hacer reclamaciones en la unidad de becas hasta la dirección de la tesis. Pese a todos los temas que llevas en marcha siempre había momentos para hacer reunión. Esas reuniones para estructurar y dirigir la tesis, sugiriendo experimentos y compartiendo conocimientos. Y, después de hablar de ciencia, ese tiempo para interesarte de cómo va todo y de enseñarme cómo funciona el sistema, ha sido una experiencia muy gratificante. Agradecer

también todo el esfuerzo realizado por mantener a todo el grupo, ha quedado claro que no es fácil.

Seguidamente, me gustaría agradecer al codirector de la tesis, Mário Hüttener. Muchas gracias por todos los consejos que me has dado, por darle explicación a esos resultados que no tenían sentido y por aportar ese toque de humor a cada situación, acompañado de un buen café. Gracias por enseñarme desde lo más básico a lo más complicado, me parece que los plásmidos ya no me van a dar problemas. Siempre me acordaré del viaje a Glasgow, aunque yo no me llevara el patito de goma.

A Sonia Aznar, por su ayuda durante estos años, por el trabajo que hace cada día y su gestión, imprescindible en el grupo. Gracias por tus consejos, por preguntar y, sobre todo, por tus risas cómplices. A David, vaya madrugones que te pegas. El que menos tiempo lleva en el grupo, pero con el que comparto horas de poyata y me ha facilitado todo el trabajo este último año. Gracias por toda tu ayuda, por entrar tan rápido en nuestra dinámica y por hacernos reír tanto. A Mari, por ser tan agradable cada vez que aparecía por el laboratorio del Parc.

A Ale, porque cada mañana, desde el primer día, nos hemos mandado un mensaje con un *ayayayayay*. No se puede tener mejor compañero, desde que empecé el TFM hasta que he acabado la tesis. Desde que te vi tirando las muestras de ARN por darme una cuchara que se me había caído hasta decir *aparcao* cuando escribí el punto final de la tesis. Horas y horas juntos, a saber cuántas, pero todas llevaderas compartiéndolas contigo. Eres un gran científico y una gran persona. Gracias por aportar ese toque de caos que también hace falta, por repetir la misma tontería para ver cuándo me voy a picar, por ser el primero en ofrecerse a ayudar y estar para todos. Al principio te venías conmigo a revelar cualquier gel para ver cómo había salido, al final casi que no

lo revelaba si no estabas tú. Me faltan muchas palabras para agradecerte todo lo que hemos compartido, has sido una pieza clave en esta etapa.

A la otra parte del laboratorio 3, Gemma y Edu, por tantas risas y cotilleos, y un “vamos a cerrar la puerta, a ver qué pasa”. Algo para picar, dulce o salado, mejor salado que si no se lo come Ale. Y a los que ya se han ido, Montse, Aurora y Júnior, gracias compañeros.

A todos los del departamento, de fase 1 y de fase 2, que nos cruzamos todos los días y que esto no funcionaría si no estuviéramos. A todo el personal de secretaría, Bea, Rosario, Manolo, Mónica y Susana, gracias por ser siempre tan agradables y ayudar con cada papel que hay que rellenar.

A los Hulios, Cris, Ale, Genoveva, Sandra, Álvaro, Sara, Óscar, Gala, Sergio, Pedro y Eva, cómo me gusta el grupo que hemos hecho. Tan diverso, con su grupo de intolerancias, con su grupo de sensatos, pero todos siempre juntos. Gracias por todo lo que hemos hecho, cada comida, cena, casas rurales y todos los juegos de mesa (incluso los cooperativos, aunque sean para perdedores). Es que no puedo acabar llorando de la risa cada vez que nos juntamos, gracias por tantos momentos y por aportar tantos puntos de vista. Me reservo un espacio para Sandra, gracias por todas las conversaciones con un café y estar para todo, no se puede ser mejor persona.

Al grupo GEMAT, por acogerme cada vez que me paso por el IQS, gracias por recibirme con los brazos abiertos desde el primer día. Qué buenos momentos en la Torreta y, por supuesto, *Firaso!* No me puedo olvidar de Tractivus, Joan, Robert, Elena y Cris (bueno va, Coral también). Vuestra dedicación es increíble, os deseo lo mejor en vuestros proyectos. Gracias por acordaros de mí y de cuidar tanto de Cris.

Al Club Handbol Sant Andreu, handbol, barri i persones. Gracias por dejarme ir a entrenar con vosotros y hacerme ahora segundo capitán. Pedro,

presidente, por darme esas ganas de competir y ver la ilusión por un proyecto. Lluís, compañero de portería, gran persona y, casualidad, también científico. Va a ser verdad que solo nos metemos en la portería los locos, aunque seamos porteros de mentira. A todo el equipo, por tomarnos esa cerveza en el David después de entrenar y liarnos hasta comernos pizza en Dieguito, o la copa en el Livingstone. Desde los másters al prebenjamín y, por supuesto, el sénior, somos un gran club.

A los que llevamos juntos desde Valencia, los Xixos, sobran las palabras con vosotros y nunca serán suficientes. Cada momento que hemos compartido es para recordar y, con la tontería, ya han pasado 10 años desde que nos juntamos en la carrera. Porque somos capaces de sacar tiempo para saber de los demás, aunque cada uno esté en una punta y, lo mejor, de juntarnos cada año y quedarnos con ganas de más. Sergio, qué capacidad tienes para involucrarte y preocuparte por todos. Bea, compañera, murciana y brasileira de adopción, cada vez que hago palomitas me acuerdo de ti. Sara, la más loca, arroset y ginebra de la bona, se nota cuando no estás. Luis, la Lusa o Luisaco, por ser un gran amigo y compañero de pádel, por hacerme reír tan fuerte cada vez que nos juntamos, aunque no sepas cuándo cortar la broma (venga, arterial!). Y, por supuesto, a Antón, porque eres el primero que me dio el empujón para meterme en Biotecnología y ahora mira, vamos a acabar siendo doctores. Sois todos únicos, no cambiéis.

A los de mi tierra, Torrevieja. Esos amigos que, cuando vuelvo, están los primeros para quedar conmigo. Porque hacéis que, aunque sea solo para un fin de semana, merezca la pena bajar y hacerme 12 horas de trayecto, 6 de ida y 6 de vuelta. Mientras tanto, vamos a conectarnos para echar una partida y hablamos un rato. Intentaré no picarme, pero no lo puedo remediar, no me gusta perder y lo sabéis. Gracias por estar cada vez, a pesar de la distancia.

A mi familia, por haberme enseñado a valerme por mí mismo y no ponerme límites. Por dejarme intentarlo todo después de decir *a veh yo*. Siempre habéis sido un ejemplo, gracias por apoyarme con cada decisión. Aunque estemos cada uno en un sitio cuento siempre los días para volver a juntarnos.

Por último, y más importante, a Cris, mi compañera de vida. Gracias por sacarme una sonrisa cada día, por cogerme y darme un abrazo cuando más lo necesito y, si no, también, nunca está de más. Gracias por ser tan buena, no sabes cuánto me has enseñado. Gracias por todos los momentos que hemos compartido, del primero al último, apoyándonos y sacando todo adelante. Eres una persona increíble, el empeño que pones en todo lo que te propones es ejemplar, eres capaz de todo. Contigo al lado todo es más sencillo.

Sinopsis

The pathogenic strains of *Escherichia coli* can be divided into two large groups: those that cause intestinal infections and those that cause extra-intestinal infections. The strains that belong to the first group are diarrheic and are grouped into pathotypes depending on the virulence factors they present and the strategy they follow to cause an infection. Among the extraintestinal strains it is possible to differentiate those that cause urinary tract infections (UPEC), those that cause neonatal meningitis (NMEC) or those that cause sepsis (SEPEC).

The enteroaggregative strains of *Escherichia coli* (EAEC) represent one of the groups of diarrheal pathogens. EAECs can be distinguished from enteropathogenic strains (EPECs) by their adherence pattern to HEp-2 cells, whereas EPECs present an adherence pattern in the form of microcolonies, EAECs follow a "brick stacking" pattern. This is because enteroaggregative strains express a specific type of adhesive fimbriae that allow adhesion to intestinal cells. Epidemiological studies show that the EAEC strains are genetically heterogeneous, having identified numerous virulence factors in strains of this pathotype. An example of this heterogeneity is the strain O104:H4, which caused an outbreak of bloody diarrhea in Germany in 2011. In addition to the typical characteristics of enteroaggregative strains, the strain O104:H4 exhibits a prophage encoding the Shiga toxin, a well-studied virulence factor characteristic of enterohemorrhagic strains (EHEC).

Among the enteroaggregative strains, strain 042 is used as a model for this pathotype, since its genome was sequenced and its virulence factors well characterized. Analyzing the genome of strain 042 we could see that, unlike other strains of *Escherichia coli*, whose genomes code for the *hha* gene and its *ydgT* paralog, its chromosome encodes four paralogs of the *hha* gene, which are *hha*, *ydgT*, and the new *hha2* and *hha3* that had not been described until now.

We hypothesize that the presence of these new paralogs *hha2* and *hha3* in strain 042 could be associated with the presence of genes related to virulence. Through genomic analysis it was possible to establish the relationship between the presence of new paralogs of the Hha protein with the *yeeR* and *irmA* genes. Furthermore, due to the presence of a higher number of copies of the Hha regulator and other genes such as, for example, *flu*, we decided to study, using bioinformatic tools, the existence of duplicate genes and their distribution in different pathotypes. In this way, we establish gene duplication patterns in all pathotypes and possible relationships between them.

Owing to the *irmA* gene is duplicated in strain 042 and associated with the presence of the *hha2/hha3* genes, the expression of both copies (*irmA2244* and *irmA4509*) was characterized under different conditions. As a result, the expression of the *IrmA4509* copy is higher in minimal medium (M9) at 37 °C. Furthermore, the expression of this copy increases under conditions of limitation of the carbon source (glucose) but is reduced in the absence of oxygen. On the other hand, the expression of the *IrmA2244* copy is very low or null in all the conditions studied. Next, the regulation of both copies was studied, determining the role of the H-NS/Hha system and the OxyR/Dam system. In this case, the *IrmA4509* copy is overexpressed in the absence of *hha* and, above all, in the absence of *hha* and *hha2* (*hhanull*), and in the absence of the OxyR regulator. However, no expression could be detected from the copy *IrmA2244*. Finally, the position of the promoter region of both copies was determined by 5' RACE, observing nucleotide changes between them. A directed mutagenesis assay was carried out, by which the importance of these nucleotide changes in the expression of both copies was determined.

On the other hand, the possible role in virulence of a region of duplicated genes widely distributed among all pathotypes was studied. For this, biofilm formation and infectivity tests were performed in the *Galleria mellonella* model organism, obtaining relevant results.

Índice

Lista de abreviaturas.....	I
1. Introducción.....	1
1.1 <i>Escherichia coli</i> y sus patotipos.....	3
1.1.1 Patotipo uropatógeno (UPEC).....	7
1.1.2 Patotipo enteropatógeno (EPEC).....	9
1.1.3 Patotipo enterohemorrágico (EHEC).....	10
1.1.4 Patotipo enterotoxigénico (ETEC).....	12
1.1.5 Patotipo enteroinvasivo (EIEC).....	13
1.1.6 Patotipo de adherencia difusa (DAEC).....	15
1.1.7 Patotipo enteroagregativo (EAEC).....	16
1.2 La cepa EAEC 042.....	20
1.3 Proteínas asociadas al nucleóide.....	22
1.3.1 La familia de proteínas H-NS.....	23
1.3.2 La familia de proteínas Hha.....	29
1.4 Duplicaciones génicas en bacterias.....	34
1.5 Operón Agn43-YeeR-Aec69.....	40
2. Objetivos.....	49
3. Materiales y métodos.....	53
3.1 Cepas bacterianas y plásmidos.....	55
3.2 Oligonucleótidos.....	59
3.3 Medios de cultivo y antibióticos.....	63
3.3.1 Medios de cultivo.....	63

3.3.2 Antibióticos	65
3.4 Métodos microbiológicos	65
3.4.1 Esterilización	65
3.4.2 Mantenimiento y crecimiento de los microorganismos	66
3.5 Técnicas experimentales con ADN	67
3.5.1 Aislamiento de ADN plasmídico	67
3.5.2 Transformación bacteriana	67
3.5.2.1 Transformación de células competentes obtenidas por tratamiento con CaCl ₂	67
3.5.2.2 Electroporación	68
3.5.3 Uso de enzimas	69
3.5.4 Amplificación de fragmentos de ADN mediante la reacción en cadena de la polimerasa (PCR)	70
3.5.5 Purificación de productos de PCR	71
3.5.6 Electroforesis de ADN en geles de agarosa	72
3.5.7 Aislamientos de fragmentos de ADN a partir de geles de agarosa	73
3.5.8 Secuenciación de fragmentos de ADN	73
3.6 Métodos de mutagénesis bacteriana	74
3.6.1 Inactivación de genes cromosómicos utilizando fragmentos de PCR	74
3.6.1.1 Generación del fragmento que contiene el gen de resistencia al antibiótico mediante PCR	76
3.6.1.2 Reemplazamiento alélico	76
3.6.1.3 Eliminación de la resistencia a antibiótico	77

3.6.2 Construcción de fusiones génicas con la secuencia codificante 3XFLAG mediante la recombinasa Red.....	78
3.6.3 Construcción de fusiones génicas con el gen <i>lacZ</i> mediante la recombinasa FLP.....	80
3.6.4 Construcción de fusiones transcripcionales en el plásmido pUJ8..	81
3.6.5 Mutagénesis Sitio-Dirigida.....	82
3.7 Técnicas experimentales con ARN.....	84
3.7.1 Extracción de ARN total y cuantificación.....	84
3.7.2 RT-PCR.....	85
3.7.3 <i>Walking RT-PCR</i>	86
3.7.4 <i>Rapid Amplification of cDNA ends (5' RACE)</i>	86
3.8 Técnicas experimentales con proteínas.....	87
3.8.1 Obtención de extractos totales.....	87
3.8.2 Electroforesis en geles de poliacrilamida en presencia de SDS	88
3.8.3 Tinción con azul de Coomassie	90
3.8.4 Inmunodetección de proteínas (<i>Western Blot</i>).....	91
3.8.4.1 Transferencia de proteínas a membrana por sistema semi-seco.....	91
3.8.4.2 Revelado quimioluminiscente.....	91
3.8.5 Obtención de anticuerpo anti-IrmA.....	92
3.8.5.1 Purificación de la proteína IrmA.....	92
3.8.5.2 Inmunización de conejos y obtención del anticuerpo policlonal	95
3.9 Valoración de la actividad enzimática	95

3.9.1 Determinación de la actividad β -galactosidasa.....	95
3.10 Estudios de formación de biofilm	97
3.11 Ensayos de infección en <i>Galleria mellonella</i>	98
3.12 Análisis bioinformáticos	99
3.12.1 Comparación genómica entre cepas con <i>hha2/hha3</i> y cepas que no presentan estos reguladores.....	99
3.12.2 Identificación de genes duplicados	100
3.12.3 Distribución de genes duplicados en diferentes patotipos	100
3.12.4 Predicción de operones.....	102
3.12.5 Predicción de promotores.....	102
3.13 Análisis estadístico.....	102
4. Resultados.....	105
4.1 Duplicaciones génicas en <i>E. coli</i> y su regulación por el sistema H- NS/Hha.....	107
4.1.1 Relación de la presencia de los parálogos <i>hha2/hha3</i> con otros genes en <i>Escherichia coli</i>	107
4.1.2 Duplicaciones génicas en la cepa 042 de <i>E. coli</i>	114
4.1.3 Papel de las proteínas H-NS y Hha en la regulación de genes duplicados en la cepa 042	121
4.1.4 Distribución de las duplicaciones de la cepa EAEC 042 en otras cepas de <i>E. coli</i>	124
4.1.5 Presencia y distribución de las duplicaciones de las cepas UPEC CFT073 y EHEC O145:H28 RM13516	128
4.2 El gen <i>aec69</i> en la cepa EAEC 042	141

4.2.1 Estudio de la expresión de la proteína IrmA en la cepa 042 por <i>Western Blot</i>	142
4.2.2 Estudio de la transcripción de las dos copias del gen <i>irmA</i> en la cepa 042 mediante las fusiones cromosómicas <i>irmA2244::lacZ</i> e <i>irmA4509::lacZ</i>	147
4.2.3 Regulación de la expresión de las dos copias del gen <i>irmA</i> en la cepa EAEC 042	149
4.2.3.1 Efecto de la temperatura y del medio de cultivo sobre la expresión de los genes <i>irmA2244</i> e <i>irmA4509</i>	149
4.2.3.2 Efecto de la presencia de oxígeno sobre la expresión de los genes <i>irmA2244</i> e <i>irmA4509</i>	154
4.2.3.3 Papel de las proteínas H-NS y Hha en la expresión de los genes <i>irmA2244</i> e <i>irmA4509</i>	155
4.2.3.4 Papel de las proteínas OxyR y Dam en la expresión de los genes <i>irmA2244</i> e <i>irmA4509</i>	157
4.2.3.5 Regulación cruzada entre las dos copias del gen <i>irmA</i>	160
4.2.4 Identificación del promotor de los genes <i>irmA</i> de la cepa 042.....	162
4.2.4.1 Expresión del gen <i>lacZ</i> bajo el control del promotor del gen <i>agn43</i>	162
4.2.4.2 Determinación del inicio de transcripción del gen <i>irmA</i> en la cepa 042 por <i>Walking RT-PCR</i>	165
4.2.4.3 Búsqueda del promotor del gen <i>irmA</i> mediante análisis bioinformáticos y mediante la determinación del inicio de transcripción por 5' RACE	167
4.2.4.4 Actividad del promotor del gen <i>irmA</i>	169

4.2.4.5 Análisis comparativo de la región promotora del gen <i>irmA</i> entre la cepa 042 y otras cepas patógenas de <i>E. coli</i>	174
4.3 Estudio de la región 55-60, presente en cuatro copias en el genoma de la cepa 042.....	178
4.3.1 Deleción de las cuatro copias del clúster 55-60 de la cepa 042 y estudio del <i>fitness</i> celular.....	179
4.3.2 Estudio de la formación de biofilm en la cepa 042.....	182
4.3.3 Ensayo de infectividad usando como organismo modelo <i>Galleria mellonella</i>	183
5. Discusión.....	187
5.1 Distribución de genes duplicados en <i>E. coli</i> y su relación entre patotipos.....	189
5.2 Papel del sistema H-NS/Hha en la regulación de genes duplicados	193
5.3 Expresión y regulación del gen <i>irmA</i>	194
5.4 Papel de la región de genes duplicados ampliamente distribuida entre los patotipos de <i>E. coli</i>	197
6. Conclusiones.....	201
7. Bibliografía.....	205
8. Anexos.....	247
9. Publicaciones.....	265

Lista de abreviaturas

the same way as the other two, but the model is more complex because the interaction between the two variables is not linear. The model is written as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_1^2 x_2 + \beta_7 x_1 x_2^2 + \beta_8 x_1^2 x_2^2 + \epsilon$$

where β_0 to β_8 are the regression coefficients and ϵ is the error term. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

The model is fitted to the data using the least squares method. The model is fitted to the data using the least squares method.

- A/E – *Attaching and effacing*
- AAF – Fimbrias de adherencia difusa
- AT – Adenina-timina
- BLASTn – *Basic Local Alignment Search Tool for Nucleotides*
- BLASTp – *Basic Local Alignment Search Tool for Proteins*
- Cb – Carbenicilina
- CDS – Regiones codificantes
- CF – Factor de colonización
- CFA – Antígeno del factor de colonización
- Cm – Cloranfenicol
- CRP – *cAMP-activated global transcriptional regulator*
- DAEC – *Escherichia coli* de adherencia difusa
- DAF – Factor acelerador de la descomposición del complement
- DMEM – *Dulbecco's Minimum Essential Medium*
- dNTPs – Nucleótidos desoxirribonucleicos
- DO_{600nm} – Densidad óptica a una longitud de onda de 600 nanómetros
- EAEC – *Escherichia coli* enteroagregativa
- EAST1 – *EAEC heat-stable enterotoxin 1*
- EHEC – *Escherichia coli* enterohemorrágica
- EIEC – *Escherichia coli* enteroinvasiva
- EPEC – *Escherichia coli* enteropatógena
- ETEC – *Escherichia coli* enterotoxigénica
- ExPEC – *Escherichia coli* extraintestinal
- Fis – *Factor Inversion Stimulation*
- Gb3 – Globotriaosilceramida
- GC – Guanina-citosina
- GDA – Procesos de duplicación y amplificación génica
- HC – Colitis hemorrágica
- HGT – Transferencia genética horizontal
- Hha – *High hemolytic activity*

- HlyE – Hemolisina E
- H-NS – *Histone-like nucleoid structuring protein*
- HU – *Heat-Unstable Nucleoid Protein*
- HUS – Síndrome urémico hemolítico
- IgA – Inmunoglobulina A
- IHF – *Integration Host Factor*
- IrmA – *Interleukin receptor mimic protein A*
- IS – Secuencia de inserción
- Km – Kanamicina
- LB – *Lysogeny Broth*
- LEE – *Locus of enterocyte effacement*
- Ler – Regulador de la región LEE
- Lrp – *Leucine-responsive regulatory protein*
- LT – Toxina termolábil
- NAPs – Proteínas asociadas al nucleóide
- NMEC – *Escherichia coli* productoras de meningitis neonatales
- ONPG – O-nitrofenil-β-galactósido
- PAI – Isla de patogenicidad
- P:C:I – Fenol:cloroformo:isopropanol
- PCR – Reacción en cadena de la polimerasa
- Pet – *Plasmid-encoded toxin*
- Pic – *Protein involved in intestinal colonization*
- PLGR – Receptor de globulina polimérica
- PVDF – Polifluoruro de vinilideno
- RACE – *Rapid Amplification of cDNA ends*
- ROD – *Regions of difference*
- SDS – Dodecilsulfato sódico
- ShET1 – *Shigella enterotoxin 1*
- SNPs – Polimorfismos en un único nucleótido
- SOB – *Super Optimal Broth*

- SOC – *Super Optimal Broth with Catabolite repression*
- SP – Péptido señal
- ST – Toxina termoestable
- Stx – Toxina Shiga
- T3SS – Sistema de secreción tipo 3
- T4SS – Sistema de secreción tipo 4
- T5SS – Sistema de secreción tipo 5
- TAM – Sistema de translocación y ensamblaje
- TLR4 – Receptores tipo *toll* 4
- UPEC – *Escherichia coli* uropatógena
- UTI – Infección del tracto urinario
- VTEC – Verotoxina

1. Introducción

the 1990s, the number of people in the world who are poor has increased. The number of people who are poor in the United States has also increased.

There are many reasons for this. One reason is that the world population has increased. There are now over 6 billion people in the world, and the number is still increasing. Another reason is that the cost of living has increased. The price of food, clothing, and housing has gone up, and this has made it harder for people to afford these things. A third reason is that there has been a lot of technological change. This has created new jobs, but it has also destroyed old jobs. Many people who have lost their jobs have not been able to find new ones.

There are many things that we can do to help people who are poor. One thing is to give them money. This can help them to buy food, clothing, and housing. Another thing is to give them education. This can help them to get better jobs. A third thing is to give them training. This can help them to learn new skills. There are many other things that we can do to help people who are poor. We need to work together to find solutions to these problems.

There are many people who are poor in the United States. There are about 30 million people who live in poverty. This is a very large number of people. There are many reasons why people are poor in the United States. One reason is that they have lost their jobs. Another reason is that they have a disability. A third reason is that they have a low level of education. There are many other reasons why people are poor in the United States. We need to find ways to help these people.

There are many things that we can do to help people who are poor in the United States. One thing is to give them money. This can help them to buy food, clothing, and housing. Another thing is to give them education. This can help them to get better jobs. A third thing is to give them training. This can help them to learn new skills. There are many other things that we can do to help people who are poor in the United States. We need to work together to find solutions to these problems.

There are many people who are poor in the United States. There are about 30 million people who live in poverty. This is a very large number of people. There are many reasons why people are poor in the United States. One reason is that they have lost their jobs. Another reason is that they have a disability. A third reason is that they have a low level of education. There are many other reasons why people are poor in the United States. We need to find ways to help these people.

There are many things that we can do to help people who are poor in the United States. One thing is to give them money. This can help them to buy food, clothing, and housing. Another thing is to give them education. This can help them to get better jobs. A third thing is to give them training. This can help them to learn new skills. There are many other things that we can do to help people who are poor in the United States. We need to work together to find solutions to these problems.

There are many people who are poor in the United States. There are about 30 million people who live in poverty. This is a very large number of people. There are many reasons why people are poor in the United States. One reason is that they have lost their jobs. Another reason is that they have a disability. A third reason is that they have a low level of education. There are many other reasons why people are poor in the United States. We need to find ways to help these people.

There are many things that we can do to help people who are poor in the United States. One thing is to give them money. This can help them to buy food, clothing, and housing. Another thing is to give them education. This can help them to get better jobs. A third thing is to give them training. This can help them to learn new skills. There are many other things that we can do to help people who are poor in the United States. We need to work together to find solutions to these problems.

There are many people who are poor in the United States. There are about 30 million people who live in poverty. This is a very large number of people. There are many reasons why people are poor in the United States. One reason is that they have lost their jobs. Another reason is that they have a disability. A third reason is that they have a low level of education. There are many other reasons why people are poor in the United States. We need to find ways to help these people.

There are many things that we can do to help people who are poor in the United States. One thing is to give them money. This can help them to buy food, clothing, and housing. Another thing is to give them education. This can help them to get better jobs. A third thing is to give them training. This can help them to learn new skills. There are many other things that we can do to help people who are poor in the United States. We need to work together to find solutions to these problems.

There are many people who are poor in the United States. There are about 30 million people who live in poverty. This is a very large number of people. There are many reasons why people are poor in the United States. One reason is that they have lost their jobs. Another reason is that they have a disability. A third reason is that they have a low level of education. There are many other reasons why people are poor in the United States. We need to find ways to help these people.

1.1 *Escherichia coli* y sus patotipos

El microorganismo *Escherichia coli* (*E. coli*) es una de las bacterias que primero coloniza el tracto gastrointestinal humano, siendo el anaerobio facultativo más abundante de la microbiota intestinal¹. Una característica importante de los aislamientos tanto ambientales como clínicos que se identifican como *E. coli* es el alto grado de diversidad genómica que presentan. Si bien todos ellos comparten un grupo de genes (*core-genome*), también presentan un elevado número de genes adquiridos por transferencia genética horizontal, los cuales pueden variar mucho entre diferentes aislamientos. Algunos de estos genes adquiridos confieren la capacidad de colonizar diferentes organismos. Por ello, y en relación con la especie humana, la especie *E. coli* no solo incluye cepas comensales, sino también un número notable de cepas patógenas. Estas cepas patógenas generan infecciones, tanto en individuos sanos como inmunodeprimidos, con una tasa importante de morbilidad y, en algunos casos, de mortalidad¹⁻⁴. Los estudios que contemplan los mecanismos que han permitido que el genoma de *E. coli* evolucione hacia genotipos que confieren virulencia consideran, al igual que en muchos otros patógenos bacterianos, principalmente dos mecanismos: la adquisición de material genético por transferencia horizontal (HGT) como parte de plásmidos, fagos, transposones o islas de patogenicidad (PAI)^{5,6}, y/o la pérdida o modificación de genes del *core-genome*⁷.

Los procesos HGT tienen lugar a través de tres mecanismos principales bien diferenciados: transformación, transducción y conjugación⁸. La transformación se basa en la adquisición de ADN libre en el ambiente por parte de bacterias competentes y su posterior integración en el genoma por recombinación homóloga^{8,9}. En cambio, la transducción tiene lugar por la acción de bacteriófagos, los cuales encapsidan material genético de la bacteria huésped durante su ciclo lítico y lo transfieren a otra bacteria al infectarla^{8,10}.

Por último, la conjugación requiere que la bacteria dadora y la receptora entren en contacto mediante el pili sexual, el cual depende de un sistema de secreción de tipo 4 (T4SS) para su formación^{8,11}.

El mecanismo alternativo a la adquisición de ADN HGT, mediante el cual las bacterias pueden aumentar su capacidad infecciosa, es a través de mutaciones en genes que ya se encuentran en el genoma, llamadas mutaciones patoadaptativas^{7,12}. Este tipo de mutaciones se pueden dividir, de manera general, en tres tipos: inactivación, variación y amplificación¹². La inactivación génica o pérdida de función puede tener lugar por mutaciones que generen sustituciones puntuales de aminoácidos en el sitio activo de la proteína, o por la aparición de un codón de terminación prematuro que resulta en una proteína truncada^{7,12}. Estas mutaciones son frecuentes en genes cuya expresión es incompatible con la virulencia de la bacteria, denominados también genes de antivirulencia¹³. Otro tipo de mutación es la variación genética, donde polimorfismos en un único nucleótido (SNPs) no dan como resultado una proteína truncada pero sí que provocan una modificación en la estructura y función final de la proteína¹². Finalmente, la amplificación génica aumenta el número de copias de un gen, lo que puede conllevar una mayor expresión en sí o la capacidad de evadir una regulación negativa de ese producto por parte de un inhibidor¹².

Respecto a la patogenicidad de *E. coli*, las cepas patógenas se pueden diferenciar por provocar tres cuadros clínicos principalmente: (i) aquellas que llevan a cabo una infección en el tracto urinario, (ii) las que provocan meningitis y sepsis y (iii) las que provocan una infección intestinal o diarreas. En algunos casos, algunas cepas pueden provocar más de uno de estos cuadros clínicos, como se ha visto en cepas que infectan el tracto urinario y también causan diarreas¹⁴⁻¹⁶. En función tanto del tipo de cuadro clínico y de los factores de virulencia que presentan, como de los perfiles epidemiológicos y filogenéticos, las cepas patógenas de *E. coli* se agrupan en patotipos^{1,2,4,17}. Las

cepas extraintestinales (ExPEC) comprenden aquellas que provocan infecciones en el tracto urinario (UPEC), meningitis neonatales (NMEC) o las que producen sepsis (SEPEC)¹⁸. Las cepas que son causa de infecciones intestinales, o cepas diarreicas, se agrupan en los siguientes patotipos: enteropatógeno (EPEC), enterohemorrágico (EHEC), enterotoxigénico (ETEC), enteroinvasivo (EIEC), de adherencia difusa (DAEC) y enteroagregativo (EAEC)^{1,2,4,17}. La presencia de tantos patotipos diferentes ejemplifica la notable plasticidad del genoma de *E. coli*, el cual se caracteriza por poseer un gran pangenoma de aproximadamente 18.000 genes, en contraste con un *core-genome* de aproximadamente 2.000 genes^{19,20}. Los genes de virulencia varían entre patotipos, y pueden encontrarse frecuentemente en elementos genéticos móviles adquiridos por procesos HGT^{5,20,21}, pero también pueden corresponder a aquellos en los que han podido tener lugar mutaciones patoadaptativas (Fig. 1.1)^{12,22}.

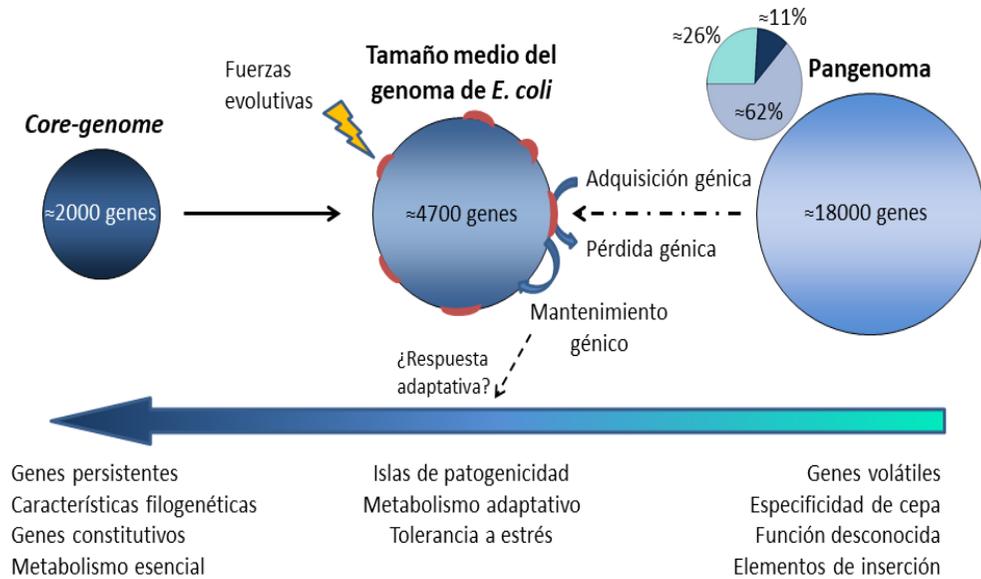


Fig. 1.1. Esquema de la evolución del genoma de *E. coli*. El genoma se ve influenciado tanto por una multitud de fuerzas evolutivas bióticas (competidores, mecanismos de defensa del huésped) como abióticas (pH, temperatura, UV, disponibilidad de nutrientes, etc.). Las cepas de *E. coli* poseen un núcleo (*core-genome*) de aproximadamente 2.000 genes que codifica, entre otros, factores para un metabolismo versátil. El pangenoma de *E. coli* consta de unos 18.000 genes, de los cuales el 11 % pertenecen al *core-genome* (azul oscuro), una gran parte (62 %, azul) está compuesta por los llamados genes “persistentes”, y el 26 % pueden ser considerados como genes “volátiles” (azul pálido)²⁰. Los procesos de ganancia y pérdida de genes están constantemente relacionados con los puntos críticos de inserción/delección (rojo) y son seleccionados por las fuerzas evolutivas. Estos procesos pueden dar lugar a la evolución de grupos de genes que definen patotipos específicos de *E. coli*. La flecha representa un gradiente hipotético en el que se muestran características genómicas de *E. coli* que, probablemente, están asociadas con genes volátiles, persistentes y centrales. Adaptado de van Elsas y colaboradores²¹.

1.1.1 Patotipo uropatógeno (UPEC)

Las infecciones llevadas a cabo por cepas pertenecientes a este patotipo representan aproximadamente el 80 % de todas las infecciones urinarias^{23,24}, causando cistitis en la vejiga y pielonefritis aguda en los riñones^{25,26}. Se cree que el reservorio principal de las cepas UPEC se encuentra dentro del tracto intestinal humano, ya que el aislamiento causante de una infección urinaria en un individuo a menudo coincide con los aislamientos rectales de esa misma persona^{27,28}. En algunos casos, la diseminación de un solo grupo clonal de aislados de cepas UPEC puede ocurrir dentro de una comunidad, a través de alimentos contaminados²⁹. Adicionalmente, las cepas UPEC aisladas de pacientes sexualmente activos coinciden, frecuentemente, con los aislamientos fecales de sus parejas, lo que indica que las infecciones urinarias pueden transmitirse sexualmente³⁰⁻³².

Los genomas de cepas UPEC son, generalmente, más grandes que los genomas de cepas comensales. Esto es debido a la presencia de genes necesarios para la supervivencia fuera del tracto intestinal³³⁻³⁵, los cuales se encuentran en PAI adquiridas a través de HGT³⁶. Las PAI son regiones cromosómicas de gran tamaño (entre 10 y 200 kb) presentes en cepas patógenas, en las cuales se encuentran codificados factores de virulencia. El porcentaje GC que presentan es inferior a la del resto del genoma y suelen contener elementos genéticos móviles asociados con procesos HGT, como, por ejemplo, integrasas o transposasas^{36,37}. Los patotipos diarreicos de *E. coli* también codifican numerosos factores de virulencia dentro de las PAI, pero la variedad de factores de virulencia que expresan las cepas diarreicas respecto a los que expresan las cepas UPEC puede diferir notablemente^{1,35}. Por ejemplo, las cepas UPEC expresan una serie de adhesinas características y sideróforos. Sin embargo, generalmente, estas cepas carecen del sistema de secreción tipo 3

utilizado por muchos aislamientos diarreicos de *E. coli* para inyectar factores de virulencia en la célula huésped^{26,33,35}.

Las células pertenecientes al patotipo UPEC son capaces de desplazarse desde el tracto intestinal hasta el tracto urinario para causar una infección^{26,28}. La capacidad de ascender por el tracto urinario desde la uretra hasta la vejiga y, posteriormente, a los riñones, refleja una capacidad excepcional para evadir la inmunidad innata y la expulsión por micción^{26,28,38}. El primer paso en la infección urinaria ascendente es la colonización de la uretra distal por parte de las bacterias patógenas³⁹. Una vez que alcanzan el área periuretral, son capaces de introducirse fácilmente en la vejiga, hecho que se ve favorecido con la práctica de relaciones sexuales³⁰⁻³². Las bacterias se replican rápidamente en la vejiga y, finalmente, mediante la unión a receptores presentes en las células epiteliales, consiguen una colonización estable⁴⁰⁻⁴². De forma general, las bacterias son incapaces de entrar en los uréteres y ascender hasta los riñones, debido a la acción de la válvula vesicouretral que resiste la presión de la vejiga. Sin embargo, debido a la infección y la inflamación que se desarrolla en la vejiga, puede tener lugar un reflujo que permite que las células UPEC pueden ascender por los uréteres y extenderse por los riñones, desarrollando una pielonefritis aguda^{24,43}.

Todo este proceso de patogénesis tiene lugar gracias a diversos factores de virulencia, como son diferentes tipos de pili o fimbrias, toxinas secretadas tales como la hemolisina^{44,45} y sistemas de adquisición de hierro o sideróforos^{26,28,38}. Dentro de las fimbrias de las cepas UPEC destacan las de tipo 1, tipo P, tipo S y F1C, codificadas por los genes *fim*, *pap*, *sfa* y *foc*, respectivamente. Es reseñable que, dentro del genoma de un mismo aislado UPEC, pueden llegar a encontrarse hasta más de 10 operones de fimbrias^{46,47}. Además, la regulación cruzada entre operones de fimbrias de una célula bacteriana, probablemente provocada por señales ambientales, induce cambios en la expresión de un tipo de fimbrias a otro en un proceso llamado variación de fase⁴⁸⁻⁵⁰. La variación de fase en una

misma cepa puede dar lugar a subpoblaciones que expresan pili funcionalmente distintos, aumentando la probabilidad de adherencia e invasión de las células del hospedador⁵¹. Los pili de tipo 1 están muy conservados en las cepas UPEC, siendo considerados como uno de los factores de virulencia más importantes involucrados en el establecimiento de una infección urinaria (UTI) en la vejiga como, por ejemplo, la adhesina FimH^{40,41}. La adhesina FimH está involucrada tanto en la adhesión bacteriana como en la invasión de células del hospedador^{40,41}, contribuyendo a la formación de biofilm intracelulares y provocando que las UTI tengan lugar de forma recurrente^{24,26,38}. Por otra parte, la colonización de los riñones, dando lugar a una pielonefritis aguda, depende de la expresión de fimbrias de tipo P, las cuales se unen a los glicolípidos que recubren el tejido renal⁵²⁻⁵⁴. Entre las adhesinas de tipo P, PapG también es capaz de interactuar con receptores tipo *toll 4* (TLR4), reduciendo la expresión del receptor de inmunoglobulina polimérica (PIGR)⁵⁵. Esto da como resultado un transporte deteriorado de la inmunoglobulina A (IgA) a través del epitelio, reduciendo así la respuesta inmune y evitando la opsonización^{24,55}.

1.1.2 Patotipo enteropatógeno (EPEC)

Entre los patógenos intestinales, el patotipo EPEC fue el primero en ser descrito. Esto ocurrió en 1945, cuando se determinó la presencia de un grupo de cepas de *E. coli* en niños con diarrea, las cuales eran serológicamente distintas a las que presentaban niños sanos⁵⁶. En la actualidad, las cepas EPEC no son causantes de grandes brotes infecciosos en países desarrollados, pero causan entre el 30-40 % de los casos de diarrea infantil en países en desarrollo que conllevan miles de muertes^{17,57-62}.

Las infecciones causadas por cepas del patotipo EPEC están asociadas con una histopatología intestinal localizada característica, denominada *attaching-and-effacing* (A/E), la cual puede ser observada en biopsias intestinales de

pacientes infectados y puede ser reproducida en cultivos celulares^{63,64}. Este fenotipo consiste en la eliminación localizada de microvellosidades del intestino delgado y la unión de la membrana bacteriana con la del epitelio intestinal. Esta unión induce un incremento en los niveles de Ca^{+2} intracelular, lo que conlleva alteraciones en el citoesqueleto y la formación de la lesión A/E⁶⁵⁻⁶⁸. Estas alteraciones en el citoesqueleto afectan principalmente a la actina, provocando que se formen estructuras similares a un pedestal en la zona de contacto entre las bacterias y los enterocitos^{64,69-71}. La unión de las membranas está mediada por dos efectores del sistema de secreción tipo 3 que poseen las cepas EPEC⁷². Se trata de una proteína de alto peso molecular llamada intimina y su receptor Tir (*translocated intimin receptor*), el cual es translocado por la bacteria dentro de los enterocitos y expuesto en su superficie (Fig. 1.2a)^{68,71,73,74}. Este sistema de secreción tipo 3 (T3SS) y sus efectores se encuentran juntos en el cromosoma en una PAI de aproximadamente 35 Kb, denominada LEE (*locus of enterocyte effacement*)⁷⁵.

Hay que destacar que este tipo de lesiones A/E también se han observado en infecciones debidas al patotipo EHEC⁷⁶. Comparando la intimina que expresa cada patotipo, se puede ver cómo guardan un alto porcentaje de homología⁷⁷. Muchas cepas EHEC también contienen la región LEE^{78,79}, pero presentan diferencias que se explican a continuación.

1.1.3 Patotipo enterohemorrágico (EHEC)

Las cepas EHEC fueron reconocidas por primera vez como una clase distinta de patógenos de *E. coli* en 1983, tras dos observaciones epidemiológicas clave. La primera tuvo lugar en dos casos de afectación gastrointestinal que presentaban dolor estomacal intenso, diarrea sanguinolenta y fiebre⁸⁰. Esta enfermedad fue llamada colitis hemorrágica (HC) y fue asociada al serotipo de *E. coli* que se pudo aislar de las heces, O157:H7, un

serotipo poco conocido en ese momento. La segunda observación, asoció casos esporádicos de síndrome urémico hemolítico (HUS) con la presencia de *E. coli* productoras de citotoxinas en las heces⁸¹. El HUS (definido por un fallo renal, trombocitopenia y anemia hemolítica) está precedido normalmente por diarrea sanguinolenta, por lo que estas dos observaciones dieron lugar al reconocimiento de un mismo patógeno entérico responsable de enfermedades intestinales y renales. Mientras las cepas EPEC siguen siendo una importante causa de diarrea en países en desarrollo, las cepas EHEC son más comunes en contaminaciones alimentarias en países industrializados⁸²⁻⁸⁵. Además, las cepas EPEC típicas tienen su único reservorio en humanos a diferencia de las cepas EHEC, las cuales son comensales en ganado bovino, ovino y equino, entre otros⁸⁶⁻⁸⁸. El serotipo O157:H7 es el más importante entre los patógenos EHEC en Estados Unidos, Reino Unido y Japón^{82,85}, pero otros serotipos tales como O26 y O111 son más predominantes en otros países⁸⁹⁻⁹². La cepa de *E. coli* O157:H7 es muy cercana genéticamente a la cepa *E. coli* O55:H7⁹³⁻⁹⁵, una cepa EPEC que causa diarreas infantiles en todo el mundo. Los factores de virulencia adicionales y el antígeno O157 fueron adquiridos por transferencia horizontal y recombinación, lo que se vio favorecido por las altas tasas intrínsecas de error que tiene esta cepa para reparar el ADN⁹⁶.

En cuanto a la patogénesis de las cepas EHEC, hay que resaltar, en primer lugar, que las mismas pueden seguir tres rutas de transmisión: mediante contaminación alimentaria, por contacto entre personas o por contacto con animales o zoonosis^{89,97-99}. El factor de virulencia más importante de las cepas EHEC es la toxina Shiga (Stx), también conocida como verotoxina (VT)^{1,2,17,25}. Las cepas EHEC expresan la toxina Stx cuando alcanzan el colon, la cual se une a la globotriaosilceramida (Gb3), un esfingolípidio que se expresa en células endoteliales de intestino, riñón y cerebro^{89,100-103}. Tras la unión al receptor en el riñón, daña el endotelio renal mediante la combinación de una toxicidad directa y una producción local de citoquinas y quimiocinas que resultan en una

inflamación, provocando finalmente HUS (Fig. 1.b)¹⁰⁴⁻¹⁰⁷. Por otra parte, al unirse a los receptores presentes en el intestino produce daños en las células del colon, resultando en HC y necrosis^{1,4,89}.

En las cepas EHEC, además de las toxinas, hay otros elementos importantes para la virulencia, como una región homóloga a la isla LEE de las cepas EPEC que muchas cepas EHEC contienen^{78,79}. La unión a enterocitos y la consecuente histopatología A/E observada es llevada cabo gracias a la unión de la intimina y su receptor Tir, tal y como se ha descrito anteriormente en las cepas EPEC⁶³⁻⁶⁸. Sin embargo, a diferencia de las cepas EPEC, las cepas EHEC provocan esta lesión A/E en el colon, en vez de en el intestino delgado. En cuanto a la formación del pedestal, en el caso de las cepas EHEC sigue un mecanismo diferente, aunque la estructura final es muy semejante¹⁰⁸⁻¹¹².

1.1.4 Patotipo enterotoxigénico (ETEC)

Este patotipo de *E. coli* engloba a las cepas que expresan al menos uno de los dos grandes grupos definidos de enterotoxinas: termoestables (ST) y termolábiles (LT)^{1,2,17,25}. Las cepas ETEC representan una de las causas más comunes de diarrea infantil en países en desarrollo^{3,113-115}, así como de diarrea del viajero en estos mismos países^{116,117}, provocando anualmente millones de casos y siendo causa de miles de muertes infantiles^{118,119}. Las cepas ETEC también representan una carga económica en el sector ganadero, ya que muchas de ellas son patógenos importantes para los animales de granja, sobre todo de cerdos¹²⁰⁻¹²².

La patogénesis de las cepas ETEC se caracteriza por la colonización de la mucosa del intestino delgado. Una vez que colonizan el intestino, las cepas ETEC liberan enterotoxinas, bien ST o LT^{123,124}. Ambas toxinas provocan una desregulación de los canales iónicos presentes en las membranas del epitelio.

Como consecuencia a esta desregulación, tiene lugar una pérdida de iones y una salida masiva de agua, siendo éste el síntoma característico de las diarreas que causan estas bacterias (Fig. 1.2c)¹²⁴⁻¹²⁸.

La colonización del intestino es posible gracias a una gran variedad de elementos de adhesión, codificados en plásmidos. Estas estructuras son denominadas, de forma genérica, factores de colonización (CF). Más de 20 CFs han sido identificados y caracterizados, clasificados en base a su antígeno (CFA), peso molecular y morfología estructural^{127,129-132}. Estudios epidemiológicos indican que la mayoría de los CF de cepas ETEC humanas se encuentran representados en los grupos CFA/I, CFA/II y CFA/IV¹³³.

1.1.5 Patotipo enteroinvasivo (EIEC)

Las cepas EIEC son un agente causante de disentería en humanos, especialmente en países en desarrollo^{2,25,134,135}. Los casos registrados en países industrializados han tenido lugar en pacientes que viajaron a alguno de los países donde presenta una gran incidencia estos patógenos^{136,137}. Sin embargo, los recientes brotes infecciosos en Europa, causados por la cepa EIEC O96:H19, han provocado que la comunidad científica reconsidere el papel de este tipo de patógeno en los países del primer mundo¹³⁸⁻¹⁴⁰.

Las cepas EIEC y las bacterias del género *Shigella* comparten muchas características fenotípicas y genotípicas, incluyendo la presencia del plásmido pINV, lo que dificulta la discriminación entre ambos tipos de bacterias¹⁴¹⁻¹⁴³. Esta dificultad para poder diferenciarlas sesga la interpretación de la información epidemiológica disponible, evaluando de forma incorrecta el número real de infecciones debidas a EIEC². Debido a la gran similitud entre EIEC y *Shigella*, se asume que las dos comparten un mismo ancestro y que, en un momento dado en la evolución, se produjo la separación^{141,144,145}. A pesar de

todas estas similitudes, se ha comprobado que la dosis infecciosa inicial de las cepas EIEC tiene que ser mucho más alta que la de *Shigella*¹⁴⁶ y que la expresión de genes de virulencia, la capacidad para escapar del fagosoma y la capacidad para proliferar y provocar daño celular es comparativamente menor^{2,147}.

La patogénesis de las cepas EIEC está determinada por la presencia del plásmido de gran tamaño de tipo F, llamado pINV^{148,149}. Este plásmido codifica la maquinaria necesaria para llevar a cabo la invasión, supervivencia y difusión de las bacterias dentro de la célula hospedadora^{145,150}. Esta familia de plásmidos solo se encuentra en este patotipo de *E. coli*, y la pérdida del mismo es muy poco frecuente, hecho que además conlleva la pérdida de la virulencia¹³⁴.

Las cepas EIEC se caracterizan por su habilidad para invadir la mucosa de los enterocitos del colon y llegar hasta la submucosa, utilizando como ruta de entrada las células M¹⁵¹. Tras un proceso de transcitosis de las células M, las bacterias son fagocitadas por macrófagos y células dendríticas. Para escapar de estas células del sistema inmune, las células EIEC secretan varios efectores a través de T3SS (presentes en el plásmido pINV) e inducen un proceso de piroptosis en los macrófagos^{152,153}, una forma de muerte celular que conlleva una gran respuesta inflamatoria^{134,154}. Una vez dentro del citosol de los enterocitos, las células EIEC se replican y reorganizan el citoesqueleto celular. Esta reorganización del citoesqueleto les permite infectar células adyacentes^{134,153}, provocando un proceso de inflamación y la destrucción de la barrera epitelial (Fig. 1.2e). Todo este proceso de infección conlleva, finalmente, disentería, caracterizada por la presencia de sangre, secreciones y leucocitos en las heces^{155,156}.

1.1.6 Patotipo de adherencia difusa (DAEC)

El patotipo de adherencia difusa (DAEC) incluye a las cepas diarreas de *E. coli* que se unen a las células del epitelio intestinal siguiendo un patrón de adherencia diferente al localizado (A/E), característico del patotipo EPEC^{1,4,17}. Se considera un grupo único y distinto de otros patotipos, pero debido a las dificultades de clasificación e identificación, la designación de DAEC como un patotipo entérico de *E. coli* requiere más estudios epidemiológicos¹⁵⁷. La clasificación de una cepa de *E. coli* dentro del patotipo DAEC tiene lugar en función de su adherencia a las células epiteliales cultivadas de la línea celular Hep-2. La adherencia de estas bacterias se produce en toda la superficie de las células Hep-2, siguiendo un patrón disperso característico¹⁵⁸.

Las cepas DAEC se han relacionado con diarrea acuosa en muchos países¹⁵⁹⁻¹⁶², pudiendo volverse persistente en niños pequeños, con un aumento en la gravedad de la enfermedad desde los 18 meses hasta los 5 años^{163,164}. Se ha sugerido que los adultos pueden convertirse en portadores asintomáticos, y que la presencia de cepas DAEC puede conducir a enfermedades intestinales inflamatorias crónicas e, incluso, a infecciones en el tracto urinario^{163,164}.

Uno de los factores de virulencia principales de las cepas típicas de este patotipo son las adhesinas Afa/Dr^{163,164}. Estas adhesinas se unen al factor acelerador de la descomposición del complemento (DAF), una molécula con una elevada expresión en la superficie de las células epiteliales intestinales y del tracto urinario, encargada de evitar la cascada de señalización del sistema de complemento. La unión a DAF induce el reordenamiento del citoesqueleto, destruyendo o reorganizando las vellosidades intestinales^{163,164}. El reordenamiento se centra en el desmontaje de la actina F y de la villina, dando como resultado lesiones en los microvilli (Fig. 1.2f). Estas lesiones conducen tanto a una pérdida como a una elongación atípica de las microvellosidades, debido a la expresión defectuosa de las proteínas intestinales funcionales

asociadas a la membrana de los enterocitos^{163,165}. Como consecuencia de la unión de las adhesinas Afa/Dr con DAF, también se induce el reclutamiento de neutrófilos, lo que provoca una sobreexpresión de DAF y, por lo tanto, un mayor número de receptores disponibles para las células DAEC^{164,166}. Por último, la liberación de la toxina Sat (*Secreted autotransporter toxin*) induce la reordenación de las proteínas asociadas a la unión estrecha entre células, provocando un aumento de la permeabilidad celular y la pérdida de agua¹⁶⁷.

1.1.7 Patotipo enteroagregativo (EAEC)

El patotipo EAEC fue asociado por primera vez con diarreas en niños en Chile en 1987¹⁶⁸ y, desde entonces, se ha podido constatar que las cepas enteroagregativas son causantes de diarreas infantiles en países desarrollados, diarreas persistentes en pacientes seropositivos, diarrea del viajero, contaminaciones de agua o alimentos y casos de diarrea esporádicos¹⁶⁹⁻¹⁷⁴. Es por ello que se ha convertido en el patógeno más común encontrado en las heces de pacientes con cuadros diarreicos^{175,176}, adquiriendo mayor importancia en los últimos años en comparación con las cepas EPEC¹⁷².

Las cepas pertenecientes al patotipo EAEC se pueden distinguir de los patotipos EPEC y DAEC por su patrón de adhesión a células intestinales HEp-2. Mientras que las células EPEC se adhieren de forma localizada formando microcolonias y las DAEC se adhieren de forma difusa, las células de cepas EAEC se disponen con un característico apilado en ladrillo, del inglés *stacked brick-like pattern*, que permite diferenciarlas¹⁶⁸. Esta adhesión está mediada por un tipo específico de fimbrias, denominadas fimbrias de adherencia agregativa (AAFs), similares a las fimbrias Dr que se encuentran en los patotipos UPEC y DAEC¹⁷². Los genes que codifican para dichos tipos de fimbrias se localizan en un grupo de plásmidos, de aproximadamente 55-65 megadaltons (MDa) de tamaño, denominado consensualmente pAA^{177,178}. En

estos mismos plásmidos se encuentra codificado un regulador clave para la expresión de determinantes de virulencia de las cepas EAEC, denominado AggR. La proteína AggR es un miembro de la familia de reguladores transcripcionales AraC/XylS^{179,180}, y se encuentra presente en todas las cepas EAEC típicas. Se trata de un activador transcripcional responsable de la regulación de factores de virulencia, presentes tanto en el plásmido pAA como en el cromosoma de las cepas EAEC¹⁸¹.

Estudios epidemiológicos demuestran la gran heterogeneidad genética que presentan las cepas EAEC, ya que se han identificado múltiples y diferentes factores de virulencia encontrados en aislados clínicos^{169,171,174,182}. Un ejemplo bien conocido de la variabilidad genéticas de las cepas EAEC es la cepa O104:H4, la cual causó un brote de diarrea hemorrágica en Alemania en 2011, con un total de 3816 casos, incluyendo 54 muertes¹⁸³. Esta cepa contenía un plásmido de la familia pAA característico de las cepas EAEC^{177,178}, pero también poseía un fago codificante para la toxina Shiga^{97,184,185}, el cual es un factor de virulencia característico de las cepas EHEC^{1,4,25,184}. No obstante, la aparición de cepas enteroagregativas con características de diferentes patotipos ya había sido descrita previamente. Al principio de los años 90, hubo un brote infeccioso en Francia que provocó el síndrome urémico hemolítico (HUS). El agente etiológico fue la cepa O111:H2, la cual mostró la capacidad de unirse a células epiteliales siguiendo el patrón de apilado en ladrillo típico de las cepas del patotipo EAEC, y era capaz de producir toxina Shiga¹⁸⁶. De la misma manera, otros casos esporádicos de infección por cepas EAEC productoras de toxina Stx ocurrieron antes del brote de 2011 en Alemania¹⁸⁷⁻¹⁸⁹. Últimamente, también ha tenido lugar un pequeño brote en 2012 en Irlanda del Norte¹⁹⁰ y en Italia en 2013¹⁹¹, ambos debidos a cepas EAEC productoras de la toxina Stx.

Es importante mencionar que, previamente, se ha descrito la relación entre cepas EAEC y cepas UPEC. En diversas cepas EAEC se ha visto la presencia de factores de virulencia característicos del patotipo UPEC¹⁴⁻¹⁶. Asimismo, en

cepas UPEC se ha detectado la presencia de marcadores característicos de cepas EAEC^{192,193}. Estos hechos demuestran la existencia de híbridos UPEC/EAEC, lo que sugiere que algunas cepas enteroagregativas pueden ser potencialmente uropatógenas, como ocurrió en el brote de infecciones urinarias en 1991 en Copenhage¹⁹⁴.

Debido a la heterogeneidad que presentan las cepas EAEC, resulta complicado definir los mecanismos precisos de virulencia de este patotipo^{172,195}. De forma general, la infección típica se inicia con la adhesión de las bacterias al epitelio del íleon y el colon. Este paso es posible gracias a las fimbrias AAF y la acción de la dispersina, que induce cambios electrostáticos en la superficie de la membrana bacteriana^{195,196}. Tras la adhesión, se produce una mayor secreción de mucus por parte de los enterocitos, provocando la formación de biofilm por parte de EAEC^{172,196}. La formación de biofilm es una de las características principales de la patogénesis de cepas EAEC, el cual les permite evadir la acción del sistema inmune y de agentes antibacterianos, como los antibióticos, y llevar a cabo infecciones recurrentes¹⁹⁶. Una vez establecido el biofilm, las células EAEC secretan diferentes toxinas que dañan los enterocitos, como es el caso de las toxinas Pet (*plasmid-encoded toxin*), que rompe la espectrina provocando daños en el citoesqueleto, o Pic (*protein involved in intestinal colonization*), una mucinasa que afecta a la integridad de la mucosa intestinal. Además, secretan las toxinas EAST1 (*EAEC heat-stable enterotoxin*), semejante a las toxinas ST de ETEC que desestabilizan el equilibrio iónico y provocan la pérdida de agua por parte de las células, y ShET1 (*Shigella enterotoxin 1*), que daña el epitelio intestinal e induce un acortamiento del microvilli¹⁹⁶ (Fig. 1.2d).

Dentro del patotipo EAEC, la cepa 042 es la mejor caracterizada, utilizándose como prototipo desde los estudios llevados a cabo por Nataro y colaboradores¹⁶⁹. El genoma de esta cepa está secuenciado y sus factores de

virulencia bien caracterizados¹⁹⁷ y, debido a la gran relevancia en esta Memoria, se va a explicar en profundidad en el próximo apartado.

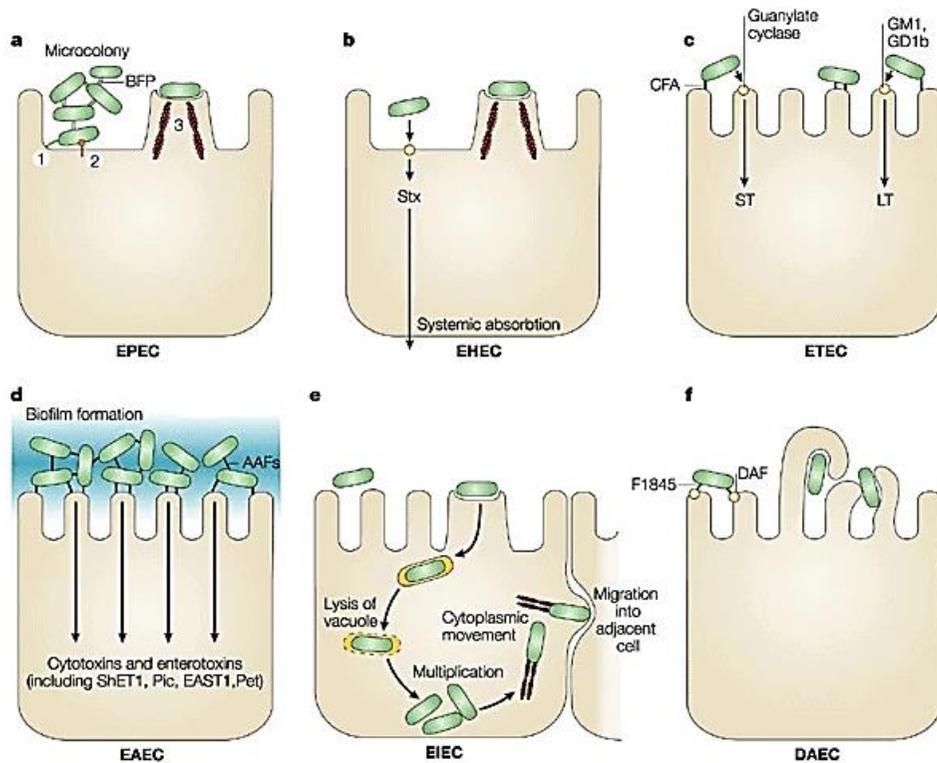


Fig. 1.2. Ilustración esquemática de la infección que lleva a cabo cada uno de los patotipos diarreicos descritos. a) EPEC se adhiere a los enterocitos, destruyendo las microvellosidades y provocando una lesión A/E característica. 1. Adhesión inicial, 2. Translocación de Tir a través de T3SS, 3. Modificación del citoesqueleto y formación del pedestal. b) EHEC provoca una lesión A/E pero en el colon, donde libera la toxina Stx, pudiéndose llegar a producir una absorción sistémica que conlleva graves complicaciones. c) ETEC se adhiere a enterocitos del intestino delgado e induce diarrea acuosa mediante la secreción de toxinas LT y ST. d) EAEC se adhiere tanto a enterocitos del intestino delgado como del grueso y forma un biofilm gracias a las fimbrias AAF. Luego libera toxinas que provocan daño celular. e) EIEC invade enterocitos del colon, lisa el fagosoma y es capaz de desplazarse intracelularmente reordenando microfilamentos de actina, propagándose a células adyacentes. f) DAEC induce un característico alargamiento atípico de las microvellosidades intestinales. Adaptado de Kaper y colaboradores¹.

1.2 La cepa EAEC 042

La cepa 042 se ha convertido en la cepa prototipo para estudiar los factores de virulencia y patogenicidad de las cepas EAEC¹⁷³. Su genoma consiste en un cromosoma circular de 5.241.977 pares de bases (pb) y un plásmido tipo pAA (denominado específicamente pAA2) de 113.346 pb, el cual pertenece al grupo de incompatibilidad IncFIIA. Este plásmido incluye 152 regiones codificantes (CDS), entre las cuales se encuentran las que codifican para las fimbrias agregativas AAF/II y el regulador transcripcional AggR¹⁹⁷.

El cromosoma de la cepa 042 incluye un total de 4.886 genes, de los cuales el 11 % codifican para proteínas hipotéticas conservadas de función desconocida y solo el 10 % corresponde a genes de elementos móviles tales como integrasas, transposasas o relacionados con bacteriófagos¹⁹⁷. La cepa 042 presenta un *core-genome* de 2.356 genes y un amplio pangenoma, hecho que concuerda con las características de los genomas de *E. coli*. También, se han identificado 78 islas genómicas que se encuentran distribuidas de forma variable entre los genomas secuenciados disponibles de *E. coli*. Estas islas genómicas se llaman ROD (*regions of difference*) y en conjunto tienen un tamaño de 1,26 Mb, lo que representa un 24 % del cromosoma. En las mismas se encuentran factores de virulencia, genes relacionados con el metabolismo, genes de función desconocida y elementos móviles como el transposón Tn2411¹⁹⁷. Este transposón es muy similar al transposón Tn21¹⁹⁸ y contiene genes que confieren resistencia a antibióticos¹⁹⁷. Tras la descripción del patotipo EAEC, se comprobó la resistencia a varios antibióticos que presentaban muchos aislamientos clínicos del mismo¹⁹⁹⁻²⁰¹. Este hecho es más frecuente que en otros patotipos diarreicos, aunque ello es quizás debido a un número mayor de aislamientos de EAEC en estudios epidemiológicos²⁰². En el caso de la cepa 042, la misma presenta resistencia a las sulfonamidas, al cloranfenicol, a los aminoglicósidos y a las tetraciclinas¹⁹⁷. Estas resistencias

están relacionadas con la presencia del transposón Tn2411, ya que contiene los genes de resistencia a cloranfenicol (*cat*) y tetraciclinas (*tetA*), y un integrón de clase 1 que presenta las resistencias a estreptomycin y espectomicina (*aadA1*), sulfonamidas (*sul*) y bromuro de etidio (*emrE*) (Fig. 1.3)¹⁹⁷.

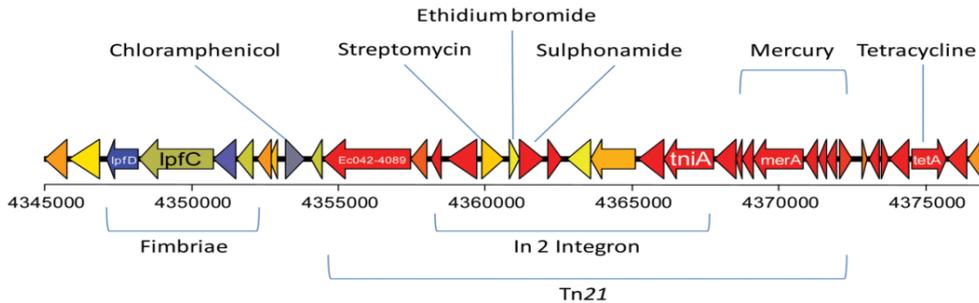


Fig. 1.3. Organización genética del transposón Tn2411 presente en el genoma de la cepa EAEC 042, el cual codifica para varias resistencias a antibióticos. Adaptado de Chaudhuri y colaboradores (2010)¹⁹⁷.

Además de la familia de fimbrias AAF que se encuentran en el plásmido pAA2, la cepa 042 codifica en su cromosoma 11 familias más de fimbrias. No hay una explicación definitiva de por qué presenta esta gran variedad de fimbrias, pero se sugiere que sirven tanto para adaptarse y colonizar diferentes hospedadores como a diferentes nichos dentro de un mismo hospedador¹⁹⁷. Por lo que hace referencia a las toxinas típicas de las cepas EAEC relacionadas con la patogénesis, en el plásmido pAA2 se encuentran los genes que codifican para las toxinas Pet y EAST1, descritas anteriormente. Es importante mencionar que otras toxinas se encuentran codificadas en el cromosoma, tal y como es el caso de ShET1 y HlyE (hemolisina E)¹⁹⁷. Esta última oligomeriza y forma poros en las membranas de las células epiteliales²⁰³.

Un análisis bioinformático del genoma de esta cepa, realizado en nuestro grupo de investigación, permitió comprobar que el mismo codifica para tres parálogos de la familia de reguladores H-NS (ver el siguiente apartado): la propia proteína H-NS, su parálogo StpA y un nuevo parálogo llamado H-NS2²⁰⁴.

En las cepas comensales de *E. coli* se encuentran solamente los genes *hns* y *stpA*. El mismo estudio permitió determinar que en dicho genoma se encuentran genes que codifican para cuatro miembros de la familia de reguladores Hha (ver apartado 1.3.2): la proteína reguladora Hha y su parálogo presente en la mayoría de cepas de *E. coli*, YdgT, y dos nuevos parálogos, llamados Hha2 y Hha3²⁰⁵. Un estudio posterior puso de manifiesto que estos nuevos parálogos de Hha se encuentran distribuidos, principalmente, en cepas virulentas de los patotipos EAEC, EHEC y UPEC. Por ello, se ha sugerido que pueden ser usados como marcadores de patogenicidad, debido a que también se encuentran presentes en algunos de los brotes infecciosos recientes más severos, como el que tuvo lugar en Alemania en 2011¹⁸⁷ o en las cepas UPEC multirresistentes pertenecientes al grupo clonal ST131, diseminado globalmente²⁰⁵. La localización de los genes que codifican estos parálogos de Hha dentro del cromosoma de las diferentes cepas de *E. coli* es variable, por lo que se ha sugerido que se han obtenido mediante mecanismos HGT²⁰⁵. Curiosamente, ninguno de estos parálogos se encuentra en plásmidos, por lo que su transmisión no ha podido ocurrir por conjugación y es razonable pensar que la obtención de estos reguladores puede estar ligada a la obtención de otros genes o factores de virulencia²⁰⁵.

1.3 Proteínas asociadas al nucleoide

El ADN bacteriano y las macromoléculas asociadas se encuentran formando un complejo que se denomina nucleoide bacteriano, característico de los organismos procariontes donde no hay un núcleo definido²⁰⁶. Este nucleoide necesita ser compactado y organizado, de manera que sea compatible con el tamaño de la bacteria²⁰⁷⁻²⁰⁹. La organización del nucleoide tiene lugar mediante la acción de varios factores, entre los que se encuentran la alta densidad macromolecular (causada principalmente por la alta concentración de

proteínas globulares y ARN dentro de la célula)^{210,211}, la flexibilidad del ADN²¹², su superenrollamiento^{213,214} y la interacción con proteínas asociadas al nucleóide (NAPs)^{207,208}.

Las NAPs tienen capacidad para unirse al ADN y alterar su conformación, ya sea doblándolo, envolviéndolo o creando puentes intramoleculares^{207,208,215}. Además de estructurar el cromosoma bacteriano, estas proteínas pueden regular la expresión de múltiples genes, interfiriendo en los procesos de replicación, transcripción y traducción^{207,208}. Por lo tanto, se puede decir que tienen una doble función, ya que por una parte contribuyen al mantenimiento de la estructura del nucleóide y, por otra, actúan como reguladores globales de la expresión génica^{207,208}. Una de las características más importantes de las bacterias es su capacidad para adaptarse rápidamente a los cambios que se producen en el ambiente, gracias a una modificación del patrón de expresión de sus genes^{208,216}. Las NAPs participan en estos procesos de alteración del patrón transcripcional en respuesta a cambios ambientales²¹⁷. Las más estudiadas en bacterias Gram-negativas son HU (*Heat-Unstable Nucleoid Protein*), IHF (*Integration Host Factor*), Fis (*Factor Inversion Stimulation*), Lrp (*Leucine-responsive regulatory protein*), H-NS (*Histone-like nucleoid structuring protein*) y Hha (*High hemolytic activity*)^{208,215}.

1.3.1 La familia de proteínas H-NS

Las proteínas de tipo H-NS se encuentran distribuidas dentro de la familia Enterobacteriaceae y otros grupos de bacterias Gram-negativas, tales como *Vibrio* y *Pseudomonas*^{207,208,218}. La proteína H-NS fue descrita como uno de los componentes mayoritarios del nucleóide bacteriano de *E. coli*^{219,220} y ha sido ampliamente estudiada, tanto en *E. coli* como en *Salmonella*, donde interviene en la organización del nucleóide y en la regulación de la expresión génica^{207,208,221,222}.

El gen *hns* codifica para una proteína de bajo peso molecular (15,6 kDa) con un total de 137 aminoácidos. Como ya se ha dicho, la proteína H-NS es muy abundante en el nucleóide bacteriano, llegando a alcanzar hasta 20.000 copias por célula^{221,223}. Este regulador afecta a la expresión de un elevado número de genes, ya que, dependiendo de la cepa de *E. coli*, aproximadamente el 5 % de los genes cromosómicos son diana de H-NS²²⁴⁻²²⁶. Por ello, los niveles de expresión de la proteína H-NS dentro de la célula son muy importantes para su función, manteniéndose relativamente constantes a lo largo del crecimiento^{221,223}. H-NS se expresa a un nivel similar tanto a 37 °C como a 25 °C, pero interviene en la regulación de genes de una forma diferente²²⁷⁻²²⁹. A diferencia de otros reguladores transcripcionales que reconocen una secuencia específica, H-NS presenta afinidad por regiones ricas en adenina y timina (AT) con cierta curvatura^{221,230}. Debido a que en vez de una secuencia reconoce un motivo estructural, H-NS es capaz de regular la expresión génica de una gran cantidad de genes, actuando como represor de forma global^{208,218,222,231}. Como se ha comentado anteriormente, la incorporación de ADN mediante mecanismos HGT permite a las bacterias adaptarse a nuevas condiciones ambientales²¹⁷, desarrollar la capacidad de colonizar nuevos hospedadores^{232,233} o adquirir la resistencia a agentes antimicrobianos^{234,235}. Sin embargo, la adquisición de forma súbita de este nuevo material genético puede presentar efectos perjudiciales para la propia bacteria, ya que incorpora un importante número de genes cuya expresión no está integrada en las redes reguladoras de la célula^{221,232}. Es aquí donde la proteína H-NS tiene un papel esencial, reconociendo aquellas secuencias con un contenido en AT superior a las del cromosoma de la célula receptora. Su unión provoca la represión de la expresión de los genes situados aguas abajo de los lugares de unión de H-NS. Este mecanismo, conocido como silenciamiento xenogénico, posibilita que las bacterias puedan incorporar nuevas funciones sin que el *fitness* celular se vea comprometido^{236,237}. A su vez, los genes silenciados por H-NS pueden expresarse en condiciones en las que esta proteína no ejerza su papel regulador

como, por ejemplo, condiciones de elevada temperatura, de estrés o por la presencia de moléculas que interfieren en su función tales como CRP (*cAMP-activated global transcriptional regulator*) o Ler (regulador de la región LEE)^{238,239}.

Por lo que hace referencia a su estructura, H-NS presenta dos dominios independientes que se corresponden con sus dos dominios funcionales. El extremo N-terminal es responsable de la capacidad de la proteína para dimerizar y oligomerizar, tanto entre monómeros de H-NS como con otras proteínas^{240,241}. El extremo C-terminal interviene fundamentalmente en la interacción de la proteína con el ADN^{240,242}. Ambos dominios se encuentran unidos a través de una región flexible (*linker*), la cual les permite actuar de manera independiente (Fig. 1.4)^{241,243}. El dominio N-terminal es fundamental para la oligomerización, mediante la cual forma un complejo nucleoproteico esencial para llevar a cabo la represión transcripcional, ya sea impidiendo la unión de la ARN polimerasa al ARN o secuestrándola²⁴⁴. Además, residuos concretos de este dominio también están involucrados en el reconocimiento de las regiones curvadas de ADN y, por tanto, en la unión a éstas conjuntamente con el dominio C-terminal^{221,241}. El dominio N-terminal presenta un buen porcentaje de conservación entre los distintos miembros de la familia de proteínas H-NS, aunque su grado de conservación es menor que la del dominio de unión al ADN presente en el extremo C-terminal²²⁶. Estas diferencias del dominio N-terminal pueden estar relacionadas con la adaptación de las bacterias como respuesta a diferentes condiciones, lo que puede conllevar variaciones en la afinidad con la que se unen a otras proteínas o al ADN^{218,245}. El dominio de dimerización permite a H-NS interactuar con otras proteínas y, mediante la formación de estos oligómeros, es capaz de regular la expresión génica de genes diferentes a los que regula de manera individual^{245,246}. Estos oligómeros se forman por la interacción de H-NS con proteínas de la misma

familia, como es el caso de su parólogo StpA, o con otras proteínas, como aquellas pertenecientes a la familia Hha^{208,221}.

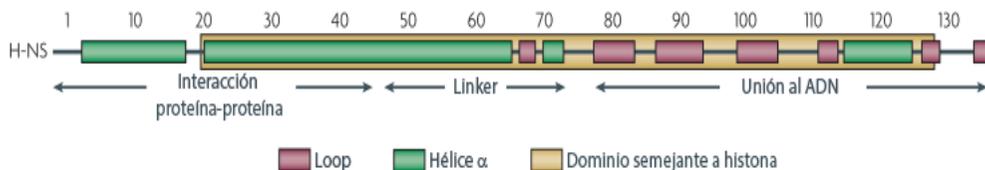


Fig. 1.4. Organización de los dominios y estructura secundaria de la proteína H-NS. En el extremo N-terminal se indica el dominio de interacción proteína-proteína y en el C-terminal el de unión a ADN, unidos por la región flexible o *linker*. En morado se destacan los bucles mientras que en verde las hélices alfa, presentes en la estructura secundaria de la proteína H-NS. En amarillo se resalta el dominio semejante a histonas, característico de la familia de proteínas H-NS. Adaptado de Dillon y Dorman²⁰⁸.

Han sido descritos diferentes parálogos y ortólogos del gen *hns*. El primero de ellos fue el gen parólogo *stpA*, identificado como una proteína supresora del fenotipo Td⁻ de fagos T4²⁴⁷. Se encuentra codificado en la mayoría de genomas de las enterobacterias y tiene un 67 % de similitud, a nivel de secuencia amonoacídica, con H-NS²⁴⁸. La estructura de la proteína StpA es similar a la de la proteína H-NS, como se puede ver en la Fig. 1.5. Es capaz de unirse tanto al ADN como al ARN, y se ha visto que tiene capacidad para actuar como chaperona²⁴⁹. Posee un dominio de oligomerización similar al de H-NS, por lo que es capaz de interaccionar y formar oligómeros tanto con H-NS^{208,221} como con proteínas de la familia Hha²⁵⁰, aunque tienen una diferencia clave en el aminoácido 21 (Fig. 1.5). En el caso de H-NS, en la posición 21 se encuentra una cisteína, mientras que en StpA hay una fenilalanina, lo cual provoca que StpA sea sensible a la degradación por parte de la proteasa Lon²⁵¹. A pesar de la alta homología que presentan las proteínas H-NS y StpA, se ha determinado que, bajo ciertas condiciones, ambas proteínas pueden desempeñar funciones biológicas distintas y unirse al ADN mediante mecanismos diferentes^{252,253}.

Por lo que hace referencia a su expresión, StpA presenta una expresión baja en un medio rico. No obstante, la misma puede aumentar cuando las células crecen en medio mínimo o se encuentran en condiciones de estrés²⁵⁴. También se ha descrito que estos niveles de expresión dependen de temperatura, ya que a 37 °C tiene mayor expresión que a 25 °C²⁵⁵. De manera semejante, estas diferencias en la expresión se encuentran durante el crecimiento, aumentando la misma a mitad de la fase exponencial y reduciéndose al entrar en fase estacionaria. Por el contrario, los niveles de H-NS se mantienen constantes a lo largo del crecimiento²⁵⁶. Estas diferencias en los niveles intracelulares de ambas proteínas pueden estar debidas a su propia autorregulación, a la regulación cruzada negativa que presentan, donde la represión que ejerce H-NS sobre el promotor de *stpA* es mayor que la de StpA sobre el promotor de *hns*²⁵⁷, o al hecho de que StpA es sensible a la actividad proteolítica de la proteasa Lon. Sin embargo, la delección del gen *hns* provoca que el gen *stpA* acumule mutaciones para evitar la degradación por la proteasa Lon y aumentar los niveles de expresión, compensando parcialmente la ausencia de H-NS y actuando como *backup*²⁵⁷.

La función que desempeña H-NS silenciando el ADN adquirido por transferencia horizontal, tal y como sucede en las PAI, ha sido ampliamente estudiada^{236,237}. En algunos casos, en estas PAI se encuentran codificados reguladores semejantes a H-NS. Un ejemplo de ello es la proteína Hfp, la cual se encuentra en la cepa UPEC 536²⁵⁸. La proteína Hfp es capaz de formar heterodímeros con las proteínas H-NS y StpA y sus redes de regulación están conectadas, ya que los tres están sujetos a regulación cruzada. El patrón de expresión que sigue el gen *hfp* es complementario al que presenta *stpA*, consiguiendo de esta manera mantener la actividad reguladora tanto a bajas como a elevadas temperaturas. Como consecuencia, el rango de temperaturas donde tiene efecto la red reguladora de las proteínas de la familia H-NS es mucho más amplio. Esto resulta de gran interés para las bacterias a la hora de

infectar, debido a que la transición del ambiente externo al huésped requiere una rápida adaptación a condiciones de estrés, como es el caso de la temperatura²⁵⁸.

Además de los parálogos de H-NS codificados en el genoma, como StpA o aquellos que se encuentran en PAIs, existen ortólogos presentes en plásmidos. Entre estos, los plásmidos IncHI1 R27 de *Salmonella* y el plásmido pSFR27 de *Shigella* codifican las proteínas H-NS_{R27} y Sfh, respectivamente^{259,260}. De manera semejante al parálogo Hfp encontrado en la cepa UPEC, estos reguladores codificados en plásmidos son capaces de formar heterodímeros con H-NS y StpA, participando en la regulación cruzada de estos genes. Aunque estas proteínas muestran similitud con H-NS, su patrón de expresión a lo largo del crecimiento es diferente y ejercen su función reguladora en otro conjunto de genes^{260,261}. Por ejemplo, la proteína H-NS_{R27} se ha especializado en la regulación de genes adquiridos por procesos HGT, mientras que H-NS lleva a cabo su función tanto en genes HGT como en genes del *core-genome*^{222,259}. En el caso de Sfh, la proteína se encarga de regular los genes del plásmido pSFR27 donde se encuentra codificado, evitando de esta manera que la concentración intracelular de la proteína H-NS citoplasmática se diluya al incorporar un plásmido de elevado peso²⁶⁰.

En cuanto al nuevo parálogo H-NS2 recientemente descrito en la cepa 042, un análisis transcriptómico ha puesto de manifiesto que este parálogo se ha especializado en la regulación de genes HGT junto a la proteína Hha²⁰⁴. La expresión de H-NS2 está regulada por temperatura y por la disponibilidad de nutrientes, ya que tiene una mayor expresión a 37 °C que a 25 °C y una mayor expresión en medio mínimo frente a un medio rico²⁰⁴. Por otra parte, el análisis comparativo de su estructura refleja la gran similitud que tiene el dominio N-terminal de oligomerización de H-NS2 con el de H-NS, donde se mantiene conservado el sitio de unión con Hha (aminoácidos 9-16). Sin embargo, el dominio C-terminal tiene una menor similitud, hecho que puede estar

relacionado con las diferencias en el patrón de genes que regula cada NAP de esta familia (Fig. 1.5)²⁰⁴.

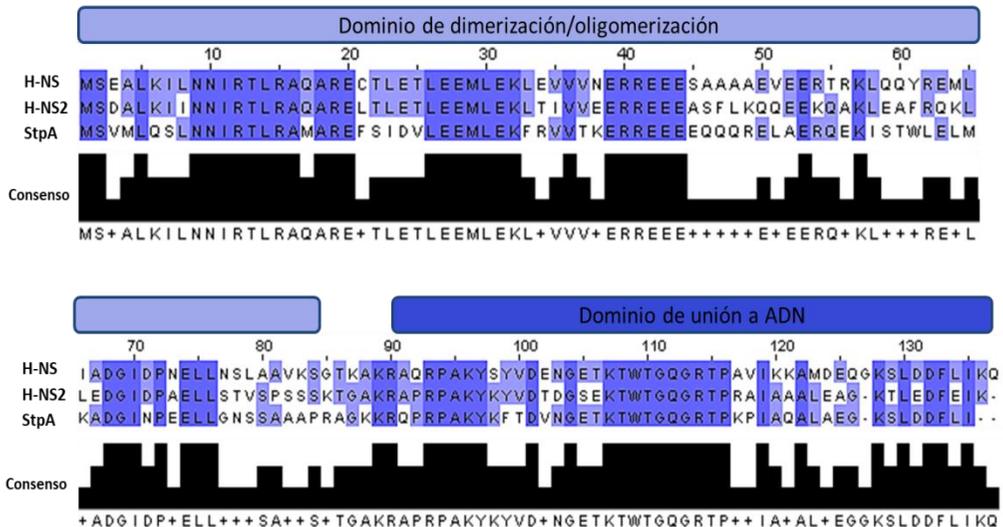


Fig. 1.5. Alineamiento de la secuencia aminoacídica de las proteínas H-NS, H-NS2 y StpA. Los aminoácidos conservados están en dos tonalidades de azul y los no conservados en blanco. En la parte inferior se muestra la secuencia consenso y las barras azul claro y azul oscuro los dominios N-terminal y C-terminal, respectivamente. Se puede ver la diferencia en el aminoácido 21, sensible a la acción de la proteasa Lon. Adaptado de Prieto y colaboradores²⁰⁴.

1.3.2 La familia de proteínas Hha

La familia de proteínas Hha está formada por un grupo de proteínas de bajo peso molecular (aproximadamente 8 kDa), involucradas en la regulación de la expresión génica en diferentes géneros de enterobacterias²⁶². Los primeros miembros de la familia de proteínas de tipo Hha descritos fueron la proteína Hha de *E. coli*²⁶³ y la proteína YmoA de *Yersinia enterocolitica*²⁶⁴. Los mutantes *hha* se identificaron como sobreproductores de la toxina alfa-hemolisina en cepas uropatógenas de *E. coli*²⁶³. Además, dichos mutantes exhibían un fenotipo pleiotrópico, afectando a la compactación del ADN plasmídico²⁶⁵ y al aumento de la transposición de secuencias de inserción

(IS)²⁶⁶. Estos cambios en el fenotipo son semejantes a los efectos que se observan en un mutante *hns* de *E. coli*.

Inicialmente, se estudió la unión de la proteína Hha al ADN, utilizando el modelo del operón hemolítico *hly* de la cepa *E. coli* 536. A pesar de utilizar diferentes tipos de aproximaciones, no pudo demostrarse que la proteína Hha fuese capaz de unirse a secuencias específicas de este operón ni a otras. Como alternativa, se estudió la capacidad de la proteína Hha para interactuar con otras proteínas, encontrándose que interactúa con la proteína H-NS, la cual se une de manera específica a secuencias de la región reguladora del operón hemolítico²⁶⁶. Por lo tanto, los dos reguladores, junto a la secuencia de ADN a la que se unen, forman un complejo nucleoproteico Hha/H-NS/ADN²⁶⁷. Este hecho concuerda con la sobreexpresión de la toxina alfa-hemolisina que tiene lugar tanto en un mutante sencillo *hha* o *hns* como en un doble mutante *hha hns*²⁶⁸, justificando que mutantes de los reguladores Hha y H-NS presenten los mismos fenotipos pleiotrópicos. Investigaciones posteriores han justificado la importancia de la interacción que hay entre ambos reguladores, ya que la unión de la proteína Hha puede alterar la especificidad de H-NS por un grupo de genes, modulando su actividad. Esta modulación de la especificidad de H-NS ocurre en el silenciamiento xenogénico²²², resaltando la importancia que tiene la familia de proteínas Hha en el mantenimiento del *fitness* celular cuando se produce la adquisición de ADN foráneo²⁶⁹.

Observando la estructura, la proteína Hha presenta motivos de hélice-giro-hélice, mostrando además pequeños bloques con una alta identidad con el dominio N-terminal de H-NS²⁷⁰. Pese a que Hha no tiene dominios definidos, sí que muestra una disposición dipolar, situándose los residuos con carga positiva en el lado opuesto a los cargados negativamente^{271,272}. Debido a esta disposición, es posible que cada una de las superficies lleve a cabo funciones específicas diferentes, como es el caso de los residuos de carga positiva que se distribuyen de manera similar a la del dominio de unión a ADN de H-NS²⁷³.

Gracias a estudios de cristalización con rayos X, se ha podido determinar la estructura de los complejos de Hha con el dominio de dimerización de H-NS, estableciendo un modelo que explica la interacción que tiene lugar en la formación del complejo nucleoproteico. El modelo describe la unión de dos monómeros de Hha de manera simétrica a ambos lados de un dímero de H-NS, aportando interacciones electrostáticas adicionales y estabilizando el complejo formado (Fig. 1.6)²⁷².

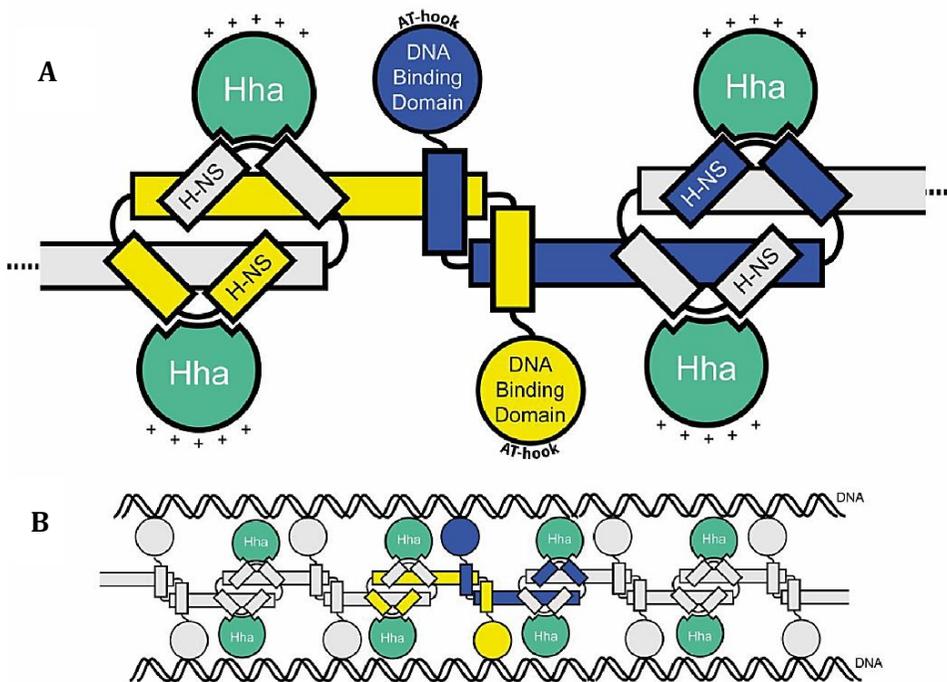


Fig. 1.6. Modelo de interacción de la proteína Hha con H-NS. A) Esquema de la unión de 4 moléculas de Hha con 2 dímeros de H-NS. Las moléculas de Hha se representan en verde y los dímeros de H-NS en amarillo y azul. Los motivos de unión a zonas ricas en AT del dominio de unión de H-NS (extremo C-terminal) interactúan con el ADN. Los residuos cargados positivamente de la proteína Hha están señalados con signos de más (+) y quedan orientados de la misma forma que el dominio de unión de H-NS. B) Modelo de la formación del complejo nucleoproteico Hha/H-NS/ADN. Adaptado de Ali y colaboradores²⁷².

En *S. Typhimurium*, se ha descrito que la proteína Hha regula conjuntamente con la proteína H-NS un grupo de genes, muchos de los cuales son HGT. Entre estos genes se incluyen PAI, proteínas de membrana celular y relacionados con la síntesis y transporte de proteínas, destacando la regulación de genes asociados con el transporte de hierro y con la motilidad²⁷⁴. Como ejemplo se encuentra la regulación del gen *hilA*, el cual es un activador transcripcional de genes implicados en procesos infecciosos de invasión localizado en la isla de patogenicidad 1 (SPI-1). Hha, juntamente con H-NS, actúa como represor de la expresión de *hilA* en condiciones de baja temperatura. Por lo tanto, ambas proteínas se encuentran implicadas en la cascada de regulación que se encarga de modular la expresión de los efectores de T3SS presente en la SPI-1 de *Salmonella*^{275,276}.

La distribución de las proteínas que pertenecen a la familia Hha está restringida a las enterobacterias, pudiendo estar presentes en una o más copias por cromosoma o en elementos transmisibles, como los plásmidos^{262,271,277}. En el 55 % de los plásmidos que codifican un miembro de la familia de proteínas de tipo H-NS se ha visto que también codifican para un ortólogo de Hha²⁷⁷. De forma similar, se encuentra presente una copia de un gen tipo *hns* en el 31 % de los plásmidos que codifican al menos una copia de genes tipo *hha*²⁷¹. Esto refuerza el papel de la relación existente entre estos dos reguladores. Los plásmidos que presentan una copia de *hha*, pero no contienen la correspondiente copia del gen *hns*, pertenecen al grupo de incompatibilidad IncF, distribuidos en cepas de *E. coli* y *Shigella* exclusivamente. Dentro de la amplia variedad de plásmidos que pueden encontrarse en las cepas de *E. coli*, los plásmidos IncF destacan por codificar un gran número de factores de virulencia^{278,279}. Por lo tanto, es posible que estos factores de virulencia estén regulados por las proteínas tipo Hha, demostrando la importancia de este regulador en el control de la expresión de la virulencia de cepas patógenas de *E. coli*²⁷¹.

Como regla general en el cromosoma de las enterobacterias, se encuentran sendos parálogos de los genes *hns* y *hha* (los genes *stpA* e *ydgT*, respectivamente), pudiendo presentarse también ortólogos de estas dos genes codificados en plásmidos conjugativos²⁸⁰. Tanto la proteína StpA como YdgT aumentan respectivamente su expresión en mutantes que carecen de H-NS o Hha, atenuando los efectos pleiotrópicos derivados^{262,281}. Al margen de actuar como un reservorio de H-NS/Hha, la función reguladora de estos parálogos sigue sin estar totalmente clara en *E. coli*.

Tal y como se comentó anteriormente, un análisis genómico reciente realizado en la cepa EAEC 042 ha puesto de manifiesto la presencia, a diferencia de muchas otras cepas de *E. coli*, de tres parálogos del gen *hns* (*hns*, *stpA* y el nuevo parálogo *hns2*) y hasta cuatro parálogos de *hha* (*hha*, *ydgT* y los nuevos alelos *hha2* y *hha3*) (Fig. 1.7)^{204,205}. Al estudiar la distribución de estos nuevos parálogos de *hha* en una gran cantidad de cepas de *E. coli* comensales y patógenas, se pudo comprobar cómo su presencia está ligada a cepas virulentas de *E. coli*, por lo que se ha propuesto que estos nuevos parálogos podrían usarse como marcadores para identificar y caracterizar aislamientos patógenos de este microorganismo²⁰⁵. *E. coli* tiene una elevada plasticidad genómica, debida en buena parte a procesos de incorporación de genes por mecanismos HGT. Esto se traduce en que incluso los aislamientos que pertenecen a un patotipo específico muestran combinaciones diferentes de factores de virulencia^{183,194}, los cuales pueden presentarse incluso en algunas cepas comensales²⁸². Por lo tanto, el uso de estos nuevos marcadores moleculares, diferentes a los factores de virulencia clásicos, puede ayudar a identificar los aislamientos de cepas patógenas de *E. coli*.

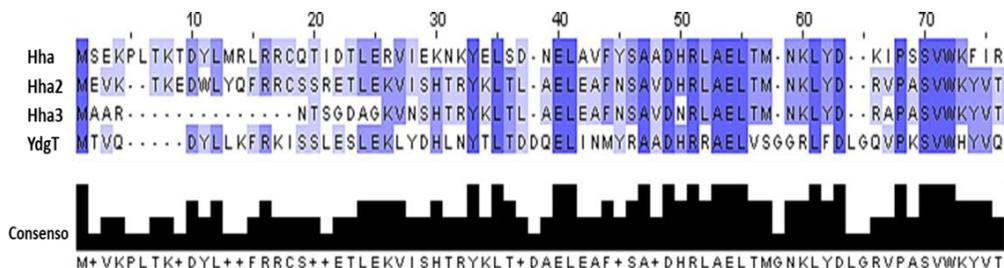


Fig. 1.7. Alineamiento de la proteína Hha de la cepa EAEC 042 con todos los parálogos codificados en el genoma de dicha cepa. Se representa la conservación de los aminoácidos en una escala de colores azules, siendo los más oscuros los más conservados y los claros los menos. Adaptado de Prieto y colaboradores²⁰⁵.

Si bien se ha estudiado en mucho detalle la función de reguladores globales tales como H-NS, Hha, CRP o Fis, no se ha realizado ningún estudio a fondo del papel de sus correspondientes parálogos y ortólogos, muchos de los cuales se encuentran frecuentemente en plásmidos²⁸⁰. Consecuentemente, el estudio de parálogos u ortólogos de moduladores globales es necesario, pudiendo proporcionar información importante para comprender mejor la biología de las células bacterianas. En este contexto, la posible correlación entre la presencia de un nuevo regulador global con la presencia de nuevos factores de virulencia, regulados por el mismo y que definan el estilo de vida del aislado patógeno, se debe tener en cuenta. Por lo tanto, realizar un análisis comparativo entre genomas de cepas que codifican estos nuevos parálogos *hha2/hha3* frente a cepas que no los tienen resulta de gran interés.

1.4 Duplicaciones génicas en bacterias

Las duplicaciones génicas tienen lugar tanto en eucariotas como en procariotas y son consideradas uno de los mecanismos evolutivos más importantes²⁸³. Las familias de genes y proteínas relacionadas entre sí son muy comunes en los genomas de bacterias y, en *E. coli*, llegan a representar

alrededor de la mitad de todo el genoma. Este hecho sugiere que ha habido un posible proceso de duplicación génica y diversificación a lo largo de la evolución, dando como resultado una gran diversidad y complejidad genómica²⁸⁴⁻²⁸⁷.

Los procesos de duplicación y amplificación génica (GDA) juegan un papel realmente importante en la evolución, pudiendo dar lugar a la aparición de nuevos fenotipos, aumentando la adaptabilidad a condiciones ambientales adversas e incrementando el *fitness* celular. Según el mecanismo mediante el cual se hayan adquirido estas nuevas funciones, se presentan tres escenarios evolutivos diferentes (Fig. 1.8)^{285,288}. En primer lugar, hay que hablar de la neofuncionalización, proceso en el cual una de las copias mantiene las funciones ancestrales mientras que, la otra, sufre mutaciones y adquiere una nueva función dentro del organismo. Sin embargo, este proceso es extremadamente complicado que tenga lugar, debido a la complejidad que conlleva obtener una función totalmente nueva a partir de otro gen preexistente^{285,288}. Por otra parte, en el caso que el gen ancestral esté implicado en diversos procesos, puede tener lugar un proceso de subfuncionalización. En este proceso, cada copia que se ha generado del gen se especializa en una de las funciones que llevaba a cabo el gen ancestral bifuncional, mediante la selección de mutaciones que optimizan cada una de las funciones^{289,290}. La subfuncionalización es típica en proteínas que llevan a cabo una función principal pero también son capaces de realizar funciones secundarias, las cuales no representan ni un beneficio ni un coste para la célula. En el momento en el que las condiciones del medio o el nicho varían, es posible que alguna de estas actividades secundarias sea realmente importante para la supervivencia, llevando a cabo su selección. Como respuesta, la generación de duplicaciones y amplificaciones génicas aumentan la disponibilidad de esta proteína, pudiendo ejercer la función secundaria de interés con más frecuencia. Posteriormente, a través de mutaciones en las nuevas copias generadas, se optimiza esta

función²⁹¹. Por último, en algunos casos ambas copias se mantienen inalterables y llevan a cabo las mismas funciones, ya que el aumento de la dosis génica y la actividad de dicho gen resultan de gran interés por sí mismo. Un ejemplo claro de aumento de la dosis génica tienen lugar en los genes implicados en la resistencia a agentes antimicrobianos, como son los antibióticos o las toxinas^{292,293}. Las resistencias a agentes antimicrobianos presentes en el genoma de bacterias muchas veces han sido adquiridas por HGT y sufren procesos GDA, por lo que estos dos mecanismos están íntimamente relacionados entre sí, siendo una muestra más de la gran adaptabilidad que muestran las bacterias²⁹¹. En muchas ocasiones, el material genético adquirido se expresa a niveles subóptimos o no lleva a cabo su función de manera eficiente en el nuevo hospedador, por lo que la duplicación de estos genes resulta muy importante para su fijación en el genoma y para alcanzar los niveles de expresión necesarios²⁹¹. Este hecho concuerda con las observaciones de Hooper y Berg, donde concluyen que la duplicación génica tiene lugar con más frecuencia en elementos HGT²⁹⁴. Hay que decir que existe un proceso directamente derivado de los mecanismos GDA descritos, llamado pseudogenización. Este proceso se caracteriza por la acumulación de mutaciones que alteran la estructura y la función de uno de los dos genes, pero no son perjudiciales, por lo que no llegan a eliminarse del genoma inmediatamente, convirtiéndose en un pseudogen. Estos pseudogenes no se expresan o no son funcionales y, posteriormente, se eliminarán del genoma o se desviarán tanto del gen parental que ya no serán identificables²⁸⁵.

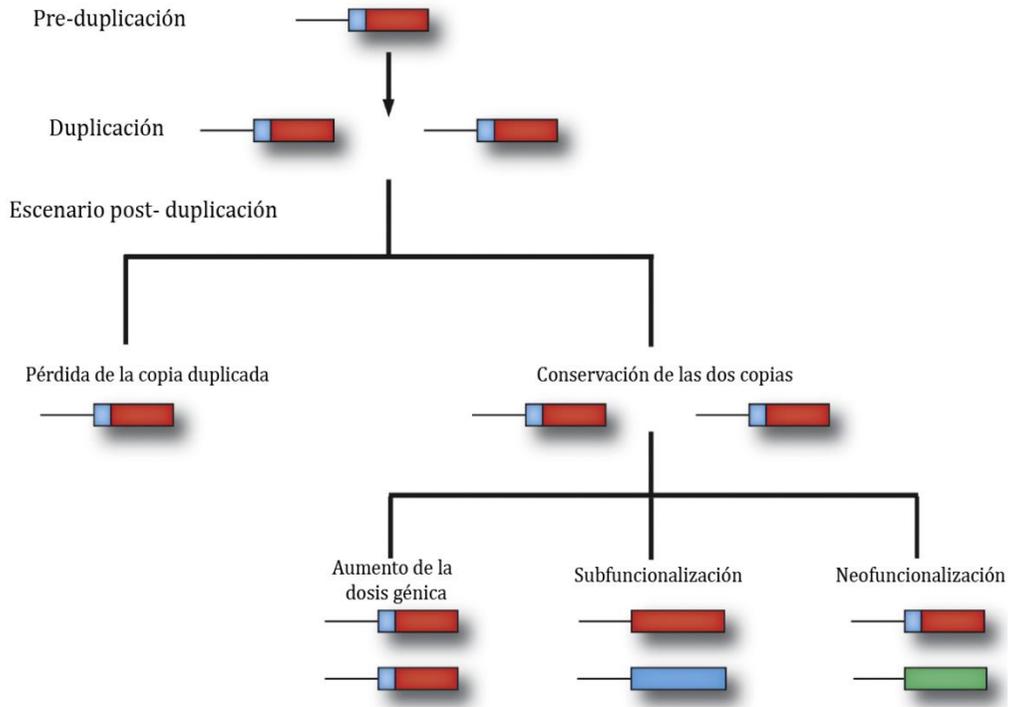


Fig. 1.8. Posibles escenarios tras un proceso de duplicación. El escenario más común es la pérdida de las copias duplicadas. Sin embargo, si la copia duplicada se conserva en el genoma, pueden tener lugar tres procesos diferentes: aumento de la dosis génica, subfuncionalización y neofuncionalización. Las funciones diferentes de los genes se representan con los colores rojo, azul y verde. Adaptado de Voordeckers y Verstrepen²⁸⁸.

Como se ha descrito anteriormente, las bacterias pueden utilizar dos métodos para modificar su genoma. Bien incorporando nuevo material genético mediante mecanismos HGT^{5,6} o bien alterando su dotación genómica^{7,12}. En este último caso, los mecanismos de duplicación génica juegan un papel importante, siendo estudiados como modelo evolutivo²⁸³. En los mismos, las duplicaciones accidentales de genes pueden seleccionarse positivamente por conferir a la bacteria nuevas propiedades fenotípicas, entre las que cabe resaltar tanto el incremento en la expresión del producto del gen duplicado, como la regulación

diferencial de las dos copias, lo que conlleva a su vez potenciar la capacidad de respuesta de las células a diferentes condiciones ambientales. Por ello, estos procesos GDA resultan de gran relevancia biológica, ya que son capaces de generar una mayor diversidad de genes, sobre los que actuará la selección natural debida a las necesidades adaptativas^{284,294,295}. De forma general, el aumento de la dosis génica implica la existencia de un mayor número de dianas donde pueden tener lugar mutaciones puntuales concretas que impliquen efectos pleiotrópicos, permitiendo superar las limitaciones en el crecimiento que se pueden presentar en determinados ambientes^{283,296,297}. La selección positiva de las duplicaciones generadas tiene lugar cuando el beneficio que produce la duplicación es mayor que el coste energético que conlleva. Debido a esto, la frecuencia de las duplicaciones que implican un menor coste aumentará en detrimento de aquellas que provocan una tasa de crecimiento más lenta. Sin embargo, la presencia de una condición o un factor limitante para la célula provoca un aumento de la frecuencia de duplicación, con la finalidad de mejorar el crecimiento, ya que el resultado final compensa el gasto que conlleva el proceso^{298,299}. Una vez que la célula consigue una mutación óptima en uno de estos genes duplicados que le permite superar el factor limitante del crecimiento, se reduce la tasa de mutación y el resto de mutaciones subóptimas generadas en el proceso se van perdiendo, posteriormente, por segregación^{298,299}. Algunas de las condiciones limitantes que incrementan la aparición de duplicaciones en el genoma son el uso de nuevos metabolitos existentes en el medio, la ausencia de algún nutriente esencial, la presencia de algún agente antimicrobiano o metales pesados en el medio, así como la supervivencia a unas condiciones ambientales extremas, donde la necesidad de nuevos genes o el desarrollo de nuevas funciones en genes preexistentes para superar estos problemas es vital^{291,293}. No obstante, es importante citar los datos obtenidos en estudios recientes, en los cuales se resalta que en una población bacteriana donde no existe ninguna condición de estrés más del 20 % de las bacterias pueden llegar a contener una duplicación en alguna región del

genoma. Por lo tanto, los mecanismos GDA no están únicamente relacionados con la presencia de condiciones de estrés³⁰⁰.

Es evidente la importancia que representan los mecanismos GDA para la evolución de los genomas, pero la investigación de los mismos ha resultado un desafío a lo largo de la historia por la falta de herramientas o técnicas eficientes para detectarlas. Esta complejidad para su estudio también es debida a que muchas de ellas se generan como respuesta a un estrés y se pierden cuando desaparece la causa, sin llegar a fijarse en el genoma de forma estable. En la actualidad, el desarrollo de las técnicas de secuenciación masiva ha permitido secuenciar un número cada vez mayor de genomas, los cuales se encuentran disponibles en las bases de datos. Todos estos genomas, junto a las herramientas bioinformáticas, representan la mejor herramienta disponible para llevar a cabo análisis genómicos precisos²⁹¹. Como ejemplo del uso de estas nuevas herramientas disponibles se encuentra el estudio realizado por Arun y colaboradores en la cepa comensal *E. coli* MG1655, en el cual categorizan todos los genes según su estado de conservación, su expresión o su duplicación, a través de la comparación con un gran número de genomas secuenciados disponibles³⁰¹. Esta comparación les permite concluir que los genes que presentan una mayor conservación y un menor grado de duplicación son aquellos implicados en funciones esenciales, como la replicación del ADN, la transcripción, la traducción o el transporte de proteínas. Sin embargo, aquellos genes involucrados en mecanismos que ayudan a la adaptación a las condiciones del medio sí que son más propensos a sufrir algún proceso de duplicación, como es el caso de las proteínas de membrana o los reguladores transcripcionales de estos genes³⁰¹. Además, se ha comprobado que las diferencias entre genes duplicados residen principalmente en su regulación más que en su función³⁰², por lo que el estudio de la regulación de estos genes duplicados es importante para conocer mejor la función que desempeñan en la biología de los microorganismos.

La existencia de genes duplicados en genomas de *E. coli* ha sido puesta de manifiesto anteriormente en diferentes trabajos. Un ejemplo de ello es el caso del gen *flu* en muchas cepas, incluida la cepa 042^{197,303}, el cual codifica para la adhesina Agn43 (ver siguiente apartado). Otro ejemplo es el estudio citado más arriba (llevado a cabo en la cepa MG1655), donde comparan y categorizan el conjunto de genes duplicados en *E. coli*³⁰¹. Pese a la disponibilidad de un gran número de genomas secuenciados, no se ha realizado hasta la fecha una investigación en profundidad sobre la distribución de las duplicaciones y amplificaciones en cepas patógenas de *E. coli* pertenecientes a los diferentes patotipos. Los datos que se pueden generar en este tipo de estudio resultarían de gran interés y aportarían más información sobre la biología de los diferentes patotipos, incluso con potencial para aplicarse en campos de conocimiento muy diversos, como es la búsqueda de nuevas dianas terapéuticas. Teniendo en cuenta la gran flexibilidad y plasticidad del genoma de *E. coli*, es más que probable que los procesos GDA no solo se hayan restringido a su *core-genome*, sino también al amplio pangenoma que presenta esta bacteria, por lo que se trata de un estudio muy extenso y necesario.

1.5 Operón Agn43-YeeR-Aec69

La colonización de múltiples ambientes requiere tanto de unas importantes capacidades adaptativas como de una gran capacidad de adhesión a diferentes superficies³⁰⁴. Numerosos estudios han demostrado que *E. coli* codifica en su genoma un importante número de adhesinas con diferente especificidad³⁰⁵⁻³⁰⁷, las cuales se pueden agrupar en dos grandes familias. Por una parte, se encuentran las adhesinas asociadas a chaperonas que reconocen proteínas glicosiladas o lípidos^{308,309}, mientras que por otra parte se encuentran los autotransportadores relacionados con el sistema de secreción de tipo 5 (T5SS)^{310,311}. La expresión de estos autotransportadores se asocia muchas veces

con la virulencia, y los mismos se caracterizan por reconocer receptores específicos y por estar implicados en las interacciones entre células bacterianas, provocando agregación celular^{312,313}. Entre todos estos autotransportadores con capacidad de adherencia hay que destacar la proteína Agn43, la cual se ha relacionado con la formación de biofilm³¹⁴ o la colonización de la vejiga en infecciones urinarias³⁰⁵.

La proteína Agn43 está codificada por un gen denominado *agn43* o *flu*, debido a la capacidad para agregarse (*fluffing*) que presentaban las células que la expresaban³¹⁵, llegándose a correlacionar su presencia con la patogenicidad de cepas de *E. coli*³¹⁶. La estructura de la proteína presenta las características típicas de este tipo de adhesinas, donde cada extremo tiene funciones bien diferenciadas. En el extremo N-terminal se encuentra un péptido señal y un dominio que se secreta, llamado subunidad α , mientras que en el extremo C-terminal la proteína adquiere forma de barril y es capaz de anclarse a la membrana externa, actuando como canal para la subunidad α (Fig. 1.9)^{303,317}. Además, se ha visto que para alcanzar la membrana externa depende de un sistema de translocación y ensamblaje (TAM), recientemente descrito³¹⁸.

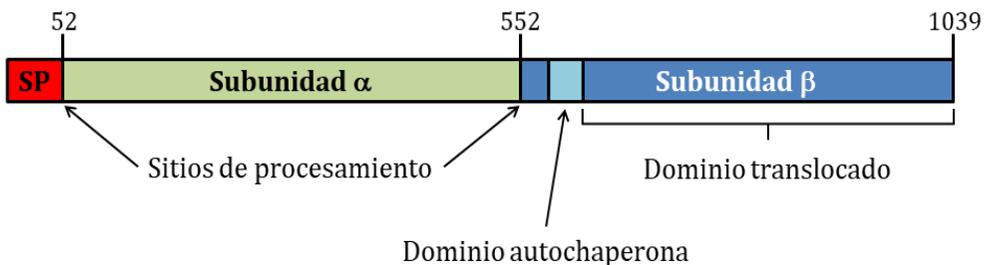


Fig. 1.9. Estructura de la proteína Agn43. En rojo se señala el péptido señal (SP) en el extremo N-terminal, en verde la subunidad α secretada y en azul oscuro la subunidad β . En azul claro se resalta el dominio autochaperona. Se indica el sitio de procesamiento de la subunidad α . Adaptado de van der Woude y Henderson³⁰³.

En *E. coli*, el gen que codifica para esta proteína se encuentra dentro del elemento génico CP4-44, el cual deriva del profago CP4. Su localización en el cromosoma varía según la cepa, lo cual sugiere que dicho elemento génico ha sido adquirido por procesos HGT³⁰³. Un aspecto resaltable es que se ha descrito la existencia de múltiples copias del gen *agn43* dentro de una misma cepa, variando de 0 a 3³⁰³, siendo este último el número de copias del gen *agn43* en la cepa EAEC 042¹⁹⁷. Mediante análisis filogenéticos, las copias de este gen se han dividido en dos subfamilias, atendiendo a pequeñas variaciones en su tamaño^{303,319,320}. En el genoma de la cepa 042 se pueden encontrar tres copias del gen *agn43* (EC042_2242, 4511 y 4803)¹⁹⁷, donde dos de estas copias (EC042_2242 y 4511) tienen un 87 % de identidad entre ellas mientras que con la copia EC042_4803 tienen un porcentaje menor (72 y 65 % con las copias 2242 y 4511, respectivamente). Atendiendo al estudio filogenético, las copias EC042_2242 y 4511 se encuentran en la subfamilia I, mientras que la copia EC042_4805 pertenece a la subfamilia II, lo que demuestra que en una misma cepa puede haber varias copias de este gen e, incluso, ser de subfamilias distintas^{303,316}. Esta variedad puede ser explicada por la diversidad que presenta el punto de corte entre los dos dominios del autotransportador, lo que da como resultado variaciones en el tamaño del dominio secretado (subunidad α)³⁰³. Por otra parte, en estudios recientes se ha llevado a cabo una clasificación diferente focalizada en la estructura de esta proteína, dividiendo las copias en cuatro clases diferentes según la combinación de sus dominios³²¹.

La presencia de múltiples copias de *agn43* se ha visto en varios aislados de *E. coli* y no está limitada a un único patotipo^{303,319}. Sin embargo, no hay una correlación absoluta entre la presencia o ausencia de uno de las copias de *agn43* y la virulencia de las correspondientes cepas de *E. coli*³⁰³. Estos estudios se han llevado a cabo principalmente en cepas UPEC^{305,322}, pero utilizando diferentes cepas modelo, como es el caso de CFT073 y UTI89. Estas dos cepas UPEC codifican copias diferentes de la proteína Agn43, y se ha llegado a

conclusiones contradictorias respecto a la función de la misma en la formación de biofilm en las infecciones urinarias³⁰⁵.

La regulación de la expresión de la proteína Agn43 tiene lugar mediante un mecanismo de variación de fase³¹⁵. La variación de fase es un mecanismo que dota de variabilidad genética a una población homogénea, lo que puede suponer una gran ventaja bajo ciertas condiciones adaptativas³²³. En el caso de Agn43, la variación de fase es epigenética, debido a que es causada por un cambio en el estado de metilación del ADN. Estas diferencias en el estado de metilación tienen lugar como consecuencia de la competición entre la metilasa Dam y el regulador OxyR^{324,325}. En la región promotora del gen *agn43* en *E. coli* se encuentran tres sitios de metilación de Dam (GATC), solapados con los sitios de unión del regulador OxyR. Cuando el ADN no está metilado, OxyR es capaz de unirse y bloquea la transcripción del gen, lo que da como resultado la fase *OFF*. No obstante, cuando la metilasa Dam metila los tres sitios GATC, la transcripción del gen tiene lugar y OxyR es incapaz de unirse, encontrándose en este caso en la fase *ON* (Fig. 1.10)^{324,326,327}. Por lo tanto, para generalizar conclusiones respecto a la proteína Agn43, es preciso tener en cuenta tanto la copia estudiada como su estado *ON/OFF*, el nivel de expresión y el contexto genético donde se encuentra^{303,304}.

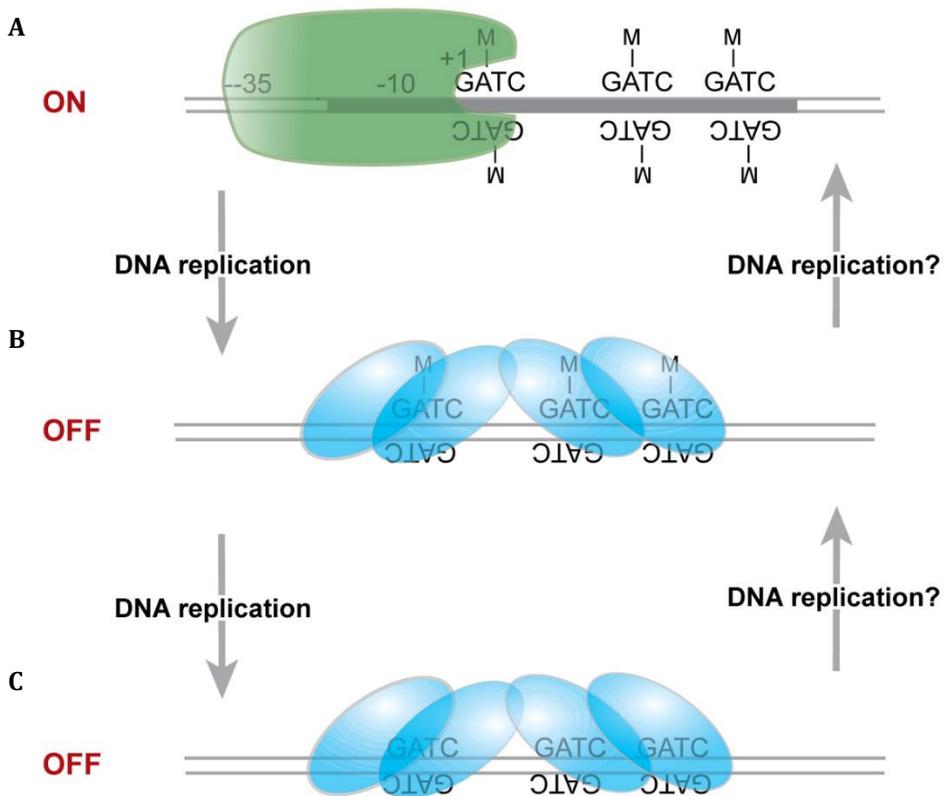


Fig. 1.10. Esquema de la regulación por variación de fase del gen *agn43* por la acción de OxyR y Dam. Se muestran las cadenas de ADN con los sitios diana GATC de la metilasa Dam. La metilación se representa con la letra M, la ARN polimerasa en verde y los dímeros de OxyR en azul. Se indican los sitios -10 y -35 del promotor y el inicio de transcripción como +1. A) Fase *ON* caracterizada por la metilación de las secuencias GATC de ambas cadenas que impiden la unión de OxyR. B) Intermediario de la fase *OFF* con los sitios GATC de una sola cadena metilados y OxyR unido. C) Fase *OFF* con todos los sitios GATC desmetilados y OxyR unido con gran afinidad. La variación de fase está relacionada con la replicación del ADN, proceso en el cual quedan los sitios GATC disponibles tanto para Dam como para OxyR. Adaptado de Van der Woude y Henderson³⁰³.

Recientemente, se ha descrito que el gen *agn43* se cotranscribe junto a dos genes más situados aguas abajo: *yeeR* y *aec69*³²⁸. Mientras que se desconoce totalmente el papel que desempeña la proteína YeeR, se ha puesto de

manifiesto que el producto del gen *aec69* es una proteína de bajo peso molecular (13 kDa) identificada por primera vez mediante el uso de vacunología reversa, resultando ser un buen candidato para su uso como antígeno en nuevas vacunas^{329,330}. De la misma forma que ocurre con Agn43, la expresión de la proteína Aec69 está regulada por el regulador OxyR, ya que la ausencia de este regulador provoca una sobreexpresión de Aec69³²⁸. Se ha podido comprobar su presencia en el medio extracelular, por lo que se trata de una proteína secretada. Sin embargo, el mecanismo mediante el cual es secretada es desconocido, porque no depende ni de Agn43 ni de YeeR, ni tampoco del sistema de translocación TAM que utiliza Agn43³²⁸.

En relación con su estructura, mediante cristalografía se ha podido determinar que la proteína Aec69 contiene un dominio semejante al de la fibronectina III³²⁸. La estructura de la proteína Aec69 consiste en siete cadenas antiparalelas unidas formando dos láminas, estabilizadas por interacciones hidrofóbicas. Esta conformación le confiere la capacidad de formar dímeros (Fig. 1.11)³²⁸. Pese a que su secuencia no guarda una gran similitud con ninguna proteína de función conocida, su estructura sí que es característica de proteínas semejantes a inmunoglobulinas, como es el caso de los dominios extracelulares de los receptores de interleuquinas humanos IL-2R, IL-4R y IL-10R³²⁸. Además, no solo su estructura es parecida a la de estos receptores, sino que también es capaz de unirse a sus ligandos (IL-2, IL-4 y IL-10, respectivamente). Estas interleuquinas están relacionadas con el crecimiento, diferenciación y proliferación de células inmunes³³¹. Debido a esto, se ha denominado a esta proteína IrmA (*interleukin receptor mimic protein A*)³²⁸.

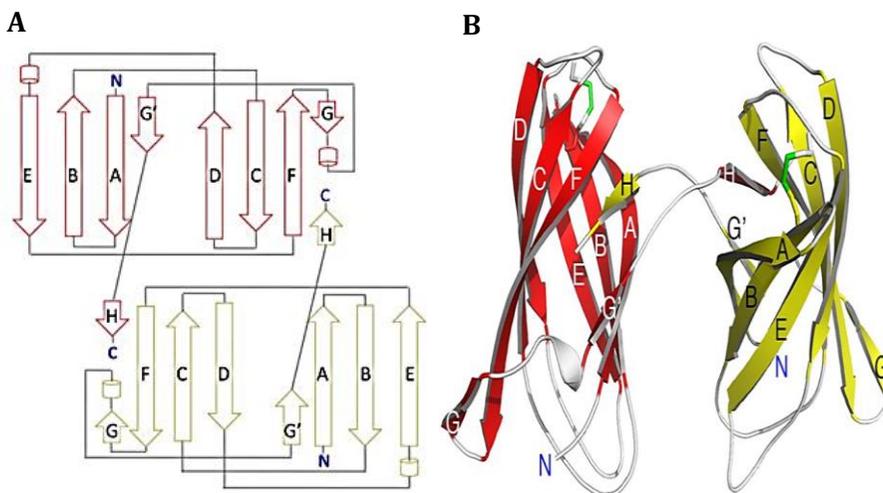


Fig. 1.11. Estructura obtenida por cristalografía de la proteína IrmA. A) Diagrama de la unión de dos monómeros de IrmA. B) Representación en tres dimensiones de un dímero de IrmA. Los monómeros se representan en rojo y en amarillo. Las láminas β están representadas como flechas y etiquetadas de la A a la G. Adaptado de Moriel y colaboradores³²⁸.

El estudio realizado por Moriel y colaboradores sobre la proteína Aec69 fue llevado a cabo en las cepas UPEC EC958 y CFT073. En la cepa CFT073 la región *agn43-yeeR-aec69* está presente en dos copias³²⁸, no obstante, es importante señalar que en una de las regiones ha tenido lugar un evento de inserción, como se puede ver en la figura 1.12. Esta inserción provoca la disrupción de la pauta de lectura del operón, de manera que los resultados mostrados en el estudio sobre la proteína Aec69 en la cepa CFT073 corresponden únicamente a una de las copias de este gen (*c1275*).

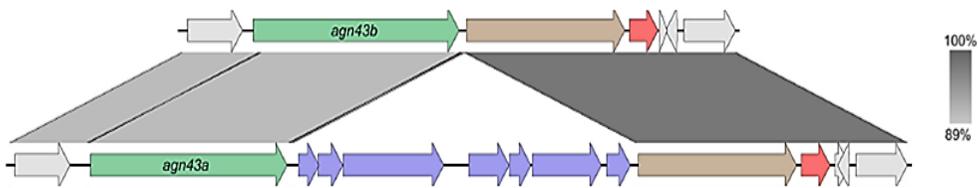


Fig. 1.12. Alineamiento de las dos regiones genómicas *agn43-yeeR-aec69* de la cepa UPEC CFT073. En verde se muestra el gen *agn43*, en marrón el gen *yeeR* y en rojo el gen *aec69*. Los genes insertados en la región causantes de la disrupción del operón se han señalado en azul. Adaptado de Moriel y colaboradores³²⁸.

En el caso de la cepa EAEC 042, la región *agn43-yeeR-aec69* también se encuentra duplicada. A diferencia de la cepa CFT073, ambas regiones son idénticas en esta cepa, como se aprecia en la figura 1.13. Es posible que en la cepa 042 las dos copias sean funcionales, debido a que no se ha producido la disrupción de la pauta de lectura en ninguna de ellas.

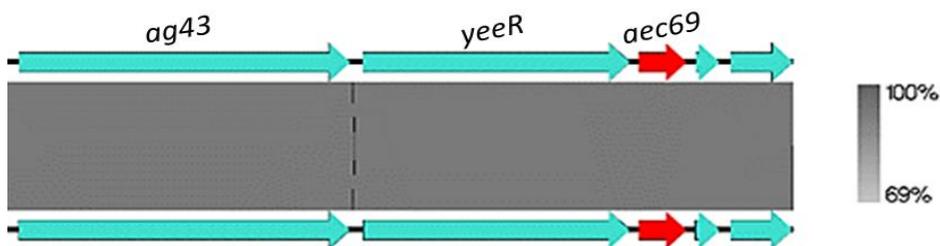


Fig. 1.13. Alineamiento de las dos copias del operón *agn43-yeeR-aec69* en la cepa 042. En rojo se muestra el gen *aec69*. Figura realizada con Easyfig³³².

La expresión de las dos copias génicas puede ser diferente según las condiciones del medio o ser modulada por factores distintos. Por lo tanto, el estudio de ambas copias y, sobre todo, el estudio de la regulación de la expresión génica de las mismas, puede resultar interesante para conocer la función que desempeñan y su posible papel en la virulencia de la cepa 042.

2. Objetivos

The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that proper record-keeping is essential for ensuring transparency and accountability in financial reporting. This section also outlines the various methods and tools used to collect and analyze data, highlighting the need for consistency and reliability in the information gathered.

The second part of the document focuses on the analysis and interpretation of the collected data. It provides a detailed overview of the statistical techniques and models employed to identify trends, patterns, and correlations within the data set. This section also addresses the challenges associated with data analysis, such as handling missing values and outliers, and offers practical solutions to overcome these obstacles.

The final part of the document presents the conclusions and recommendations derived from the analysis. It summarizes the key findings and discusses their implications for the organization or industry. This section also provides actionable insights and suggestions for future research or improvements, based on the results of the study.

En el momento del inicio de esta tesis, en el grupo de investigación se había puesto de manifiesto la presencia de múltiples copias del gen *hha* en la cepa 042 de *E. coli* y también en otras cepas patógenas de *E. coli*²⁰⁵.

El desarrollo de esta tesis parte del intento de contestar a una pregunta sencilla. Teniendo en cuenta que la proteína Hha juega un papel regulador en *E. coli*, la presencia de los nuevos alelos *hha2* y *hha3* en algunas cepas ¿podía asociarse a la presencia en las mismas de genes específicos regulados por los productos de dichos alelos?

Para responder a esta pregunta se utilizaron estrategias bioinformáticas que han llevado a correlacionar la presencia de los alelos *hha2* y *hha3* con la presencia de múltiples copias del clúster de genes *yeeR aec69*. Ello puso de manifiesto que la cepa 042 presenta un número de duplicaciones importantes en su genoma. Como hasta el momento se desconocía en detalle el impacto de las duplicaciones en el genoma de *E. coli*, los estudios continuaron utilizando herramientas bioinformáticas para establecer un catálogo preciso de duplicaciones en el genoma de los diferentes patotipos de *E. coli*. Ello, a su vez, puso de manifiesto la existencia de un conjunto de genes, la mayoría de función desconocida, presentes en múltiples copias que son comunes a prácticamente todos los patotipos, cosa insólita teniendo en cuenta la amplia variabilidad genómica existente entre los mismos.

El capítulo siguiente de la tesis tuvo como objetivo analizar la regulación de las dos copias diferentes del gen *aec69* en la cepa 042, a fin de tratar de dar sentido biológico a la duplicación de dicho gen.

Finalmente, partiendo de la identificación anteriormente mencionada de un grupo de genes amplificado cuatro veces en la cepa 042, y que está presente en prácticamente todas las cepas virulentas de *E. coli*, decidimos investigar si dichos genes juegan un papel en la virulencia de la cepa 042.

3. Materiales y métodos

3.1 Cepas bacterianas y plásmidos

Las cepas de *E. coli* utilizadas en este trabajo y su genotipo se detallan en la Tabla 3.1.

Tabla 3.1. Cepas utilizadas en este trabajo y sus características.

Cepa	Genotipo	Referencia u origen
DH5 α	<i>fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	Taylor <i>et al.</i> , 1993 ³³³
BL21 (DE3)	<i>hsdS gal (λclts857 ind1 Sam7 nin5 lac-UV5-T7)</i>	Studier y Moffatt, 1986 ³³⁴
EAEC 042	<i>E. coli</i> 042, aislamiento clínico: Cm ^r Sm ^r Tc ^r	Prof. I. Henderson
EAEC 042 pAA-	Cepa 042 con el plásmido pAA2 curado	Stock de laboratorio
EAEC 042 <i>irmA2244</i>	<i>E. coli</i> 042 mutante sencillo para el gen <i>irmA2244</i>	Este trabajo
EAEC 042 2244::3X	<i>E. coli</i> 042 con el epítipo 3XFLAG insertado en el extremo C-terminal de la proteína IrmA2244	Este trabajo
EAEC 042 2244::3X <i>hha</i>	<i>E. coli</i> 042 2244::3X mutante sencillo para el gen <i>hha</i>	Este trabajo
EAEC 042 2244::3X <i>hhanull</i>	<i>E. coli</i> 042 2244::3X mutante doble para los genes <i>hha</i> y <i>hha2</i>	Este trabajo
EAEC 042 2244::3X <i>hns</i>	<i>E. coli</i> 042 2244::3X mutante sencillo para el <i>hns</i>	Este trabajo
EAEC 042 2244::3X <i>oxyR</i>	<i>E. coli</i> 042 2244::3X mutante sencillo para el gen <i>oxyR</i>	Este trabajo
EAEC 042 2244::3X <i>dam</i>	<i>E. coli</i> 042 2244::3X mutante sencillo para el gen <i>dam</i>	Este trabajo
EAEC 042 2244::3X <i>irmA4509</i>	<i>E. coli</i> 042 2244::3X mutante sencillo para el gen <i>irmA4509</i>	Este trabajo
EAEC 042 <i>irmA4509</i>	<i>E. coli</i> 042 mutante sencillo para el gen <i>irmA4509</i>	Este trabajo

Cepa	Genotipo	Referencia u origen
EAEC 042 4509::3X	<i>E. coli</i> 042 con el epítipo 3XFLAG insertado en el extremo C-terminal de la proteína IrmA4509	Este trabajo
EAEC 042 4509::3X <i>hha</i>	<i>E. coli</i> 042 4509::3X mutante sencillo para el gen <i>hha</i>	Este trabajo
EAEC 042 4509::3X <i>hhanull</i>	<i>E. coli</i> 042 4509::3X mutante doble para los genes <i>hha</i> y <i>hha2</i>	Este trabajo
EAEC 042 4509::3X <i>hns</i>	<i>E. coli</i> 042 4509::3X mutante sencillo para el gen <i>hns</i>	Este trabajo
EAEC 042 4509::3X <i>oxyR</i>	<i>E. coli</i> 042 4509::3X mutante sencillo para el gen <i>oxyR</i>	Este trabajo
EAEC 042 4509::3X <i>dam</i>	<i>E. coli</i> 042 4509::3X mutante sencillo para el gen <i>dam</i>	Este trabajo
EAEC 042 4509::3X <i>irmA2244</i>	<i>E. coli</i> 042 4509::3X mutante sencillo para el gen <i>irmA2244</i>	Este trabajo
EAEC 042 4509::3X <i>prom4509</i>	<i>E. coli</i> 042 4509::3X mutante para la región promotora del gen <i>irmA4509</i>	Este trabajo
EAEC 042 4509::3X <i>prom4511</i>	<i>E. coli</i> 042 4509::3X mutante para la región promotora del gen <i>agn43_4511</i>	Este trabajo
EAEC 042 <i>irmAnull</i>	<i>E. coli</i> 042 mutante doble para los genes <i>irmA2244</i> y <i>irmA4509</i>	Este trabajo
EAEC 042 ΔLC	<i>E. coli</i> 042 mutante doble para los genes <i>lacZ</i> y <i>cat</i>	Stock de laboratorio
EAEC 042 ΔLC 2244:: <i>lacZ</i>	<i>E. coli</i> 042 ΔLC con la fusión transcripcional cromosómica del gen reportero <i>lacZ</i> bajo el control del promotor del gen <i>irmA2244</i>	Este trabajo
EAEC 042 ΔLC 2244:: <i>lacZ hha</i>	<i>E. coli</i> 042 ΔLC 2244:: <i>lacZ</i> mutante simple para el gen <i>hha</i>	Este trabajo
EAEC 042 ΔLC 2244:: <i>lacZ oxyR</i>	<i>E. coli</i> 042 ΔLC 2244:: <i>lacZ</i> mutante simple para el gen <i>oxyR</i>	Este trabajo
EAEC 042 ΔLC 2244:: <i>lacZ dam</i>	<i>E. coli</i> 042 ΔLC 2244:: <i>lacZ</i> mutante simple para el gen <i>dam</i>	Este trabajo
EAEC 042 ΔLC 4509:: <i>lacZ</i>	<i>E. coli</i> 042 ΔLC con la fusión transcripcional cromosómica del gen reportero <i>lacZ</i> bajo el control del promotor del gen <i>irmA4509</i>	Este trabajo

Cepa	Genotipo	Referencia u origen
EAEC 042 ΔLC 4509:: <i>lacZ hha</i>	<i>E. coli</i> 042 ΔLC 4509:: <i>lacZ</i> mutante simple para el gen <i>hha</i>	Este trabajo
EAEC 042 ΔLC 4509:: <i>lacZ oxyR</i>	<i>E. coli</i> 042 ΔLC 4509:: <i>lacZ</i> mutante simple para el gen <i>oxyR</i>	Este trabajo
EAEC 042 <i>hha</i>	<i>E. coli</i> 042 mutante simple para el gen <i>hha</i>	Este trabajo
EAEC 042 <i>hhanull</i>	<i>E. coli</i> 042 mutante doble para los genes <i>hha</i> y <i>hha2</i>	Este trabajo
EAEC 042 <i>hns</i>	<i>E. coli</i> 042 mutante simple para el gen <i>hns</i>	Este trabajo
EAEC 042 <i>oxyR</i>	<i>E. coli</i> 042 mutante simple para el gen <i>oxyR</i>	Este trabajo
EAEC 042 <i>irmA2244 oxyR</i>	<i>E. coli</i> 042 mutante doble para los genes <i>irmA2244</i> y <i>oxyR</i>	Este trabajo
EAEC 042 <i>irmA4509 oxyR</i>	<i>E. coli</i> 042 mutante doble para los genes <i>irmA4509</i> y <i>oxyR</i>	Este trabajo
EAEC 042 1 Δ	<i>E. coli</i> 042 mutante para la región génica EC042_2245-49	Este trabajo
EAEC 042 2 Δ	<i>E. coli</i> 042 mutante para las regiones génicas EC042_2245-49 y EC042_3221-26	Este trabajo
EAEC 042 3 Δ	<i>E. coli</i> 042 mutante para las regiones génicas EC042_2245-49, EC042_3221-26 y EC042_4507-02	Este trabajo
EAEC 042 4 Δ	<i>E. coli</i> 042 mutante para las regiones génicas EC042_2245-49, EC042_3221-26, EC042_4507-02 y EC042_4805-10	Este trabajo

Respecto a los plásmidos utilizados a lo largo de este trabajo, en la Tabla 3.2 se encuentran detallados.

Tabla 3.2. Plásmidos utilizados.

Plásmido	Características	Referencia u origen
pKD3	<i>bla</i> FRT <i>cat</i> FRT PS1 PS2 oriR6K Cm ^r Cb ^r	Datsenko y Wanner, 2000 ³³⁵
pKD4	<i>bla</i> FRT <i>ahp</i> FRT PS1 PS2 oriR6K Km ^r Cb ^r	Datsenko y Wanner, 2000 ³³⁵
pKD46	<i>bla</i> P _{BAD} <i>gam</i> <i>bet</i> <i>exo</i> pSC101 oriTS Cb ^r	Datsenko y Wanner, 2000 ³³⁵
pCP20	<i>bla</i> <i>cat</i> <i>cl857</i> <i>IP_Rflp</i> pSC101 oriTS Cb ^r	Cherepanov y Wackernagel, 1995 ³³⁶
pSUB11	Vector utilizado como molde de ADN para obtener el epítipo 3XFLAG	Uzzau <i>et al.</i> , 2001 ³³⁷
pLATE31	P _{T7} <i>lacO</i> <i>rep</i> (pMB1) <i>bla</i> (Cb ^r) 6xHis C-terminal	aLICator LIC Cloning & Expression System (<i>ThermoScientific</i>)
pLATE31-IrmA	Gen <i>irmA</i> clonado con el epítipo 6Histidinas en el extremo C-terminal bajo el promotor de la ARN polimerasa T7 en el plásmido pLATE31	Este trabajo
pKG136	<i>ahp</i> <i>km</i> FRT <i>lacZY+</i> t _{his} oriR6K	Ellermeier <i>et al.</i> , 2002 ³³⁸
pUJ8	<i>trp'</i> - <i>lacZ</i> vector sin promotor, fusiones <i>lacZ</i> tipo I	de Lorenzo <i>et al.</i> , 1990 ³³⁹
pUJ8-prom2242	Región promotora del gen <i>agn43_2242</i> clonada en el plásmido pUJ8	Este trabajo
pUJ8-prom4511	Región promotora del gen <i>agn43_4511</i> clonada en el plásmido pUJ8	Este trabajo
pUJ8-prom2244	Región promotora del gen <i>irmA2244</i> clonada en el plásmido pUJ8	Este trabajo
pUJ8-prom4509	Región promotora del gen <i>irmA4509</i> clonada en el plásmido pUJ8	Este trabajo
pUJ8-mut*prom4509	Región promotora del gen <i>irmA4509</i> con sustituciones puntuales clonada en el plásmido pUJ8	Este trabajo

3.2 Oligonucleótidos

Los oligonucleótidos utilizados a lo largo de este trabajo se encuentran recogidos en la siguiente tabla.

Tabla 3.3. Descripción de los oligonucleótidos utilizados

Oligonucleótido	Secuencia 5' → 3'	Uso
<i>irmA</i> pLATE31CT fw	AGAAGGAGATATAACTATGATTACAC CTGTTCAAAAACCTGCATGATTACC	Clonaje del gen <i>irmA</i> en el plásmido pLATE31
<i>irmA</i> pLATE31CT rv	GTGGTGGTGATGGTGGTGGCCGTTA ACGTTTTTTCCGGAAACGGTAAT	Clonaje del gen <i>irmA</i> en el plásmido pLATE31
LIC forward sequencing primer	TAATACGACTCACTATAGGG	Secuenciación del plásmido pLATE31
LIC reverse sequencing primer	GAGCGGATAACAATTTACACAGG	Secuenciación del plásmido pLATE31
2244 P1	GTGTCATCGTCATGACTCAGAGAGG TAAATACCATGATTACCGTGTAGGC TGGAGCTGCTTC	Mutagénesis
4509 P1	GTGTCATCGTCATGATTACAGAGAGG TAAATACCATGATTACCGTGTAGGC TGGAGCTGCTTC	Mutagénesis
<i>irmA</i> P2	ACAGAGGCTGAAGCATGAAACTGAC TTCAGGGATCAGTTAACCATATGAA TATCCTCCTTAGT	Mutagénesis
2244 P1up	CCGGCGCAACGGATCTTCAAC	Comprobación mutagénesis
2244 P2down	GTTTTTCAGGAACGAAAGCTG	Comprobación mutagénesis
4509P1up	GACGGGGGAAAACATTTTCAGT	Comprobación mutagénesis
4509P2down	GTCATCATCAGAAATGGAACG	Comprobación mutagénesis
IrmA:3X P1	CCATCCACTGCTAATTACCGTTTCC GGAAAAACGTTAACGACTACAAAG ACCATGACGG	Inserción epítipo 3XFLAG

Oligonucleótido	Secuencia 5' → 3'	Uso
IrmA:3X P2	ACAGAGGCTGAAGCATGAAACTGAC TTCAGGGATCAGTTACATATGAATA TCCTCCTTAG	Inserción epítipo 3XFLAG
<i>hns</i> P1	ATGAGCGAAGCACTTAAAAATTCTGA ACAACATCCGTA CTCTTGTGTAGGC TGGAGCTGCTTC	Mutagénesis
<i>hns</i> P2	TTATTGCTTGATCAGGAAATCGTCG AGGGATTTACCTTGCTCCATATGAA TATCCTCCTTAGT	Mutagénesis
<i>hns</i> P1up	CCACCCCAATATAAGTTTGAG	Comprobación mutagénesis
<i>hns</i> P2down	GGGATTTTAAGCAAGTGCAATC	Comprobación mutagénesis
<i>hha</i> P1	ATGTCCGAAAAACCTTTAACGAAAA CCGATTATTTAATGCGTGTGTAGGC TGGAGCTGTCTTC	Mutagénesis
<i>hha</i> P2	TTAGCGAATAAAATTTCCATACTGAG GAAGGGATCTTGTCTGACATATGAA TATCCTCCTTAGT	Mutagénesis
<i>hha</i> P1up	GTTAGAATTATTACAACCATGGG	Comprobación mutagénesis
<i>hha2</i> P1	ATCACGTTAATTGCAGCATAAAGTG ATGAGAGGCTAATGGAAGTGTAGCT GGAGCTGCTTC	Mutagénesis
<i>hha2</i> P2	AACGCCC GG CACAATACCGCTGTGA TTAACGAACGTT CAGGTCATATGAA TATCCTCCTTAGT	Mutagénesis
<i>hha2</i> P1up	GTCTGAGCGGTAAACGTATC	Comprobación mutagénesis
KT	CGGCCACAGTCGATGAATCC	Comprobación mutagénesis
<i>oxyR</i> P1	GCCATGAACTATCGTGGCGATGGAG GATGGATAATGAATATTGTGTAGGC TGGAGCTGCTTC	Mutagénesis
<i>oxyR</i> P2	GCGGAAGCCTATCGGGTAGCTGCGC TAAATGGCTTAAACCGCCATATGAA TATCCTCCTTAGT	Mutagénesis
<i>oxyR</i> P1up	GATAGGGATAATCGTTCATTG	Comprobación mutagénesis
<i>oxyR</i> P2down	GCAATACTATTGAGTACTTCG	Comprobación mutagénesis

Oligonucleótido	Secuencia 5' → 3'	Uso
<i>dam</i> P1	CACAGCCGGAGAAGGTGTAATTAGT TAGTCAGCATGAAGAAAGTGTAGGC TGGAGCTGCTTC	Mutagénesis
<i>dam</i> P2	TCAAATACTGTTTCATCCGCTTCTC CTTGAGAATTATTTTTTCATATGAA TATCCTCCTTAGT	Mutagénesis
<i>dam</i> P1up	CCTGCAATGGTTTATCATCG	Comprobación mutagénesis
<i>dam</i> P2down	TAGTGGTTATCCATGACGTC	Comprobación mutagénesis
A3 adaptador	GAUAUGGCGGAAUUCUGUAGAAC GAACACUAGAAGAAA	5' RACE
B6	GCGCGAATTCCTGTAGA	5' RACE
<i>irmA</i> rv	GCGACATTTTCCACGCACTC	<i>Walking RT-PCR</i>
<i>yeeR</i> fw	GAAATATCAGTGCCCTGAAC	<i>Walking RT-PCR</i>
Int fw	TTGTCAGTGTCATCGTCATG	<i>Walking RT-PCR</i>
<i>irmA</i> fw	GCATGATTACCGCCTTCATT	<i>Walking RT-PCR</i>
prom2242 fw <i>EcoRI</i>	CGGAATTCACCTGTCGTGACTGATG CCCT	Clonaje del promotor del gen <i>agn43_2242</i> en el plásmido pUJ8
prom2242 rv <i>BamHI</i>	CGGGATCCCAGCTTTTCCCTTAGATT GAGG	Clonaje del promotor del gen <i>agn43_2242</i> en el plásmido pUJ8
prom4511 fw <i>EcoRI</i>	CGGAATTCACCTGTTGTGGATGATG TCCT	Clonaje del promotor del gen <i>agn43_4511</i> en el plásmido pUJ8
prom4511 rv <i>BamHI</i>	CGGGATCCTTCCTTAATCAGAGTGA GGGT	Clonaje del promotor del gen <i>agn43_4511</i> en el plásmido pUJ8
promirmA fw <i>EcoRI</i>	CGGAATTC AAGAAGCATTACCCCGG GAGT	Clonaje del promotor del gen <i>irmA</i> en el plásmido pUJ8
promirmA rv <i>BamHI</i>	CGGGATCCTCATGACGATGACACTG ACAA	Clonaje del promotor del gen <i>irmA</i> en el plásmido pUJ8
pUJ8 p1up	CTGTAAGCGGATGCCGGGAGCA	Secuenciación del plásmido pUJ8

Oligonucleótido	Secuencia 5' → 3'	Uso
lacZR	GATGACCTGCAAGGCGATTA	Secuenciación del plásmido pUJ8
promirmA P1	ATATGGTTAACCGGCTGGAAGAAGC ATTACCCCGGAGTTACGTGTAGGC TGGAGCTGCTTC	Mutagénesis
promirmA P2	TTACCTCTCTGAATCATGACGATGA CACTGACAAGTCAGGTGCATATGAA TATCCTCCTTAGT	Mutagénesis
promirmA P1up	CACTGGTGGTCAGTATCATC	Comprobación mutagénesis
prom4511 P1	GTTGCCCGCGCGGTATGGAAGTGG TCTTCTTCTGAACCTGTGTGTAGGC TGGAGCTGCTTC	Mutagénesis
prom4511 P2	CCTGTAGCAGGTATTTCAGATGTCGT TTCATCAGCATTTCCTTCATATGAA TATCCTCCTTAGT	Mutagénesis
prom4511 P1up	GTGGACCGGATATTTGACAC	Comprobación mutagénesis
prom4511 P2down	TGCCCGCCATGTTTAATGGT	Comprobación mutagénesis
Mut_prom_4509 fw	ACCGGCATAACGGCTATTCTGCGCC TTTTACGCTTTATCACCTGA	Mutagénesis dirigida
Mut_prom_4509 rv	TCAGGTGATAAAGCGTAAAAGGCGC AGAATAGCCGTTATGCCGGT	Mutagénesis dirigida
4507-02/2245-49/4805 P1	CAGTAAATCTTAACCACCGATAAGG AGCAAATTATGCGACTAGTGTAGGC TGGAGCTGCTTC	Mutagénesis
3221-26 P1	TTATACCGCTAATTCATTCATAAGG AGCAAAGTATGCAATTAGTGTAGGC TGGAGCTGCTTC	Mutagénesis
4507-02/3221-26 P2	TAATGCTGTCGGCTTCCAGCGTCAG GGCCAGTTTCATTGTTTCATATGAA TATCCTCCTTAGT	Mutagénesis
2245-49 P2	CCTGTACGTTAACGCTGTCGGCTTC CAGTGTATCATTTTCGCCATATGAA TATCCTCCTTAGT	Mutagénesis
4805-10 P2	CACTGTCGGCTTCCAGTGTCCAGGGC CAGTTTCATCATTTTCGCCATATGAA TATCCTCCTTAGT	Mutagénesis
2245-49/4507-02 P1up	CGCCAGTGACGGTCTTCTCAG	Comprobación mutagénesis

Oligonucleótido	Secuencia 5' → 3'	Uso
2245-49 P2down	TGGTGAGGGCTGGCATAAGT	Comprobación mutagénesis
4507-02 P2down	TTGTCTTCCGGCAACTGATG	Comprobación mutagénesis
3221-26 P1up	CAAACAGGCAATGGCTGCCA	Comprobación mutagénesis
3221-26 P2down	CGGCAAGGGTTGCAAACGGT	Comprobación mutagénesis
4805-10 P1up	GTTGCGCTGGTGATGCCGGA	Comprobación mutagénesis
4805-10 P2down	TATCTACGGCAGCAAGCCAG	Comprobación mutagénesis

3.3 Medios de cultivo y antibióticos

3.3.1 Medios de cultivo

Durante el trabajo se han utilizado diferentes medios de cultivo cuya composición se especifica a continuación:

- LB (*Lysogeny Broth*): se trata de un medio de cultivo líquido complejo para bacterias donde se ha crecido *E. coli* de forma rutinaria. Está compuesto por 10 g/L de peptona trípica de caseína, 5 g/L de extracto de levadura y 10 g/L de NaCl.
- LB agar: medio de cultivo sólido utilizado para el crecimiento bacteriano rutinario. Se obtiene mediante la adición de 15 g/L de agar para bacteriología en medio LB.
- LB agar X-Gal: medio de cultivo sólido usado para seleccionar colonias por su fenotipo blanco/azul. Su composición es igual que la del LB agar,

suplementada con 40 $\mu\text{g/mL}$ de X-Gal disuelto en N,N-dimetilformamida, tras la esterilización del medio.

- SOB (*Super Optimal Broth*): medio rico que se ha utilizado en los experimentos de inactivación de genes cromosómicos por fragmentos de PCR, explicados más adelante. Su composición es 20 g/L peptona trípica de caseína, 5 g/L extracto de levadura, 0,58 g/L NaCl, 0,19 g/L KCl y $\text{MgCl}_2/\text{MgSO}_4$ 20 mM (partiendo de una solución de 1 M de ambas sales)³⁴⁰.
- SOC (*Super Optimal Broth with Catabolite repression*): este medio se ha utilizado para la recuperación de células tras someterse a algún proceso de transferencia genética. Se obtiene suplementando el medio SOB con 20 mM de glucosa esterilizada por filtración³⁴⁰.
- M9: medio mínimo líquido de composición definida utilizado para el crecimiento bacteriano. Está compuesto por sales M9 1X, glucosa 0,4 %, MgSO_4 20 mM y 0,1 mM de CaCl_2 .
- Sales M9 10X: solución de sales utilizada para la preparación del medio M9. Su composición es 70 g/L de $\text{Na}_2\text{HPO}_4 \cdot 7 \text{H}_2\text{O}$, 30 g/L de KH_2PO_4 , 5 g/L de NaCl, 10 g/L de NH_4Cl .
- DMEM (*Dulbecco's Minimum Essential Medium*): medio de cultivo comercial (Gibco) utilizado en los ensayos de formación de biofilm en los distintos mutantes de la cepa 042 de *E. coli*. El medio fue suplementado con 0,45 % (p/v) de glucosa como fuente de carbono.

3.3.2 Antibióticos

En algunos casos ha sido necesario suplementar los medios sólidos y líquidos con antibióticos. Estos antibióticos se conservan a -20 °C y han sido preparados siguiendo las siguientes indicaciones:

- Carbenicilina (Cb): se ha preparado un stock a una concentración de 100 mg/mL en agua destilada y se ha esterilizado por filtración, mediante filtros de 0,22 µm de diámetro de poro (*Millipore*). Se ha usado a una concentración final de 50 o 100 µg/mL.
- Kanamicina (Km): el stock se ha preparado a una concentración de 50 mg/mL en agua destilada y se ha esterilizado por filtración. Su concentración de uso ha sido de 50 µg/mL.
- Cloranfenicol (Cm): la solución madre se ha preparado a una concentración de 100 mg/mL en etanol absoluto. La concentración final a la que se ha utilizado es de 25 µg/mL.

3.4 Métodos microbiológicos

3.4.1 Esterilización

El material de vidrio y plástico utilizado, como los medios de cultivo preparados, han sido esterilizados por calor húmedo en el autoclave. Se ha realizado un ciclo de 20 minutos a 121 °C a una atmósfera de presión. Sin embargo, en el caso de soluciones termolábiles, la esterilización se ha realizado mediante filtración, utilizando filtros estériles de 0,22 µm de diámetro de poro.

En las técnicas realizadas con ARN, el material se ha adquirido directamente de la casa comercial libre de nucleasas.

3.4.2 Mantenimiento y crecimiento de los microorganismos

Las cepas obtenidas y utilizadas en todo este trabajo se encuentran congeladas a $-80\text{ }^{\circ}\text{C}$, en una solución de LB con glicerol al 20 % (v/v) como crioprotector. La recuperación de las células se ha llevado a cabo mediante la siembra en placas de LB agar. Las placas se han suplementado con los antibióticos necesarios en cada caso y se han dejado crecer durante 16 horas a 30 o $37\text{ }^{\circ}\text{C}$, según la cepa descongelada.

Los experimentos realizados en este trabajo se han realizado a partir de un inóculo líquido. El inóculo se ha obtenido resuspendiendo varias colonias en medio LB líquido y creciéndolo a la temperatura indicada en agitación, durante 16 horas. Tras el crecimiento de estos cultivos, se ha inoculado en medio líquido fresco una dilución 1:100 y se ha realizado el experimento en las condiciones de temperatura y agitación correspondientes.

En los experimentos donde se ha realizado un seguimiento del crecimiento celular, se ha medido la densidad óptica de los cultivos a 600 nm ($\text{DO}_{600\text{nm}}$). Esta medida se ha realizado cada 30 minutos, haciendo uso del espectrofotómetro Genesys 10 UV-VIS de *ThermoScientific*.

En el caso de los ensayos en condiciones de anaerobiosis, se han crecido cultivos en medio LB en estático durante toda la noche, donde se ha evitado el transporte gaseoso mediante la adición de vaselina. Por otra parte, en el caso de realizar un curado de un plásmido termosensible, las células se han incubado a $42\text{ }^{\circ}\text{C}$ en cultivos sucesivos para inducir la completa pérdida del vector.

3.5 Técnicas experimentales con ADN

3.5.1 Aislamiento de ADN plasmídico

Para llevar a cabo la extracción de ADN plasmídico, se ha utilizado el kit comercial *GeneJET Plasmid Miniprep Kit* de *ThermoScientific*. Se ha seguido el protocolo indicado por el fabricante, el cual se basa en la lisis alcalina.

3.5.2 Transformación bacteriana

La transformación bacteriana se ha utilizado para la incorporación de ADN externo a la célula de interés. Se ha llevado a cabo mediante dos métodos: transformación de células competentes obtenidas por tratamiento con cloruro de calcio (CaCl_2)³⁴¹ y electroporación³⁴². La transformación se ha realizado para la incorporación de ADN plasmídico mientras que, en aquellos casos donde se requería una eficiencia mayor del proceso, se ha electroporado.

3.5.2.1 Transformación de células competentes obtenidas por tratamiento con CaCl_2

En este método se consigue la competencia de las células mediante tratamiento con CaCl_2 , el cual desestabiliza la membrana celular. En primer lugar, se ha crecido un cultivo en LB de la cepa que se quiere transformar a 37 °C en agitación hasta la fase exponencial temprana ($\text{DO}_{600\text{nm}}$ 0,2). A continuación, el cultivo se ha centrifugado a 4.000 revoluciones por minuto (rpm) durante 15 minutos a una temperatura de 4 °C para sedimentar las células, retirando el medio de cultivo. Las células se han lavado con 20 mL de una solución de 50 mM CaCl_2 autoclavada y fría, volviendo a centrifugar en las mismas condiciones. Este lavado se ha repetido 3 veces en total. Finalmente, las

células se han resuspendido con la misma solución de 50 mM CaCl₂ en el volumen adecuado para su uso (150 µL por transformación), manteniéndolas en hielo durante 30 minutos como mínimo antes de su uso.

La transformación bacteriana se ha hecho usando 150 µL de células competentes y un volumen de 1-10 µL de ADN plasmídico (10-50 ng). La mezcla se ha dejado 30 minutos adicionales en hielo. Seguidamente, se han sometido las células a un choque térmico para conseguir la formación de poros en la membrana, a través de los cuales se introducirá el ADN. El choque térmico ha consistido en introducir las células en un baño a 42 °C durante 2 minutos y dejarlas en hielo de nuevo durante otros 2 minutos. Las células se han recuperado añadiendo LB o SOC hasta un volumen de 1 mL. Luego, para permitir la expresión de los marcadores de resistencia presentes en el ADN plasmídico transformado, se ha incubado durante 1-2 horas en agitación a 30 o 37 °C, según la termosensibilidad del plásmido. Por último, se han sembrado en placas de LB suplementadas con el antibiótico correspondiente para seleccionar las células transformantes.

3.5.2.2 Electroporación

La permeabilización de las membranas se obtiene en este método gracias a la aplicación de un voltaje determinado. Se ha crecido un cultivo hasta DO_{600nm} 0,6 y se ha centrifugado durante 15 minutos a 4.000 rpm en frío. Se ha retirado el sobrenadante y el sedimento se ha resuspendido en 20 mL de glicerol al 10 % frío, volviéndolo a centrifugar en las mismas condiciones. Este lavado se ha repetido 3 veces más. Estos lavados eliminan las sales presentes en el medio de cultivo, reduciendo así la fuerza iónica del medio. Por último, se han resuspendido las células en el volumen adecuado de glicerol 10 % frío (60 µL por electroporación), manteniéndolas en hielo hasta su utilización.

La transformación bacteriana se ha realizado mediante la adición de 2-5 μL de ADN purificado (alrededor de 1 μg) a 60 μL de células competentes. La mezcla se ha introducido en una cubeta de electroporación con una amplitud de hendidura de 1 mm (*1 mm Electroporation cuvettes Molecular BioProducts*). A continuación, la cubeta ha sido sometida a una descarga eléctrica de 1.250 V con pulsos de 5,4 a 5,8 ms en el electroporador (*Electroporator 2510 Eppendorf*). Las células se han recuperado con medio SOC hasta un volumen de 1 mL. Finalmente, se han incubado durante 1-2 horas a 30 o 37 $^{\circ}\text{C}$ y se han seleccionado las células transformadas, sembrándolas en placas de LB suplementadas con el antibiótico adecuado.

3.5.3 Uso de enzimas

Las enzimas de restricción utilizadas pertenecen a la casa comercial *ThermoScientific*. Se ha hecho uso del tampón concentrado 10 veces suministrado por la casa comercial, siguiendo las indicaciones del fabricante. En el caso de llevar a cabo una digestión con dos enzimas en la misma reacción, se ha utilizado el software *DoubleDigest Calculator (ThermoScientific)*, determinado las condiciones óptimas de la misma.

Tras la digestión de plásmidos, se ha llevado a cabo una reacción de desfosforilación de sus extremos mediante la fosfatasa alcalina *FastAP*. Esta reacción evita su recircularización, optimizando su uso en reacciones de ligación. Las reacciones de ligación se han llevado a cabo mediante la ligasa de ADN T4. De forma general, se ha calculado la cantidad de ADN plasmídico y de inserto de interés necesarias para obtener una relación molar 1:3. El cálculo ha sido el siguiente:

$$\frac{x \text{ ng vector} \cdot y \text{ Kb inserto}}{z \text{ Kb vector}} \cdot \frac{3}{1}$$

3.5.4 Amplificación de fragmentos de ADN mediante la reacción en cadena de la polimerasa (PCR)

Esta reacción se ha llevado a cabo para la amplificación de fragmentos de ADN de interés. En función del uso final del producto de PCR, se ha utilizado una polimerasa diferente.

De forma rutinaria, el genotipado de cepas bacterianas o la comprobación de clonajes se han realizado con la mezcla comercial *DreamTaq Green PCR Master Mix* (*ThermoScientific*). Esta mezcla comercial está a una concentración 2X y, además de la polimerasa, contiene iones Mg^{+2} , nucleótidos desoxirribonucleicos (dNTPs) y el tampón necesario para un funcionamiento óptimo de la reacción. Las cantidades utilizadas para una reacción han sido *DreamTaq Master Mix* a una concentración 1X, 0,4 mM de cada oligonucleótido, 2 μ L de muestra y agua bidestilada hasta un volumen de 25 μ L.

Por otra parte, cuando ha sido necesaria una elevada fidelidad de la amplificación, se ha utilizado la enzima *Phusion Hot Start II DNA Polymerase* (*ThermoScientific*). Esta enzima posee la actividad polimerasa 5' -> 3' inhibida a temperatura ambiente, previniendo de esta manera la amplificación de productos inespecíficos. Por otra parte, también tiene actividad exonucleasa 3' -> 5', por lo que su tasa de error es mínima. En la reacción se ha utilizado tampón HF (incluido con la enzima *Phusion*) a una concentración 1X, 0,2 mM de una mezcla de dNTPs, 0,2 mM de cada oligonucleótido, 0,5 μ L de polimerasa (2U/ μ L), 2 μ L de muestra y agua bidestilada hasta un volumen de 50 μ L.

La determinación de la temperatura y tiempo de la etapa de hibridación ha dependido de la longitud y secuencia de los oligonucleótidos que se han usado. El tiempo de extensión ha dependido de la longitud del fragmento que se ha amplificado y de la velocidad de extensión de la polimerasa utilizada. En términos generales, el programa utilizado para la enzima *DreamTaq* ha sido 5

minutos (min) a 95 °C de desnaturalización inicial, seguidos de 30 segundos (seg) a 95 °C. A continuación, se ha realizado la hibridación durante 30 seg a X °C (según los oligonucleótidos empleados) y la extensión a 72 °C durante 1 min/Kb. Estas 3 etapas se han repetido 30 ciclos y, por último, se ha llevado a cabo una extensión final durante 10 min a 72 °C. En el caso de la enzima *Phusion*, la etapa de desnaturalización se ha realizado a una temperatura de 98 °C y el tiempo de extensión es de 15-30 seg/Kb.

Las muestras utilizadas como ADN molde en las reacciones de PCR han consistido en ADN plasmídico o una suspensión de una colonia bacteriana en 50 µl de agua bidestilada estéril.

3.5.5 Purificación de productos de PCR

Una vez realizada la amplificación del fragmento de interés, se ha llevado a cabo su purificación. La purificación permite eliminar tanto los posibles dímeros de oligonucleótidos que se hayan podido formar como los dNTPs y sales del tampón utilizados. Esta purificación se ha realizado siguiendo las instrucciones del kit comercial *GeneJET PCR Purification Kit (ThermoScientific)*.

Tras este paso, se ha cuantificado la muestra mediante el espectrofotómetro NanoDrop ND-1000 UV-Vis Spectrophotometer ND-1000 (*ThermoScientific*). En el caso de haber contaminaciones o una baja concentración de ADN, se ha precipitado el ADN mediante la adición de 3 volúmenes de etanol 100 % y 0,1 volúmenes de acetato sódico 3 M pH 5,2. La mezcla se ha incubado a -20 °C durante toda la noche. Al día siguiente, se ha centrifugado a 13.400 rpm durante 10 minutos y se ha descartado el sobrenadante. Se ha adicionado 1 volumen de etanol 70 % y se ha vuelto a centrifugar durante 10 minutos. Por último, el sedimento se ha secado al aire y resuspendido en 12 µL de agua bidestilada estéril.

En el caso de que haya sido necesaria una pureza muy elevada, tras haber realizado estos pasos de purificación y precipitación se han dializado los ácidos nucleicos. Este proceso se ha realizado con filtros de 0,025 μm , eliminando cualquier resto de sales que hubiera podido quedar remanente.

3.5.6 Electroforesis de ADN en geles de agarosa

El ADN plasmídico aislado y los productos de PCR obtenidos se han analizado mediante una electroforesis en gel de agarosa. Para ello, se ha hecho uso de una cubeta horizontal sumergida en tampón TBE 0.5X (45 mM Tris, 45 mM ácido bórico y 1 mM EDTA a un pH final de 8,3). La concentración de agarosa utilizada ha variado en función del tamaño del fragmento a separar electroforéticamente. Se han preparado los geles con una concentración de agarosa del 2 % (p/v), para aquellos fragmentos con un tamaño inferior a 1 kb y, para tamaños superiores, a una concentración del 0,8-1,2 % (p/v). La electroforesis se ha realizado a temperatura ambiente, aplicando un voltaje de 100 voltios (V) durante un tiempo de 30 min. Tras la electroforesis, se ha teñido el gel en una solución de bromuro de etidio a una concentración de 0,5 $\mu\text{g}/\text{mL}$ en TBE 0,5X durante 20 minutos. Finalmente, se han observado los geles con luz ultravioleta con el equipo Molecular Imager® Gel Doc™ XR+ (Bio-Rad), obteniendo y guardando imágenes de los geles.

En todos los casos, se ha cargado un marcador de pesos moleculares adecuado al tamaño del fragmento de interés:

- *Lambda/HindIII Marker* (ThermoScientific): recomendado para determinar el tamaño de fragmentos grandes ya que contiene bandas hasta un tamaño de 23 kb.

- *Generuler 1Kb DNA ladder (ThermoScientific)*: marcador que contiene un patrón de 14 bandas comprendido entre 250 pb y 10 kb, adecuado en el caso de fragmentos de PCR de tamaño medio-alto.
- *Generuler 100 bp DNA ladder (ThermoScientific)*: marcador utilizado para determinar el tamaño de fragmentos pequeños ya que contiene un patrón de 10 bandas comprendidas entre 100 pb y 1 kb.

3.5.7 Aislamientos de fragmentos de ADN a partir de geles de agarosa

El aislamiento de fragmentos de ADN a partir de geles de agarosa se ha llevado a cabo visualizándolo con una lámpara de luz UV. La luz UV empleada ha sido de onda larga y baja intensidad para no ocasionar daños en el ADN. Se ha cortado el trozo de gel que contenía el fragmento de interés y se ha purificado, siguiendo las instrucciones del fabricante del kit comercial *QIAEX II Gel Extraction Kit (Qiagen)*.

3.5.8 Secuenciación de fragmentos de ADN

En los casos en los que ha sido necesario confirmar la secuencia del plásmido o del fragmento de PCR, se ha realizado una secuenciación por el método de Sanger con la empresa *Eurofins Genomics*. Se han seguido las indicaciones de la empresa: 10 μL de reacción en un tubo eppendorf de 1,5 mL que contienen 5 μL de ADN molde a 100 o 12 $\text{ng}/\mu\text{L}$, según sea plásmido o producto de PCR, y 5 μL de oligonucleótido directo o reverso a una concentración de 5 μM .

3.6 Métodos de mutagénesis bacteriana

3.6.1 Inactivación de genes cromosómicos utilizando fragmentos de PCR

De manera general, las bacterias, como es el caso de *E. coli*, no se pueden transformar con ADN lineal, debido a que las exonucleasas citoplasmáticas lo degradan fácilmente. Como solución, Datsenko y Wanner, en el año 2.000, describieron un protocolo para conseguir el reemplazamiento de una secuencia cromosómica de interés por un gen de resistencia a un antibiótico, el cual se genera mediante una única reacción de PCR³³⁵.

Siguiendo el protocolo, el gen de resistencia al antibiótico se ha obtenido por PCR mediante una pareja de oligonucleótidos de aproximadamente 60 nucleótidos (llamados en este trabajo por el nombre del gen seguido de P1 o P2, (Tabla 3.3), usando como ADN molde los plásmidos pKD3 o pKD4. Los oligonucleótidos presentan homología tanto al casete de resistencia flanqueado por secuencias FRT presentes en un plásmido (kanamicina en el caso del plásmido pKD4 y cloranfenicol en el plásmido pKD3) como a las regiones adyacentes del gen cromosómico que se quiere reemplazar. Esta resistencia permite seleccionar las bacterias transformantes en placas de LB agar suplementadas con el antibiótico correspondiente. Una vez que se han seleccionado y comprobado los mutantes, puede ser eliminada la resistencia al antibiótico mediante la recombinación de las secuencias FRT que se han introducido. Las secuencias FRT se recombinan mediante la acción de la recombinasa FLP codificada en el plásmido auxiliar pCP20. Todo este proceso se encuentra resumido en la Fig. 3.1.

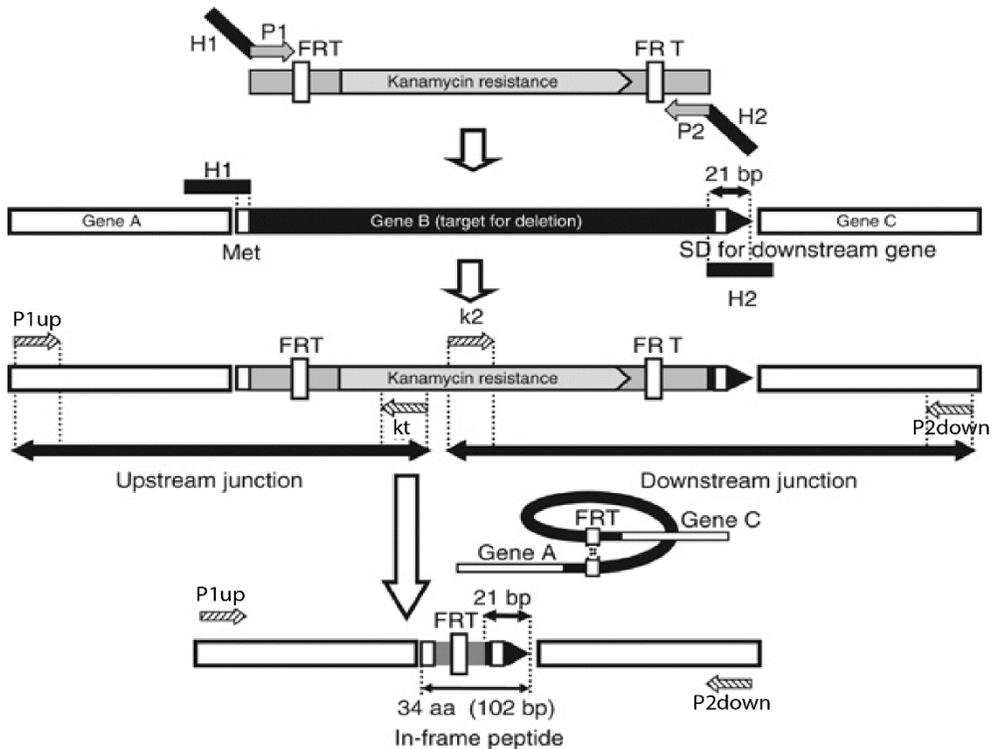


Fig. 3.1. Representación esquemática de la estrategia utilizada en el protocolo de mutagénesis bacteriana descrito por Datsenko y Wanner³³⁵. Se genera un producto de PCR mediante el uso de oligonucleótidos que presentan regiones homólogas al fragmento cromosómico a reemplazar (H1 y H2, de 40 nucleótidos) y a un plásmido (pKD3 o pKD4), mediante las colas P1 y P2 (con un tamaño de 20 nucleótidos). El producto de PCR consiste en un casete de resistencia a antibiótico (Cm^R en el caso del plásmido pKD3 o Km^R en el caso del plásmido pKD4). La recombinasa Red presente en el plásmido pKD46 cataliza la reacción de recombinación entre el producto de PCR y la región cromosómica de interés mediante las regiones de homología introducidas en las colas H1 y H2. Opcionalmente, el casete de resistencia se puede eliminar con la ayuda de la recombinasa FLP, que catalizará la reacción de recombinación entre las repeticiones directas FRT que flanquean los extremos del casete de resistencia amplificado en el producto de PCR. Adaptado de Baba y colaboradores³⁴³.

3.6.1.1 Generación del fragmento que contiene el gen de resistencia al antibiótico mediante PCR

Las resistencias utilizadas en este trabajo han sido cloranfenicol y kanamicina, las cuales se encuentran codificadas en los plásmidos pKD3 y pKD4, respectivamente. Para amplificarlas, se han diseñado oligonucleótidos (directo y reverso) que contienen 42 nucleótidos homólogos al gen de interés que se quiere reemplazar en el extremo 5', mientras que el extremo 3' contiene las secuencias complementarias al plásmido que contiene la resistencia (20 nucleótidos). De esta manera, se genera por PCR un fragmento con secuencias homólogas al gen de interés en cada extremo y, en la zona central, el gen de resistencia al antibiótico flanqueado por secuencias FRT (Fig. 3.1). Una vez obtenido este fragmento de PCR, se ha realizado un tratamiento con la enzima *DpnI*. Esta enzima digiere únicamente el ADN metilado, degradando de manera específica el ADN molde (plásmidos pKD3 o pKD4), ya que el ADN sintetizado *de novo* no se encuentra metilado. Por último, se ha llevado a cabo su purificación como se ha indicado en el apartado 3.5.5.

3.6.1.2 Reemplazamiento alélico

La cepa que se ha querido mutagenizar se ha transformado previamente con el plásmido pKD46, el cual tiene un replicón termosensible y resistencia a carbenicilina. Este plásmido también contiene la recombinasa Red del fago λ bajo el control de un promotor inducible por arabinosa (P_{BAD}), la cual permite la recombinación del fragmento que se ha generado con la región cromosómica de interés. Además, esta enzima inhibe la acción de la exonucleasa V celular, favoreciendo así la entrada de ADN a la célula sin que sea degradado.

La cepa transformada con el plásmido pKD46 se ha crecido a 30 °C en medio SOB suplementado con carbenicilina 100 $\mu\text{g}/\text{mL}$, 10 mM de L-arabinosa,

para inducir la expresión de la recombinasa Red, y 10 mM de Mg^{+2} (a partir de una solución de 1 M $MgSO_4$ y 1 M $MgCl_2$), necesario como cofactor de la recombinasa, hasta una DO_{600nm} de 0,6-0,8. Una vez alcanzada esta DO_{600nm} , se ha procedido a la obtención de células electrocompetentes, como se ha explicado en el apartado 3.5.2.2. Se ha electroporado con 1 μg del fragmento de ADN generado por PCR previamente. A continuación, se han recuperado las células en 1 mL de medio SOC y se han incubado a 37 °C durante 2 horas con una buena agitación y aireación. Tras la incubación, se han sembrado 200 μL en placas de LB agar suplementadas con el antibiótico correspondiente y el resto de la suspensión sobrante se ha incubado durante toda la noche a temperatura ambiente, sembrándola al día siguiente. Las placas se han incubado a 37 °C y las colonias resistentes resultantes se han genotipado mediante PCR (oligonucleótidos con el nombre del gen seguido de P1up o P2down y KT, Tabla 3.3).

3.6.1.3 Eliminación de la resistencia a antibiótico

Una vez comprobada la inserción cromosómica de la resistencia, se ha llevado a cabo su eliminación opcionalmente. Esto es debido a que la presencia del casete puede afectar a la expresión de genes situados aguas abajo. Se ha transformado el mutante con el plásmido pCP20, el cual también es termosensible, aporta resistencia a carbeniclina y cloranfenicol y codifica para la recombinasa FLP, cuya expresión es inducible a 42 °C³³⁶. La recombinasa FLP induce la recombinación de las secuencias repetidas FRT, las cuales flanquean al casete de resistencia. La transformación se ha sembrado en placas de LB con ampicilina 50 $\mu g/mL$, incubándolas a 30 °C durante toda la noche para seleccionar las bacterias que contienen el plásmido. Una vez crecidas, se han sembrado de nuevo los transformantes en placas de LB sin antibiótico a una temperatura de 42 °C, provocando la recombinación de los sitios FRT por la

recombinasa FLP a la vez de realizar el curado del plásmido pCP20. Se han hecho estrías sucesivas a esta temperatura y, por último, se han verificado los mutantes mediante PCR.

3.6.2 Construcción de fusiones génicas con la secuencia codificante 3XFLAG mediante la recombinasa Red

Con la finalidad de detectar proteínas de interés mediante ensayos de inmunodetección y de reducir los costes y tiempo que conlleva la utilización de antisueros específicos contra la proteína de interés, se ha insertado una etiqueta 3XFLAG en el extremo carboxi-terminal. Se ha utilizado un anticuerpo monoclonal comercial específico contra esta etiqueta (Monoclonal ANTI-FLAG M2 antibody, Sigma).

Para llevarlo a cabo, se ha seguido el protocolo descrito por Uzzau y colaboradores³³⁷. Este protocolo es semejante al anteriormente descrito en el apartado 3.7.1, realizando algunas modificaciones para obtener una fusión traduccional con la secuencia codificante del epítipo 3XFLAG en el extremo C-terminal. Se han diseñado dos oligonucleótidos (nombrados por el nombre del gen seguido de 3XP1 o 3XP2, Tabla 3.3) que amplifican la etiqueta 3XFLAG seguida de la resistencia a kanamicina. Esta PCR se ha realizado con el plásmido pSUB11 como molde. El oligonucleótido directo es complementario a los 40 nucleótidos anteriores al codón de parada del gen de interés en su extremo 5', mientras que el extremo 3' es complementario a los primeros 20 nucleótidos de la etiqueta 3XFLAG presente en el plásmido pSUB11. Por otra parte, el oligonucleótido reverso presenta en el extremo 5' complementariedad a 40 nucleótidos de la secuencia adyacente al gen de interés, empezando por 4 nucleótidos aguas arriba del codón de parada de este gen. En el extremo 3' se encuentran 20 nucleótidos complementarios al *priming site 2* del plásmido

pSUB11. De esta manera, se ha obtenido un fragmento de PCR que contiene de 5' a 3' una región homóloga al gen en el que se quiere realizar la fusión, la secuencia del epítipo 3XFLAG, la resistencia a kanamicina flanqueada por secuencias FRT y una región homóloga a la secuencia posterior al gen de interés (Fig. 3.2). Una vez hecha la PCR (apartado 3.5.4), se ha hecho un tratamiento con la enzima *DpnI* para eliminar el ADN molde y se ha purificado, tal y como se ha descrito en el apartado 3.5.5. Una vez purificado, la sustitución alélica se llevó a cabo tal y como se ha descrito en el apartado 3.6.1.2. Las fusiones se han seleccionado en placas de LB suplementadas con kanamicina y se han verificado mediante PCR. Finalmente, la resistencia a kanamicina se ha curado gracias a las secuencias FRT flanqueantes, siguiendo la metodología indicada en el apartado 3.6.1.3.

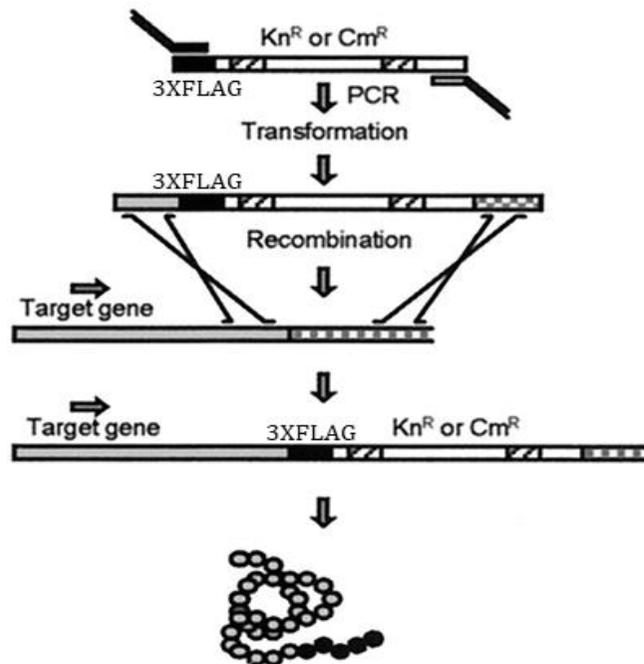


Fig. 3.2. Esquema del protocolo de fusión cromosómica de la etiqueta 3XFLAG en el extremo C-terminal de un gen determinado (*target gene*). En negro se indica el epítipo 3XFLAG. La resistencia a antibiótico (Km^R) se encuentra flanqueada por dos secuencias FRT (marcadas a rayas). El producto de PCR generado se integra en el cromosoma mediante recombinación, tal y como se ha indicado en el apartado 3.6.1. Adaptado de Uzzau y colaboradores³³⁷.

3.6.3 Construcción de fusiones génicas con el gen *lacZ* mediante la recombinasa FLP

La creación de fusiones transcripcionales con el gen reportero *lacZ* se ha realizado a partir del método descrito por Ellermeier y colaboradores³⁴⁴. Este método también es una modificación del protocolo de inactivación de genes cromosómicos por fragmentos de PCR (apartado 3.6.1).

En este caso, a los mutantes en los que se ha curado la resistencia a antibiótico se les ha vuelto a transformar el plásmido pCP20 que codifica la recombinasa FLP. A continuación, se les ha transformado también el plásmido pKG136, el cual presenta el gen *lacZ* sin promotor y con secuencias FRT seguido de un gen de resistencia a kanamicina. La acción de la recombinasa FLP permite la recombinación entre los sitios FRT existentes en el gen que se ha mutado con los sitios FRT que flanquean al gen *lacZ* y la resistencia a kanamicina presentes en el plásmido pKG136 (Fig. 3.3). Mediante la siembra en placas suplementadas con kanamicina se han seleccionado aquellos clones que han incorporado la resistencia y, por tanto, la fusión transcripcional en el gen de interés. Estas fusiones también han sido comprobadas por PCR y, posteriormente, utilizadas en ensayos de valoración de la actividad β -galactosidasa.

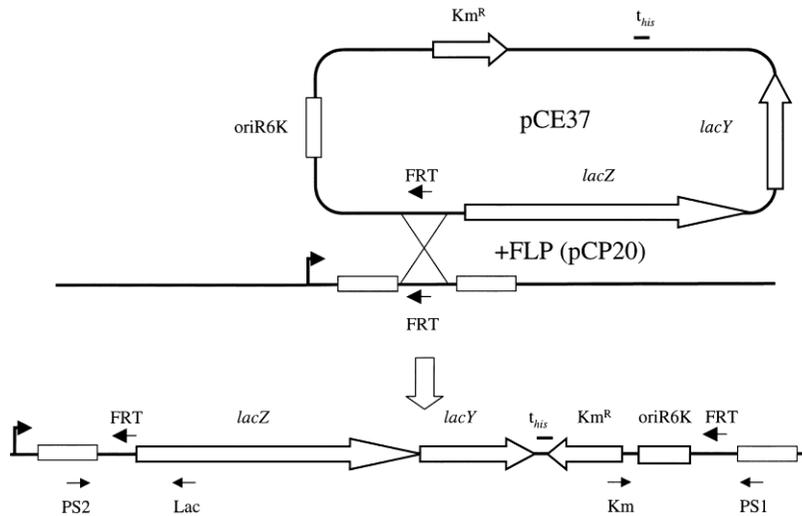


Fig. 3.3. Representación esquemática del protocolo de fusión del gen reportero *lacZ* en el cromosoma bacteriano. El gen reportero *lacZ* presente en el plásmido pKG136 se integra en la secuencia FRT generada previamente en el cromosoma bacteriano mediante el protocolo de mutagénesis basado en el reemplazamiento alélico. La recombinasa FLP presente en el plásmido pCP20 se encarga de catalizar dicha reacción que permite obtener fusiones cromosómicas con el gen reportero *lacZ*. Adaptado de Ellermeier y colaboradores³³⁸.

3.6.4 Construcción de fusiones transcripcionales en el plásmido pUJ8

Las fusiones transcripcionales plasmídicas se han generado para el estudio de las regiones reguladoras de los genes de interés utilizando el plásmido pUJ8³³⁹ (Fig. 3.4). Este plásmido presenta el gen reportero *lacZ* sin su región promotora y un casete de resistencia a ampicilina para su selección. Las regiones promotoras que se han querido estudiar se han amplificado por PCR usando la enzima *Phusion Hot Start II DNA Polymerase* (apartado 3.5.4). En los fragmentos de PCR se ha añadido la diana de restricción *EcoRI* en el extremo 5' y la diana *BamHI* en el extremo 3'. Una vez amplificado, se ha digerido tanto el inserto como el vector y se ha ligado en el sitio *EcoRI/BamHI*. La fusión

transcripcional obtenida se ha utilizado para valorar la actividad β -galactosidasa.

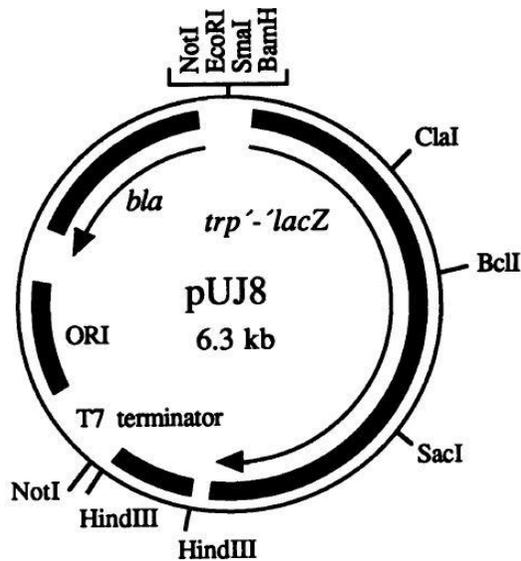


Fig. 3.4. Esquema del plásmido pUJ8. El plásmido contiene el gen *lacZ* sin promotor y codifica el gen para la resistencia a ampicilina (*bla*). Adaptado de de Lorenzo y colaboradores³³⁹.

3.6.5 Mutagénesis Sitio-Dirigida

La mutagénesis dirigida es una técnica útil para estudiar la relación estructura-función de proteínas, la expresión génica y para llevar a cabo modificaciones puntuales en vectores, como es el caso del plásmido pUJ8. Esta técnica permite introducir mutaciones puntuales en secuencias de ADN plasmídico, haciendo uso de una polimerasa robusta y altamente fiable, como es el caso de la polimerasa *Phusion Hot Start II DNA Polymerase* (*ThermoScientific*), y oligonucleótidos complementarios entre sí con la mutación deseada. La polimerasa permite la síntesis de todo el plásmido *de novo*, introduciendo las mutaciones puntuales codificadas en los

oligonucleótidos diseñados. La figura 3.5 muestra este proceso de manera esquemática.

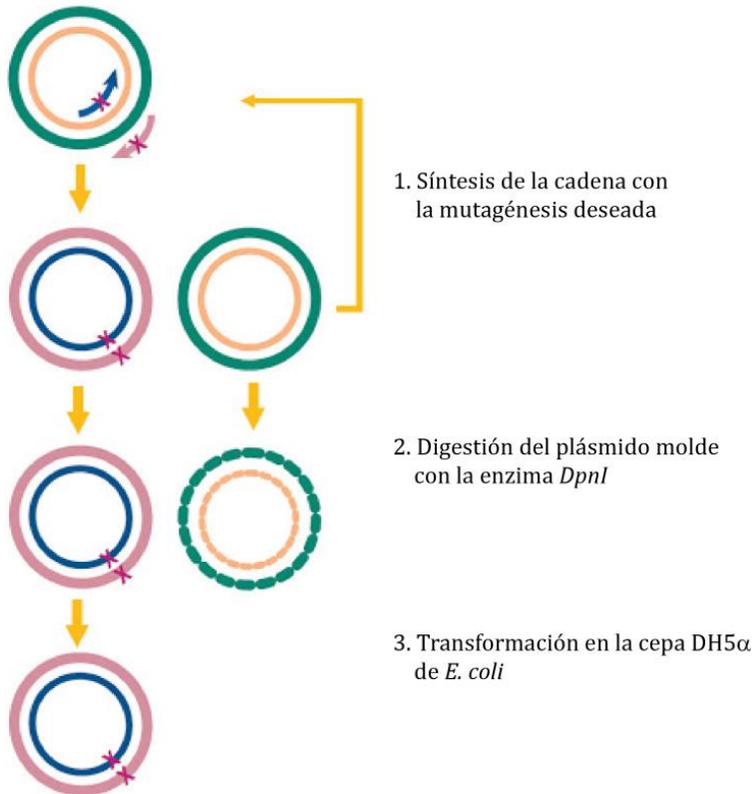


Fig. 3.5. Esquema de la metodología empleada para llevar a cabo la mutagénesis Sitio-Dirigida. En verde y naranja se indica la secuencia molde. En azul y rosa la secuencia sintetizada *de novo*. Las flechas rojas indican la modificación puntual introducida.

El fragmento génico de interés fue clonado en el vector pUJ8, tal y como se ha indicado en el apartado 3.6.4. Se diseñaron oligonucleótidos, de entre 25 y 45 nucleótidos complementarios entre sí, que contenían en el centro de su secuencia el cambio nucleotídico deseado. A partir de estos dos oligonucleótidos complementarios, se realizó la reacción de amplificación para obtener la mutación nucleotídica deseada con el siguiente programa: 3 min a

98 °C de desnaturalización inicial, seguidos de 30 seg a 98 °C. A continuación, se ha realizado la hibridación durante 30 seg a 55 °C y la extensión a 68 °C durante 30 seg/Kb. Estas 3 etapas se han repetido 18 ciclos y, por último, se ha llevado a cabo una extensión final durante 10 min a 68 °C (Fig. 3.5, Punto 1). Una vez finalizada la amplificación, el producto de PCR fue digerido con la enzima *DpnI*, eliminando el plásmido molde metilado (Fig. 3.5, Punto 2). Seguidamente, se transformó la cepa DH5 α mediante electroporación (apartado 3.5.2.2) y se sembró en placas de LB suplementadas con carbenicilina (Fig. 3.5, Punto 3). Las bacterias transformantes se confirmaron mediante la extracción de plásmido y su secuenciación. Finalmente, los plásmidos con las mutaciones deseadas fueron transformados en la cepa EAEC 042 Δ LC (Tabla 3.1).

3.7 Técnicas experimentales con ARN

3.7.1 Extracción de ARN total y cuantificación

La extracción de ARN se ha llevado a cabo con el reactivo *TriPure Isolation Reagent (Roche)* a partir de 5 mL de un cultivo bacteriano crecido hasta una DO_{600nm} de interés, momento en el que se ha parado la transcripción mediante la adición de una mezcla de etanol y fenol en una proporción 95:5. Este reactivo es una mezcla de tiocianato de guanidinio (potente agente caotrópico, desnaturalizante) y fenol. En su presencia, las células se lisan rápidamente, se solubilizan sus componentes y se inactivan las ribonucleasas. Seguidamente, se añadió cloroformo con la consiguiente generación de dos fases: una fase acuosa (con el ARN) y una fase orgánica (con las proteínas desnaturalizadas), mientras que el ADN quedaba en la interfase. A continuación, el ARN fue precipitado mediante isopropanol y la posterior adición de etanol al 75 % (v/v). Tras la extracción, para eliminar totalmente cualquier resto de ADN que pudiera

quedar en la muestra, se ha realizado un segundo tratamiento con la ADNasa comercial *Turbo DNA-free* (*ThermoScientific*). Finalmente se ha purificado con fenol, cloroformo e isopropanol (P:C:I) a una proporción 25:24:1 y se ha precipitado durante toda la noche con 3 volúmenes de etanol 100 % y 0,1 volúmenes de acetato sódico pH 5,4 a -20 °C. Al día siguiente, se ha centrifugado a 13.400 rpm durante 10 minutos y se ha descartado el sobrenadante. Se ha adicionado 1 volumen de etanol 70 % y se ha vuelto a centrifugar durante 10 minutos. Por último, el sedimento se ha secado al aire y resuspendido en 50 µL de agua libre de nucleasas.

Una vez precipitado y resuspendido el ARN total, se ha procedido a medir la concentración con el espectrofotómetro NanoDrop® ND-1000 UV-Vis Spectrophotometer ND-1000 (*ThermoScientific*).

3.7.2 RT-PCR

Esta técnica se ha llevado a cabo para la retrotranscripción del ARN a ADN complementario (ADNc). Esta reacción se ha realizado siguiendo las instrucciones del kit *High-capacity cDNA Reverse Transcription* (*Applied Biosystems*), donde se ha mezclado tampón de retrotranscriptasa a una concentración final de 1X, 4 mM de nucleótidos desoxirribonucleicos (dNTPs), oligonucleótidos degenerados a 1X, 50 unidades de retrotranscriptasa, 1 µg de ARN molde y agua libre de nucleasas hasta un volumen de 20 µL. De forma paralela, se ha preparado un control negativo de la retrotranscripción en el cual no se ha añadido la enzima retrotranscriptasa. La reacción ha tenido lugar en un ciclo de 10 minutos de hibridación a 25 °C, 120 minutos a 37 °C y 5 minutos finales a 85 °C para la desnaturalización de la enzima.

3.7.3 Walking RT-PCR

Esta técnica ha sido realizada con el objetivo de acotar la longitud del ARN mensajero del gen *irmA*. En primer lugar, se ha realizado una extracción de ARN total y se ha retrotranscrito, tal y como se ha indicado en el apartado anterior. A continuación, se han llevado a cabo diversas PCR usando como molde el ADNc obtenido. En estas reacciones se ha fijado el oligonucleótido reverso *irmA* rv (Tabla 3.3) y se ha usado junto a diversos oligonucleótidos directos (*irmA* fw, *int* fw y *yeeR* fw, Tabla 3.3). Los oligonucleótidos directos se encuentran cada vez más alejados del reverso. El tamaño del amplicón se ha confirmado usando ADN genómico, mientras que, como control negativo, se ha utilizado el ARN extraído donde no hay restos de ADN.

3.7.4 Rapid Amplification of cDNA ends (5' RACE)

El ensayo de 5' RACE se ha realizado para la determinación del extremo 5' del ARN mensajero transcrito de las dos copias del gen *irmA* (EC042_2244 y EC042_4509) de la cepa EAEC 042. Para ello, se ha seguido el protocolo descrito por Vogel y Wagner³⁴⁵, realizando algunas modificaciones.

En primer lugar, tras la extracción del ARN total y su purificación (apartado 3.7.1), se han alicuotado 12 µg. Se han desfosforilado los extremos 5' trifosfatos a monofosfatos mediante un tratamiento con 25 unidades de pirofosfatasa ácida de tabaco (TAP, *Epicentre*), tampón TAP 1X y 0,5 µL de inhibidor de ARNasa (*Ambion*) durante 1 hora a 37 °C. De manera paralela, se ha realizado un control en el que se ha incubado el ARN en las mismas condiciones, pero sin la adición de la enzima TAP. Tras la incubación, se han añadido 500 pmol del adaptador de ARN A3 (Tabla 3.3), se ha hecho una extracción con P:C:I y se ha precipitado con etanol 100 % en presencia de acetato sódico 0,3 M a pH 5,7. El ARN se ha resuspendido en agua libre de

nucleasas y se ha realizado una desnaturalización por calor a una temperatura de 90 °C, tras la cual se ha dejado en hielo durante 5 minutos. Seguidamente, se ha ligado el adaptador A3 gracias a la adición de una mezcla de ARN ligasa T4 (*ThermoScientific*) e inhibidor de ARNasa (*Ambion*) en una proporción 9:1, tampón de ligasa a una concentración final de 1X y dimetilsulfóxido (DMSO) al 10 %. Se ha incubado la reacción a 17 °C durante 16 horas, tras lo cual se ha vuelto a purificar con P:C:I y precipitado con etanol y acetato sódico.

El ARN ligado con el adaptador A3 se ha resuspendido con agua y se han retrotranscrito 5 µg con oligonucleótidos degenerados, siguiendo las instrucciones del kit *High-capacity cDNA Reverse Transcription (Applied Biosystems)* (apartado 3.7.2). Los productos de la retrotranscripción se han amplificado usando la enzima *DreamTaq Hot Start DNA polymerase (ThermoScientific)*, donde por cada 50 µL de reacción se han usado 2 µL de ADNc y 0,5 µM del oligonucleótido directo del adaptador y reverso del gen (B6 y 2244_rv5, respectivamente).

Los productos de PCR se han purificado (apartado 3.5.5) y clonado en el vector pCR2.1TOPO (*Invitrogen*), siguiendo las instrucciones del fabricante. Las ligaciones se han sembrado en placas de LB suplementadas con carbenicilina y XGal. Por último, se han seleccionado las colonias blancas, en las cuales el inserto ha roto la pauta de lectura del gen *lacZ* reportero. Estas colonias se han genotipado por PCR usando los oligonucleótidos M13 fw y M13 rv (Tabla 3.3) y se han secuenciado para verificar los insertos.

3.8 Técnicas experimentales con proteínas

3.8.1 Obtención de extractos totales

Esta técnica se ha utilizado para conseguir todas las proteínas expresadas por la cepa estudiada en unas condiciones determinadas. En primer lugar, se ha crecido un cultivo hasta la DO_{600nm} que se quería estudiar (normalmente 0,4 y 2) y se ha recogido 1 mL de la suspensión bacteriana. Seguidamente, se ha llevado a cabo una centrifugación a 13.000 rpm y se ha desechado el sobrenadante, resuspendiendo el sedimento en un volumen de tampón de proteínas 1X, según la DO_{600nm} que tuviese el cultivo. Este tampón se ha preparado diluyendo la solución madre concentrada 4X compuesta por 20 % glicerol, 10 % β -mercaptoetanol (*Sigma*), 4,6 % SDS, 0,125 M Tris-HCl pH 6,8 y 0,2 % azul de bromofenol. Una vez que se ha resuspendido el sedimento en el tampón de proteínas, se ha procedido a hervir la muestra durante 10 minutos para acabar de desnaturalizar las proteínas.

Siguiendo la relación $DO_{600nm}/4 = \text{Volumen de tampón de muestra 1X (mL)}$, se ha conseguido igualar la concentración de todos los extractos proteicos que se han recogido, pudiendo así comparar los resultados entre ellos.

3.8.2 Electroforesis en geles de poliacrilamida en presencia de SDS

Tras la obtención de los extractos proteicos, se ha llevado a cabo la separación de las proteínas mediante geles de poliacrilamida en presencia de SDS, atendiendo únicamente a su peso molecular. Se han utilizado geles de poliacrilamida trifásicos (Tris-Tricina-SDS) en vez de geles con solo dos fases (Tris-Glicina-SDS), debido a que permiten una mayor resolución de proteínas, con tamaños comprendidos entre 5 y 50 kDa³⁴⁶. Los geles trifásicos están formados por la fase de compactación, la fase espaciadora y la fase de resolución, donde las dos primeras fases ocupan la primera mitad del gel y la fase de resolución la segunda mitad. El grosor de los geles que se han utilizado

en este trabajo ha sido de 0,75 mm. La composición de estas fases ha sido la misma en todos los casos, como se detalla en la Tabla 3.4.

Tabla 3.4. Composición de las fases que forman el gel de poliacrilamida trifásico

	Fase de resolución 16.5 %T 3 %C	Fase espaciadora 10 %T 3 %C	Fase de compactación 4 %T 3 %C
Acrilamida/Bis 40 %T 3,3 %C (<i>Bio-Rad</i>)	1,2 mL	0,36 mL	0,18 mL
Tampón de gel 3X (Tabla 3.5)	1 mL	0,5 mL	0,45 mL
Glicerol 50 % (v/v)	0,8 mL	-	-
Agua bidestilada	-	0,64 mL	1,23 mL
APS 10 % (p/v)	20 µL	10 µL	15 µL
TEMED	4 µL	2 µL	3 µL

% T: Porcentaje de acrilamida + bisacrilamida total

% C: Porcentaje de bisacrilamida respecto a T

La preparación del gel se ha iniciado mediante la adición de la fase de resolución, enrasando con agua hasta el borde superior de los vidrios. Una vez que ha polimerizado la fase, se ha retirado el agua y se ha añadido la fase espaciadora, volviendo a enrasar con agua. Por último, cuando ha polimerizado esta fase espaciadora, se ha vuelto a retirar el agua y se ha añadido la fase de compactación, colocando el peine que permite la formación de los pocillos.

Una vez que el gel ha polimerizado, se ha dispuesto en la cubeta de electroforesis. Se ha llenado el espacio interior entre geles con tampón cátodo 1X (Tabla 3.5) y el espacio exterior con tampón ánodo 1X (Tabla 3.5). Se ha

procedido a la separación de las proteínas mediante electroforesis aplicando un voltaje de 50 voltios (V) hasta que han entrado en la fase espaciadora, a partir de la cual se ha aplicado un voltaje mayor de 120 V. Junto a las muestras proteicas, se ha cargado el marcador de pesos moleculares *PageRule Prestained Plus Protein Ladder* (ThermoScientific).

Tabla 3.5. Composición de los tampones utilizados para la electroforesis

	Tampón de gel 1X (pH ajustado a 8,5 con HCl)	Tampón ánodo 1X (tampón ajustado a pH 8,9 con HCl)	Tampón cátodo 1X (tampón a pH 8,25 sin ajustar)
Tris (M)	1	0,2	0,1
Tricina (M)	-	-	0,1
SDS (%)	0,1	-	0,1

3.8.3 Tinción con azul de Coomassie

Una vez realizada la electroforesis de los geles, se ha procedido a su tinción para visualizar el patrón de bandas de cada extracto. Esta tinción se ha llevado a cabo mediante el uso de la solución de tinción (*Coomassie Brilliant Blue R-250, Bio-Rad*), durante 45 min a temperatura ambiente en agitación suave. Una vez transcurrido este tiempo, se ha retirado la solución de tinción y se ha lavado el gel con ácido acético 10 %, eliminando el exceso de tinción de los geles para poder visualizar las bandas presentes en el gel.

Finalmente, se ha retirado el ácido acético y se ha añadido agua destilada para conservar el gel y realizar fotografías con el equipo *Molecular Imager® GelDoc™ XR+* (Bio-Rad).

3.8.4 Inmunodetección de proteínas (*Western Blot*)

3.8.4.1 Transferencia de proteínas a membrana por sistema semi-seco

Las proteínas que se han separado por electroforesis han sido transferidas a membranas de polifluoruro de vinilideno (PVDF) para poder realizar su inmunodetección. Las membranas de PVDF han sido tratadas con metanol durante 5 minutos debido a la hidrofobicidad que presentan, la cual es anulada mediante el tratamiento con solventes orgánicos. Tras realizar este tratamiento, se ha sumergido junto al gel de poliacrilamida en tampón de transferencia (40 mM Tris, 39 mM glicina, 20 % metanol, 1,3 mM SDS), equilibrando el pH antes de la transferencia.

Para llevar a cabo la transferencia, se ha realizado un montaje en el aparato *Trans-Blot Semy-Dry Transfer Cell* (Bio-Rad). En primer lugar, se han colocado 3 papeles de filtro (Whatmann 3 mm) hidratados con tampón de transferencia, a continuación, la membrana de PVDF y el gel encima y, por último, otros 3 papeles de filtro hidratados. Este montaje se ha colocado entre el ánodo y el cátodo del aparato, realizando la transferencia aplicando un voltaje de 15 V durante 40 minutos.

3.8.4.2 Revelado quimioluminiscente

Tras la transferencia de las proteínas a membranas de PVDF, se ha realizado el bloqueo de las mismas para evitar uniones inespecíficas. Para ello, se ha usado la solución de bloqueo, compuesta por PBS 1X (80 mM Na₂HPO₄, 20 mM NaH₂PO₄, 100 mM NaCl pH 7,3), Tritón X-100 0,2 % (v/v) y leche descremada en polvo 5 % (p/v) (*Scharlau*), durante una hora en agitación suave a temperatura ambiente. Tras el tiempo de bloqueo, se ha retirado la

solución y se ha incubado la membrana con el anticuerpo primario correspondiente de la proteína a detectar, durante toda la noche en agitación suave y a 4 °C. Los anticuerpos primarios se han conservado a -20 °C, diluidos en una solución de PBS-Tritón X-100 0,2 % (v/v) con leche descremada en polvo al 3 % (p/v). Una vez incubado con el anticuerpo primario, la membrana se ha lavado 3 veces con PBS-Tritón X-100 0,2 % (v/v) durante 15 minutos con una leve agitación a temperatura ambiente, con la finalidad de retirar el exceso de anticuerpo que haya podido quedar. Después, se han añadido 10 mL de una dilución 1:2.500 del anticuerpo secundario conjugado con peroxidasa *Anti-mouse IgG HRP Conjugate (Promega)*, o una dilución 1:2.000 del anticuerpo conjugado con peroxidasa *Anti-Rabbit IgG (whole molecule)-Peroxidase antibody (Sigma)*. Se ha incubado durante 45 minutos a temperatura ambiente en agitación. Tras esta incubación, se ha retirado de nuevo el exceso de anticuerpo mediante 3 lavados de 30 minutos con PBS-Tritón X-100 0,2 % (v/v).

Para llevar a cabo la reacción de quimioluminiscencia para revelar las proteínas, se ha incubado la membrana de PVDF durante 5 minutos con la solución de revelado, siguiendo las indicaciones del fabricante (*ECL Western Blotting, GE Healthcare*). La detección de la reacción y la captura de imágenes se han realizado gracias al equipo *Molecular Imager ChemiDoc XRS System* y el software *QuantityOne 4.6.6 (Bio-Rad)*.

3.8.5 Obtención de anticuerpo anti-Irma

3.8.5.1 Purificación de la proteína Irma

Esta técnica se ha realizado siguiendo las instrucciones del kit comercial aLICator LIC Cloning and Expression Set 1 (All-in-One) (*ThermoScientific*). Se basa en el método de clonaje LIC (Ligation Independent Cloning), en el que se

utiliza la actividad de exonucleasa 3' - 5' de la ADN polimerasa T4 para generar extremos de 14-21 nucleótidos cohesivos específicos de cadena sencilla, tanto en los insertos de ADN específicos como en los vectores de clonaje pLATE (Fig. 3.6).

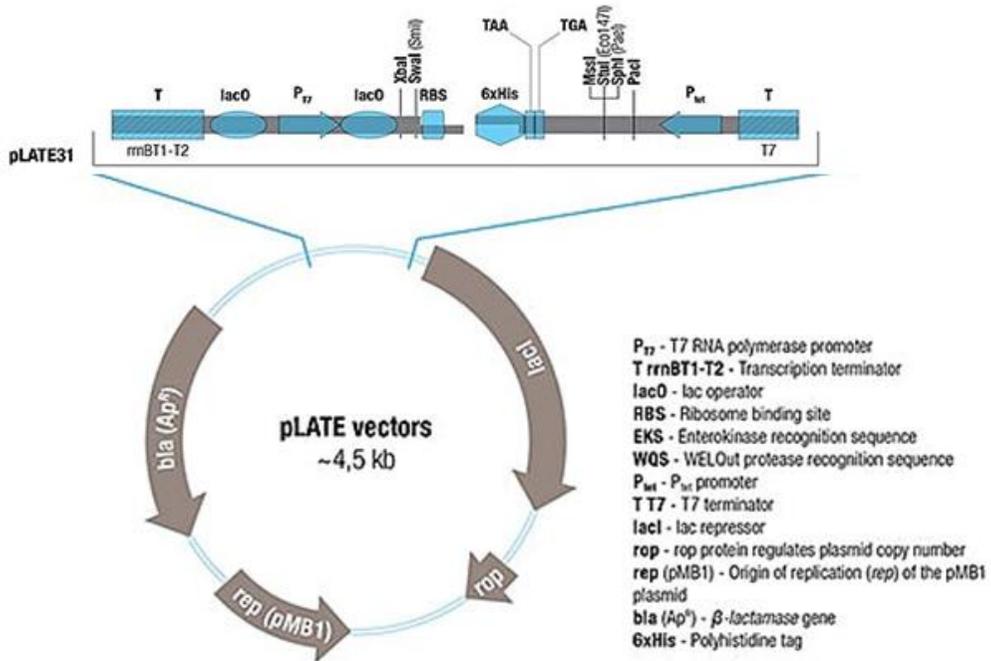


Fig. 3.6. Representación esquemática del vector pLATE31. Utilizado para clonar y sobreexpresar, de manera inducible, proteínas fusionadas con una cola de 6 histidinas en el extremo C-terminal.

En primer lugar, se amplificó el gen *irmA* mediante PCR (apartado 3.5.4), haciendo uso de los oligonucleótidos *irmA* pLATE31 fw y rv (Tabla 3.3) y la polimerasa de alta fidelidad *Phusion Hot Start II DNA Polymerase* (*ThermoScientific*). El fragmento de PCR obtenido se purificó, tal y como se ha descrito en el apartado 3.5.5, y se realizó una ligación con el plásmido pLATE31, siguiendo las instrucciones del fabricante. A continuación, la ligación se transformó en la cepa DH5α y se comprobó mediante secuenciación. Seguidamente, se realizó una extracción de plásmido con el kit comercial

GeneJET Plasmid Miniprep Kit de *ThermoScientific* y se transformó la cepa BL21 (DE3), por tratamiento con calcio (apartado 3.5.2.1).

Una vez transformada la cepa BL21 (DE3) con el clon escogido, se creció un cultivo de 20 mL en LB a 37 °C, durante toda la noche en agitación. Al día siguiente, se inoculó 1 litro de medio LB con todo el inóculo (dilución 1:50) y se creció a 37 °C en agitación. Cuando el cultivo alcanzó una DO_{600nm} de 0,5, se añadieron 10 mL de isopropil- β -D-1-tiogalactopiranosido (IPTG) 100 mM para inducir la expresión de la proteína Irma, dejándolo crecer durante 3 horas más.

Una vez pasado este tiempo, se vertió el contenido en 4 tubos y se centrifugaron a 7.500 rpm durante 45 minutos a 4 °C. Se eliminó el sobrenadante y se resuspendió el sedimento de cada uno con 5 mL de tampón A50 (Hepes 20 mM, KCl 100 mM, $MgCl_2$ 5 mM e imidazol 50 mM). Los 20 mL resuspendidos se vertieron en otro tubo, se disolvieron dos pastillas de inhibidores de proteasas (*cOmplete ULTRA Tablets, Mini, EDTA-free, EASYpack Roche*) y se conservó a -80 °C. Más tarde, se procedió a la rotura celular para la extracción de todo el extracto proteico. Para ello, se utilizó la técnica de *French Press*, la cual consigue romper las células mediante un cambio de presión entre la parte interna y la externa del pistón. Una vez realizada la rotura celular, se centrifugó el extracto a 12.000 rpm durante 45 min a 4 °C y se guardó el sobrenadante (20 mL), desechando el sedimento.

Seguidamente, se procedió a realizar la purificación de la proteína Irma mediante una cromatografía de afinidad. Para ello, se utilizó una resina de níquel-agarosa (*HisPur Ni-NTA Superflow Agarose, ThermoScientific*), aprovechando la gran afinidad que presentan las histidinas (fusionadas a la proteína) con este elemento químico. Se hicieron interaccionar 10 mL del extracto proteico con 0,5 mL de la resina, durante toda la noche a 4 °C con una agitación suave. Tras esto, se pasó el extracto por una columna (*Poly-Prep Chromatography Column, Bio-Rad*) y se hicieron 3 lavados con 10 mL del

tampón A50 (Hepes 20 mM, KCl 100 mM, MgCl₂ 5 mM e imidazol 50 mM), retirando así las interacciones inespecíficas. Finalmente, se realizaron 3 eluciones con 0,5 mL del tampón A200 (Hepes 20 mM, KCl 100 mM, MgCl₂ 5 mM e imidazol 200 mM), el cual tiene una concentración mayor de imidazol que el tampón A50.

3.8.5.2 Inmunización de conejos y obtención del anticuerpo policlonal

La purificación de proteína se realizó varias veces hasta que se obtuvo cantidad suficiente de proteína para la inmunización de conejos. Se ajustó la concentración de proteína y se inoculó a los conejos con una primera dosis de 1 mL a una concentración de 1 mg/mL, de acuerdo a los protocolos estándar (Unitat d'Experimentació Animal de Farmàcia-CCiTUB). Tras esta primera dosis, se administraron 3 dosis más sucesivas de 1 mL, a una concentración de 0,5 mg/mL. Para finalizar, se obtuvo tanto suero preinmune como suero después de la inmunización. Estos sueros se diluyeron 1:1.000 en PBS 1X, Tritón X-100 0,2 % (v/v) y leche descremada en polvo 3 % (p/v) (*Scharlau*) y se testaron para determinar si reaccionaban con la proteína IrmA purificada mediante *Western Blot*.

3.9 Valoración de la actividad enzimática

3.9.1 Determinación de la actividad β -galactosidasa

Este procedimiento, descrito por Miller³⁴⁷, se ha llevado a cabo para el estudio de la actividad transcripcional de los promotores de los genes donde se ha realizado una fusión con el gen reportero *lacZ*, tanto en el cromosoma como

en plásmidos. Se han crecido cultivos en las condiciones de interés hasta la densidad óptica que se quería estudiar, normalmente DO_{600nm} 0,4 y 2. Se han recogido alícuotas de 0,1 o 0,05 mL, respectivamente. A continuación, se han llevado hasta un volumen de 1 mL con tampón Z a pH 6,9 (60 mM Na_2HPO_4 , 40 mM NaH_2PO_4 , 10 mM KCl, 1 mM $MgSO_4$ y 50 mM β -mercaptoetanol), se han añadido 50 μ L de cloroformo y 25 μ L de SDS 0,1 % y se ha agitado en vórtex durante 15 segundos, consiguiendo así la permeabilización de las membranas. La mezcla se ha atemperado durante 5 minutos a 28 °C, tras lo cual se le ha añadido 200 μ L de O-nitrofenil- β -galactósido (ONPG) preparado en tampón Z a una concentración de 4 mg/mL, incubando todo a 28 °C. El ONPG es un análogo sintético de la lactosa que, al ser degradado por la enzima β -galactosidasa, origina o-nitrofenol, el cual dota de una coloración amarilla al medio. La incubación a 28 °C se llevó a cabo hasta la aparición de la coloración amarilla típica de la degradación del ONPG, con una DO_{420nm} entre 0,3 y 0,9, momento en el que la reacción se ha detenido mediante la adición de 500 μ L de Na_2CO_3 1 M.

Una vez detenida la reacción, se han dejado reposar las muestras durante unos minutos y se ha medido la absorbancia a las longitudes de onda de 420 y 550 nm. Con los valores obtenidos, se ha calculado la actividad β -galactosidasa con la siguiente fórmula:

$$\text{Actividad } \beta - \text{galactosidasas (U.M.)} = \frac{(DO_{420nm} - 1,75 \cdot DO_{550nm})}{t(\text{min}) \cdot V(\text{mL}) \cdot DO_{600nm}} \cdot 1.000$$

U.M.: Unidades Miller

DO_{420nm} y DO_{550nm} : absorbancias de la muestra a las longitudes de onda de 420 (coloración amarilla) y 550 nm (restos celulares), medidas tras detener la reacción.

DO_{600nm} : absorbancia a 600 nm del cultivo en el momento en el que se tomó la muestra.

t: tiempo transcurrido desde que se ha añadido el ONPG hasta que se ha detenido la reacción, en minutos.

V: volumen de muestra utilizado para llevar a cabo la reacción, en mililitros

3.10 Estudios de formación de biofilm

La capacidad de formación de biofilm de los distintos mutantes de la cepa 042 de *E. coli* fue determinada mediante el protocolo descrito por Sheikh y colaboradores³⁴⁸. Estos autores establecen el medio DMEM suplementado con un 0,45 % de glucosa como medio estándar para el estudio de formación de biofilm de cepas enteroagregativas de *E. coli*. Los cultivos de las cepas de interés se crecieron en LB a 37 °C en agitación hasta una DO_{600nm} de 2. A continuación, se inoculó la cantidad de suspensión bacteriana y de medio de cultivo que indica la Tabla 3.6, dependiendo de los pocillos de la placa de cultivo que se utilizara. Las placas fueron incubadas a 37 °C durante 16 horas.

Tabla 3.6. Condiciones utilizadas en los ensayos de formación de biofilm.

DMEM + glucosa 0,45 % (μ L/pocillo)	Tipo de placa de cultivo	μ L de suspensión bacteriana a DO_{600nm} 2
200	96 pocillos	2
800	24 pocillos	8

Tras la incubación, se retiró el medio de cultivo de cada pocillo, teniendo cuidado de no tocar la monocapa formada por las bacterias. A continuación, se lavaron los pocillos con 1 mL de PBS 1X en el caso de la placa de 24 pocillos, o con 200 μ L en el caso de la placa de 96 pocillos. Seguidamente, la monocapa formada por las bacterias se tiñó mediante 100 μ L de una solución de cristal violeta 0,5 % en las placas de 24 pocillos, o con la mitad de este volumen en el

caso de las placas de 96. Tras incubar la placa durante 5 minutos a temperatura ambiente, se retiró el colorante y se realizaron lavados consecutivos de las placas con agua destilada. Finalmente, se dejaron secar las placas a temperatura ambiente.

La cuantificación de los biofilms se realizó en las placas de 96 pocillos, mediante la determinación de la DO_{570nm} . Este procedimiento se llevó a cabo añadiendo a cada pocillo 200 μ L de etanol al 95 %, incubándolo 2 minutos a temperatura ambiente. A continuación, se resuspendió la monocapa teñida con cristal violeta de cada pocillo en el etanol añadido con ayuda de una pipeta automática, y se transfirieron 150 μ L de la solución a otro pocillo de una placa nueva. Finalmente, se realizó la lectura de la DO_{570nm} de las placas.

3.11 Ensayos de infección en *Galleria mellonella*

Los experimentos realizados con las larvas de *Galleria mellonella* fueron llevados a cabo en colaboración con el grupo de investigación del Dr. Eduard Torrents, haciendo uso de las instalaciones del Instituto de Bioingeniería de Cataluña (IBEC). Las larvas se mantuvieron a 34 °C con una dieta artificial, formada por un 15 % harina de maíz, 15 % harina de trigo, 15 % cereal infantil, 11 % de leche en polvo, 6 % de levadura de cerveza, 25 % de miel y 13 % de glicerol³⁴⁹. Para preparar el inóculo bacteriano que se iba a inyectar, se creció un cultivo durante toda la noche en medio LB a 37 °C en agitación. Se cogieron 5 mL del cultivo y se centrifugaron durante 10 minutos a 6.000 rpm. El sobrenadante se descartó y el sedimento se lavó 3 veces con PBS 1X. Seguidamente, se diluyó el cultivo con PBS 1X hasta una DO_{600nm} de 1 y se realizó un banco de diluciones. Finalmente, a las larvas de *G. mellonella* (con un peso de 200-220 mg) se les inyectó 20-40 CFU, a través de las pseudopatas de la zona superior derecha con una microjeringuilla de calibre 26 (Hamilton, Reno, NV, USA). Se inyectaron grupos de 6 larvas por cada cepa que se estudió y

se dejaron crecer a 37 °C durante el proceso de infección (51 horas). Finalmente, se calculó el porcentaje de larvas muertas respecto al número total inicial de larvas infectadas.

3.12 Análisis bioinformáticos

3.12.1 Comparación genómica entre cepas con *hha2/hha3* y cepas que no presentan estos reguladores

Se ha realizado un análisis genómico comparativo entre cepas que codifican en su genoma los nuevos parálogos de *hha* (*hha2* y *hha3*) encontrados en la cepa EAEC 042 frente a cepas que no los contienen. Este análisis se ha llevado a cabo haciendo uso de la herramienta MaGe Pan/Core, la cual se encuentra en la página web <http://www.genoscope.cns.fr/agc/microscope/compgenomics/pancoreTool.php>. La agrupación de los genes en los conjuntos *core-genome* y genes variables se ha realizado mediante el algoritmo *MicroScope gene families* (MICFAM), integrado en el software SiLiX, el cual es una plataforma que integra herramientas para la anotación, comparación y visualización de genomas procariontes³⁵⁰. Este algoritmo agrupa los genes que presentan un porcentaje de identidad y de longitud de alineamiento por encima del umbral establecido en familias de genes. Las familias génicas compartidas por todas las cepas que se analizan se clasifican como *core-genome*, mientras que las familias compartidas solo entre algunas cepas o presentes en una única cepa son clasificadas como genes variables. El pangenoma resulta de la suma de ambos conjuntos de genes. En los análisis realizados, los parámetros que se han utilizado para la clasificación de los genes de las cepas seleccionadas han sido un 80 % de identidad y un 80 % de longitud de alineamiento.

3.12.2 Identificación de genes duplicados

La identificación de genes duplicados en el genoma de una cepa de *E. coli* se ha realizado en colaboración con el grupo de investigación *Evolutionary Genomics and Bioinformatics* del Dr. Julio Rozas, en la Universidad de Barcelona. Este análisis se ha llevado a cabo, principalmente, mediante la herramienta BLASTp (*Basic Local Alignment Search Tool for Proteins*)³⁵¹. Se han comparado todas las proteínas de una cepa frente a sí mismas, filtrando los resultados con unos parámetros restrictivos, como son un 85 % de similitud, un 85 % de longitud de alineamiento y un e-value menor de 10^{-10} .

Una vez que se ha llevado a cabo este análisis, se han marcado las coordenadas que ocupa cada duplicado dentro del genoma de la cepa estudiada. La recopilación de toda la información disponible sobre los genes duplicados se ha obtenido de la base de datos *GenBank* (descargando aquellos archivos con la extensión .gff). Todo este análisis se automatizó mediante la programación de un algoritmo, mientras que la representación de los genes duplicados dentro del genoma se realizó con R.

Los algoritmos desarrollados empleados para el análisis de las duplicaciones se encuentran disponibles en el sitio web github, en el enlace <https://github.com/molevol-ub/BacterialDuplicates>.

3.12.3 Distribución de genes duplicados en diferentes patotipos

El análisis de genes duplicados entre cepas de diferentes patotipos, mediante la identificación de su presencia o ausencia en los genomas seleccionados, se ha llevado a cabo mediante una modificación del algoritmo previamente desarrollado. Se ha realizado una búsqueda de los genes duplicados en una cepa, siguiendo la metodología del apartado anterior. A

continuación, estos duplicados se han buscado en todas las secuencias codificantes de otros 28 genomas seleccionados, representativos de todos los patotipos (Tabla 3.7).

Tabla 3.7. Cepas utilizadas en los análisis bioinformáticos

Patotipo	Cepa	Accession number	Assembly Accession
Comensal	<i>E. coli</i> K-12 MG1655	AJGD00000000.1	GCF_000482265.1
EAEC	<i>E. coli</i> 042	NC_017626.1	GCF_000027125.1
	<i>E. coli</i> O104:H4 2011C-3493	NC_018658.1	GCF_000299455.1
	<i>E. coli</i> O104:H4 LB226692	AFOB00000000.2	GCA_000215685.3
	<i>E. coli</i> 55989	NC_011748.1	GCF_000026245.1
EHEC	<i>E. coli</i> O26:H11 11368	NC_013361.1	GCF_000091005.1
	<i>E. coli</i> O157:H7 Sakai	NC_002695.1	GCF_000008865.1
	<i>E. coli</i> O111:H- 11128	AP010960.1	GCF_000010765.1
	<i>E. coli</i> O145:H28 RM13516	CP006262.1	GCF_000520055.1
ETEC	<i>E. coli</i> UMN18	AGTD00000000.1	GCF_000220005.1
	<i>E. coli</i> O139:H28 E24377A	JXRF00000000.1	GCF_000017745.1
	<i>E. coli</i> O103:H2 2011C-3750	JHLL00000000.1	GCF_000616345.2
	<i>E. coli</i> H10407	FN649414.1	GCF_000210475.1
EPEC	<i>E. coli</i> O127:H6 E2348-69	NC_011601.1	GCF_000026545.1
	<i>E. coli</i> O55:H7 CB9615	CP001846.1	GCF_000025165.1
EIEC	<i>E. coli</i> O96:H19	JHNY01000124.1	GCF_001007915.1
	<i>E. coli</i> O143:H26 4608-58	JTCO01000000	GCF_000805835.1
	<i>E. coli</i> O28ac:NM 02-3404	JHNY00000000	GCF_000617165.2
	<i>E. coli</i> O124:H30 M4163	JTCN01000000	GCF_000805815.1
	<i>E. coli</i> 53638	AAKB00000000.2	GCF_000167915.2
ST131	<i>E. coli</i> JJ1886	CP006784.1	GCF_000493755.1
	<i>E. coli</i> O25b:H4 EC958	HG941718.1	GCF_000285655.3
	<i>E. coli</i> NA114	MIPU00000000.1	GCF_000214765.2

Patotipo	Cepa	Accession number	Assembly Accession
	<i>E. coli</i> 536	CP000247.1	GCF_000013305.1
	<i>E. coli</i> UMN026	CU928163.2	GCF_000026325.1
UPEC	<i>E. coli</i> CFT073	AE014075.1	GCF_000007445.1
	<i>E. coli</i> UTI89	CP000243.1	GCF_000013265.1
	<i>E. coli</i> IAI39	NC_011750.1	GCF_000026345.1

3.12.4 Predicción de operones

La predicción de operones *in silico* se ha realizado usando la herramienta en línea FGENESB (Softberry, Inc., Mount Kisco, NY) (<http://www.softberry.com/>).

3.12.5 Predicción de promotores

La predicción de promotores *in silico* se ha llevado a cabo gracias a la herramienta web BPROM (Softberry, Inc., Mount Kisco, NY) (<http://www.softberry.com/>).

3.13 Análisis estadístico

Los análisis estadísticos realizados para la comparación de grupos de cepas que codifican los reguladores *hha2/hha3* frente a las cepas que no los codifican se han llevado a cabo mediante el test de Fisher. La prueba exacta de Fisher es una prueba de significación estadística, indicada para el análisis de tablas de contingencia, donde se quiere estudiar si existe asociación entre dos variables cualitativas. Los resultados se han considerado significativos cuando el p-valor obtenido ha sido menor de 0,05.

En el caso de los ensayos de beta-galactosidasa y de formación de biofilm, el test estadístico utilizado ha sido la distribución t (de Student). Este test estadístico se ha usado para determinar si había una diferencia significativa entre las medias de dos grupos. Los resultados se han considerado significativos cuando el p-valor obtenido ha sido menor de 0,05.

Por último, en el caso de los ensayos de supervivencia en *Galleria mellonella*, las poblaciones se han comparado mediante la prueba de Mantel-Cox (logrank-test). Se trata de una técnica estadística usada para contrastar las funciones de supervivencia de dos poblaciones. Los resultados se han considerado significativos cuando el p-valor obtenido ha sido menor de 0,05.

Los test estadísticos aplicados se llevaron a cabo gracias al programa GraphPad Prism, obteniendo los valores de significancia.

4. Resultados

4.1 Duplicaciones génicas en *E. coli* y su regulación por el sistema H-NS/Hha

4.1.1 Relación de la presencia de los parálogos *hha2/hha3* con otros genes en *Escherichia coli*

Tal y como se ha referido en el apartado de Objetivos, la primera parte de este trabajo tuvo como objetivo intentar correlacionar la presencia de los parálogos *hha2* y *hha3* con la presencia de otros genes que hipotéticamente fuesen la diana de las proteínas Hha2 y Hha3. Para ello, se realizó un análisis genómico comparativo entre dos grupos de cepas de *E. coli*, comparando aquellas que codifican estos nuevos parálogos *hha2/hha3* frente a cepas que no los tienen. Para llevarlo a cabo, se seleccionaron tanto 5 cepas de *E. coli* que codifican en su genoma los genes *hha2*, *hha3* o ambos, como otras 5 cepas que no contienen ninguno de los dos genes, perteneciendo en ambos casos a diversos patotipos (Tabla 4.1).

Tabla 4.1. Cepas seleccionadas para realizar el análisis comparativo entre genomas que contienen *hha2/hha3* frente a los que no los tienen.

Patotipo	Cepa	Presencia de <i>hha2/3</i>
EAEC	<i>E. coli</i> 042	<i>hha2/3</i> ⁺
	<i>E. coli</i> O104:H4 LB226692	<i>hha2/3</i> ⁺
ETEC	<i>E. coli</i> H10407	<i>hha3</i> ⁺
UPEC	<i>E. coli</i> NA114	<i>hha2/3</i> ⁺
	<i>E. coli</i> UMN026	<i>hha2</i> ⁺
EPEC	<i>E. coli</i> O127:H6 E2348-69	<i>hha2/3</i> ⁻
UPEC	<i>E. coli</i> 53638	<i>hha2/3</i> ⁻
	<i>E. coli</i> 1A139	<i>hha2/3</i> ⁻
EHEC	<i>E. coli</i> O111:H- 11128	<i>hha2/3</i> ⁻
	<i>E. coli</i> O157:H7 Sakai	<i>hha2/3</i> ⁻

En primer lugar, se identificó tanto el *core-genome* y los genes variables de ambos grupos de cepas, como el pangenoma, que resulta de la suma de los dos conjuntos de genes. El *core-genome* representa todos los genes que son comunes en todas las cepas analizadas, por lo que los parálogos de Hha pertenecen a este conjunto en las 5 cepas seleccionadas que los codifican. La identificación de estos grupos génicos se realizó gracias al algoritmo MICFAM (ver apartado 3.12.1, Materiales y métodos), donde se seleccionaron unos parámetros restrictivos como son un 80 % de identidad y un 80 % de longitud de alineamiento. Los resultados obtenidos para los dos grupos de cepas seleccionadas se pueden ver en la Tabla 4.2.

Tabla 4.2. Pangenoma, *core-genome* y genoma variable de los dos grupos de cepas seleccionadas.

Componente	<i>hha2/3 +</i>		<i>hha2/3 -</i>	
	Familias	Genes	Familias	Genes
Pangenoma	8151	26402	8289	26847
<i>Core-genome</i>	3421	18335	3205	17024
Genoma variable	4730	8067	5084	9823

Una vez establecido el *core-genome* y el genoma variable de estos 2 grupos de cepas, se procedió a comparar ambos conjuntos de cepas, con la finalidad de identificar genes asociados a la presencia de los reguladores *hha2* y *hha3*. El análisis se realizó gracias a la herramienta genómica disponible en línea *MaGe Pan/Core* (apartado 3.12.1, Material y métodos), siguiendo una estrategia muy restrictiva. Se excluyeron todos los genes del pangenoma de las cepas que no contienen *hha2/3* del *core-genome* del conjunto de cepas que sí que codifican estos nuevos reguladores. De esta manera, se obtuvieron los genes que están asociados exclusivamente a la presencia de los nuevos parálogos de *hha* (Tabla 4.3). Esta comparación se representó con un diagrama de Venn (Fig. 4.1).

reducido a 3. Al analizar estas 3 familias, se comprobó que una era la familia *hha2/3* y las otras dos corresponden a las familias de los genes *yeeR* y *aec69*.

Tal y como se ha comentado en el apartado correspondiente a la Introducción, *yeeR* y *aec69* se encuentran junto a los genes *flu*, *aec70* y *aec71*, los cuales forman parte del profago CP4-44. Esta región génica en la cepa de *E. coli* MG1655 tiene el gen *yeeR* truncado, y los genes *aec69*, *aec70* y *aec71* se han perdido (Fig. 4.2).

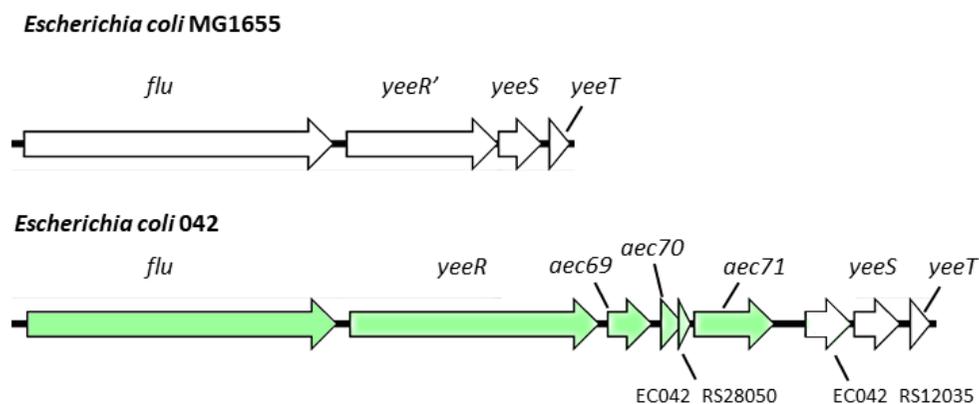


Fig. 4.2. Región del profago CP4-44 en la cepa comensal MG1655 y en la cepa EAEC 042 utilizada en el análisis genómico comparativo. En verde se muestran los genes *flu-yeeR-aec69-aec70-aec71*. El gen *yeeR* muestra un tamaño menor en la cepa MG1655 y los genes *aec69*, *aec70* y *aec71* se han perdido. Recientemente, se ha descrito que los genes *agn43*, *yeeR* y *aec69* forman una unidad transcripcional³²⁸. Figura realizada mediante EasyFig³³².

El resultado de este análisis es de gran importancia, ya que asoció la presencia de los genes *hha2* y *hha3* con dos genes más, los cuales se encuentran juntos en la misma región del genoma. Sin embargo, debido a la gran variabilidad y plasticidad característica de los genomas de *E. coli*, no se podía asegurar que en otras cepas que codificaran los reguladores *hha2/3* también estuvieran *yeeR* y *aec69*, ni tampoco se podía excluir la posibilidad de que estos

genes estuvieran presentes en el genoma de cepas que no tuvieran estos nuevos parálogos del gen *hha*. Para mejorar el análisis, se decidió realizar una búsqueda de las proteínas que codifican los genes *flu*, *yeeR*, *aec69*, *aec70* y *aec71* en 28 cepas seleccionadas, incluyendo una cepa comensal y cepas representativas de todos los patotipos (Tabla 3.6, Materiales y métodos). Entre todas estas cepas, 18 de ellas codifican alguno de los dos parálogos de la familia Hha, mientras que otras 10 cepas no presentan ninguno de los dos. Este análisis se llevó a cabo haciendo uso del algoritmo BLASTp, obteniendo los resultados presentados en la Tabla 4.4.

Tabla 4.4. Resultado de la búsqueda de las proteínas codificadas por los genes *flu*, *yeeR*, *aec69*, *aec70* y *aec71* mediante BLASTp en 28 cepas de *E. coli*. Se ha marcado con una "X" la presencia de estos genes en el genoma de las cepas.

Patotipo	Cepa/Gen	<i>flu</i>	<i>yeeR</i>	<i>aec69</i>	<i>aec70</i>	<i>aec71</i>	<i>hha2/3</i>
EAEC	<i>E. coli</i> 042	X	X	X	X	X	<i>hha2/3</i> ⁺
	<i>E. coli</i> O104:H4 2011C-3493	X	X	X	X	X	<i>hha2/3</i> ⁺
	<i>E. coli</i> O104:H4 LB226692		X	X	X	X	<i>hha2/3</i> ⁺
	<i>E. coli</i> 55989	X	X	X	X	X	<i>hha2/3</i> ⁺
EHEC	<i>E. coli</i> O26:H11 11368	X	X	X	X	X	<i>hha2/3</i> ⁺
	<i>E. coli</i> O145:H28 RM13516	X	X	X	X	X	<i>hha2</i> ⁺
ETEC	<i>E. coli</i> O139:H28 E24377A					X	<i>hha3</i> ⁺
	<i>E. coli</i> H10407		X	X	X	X	<i>hha3</i> ⁺
EPEC	<i>E. coli</i> O55:H7 CB9615					X	<i>hha2</i> ⁺
EIEC	<i>E. coli</i> O96:H19						<i>hha2</i> ⁺
	<i>E. coli</i> O143:H26 4608-58	X	X			X	<i>hha2/3</i> ⁺
	<i>E. coli</i> O28ac:NM 02-3404	X	X	X		X	<i>hha2/3</i> ⁺
	<i>E. coli</i> O124:H30 M4163	X	X	X	X	X	<i>hha2/3</i> ⁺
ST131	<i>E. coli</i> JJ1886	X		X	X	X	<i>hha2/3</i> ⁺
	<i>E. coli</i> NA114		X	X	X	X	<i>hha2/3</i> ⁺
UPEC	<i>E. coli</i> UMN026	X	X	X		X	<i>hha2</i> ⁺
	<i>E. coli</i> CFT073	X	X	X	X	X	<i>hha2/3</i> ⁺
	<i>E. coli</i> UTI89					X	<i>hha3</i> ⁺

Patotipo	Cepa/Gen	<i>flu</i>	<i>yeeR</i>	<i>aec69</i>	<i>aec70</i>	<i>aec71</i>	<i>hha2/3</i>
Comensal	<i>E. coli</i> K-12 MG1655	X					<i>hha2/3</i> ⁻
EHEC	<i>E. coli</i> O157:H7 Sakai					X	<i>hha2/3</i> ⁻
	<i>E. coli</i> O111:H- 11128					X	<i>hha2/3</i> ⁻
ETEC	<i>E. coli</i> UMNF18		X	X	X	X	<i>hha2/3</i> ⁻
	<i>E. coli</i> O103:H2 2011C-3750	X	X	X	X	X	<i>hha2/3</i> ⁻
EPEC	<i>E. coli</i> O127:H6 E2348-69					X	<i>hha2/3</i> ⁻
EIEC	<i>E. coli</i> 53638					X	<i>hha2/3</i> ⁻
ST131	<i>E. coli</i> EC958	X	X	X			<i>hha2/3</i> ⁻
UPEC	<i>E. coli</i> 536	X				X	<i>hha2/3</i> ⁻
	<i>E. coli</i> IAI39					X	<i>hha2/3</i> ⁻

A partir de estos resultados, se realizó un análisis estadístico para determinar si la relación entre la presencia de los parálogos de Hha y estos genes es significativa (apartado 3.13, Materiales y métodos). Como se puede ver, el 72 % de las cepas que contienen *hha2/hha3* también poseen los genes *yeeR* y *aec69*, con un *p-value* menor de 0,05, por lo que esta relación sí que es significativa. Además, el 61 % de estas cepas también contienen los genes *flu* y *aec70*. Sin embargo, la relación entre los reguladores y estos dos genes es menos significativa (*p-value*=0,055). Por otra parte, entre las cepas que no codifican estos reguladores, solo entre el 20-40 % poseen los genes *flu*, *yeeR*, *aec69* y *aec70*. El gen *aec71* no muestra ninguna relación con la presencia de los parálogos *hha2/3*, ya que se encuentra distribuido en casi todas las cepas.

Tras determinar la relación existente entre *hha2/3* y *yeeR-aec69*, se procedió a mapear la posición de los mismos en los genomas de 8 cepas de *E. coli* diferentes, incluyendo tanto cepas patógenas intestinales como extraintestinales. Este análisis se realizó mediante BLASTp, representando las posiciones que ocupan en el genoma cada uno de los genes que se buscaron (Fig. 4.3).

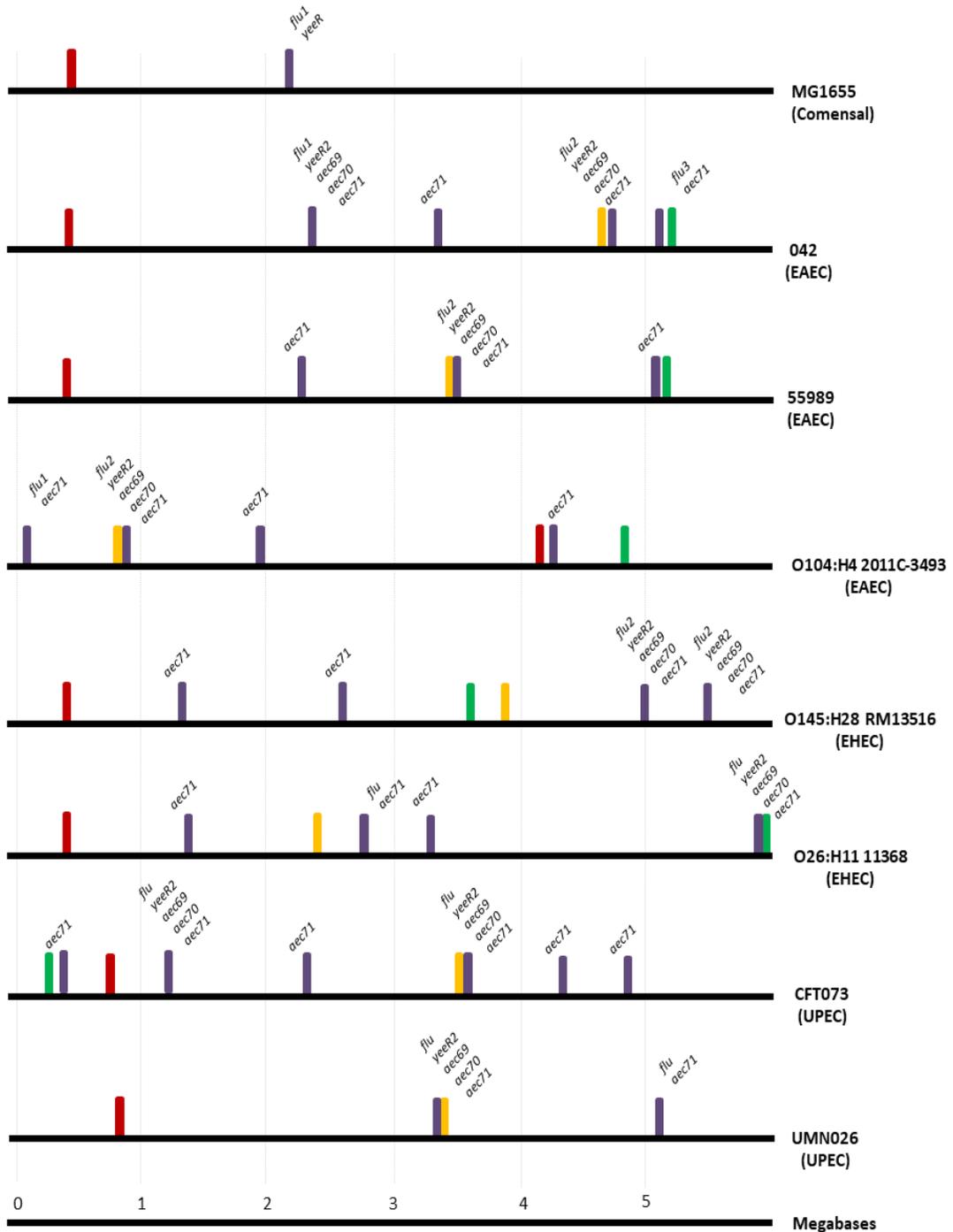


Fig. 4.3. Posición de los genes *hha* (barra roja), *hha2* (barra naranja), *hha3* (barra verde) y el conjunto de genes *flu-yeeR-aec69-aec70-aec71* (barras violetas) en los cromosomas de la cepa comensal MG1655 y 7 cepas más de diferentes patotipos indicados. Las líneas discontinuas verticales corresponden a megabases en la escala utilizada.

Observando estos resultados, se puede afirmar que la posición de los genes *hha2/3* se encuentra, en muchas ocasiones, junto al conjunto de genes a los que se ha visto que están asociados. Este estudio muestra también que, en cepas patógenas, incluyendo la cepa 042, los genes *flu*, *yeeR*, *aec69*, *aec70* y *aec71* se encuentran en dos o más copias dentro del genoma de cada cepa. En el genoma de la cepa 042 se habían encontrado previamente 3 miembros de la familia de reguladores *hns* (*hns*, *stpA* y *hns2*)²⁰⁴, 4 miembros de la familia de reguladores *hha* (*hha*, *ydgT*, *hha2* y *hha3*)²⁰⁵ y tres copias del gen *flu*¹⁹⁷. Con el análisis que aquí se ha realizado, se puede ver que el genoma de esta cepa también contiene los genes *yeeR* y *aec69* duplicados, y el gen *aec71* se encuentra hasta en 4 copias. Teniendo en cuenta todos estos resultados, y después de asociar la presencia de los parálogos *hha2/hha3* al clúster *yeeR aec69*, pareció importante evaluar el impacto de las duplicaciones génicas dentro del genoma de la propia cepa 042, independientemente de que estuvieran asociados o no a la presencia de los reguladores *hha2/3*.

4.1.2 Duplicaciones génicas en la cepa 042 de *E. coli*

El análisis necesario para la identificación de todos los genes duplicados dentro del genoma de la cepa EAEC 042 se llevó a cabo mediante el algoritmo BLASTp. Se realizó una comparación de todas las proteínas frente a sí mismas y, para poder clasificar aquellas proteínas que presentan duplicaciones dentro del genoma, se aplicaron unos parámetros restrictivos como son un 85 % de similitud y un 85 % de longitud de alineamiento (ver 3.12.2, Materiales y métodos). Tras obtener los resultados, se representó el genoma de manera lineal y se marcaron las posiciones en las que se encuentra cada proteína duplicada dentro del genoma (Fig. 4.4).

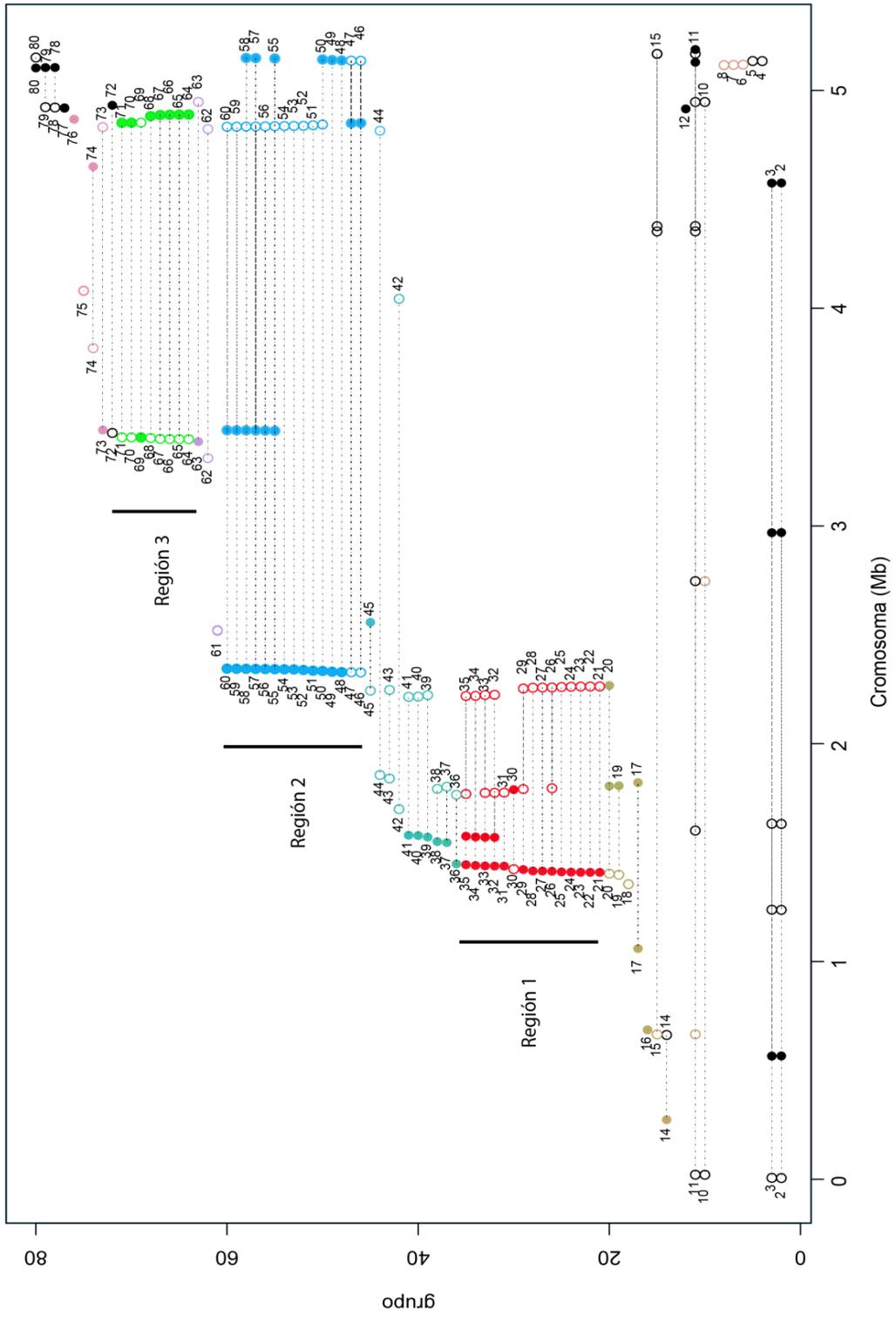


Fig. 4.4. Genes duplicados en el genoma de la cepa EAEC 042. El eje de las X representa el mapa lineal del cromosoma. Cada grupo de círculos unidos con una línea horizontal discontinua corresponde a un gen duplicado o amplificado. Los círculos indican la posición que ocupa cada duplicado en el genoma. Los círculos enteros corresponden a los genes que se encuentran en la cadena positiva, mientras que los círculos vacíos corresponden a los de la cadena negativa. Los números corresponden a los grupos de genes duplicados, numerados en orden empezando desde el origen del mapa cromosómico. Los grupos 2-5, 10-12 y 77-80 están coloreados en negro y corresponden a transposasas. Los colores (rojo, azul y verde) y las barras verticales definen las tres regiones principales que contienen genes duplicados. Los grupos 1, 9 y 13 no se muestran debido a que se encuentran en el plásmido pAA que tiene la cepa. Los grupos 4-8 tienen una copia en el cromosoma y otra en el plásmido.

Tras realizar este análisis, filtrando los resultados con los parámetros indicados, un total de 80 genes aparecen duplicados o amplificados (presencia de más de dos copias) (GDA) en el genoma de la cepa 042. Algunos de estos genes codifican para transposasas, las cuales se han marcado de color negro (todo el círculo o el contorno, según la cadena donde se encuentren) en la figura. La mayoría de los genes GDA se encuentran en tres regiones principales, las cuales se han indicado con barras verticales y numeradas del 1 al 3 de manera arbitraria. Toda la información disponible sobre estos genes duplicados se recopiló en la Tabla 4.5. Este mismo análisis, realizado con la secuencia aminoacídica de todas las proteínas de la cepa 042, también se llevó a cabo utilizando todas las secuencias de nucleótidos, con el algoritmo BLASTn. Los resultados obtenidos con este análisis son muy semejantes a los obtenidos con BLASTp (figura anexa A1), por lo que se decidió seguir con el primer análisis realizado con las proteínas, evitando así posibles sesgos debidos a la degeneración característica del código genético.

Tabla 4.5. Localizador de cada una de las copias de los genes duplicados o amplificados (Locus tag) en la cepa EAEC 042 y descripción de su función. A) Genes de la región 1: 21-35. B) Genes de la región 2: 46-60. C) Genes de la región 3: 64-71

A	Grupo	Locus tag 1	Locus tag 2	Locus tag 3	Locus tag 4	Descripción
Región 1: 21-35	21	EC042_1328	EC042_2193			Putative phage protein
	23	EC042_1330	EC042_2191			Putative phage protein
	24	EC042_1333	EC042_2189			Phage protein
	25	EC042_1336	EC042_1705	EC042_2186		Putative host cell killing modulation protein
	26	EC042_1342	EC042_2183			Phage protein
	27	EC042_1343	EC042_2182			Putative phage endodeoxyribonuclease
	28	EC042_1344	EC042_2181			Phage antitermination protein
	29	EC042_1349	EC042_1703E	EC042_2175		Putative phage lysozyme
	30	EC042_1353	EC042_1702			Putative phage protein
	31	EC042_1371	EC042_1685			Phage minor tail protein
	32	EC042_1372	EC042_1509	EC042_1684	EC042_2138	Phage minor tail protein
	33	EC042_1373	EC042_1510	EC042_1683	EC042_2137	Phage tail assembly protein
	34	EC042_1376	EC042_1512	EC042_2135		Phage host specificity protein
	35	EC042_1377	EC042_1513	EC042_1680	EC042_2134	Putative prophage encoded outer membrane protein

B	Grupo	Locus tag 1	Locus tag 2	Locus tag 3	Locus tag 4	Descripción
Región 2: 46-60	46	EC042_2236A	EC042_4519	EC042_4793		Conserved hypothetical protein
	47	EC042_2237	EC042_4518	EC042_4794		Conserved hypothetical protein
	48	EC042_2238	EC042_4795			Hypothetical protein
	49	EC042_2239	EC042_4798			Conserved hypothetical protein
	50	EC042_2241	EC042_4512	EC042_4802		Putative GTP binding protein
	51	EC042_2242	EC042_4511			Antigen 43 precursor (autotransporter)
	52	EC042_2243	EC042_4510			YeeR
	53	EC042_2244	EC042_4509			Aec69
	54	EC042_2244A	EC042_4508A			Aec70
	55	EC042_2245	EC042_3221	EC042_4507	EC042_4805	Aec71
	56	EC042_2246	EC042_3222	EC042_4506		Putative antirestriction protein
	57	EC042_2247	EC042_3223	EC042_4505	EC042_4807	Putative DNA repair protein
	58	EC042_2247A	EC042_3224	EC042_4504	EC042_4808	Conserved hypothetical protein
	59	EC042_2248	EC042_3225	EC042_4503		Conserved hypothetical protein
	60	EC042_2249	EC042_3226	EC042_4502		Conserved hypothetical protein

C	Grupo	Locus tag 1	Locus tag 2	Locus tag 3	Locus tag 4	Descripción
Región 3: 64-71	64	EC042_3180	EC042_4556			Conserved hypothetical protein
	65	EC042_3181	EC042_4555			Putative transcriptional regulator
	66	EC042_3182	EC042_4554			ParB like nuclease
	67	EC042_3183	EC042_4553			Conserved hypothetical protein
	68	EC042_3187	EC042_4548			Putative helicase
	69	EC042_3189	EC042_4523			Phage protein
	70	EC042_3190	EC042_4522			Conserved hypothetical protein
	71	EC042_3191	EC042_4521			Putative DNA binding protein

Las 3 regiones principales, donde se encuentran la mayoría de genes duplicados de esta cepa, se pueden visualizar con más detalle en la figura 4.5. Para mejorar este análisis, se llevó a cabo una predicción de unidades transcripcionales con los genes de cada región.

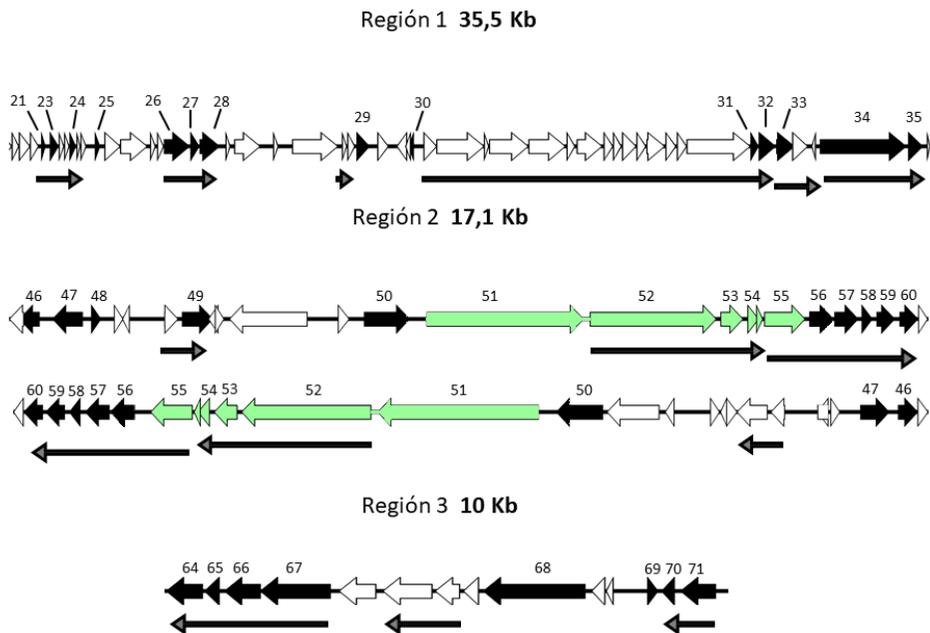


Fig. 4.5. Detalle de las 3 regiones principales que contienen genes duplicados en la cepa 042. Los genes duplicados dentro de estas regiones se muestran en negro y con su número (Tabla 4.5), mientras que los genes que no se han duplicado están en blanco. Los genes en verde corresponden a la región *flu-yeeR-aec69-aec70-aec71*. Las flechas en gris muestran la predicción *in silico* de operones realizada con el programa FGENESB (Softberry, Inc., Mount Kisco, NY). Se muestran las dos copias de la región 2 para apreciar un fenómeno de inversión.

A partir de toda la información recogida en la Tabla 4.5 y en la figura 4.5, se observó que tanto la longitud de las regiones como la función de los genes que se encuentran en las mismas eran diferentes en cada caso. La región 1 tiene una longitud aproximada de 35,5 Kb y los genes que se encuentran en la misma corresponden a genes fágicos. En el caso de la región 2, su longitud aproximada es de 17,1 kb y contiene los genes *flu*, *yeeR*, *aec69*, *aec70* y *aec71*. Esta región también incluye a los genes que codifican para un sistema toxina-antitoxina y otras proteínas de función desconocida. Cabe señalar que en una de las copias de esta región se encuentra el regulador *hha2*, lo cual concuerda con el análisis

que habíamos realizado previamente, representado en la figura 4.3. Por último, la región 3 tiene una longitud aproximada de 10 kb. Los productos génicos de los genes duplicados que mapean dicha región son de función desconocida. Los resultados obtenidos también sugieren que el origen de estas duplicaciones puede deberse a diferentes procesos. En el caso de las dos copias de la región 2, estas se encuentran invertidas en el genoma de la cepa 042. Este hecho sugiere que la duplicación de toda esta región puede estar relacionada con un reordenamiento genético, el cual ha afectado a toda la región como una única unidad. En el caso de la región 1 y 3, no ha tenido lugar un proceso de este tipo, aparentemente. Sin embargo, en el caso de la región 1, todas las proteínas provienen de un fago, por lo que el origen de estas duplicaciones puede estar relacionado con procesos HGT.

Como ya se ha mencionado, en la región 2 se encuentran los genes *flu*, *yeeR*, *aec69*, *aec70* y *aec71*. Considerando que se encuentran asociados a la presencia de los reguladores *hha2/hha3*, y que el gen *hha2* se encuentra en esta región, es posible que el sistema H-NS/Hha se encargue de la regulación de alguno de estos genes duplicados.

4.1.3 Papel de las proteínas H-NS y Hha en la regulación de genes duplicados en la cepa 042

El estudio de la regulación de los genes que se encuentran duplicados en el genoma de la cepa 042 por el sistema H-NS/Hha se llevó a cabo analizando los perfiles transcriptómicos, previamente publicados en nuestro grupo de investigación²⁰⁵, que presentan todos los genes de la cepa 042, tanto en una cepa salvaje como en los mutantes *hhanull* (mutante para los reguladores *hha* y *hha2*) y *hns*. El análisis se llevó a cabo en cultivos crecidos en medio rico LB a una temperatura de 37 °C. A partir de este análisis transcriptómico, se

obtuvieron los datos de expresión comparativa entre la cepa salvaje y mutantes *hhanull/hns* de los genes duplicados de las 3 regiones, con el fin de ver una posible regulación por el sistema H-NS/Hha (Tabla 4.6). Para eliminar la función de la proteína Hha se había suprimido en la cepa 042 solamente la función de los genes *hha* y *hha2*, ya que en esta cepa no se observó expresión del gen *hha3*.

Tabla 4.6. Expresión comparativa de los genes duplicados de las 3 regiones de la cepa 042 en el mutante *hhanull* (mutante para *hha* y *hha2*) y en el mutante *hns*. Los valores indican la diferencia de expresión de estos genes en los mutantes respecto a su expresión en la cepa salvaje (*fold-change*). A) Genes de la región 1: 21-35. B) Genes de la región 2: 46-60. C) Genes de la región 3: 64-71

A	Grupo	Locus tag	<i>hhanull</i>	<i>hns</i>
Región 1: 21-35	21	EC042_1328	2,1	6
	23	EC042_1330	2,6	5
	24	EC042_1333	2,5	4,2
	25	EC042_1336	1,8	0,6
	26	EC042_1342	3,5	4,8
	27	EC042_1343	2,8	4,8
	28	EC042_1344	1,7	4
	29	EC042_1349	2,4	3,8
	30	EC042_1353	4,3	3,3
	31	EC042_1371	1,5	6,1
	32	EC042_1372	1,5	3,8
	33	EC042_1373	1,6	3,3
	34	EC042_1376	2	3,7
	35	EC042_1377	2,4	4,2

B	Grupo	Locus tag	<i>hhanull</i>	<i>hns</i>
Región 2: 45-60	46	EC042_2236A	2,9	2,9
	47	EC042_2237	3,8	4,5
	48	EC042_2238	5,6	4,2
	49	EC042_2239	3,8	3
	50	EC042_2241	6,2	4,8
	51	EC042_2242	3,1	3,9
	52	EC042_2243	3	3,8
	53	EC042_2244	2,2	2,2
	54	EC042_2244A	4,9	5,9
	55	EC042_2245	5	5,1
	56	EC042_2246	4,6	4,8
	57	EC042_2247	5	3,9
	58	EC042_2247A	5	4,7
	59	EC042_2248	5,1	4,5
	60	EC042_2249	4,9	2,6

C	Grupo	Locus tag	<i>hhanull</i>	<i>hns</i>
Región 3: 64-71	64	EC042_3180	3	4,4
	65	EC042_3181	3,4	2,4
	66	EC042_3182	2,1	2,2
	67	EC042_3183	1,9	1,3
	68	EC042_3187	3,8	1,1
	70	EC042_3190	0,6	0,3
	71	EC042_3191	1,9	1,6

Se consideraron significativos los valores de expresión comparativa superiores a 2, es decir, aquellos donde como mínimo la expresión en los mutantes se hubiera duplicado respecto a la cepa salvaje. Como se puede ver, los genes que se hallan en la región 2 muestran valores superiores a 2, tanto en

el mutante *hhanull* como en el mutante *hns*. En la región 1, también se puede apreciar un aumento de la expresión en ambos mutantes, por lo que estos reguladores están influyendo en la expresión de dichos genes. Sin embargo, en la región 3, solo algunos genes muestran grandes diferencias en ambos mutantes.

Teniendo en cuenta estos resultados, tanto los genes de la región 1 como de la región 2 muestran una correulación por el sistema H-NS/Hha, donde el parólogo Hha2 también puede estar desempeñando una función reguladora. El estudio que inicialmente tenía como objetivo estudiar la implicación de las proteínas Hha2/Hha3 en la regulación de los genes *yeeR* y *aec69* ha llevado finalmente a determinar que las proteínas Hha y Hha2, juntamente con H-NS, no solamente regulan dichos genes, sino la mayoría de genes duplicados en la cepa 042.

4.1.4 Distribución de las duplicaciones de la cepa EAEC 042 en otras cepas de *E. coli*

Una vez demostrada la relación entre el sistema Hha/H-NS y los genes duplicados de la cepa 042, nos planteamos la pregunta de si las duplicaciones observadas eran una característica concreta de esta cepa o si, por el contrario, las mismas se habían generado en un proceso que había ocurrido anteriormente en una cepa ancestral y dichas duplicaciones se encontraban distribuidas entre los distintos patotipos de *E. coli*. Para responder a esta pregunta, se utilizaron las 28 cepas patógenas representativas de los diferentes patotipos (incluida la cepa EAEC 042) seleccionadas previamente (Tabla 3.6, Materiales y métodos), donde se había buscado la presencia de los genes *flu*, *yeeR*, *aec69*, *aec70* y *aec71*. Se llevó a cabo una búsqueda en sus genomas de los genes duplicados de las 3 regiones principales de la cepa 042, utilizando el

algoritmo BLASTp (apartado 3.12.3, Materiales y métodos). Se aplicaron los mismos parámetros restrictivos que se usaron en el análisis anterior, es decir, un 85 % de similitud y un 85 % de longitud de alineamiento. Por lo tanto, los resultados muestran aquellos genes de cada cepa que presentan un 85 % de similitud y de longitud de alineamiento con el duplicado de la cepa 042 analizado. Para cada gen que se encuentra duplicado en la cepa 042, se anotó el número de copias que cumplen con estos parámetros dentro de los genomas de las 28 cepas analizadas (Fig. 4.6). Las cepas del patotipo UPEC se han separado en dos grupos: aquellas que no pertenecen a la secuencia tipo 131, que hemos denominado UPEC, y las que sí que pertenecen a la secuencia tipo 131, denominadas ST131. Esta diferenciación se ha realizado debido a la gran importancia que muestra este grupo clonal, ya que las cepas ST131 contienen un elevado número de resistencias a antibióticos y son un patógeno emergente.

Resultados

		Región 1: 21-35														Región 2: 46-60														Región 3: 64-71													
Patotipo	Cepa	21	23	24	25	26	27	28	29	30	31	32	33	34	35	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	64	65	66	67	68	69	70	71					
Comensal	<i>E. coli</i> K-12 MG1655	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0					
	<i>E. coli</i> 042	2	2	2	2	3	2	2	3	2	2	4	4	3	4	3	3	2	2	3	2	2	2	2	2	4	3	4	4	3	3	2	2	2	2	2	2	2	2				
	<i>E. coli</i> O104:H4 2011C-3493	0	0	1	0	1	1	1	4	1	3	3	3	3	3	4	4	1	3	3	1	1	1	1	1	4	5	5	5	5	3	1	1	0	1	1	1	1	2	2			
EAEC	<i>E. coli</i> O104:H4 LB226692	0	0	1	1	0	1	1	2	1	3	2	1	1	1	3	3	1	2	2	0	1	1	1	2	4	3	5	0	3	1	1	0	1	1	0	2	2					
	<i>E. coli</i> 55989	1	0	1	1	3	3	2	3	1	2	3	4	2	3	3	3	1	2	1	1	1	1	1	3	4	4	4	4	3	1	1	1	1	1	1	1	1					
	<i>E. coli</i> O26:H11 11368	1	0	1	1	7	6	2	7	0	2	3	9	1	2	2	2	0	1	4	2	1	2	2	4	4	4	4	4	3	1	1	1	0	0	1	1	1					
EHEC	<i>E. coli</i> O157:H7 Sakai	0	0	1	0	6	4	1	9	0	1	2	4	1	4	0	0	0	0	0	0	0	0	0	2	2	2	2	2	2	1	1	1	0	1	1	1	1					
	<i>E. coli</i> O111:H- 11128	0	0	1	1	3	5	1	7	0	0	0	8	0	1	0	0	0	0	0	0	0	0	0	0	3	3	3	2	1	1	1	1	0	0	1	1	1					
	<i>E. coli</i> O145:H28 RM13516	0	0	0	1	4	5	1	9	0	0	1	7	1	3	4	4	4	2	3	2	2	2	1	4	4	4	4	4	2	0	0	0	0	0	1	1	1					
ETEC	<i>E. coli</i> UMN#28	1	0	1	2	2	3	0	0	0	0	0	0	1	0	2	2	0	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
	<i>E. coli</i> O139:H28 E24377A	0	0	0	0	2	2	0	0	1	0	0	0	0	0	2	1	1	2	2	0	0	0	0	3	3	3	3	3	3	0	0	0	0	0	0	0	0					
	<i>E. coli</i> O103:H2 2011C-3750	0	0	0	0	1	1	1	2	0	0	1	4	0	0	1	0	0	2	0	1	1	1	1	2	4	1	5	2	3	1	1	1	0	0	1	1	1					
EPEC	<i>E. coli</i> H10407	1	0	0	1	0	1	0	0	0	2	3	3	2	3	3	0	3	3	0	2	2	2	4	4	4	4	4	4	0	0	0	0	0	1	1	1						
	<i>E. coli</i> O127:H6 E2348-69	0	0	0	1	1	2	0	1	0	1	2	3	0	1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0						
	<i>E. coli</i> O55:H7 CB9615	0	0	1	1	2	3	1	5	0	4	5	8	2	2	1	0	1	2	0	0	0	0	0	2	2	2	2	2	1	0	0	1	0	0	0	0	0					
EIEC	<i>E. coli</i> O96:H19	2	0	0	0	0	1	0	3	1	1	3	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0						
	<i>E. coli</i> O143:H26 4608-58	0	0	0	0	1	2	1	2	0	0	2	2	1	2	2	2	1	1	0	1	0	0	2	2	2	2	2	2	0	0	1	1	0	0	0	0						
	<i>E. coli</i> O28ac:NM 02-3404	1	0	0	0	1	2	1	2	1	1	2	0	1	0	1	1	0	0	0	1	1	0	2	2	1	2	2	0	0	0	1	1	0	1	1	1						
	<i>E. coli</i> O124:H30 IM4163	1	0	0	0	1	1	0	2	0	0	2	2	1	1	2	2	2	2	1	1	1	1	1	2	2	2	2	2	1	0	0	1	1	0	0	0	0					
ST131	<i>E. coli</i> 53638	3	0	0	0	0	1	0	3	0	0	2	2	3	1	1	1	1	1	0	0	0	0	1	1	1	1	1	0	0	0	1	1	0	0	0	0						
	<i>E. coli</i> J11886	1	0	0	1	1	2	2	3	0	1	3	3	2	2	4	4	1	3	3	1	0	1	1	4	4	4	4	4	4	0	0	1	1	0	0	0	0					
	<i>E. coli</i> O25b:H4 EG58	1	0	0	1	1	2	2	0	1	2	3	2	1	1	4	4	1	3	3	1	1	0	4	4	4	4	4	4	0	0	1	1	0	0	0	0						
UPEC	<i>E. coli</i> NA114	1	0	0	1	1	2	2	0	1	2	3	2	1	3	3	1	2	0	0	1	1	1	4	3	1	3	4	3	0	0	1	1	0	0	1	1						
	<i>E. coli</i> 536	0	0	0	0	0	0	0	0	0	1	1	1	0	2	0	1	2	1	0	0	0	0	5	5	5	5	5	5	0	0	2	2	0	0	1	1						
	<i>E. coli</i> UMN026	0	0	0	0	2	1	0	2	1	2	3	4	3	1	2	2	0	1	2	2	1	0	2	0	2	2	1	1	0	0	0	0	0	0	0	0						
UPEC	<i>E. coli</i> CFT073	1	0	0	1	1	1	1	1	2	1	2	3	2	2	3	2	0	1	2	2	2	2	1	6	6	6	5	5	1	1	2	2	0	1	1	0						
	<i>E. coli</i> UT189	0	0	1	0	1	1	1	1	1	1	2	3	3	2	1	1	1	1	1	0	0	0	0	3	3	3	3	3	0	0	1	1	0	0	1	1						
	<i>E. coli</i> IA139	0	0	0	0	0	0	0	0	1	0	1	2	4	1	4	0	0	0	0	1	0	0	0	0	3	3	3	3	2	0	0	0	1	0	0	0	0					

Fig. 4.6. Distribución de los genes duplicados de las 3 regiones de la cepa 042 en 28 cepas seleccionadas de diferentes patotipos. En blanco se muestran los genes que no tienen ninguna copia que cumpla con los parámetros que se analizan. En gris, los genes que presentan solo una copia. En negro, los genes que presentan dos o más copias que cumplen con los parámetros. Los genes 51-54, señalados en verde, corresponden a los genes *flu-yeeR-aec69-aec70*. Los números muestran cuántas copias de cada gen de cada cepa cumplen con los parámetros establecidos.

Considerando los resultados obtenidos tras este análisis, se puede apreciar una distribución diferente de los genes duplicados de la cepa 042 entre los patotipos estudiados. En primer lugar, entre los duplicados 21 al 35, pertenecientes a la región 1 de la cepa 042, se puede ver cómo 9 de ellos (del 26 al 29 y del 31 al 35) se encuentran duplicados en muchas de las cepas analizadas. Es interesante constatar que, en algunos casos, se puede hablar de amplificación más que de duplicación, debido a que se pueden encontrar genes con más de 3 copias, llegando incluso hasta 9. Respecto a los 15 genes que engloba la región 2, llama la atención la distribución que se puede observar de los mismos en los diferentes patotipos. La mayoría de estos genes se encuentran duplicados o amplificados en muchas de las cepas analizadas. En concreto, los genes 55 al 60, que aparecen como una única unidad transcripcional (Fig. 4.5), se encuentran en la mayoría de los genomas analizados, en 4 o hasta 6 copias cada uno. Ninguno de estos 6 genes tiene una función definida y bien estudiada, ya que los genes 55 y 58-60 codifican proteínas hipotéticas conservadas, el 56 codifica una proteína antirrestricción putativa y la 57 una proteína de reparación de ADN putativa. Por último, la duplicación de los genes de la región 3 parece específica de la cepa 042, aunque algunos de los genes se encuentran duplicados en otras cepas del mismo patotipo o en cepas UPEC.

Es importante señalar la similitud entre los patrones de duplicación de las cepas EAEC y las cepas UPEC. Esta similitud representa una muestra más de la

relación evolutiva que presentan estos dos patotipos, como se ha resaltado en las cepas híbridas con características de ambos patotipos que se han encontrado en otros estudios^{16,192,193}. En cuanto a la cepa comensal MG1655 analizada, solo contiene una copia de 9 de los 37 genes duplicados en la cepa 042 que se han analizado.

Por lo tanto, como respuesta a la hipótesis inicial, algunas de las duplicaciones han ocurrido de manera específica la cepa 042, como por ejemplo la región 3. Sin embargo, otros genes, como es el caso del 55 al 60, presentan un gran número de duplicaciones y amplificaciones distribuidas en todos los patotipos, por lo que su origen pudo tener lugar en un ancestro común.

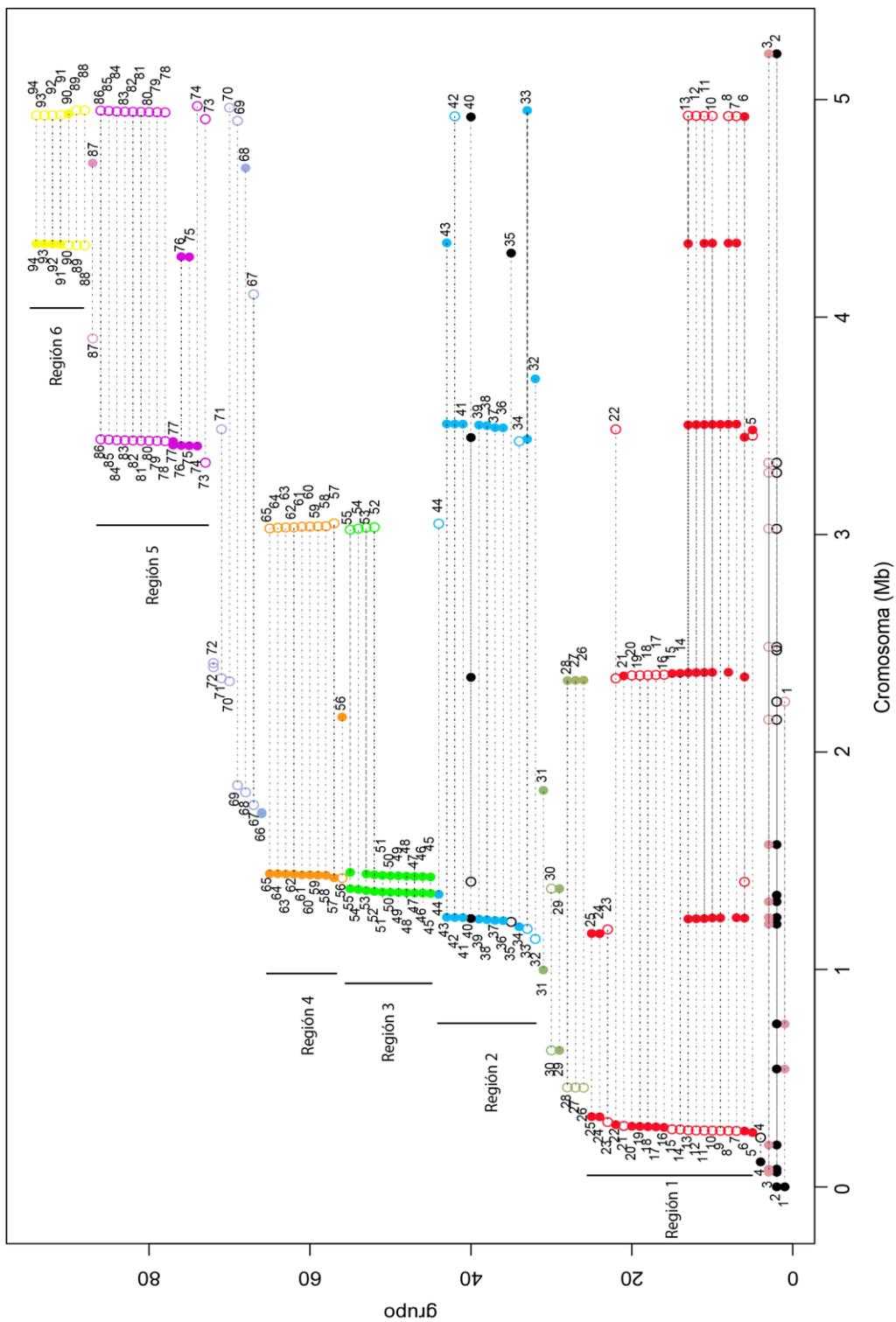
4.1.5 Presencia y distribución de las duplicaciones de las cepas UPEC CFT073 y EHEC O145:H28 RM13516

Los datos obtenidos a partir de los análisis realizados en la cepa 042 fueron de gran interés, tanto por la información que arrojaron sobre la presencia de duplicaciones y su posible regulación como por su distribución y relación entre los diferentes patotipos. Por ello, se decidió ampliar este estudio, con el objetivo de obtener un esquema más completo de la presencia de duplicaciones en *E. coli*. Para llevarlo a cabo, se analizaron las duplicaciones existentes en los genomas de dos cepas más de patotipos diferentes: la cepa UPEC CFT073 y la cepa EHEC O145:H28 RM13516.

Se realizó la búsqueda de los genes duplicados que contiene cada una de estas cepas dentro de sus genomas, siguiendo la misma estrategia utilizada en la cepa EAEC 042. Todas las proteínas codificadas en sus genomas fueron comparadas consigo mismas haciendo uso de BLASTp, filtrando los resultados de manera restrictiva aplicando los mismos parámetros descritos

anteriormente. Seguidamente, se hizo una representación lineal del cromosoma de cada una de las cepas, y se marcaron las coordenadas que ocupan los genes duplicados obtenidos a partir del análisis (Fig. 4.7).

A



B

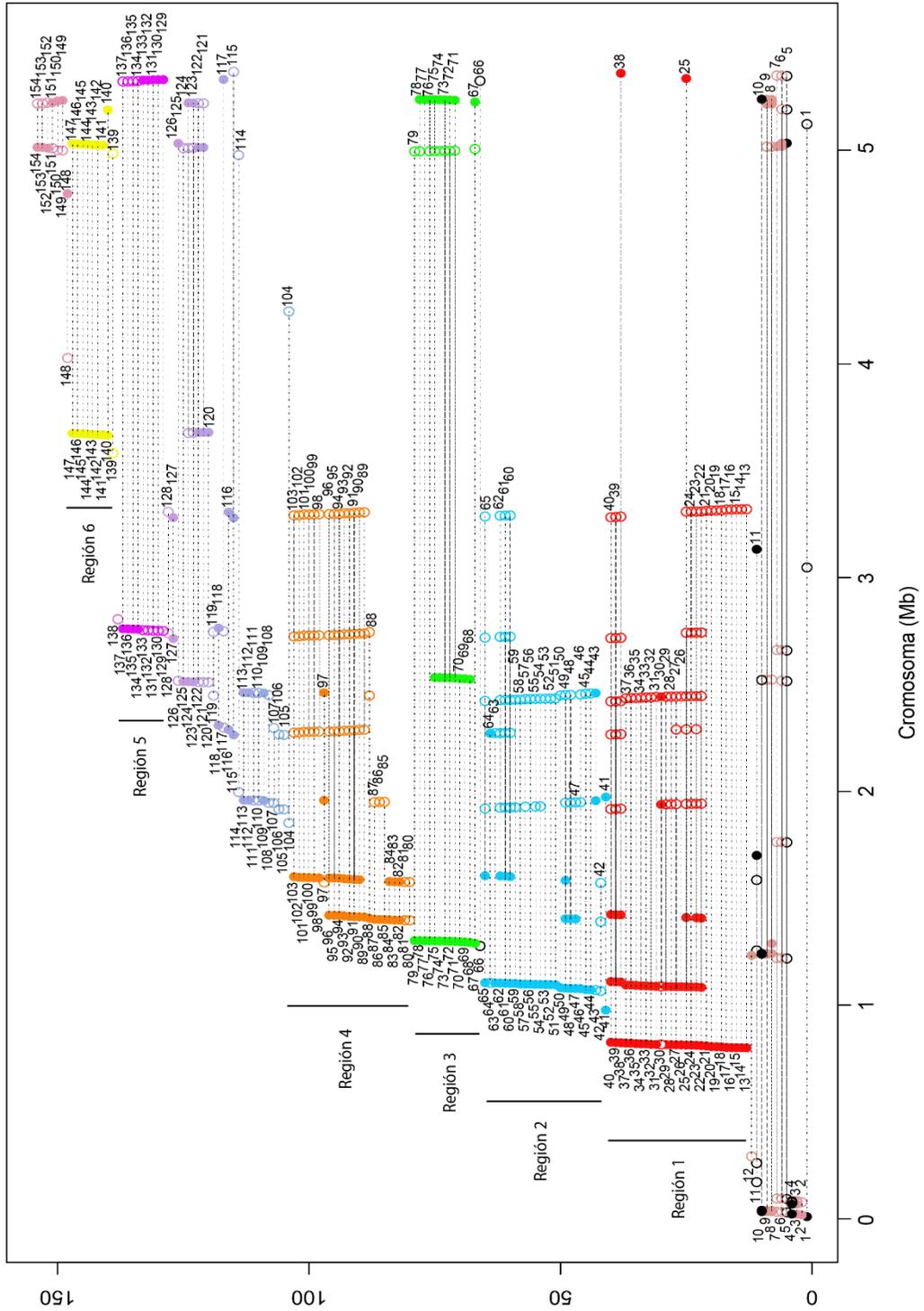


Fig. 4.7. A) Genes duplicados en la cepa UPEC CFT073. B) Genes duplicados en la cepa EHEC O145:H28 RM13516. El eje de las X representa el mapa lineal del cromosoma. Cada grupo de círculos unidos con una línea horizontal discontinua corresponde a un gen duplicado o amplificado. Los círculos indican la posición que ocupa cada duplicado en el genoma. Los círculos enteros corresponden a los genes que se encuentran en la cadena positiva, mientras que los círculos vacíos corresponden a los de la cadena negativa. Los números corresponden a los grupos de genes duplicados, numerados en orden empezando desde el origen del mapa cromosómico. Los círculos negros, enteros o vacíos, corresponden a transposasas. Las barras verticales y los colores definen las 6 regiones principales que contienen genes duplicados.

Una vez llevada a cabo este análisis y su representación gráfica, quedó patente que el número de duplicaciones que presentaban estas cepas era mayor que el que se vio en la cepa 042. En el caso de la cepa CFT073, el número de duplicados que se identificaron fue de 94. Algunos de estos genes duplicados corresponden a transposasas (marcados en negro), tal y como había ocurrido con la cepa 042. Se agruparon todos en 6 regiones principales para facilitar el estudio de los mismos, como se puede observar en la figura 4.7A, donde se han señalado con barras verticales y colores diferentes. La función de cada uno de estos genes duplicados en la cepa CFT073, como su localizador, se recogió en la Tabla anexa A1. En todas las regiones hay un gran número de genes con función desconocida, como se puede apreciar sobre todo en las regiones 2 y 6. En la región 1 de esta cepa se puede ver la duplicación de un sistema toxina-antitoxina (números 7 y 8), y los duplicados 16 al 19 son transportadores del tipo ABC relacionados con el metabolismo del hierro. Como ocurría en la región 1 de la cepa 042, las regiones 3 y 4 de la cepa CFT073 engloban genes de origen fágico, mostrando una vez más la relación existente entre genes HGT y procesos GDA. Por último, cabe señalar que un número significativo de genes presentes en la región 5 están relacionadas con fimbrias, las cuales son de gran importancia dentro de la patogénesis de las cepas uropatógenas.

Respecto a la cepa O145:H28 RM13516, el número de duplicados que se encuentran en su genoma es significativamente mayor, siendo un total de 154. Siguiendo la misma pauta, algunos de estos genes corresponden a transposasas, las cuales también se marcaron de color negro. Se definieron 6 regiones principales que contenían la mayoría de genes duplicados, definidas en la figura 4.7B, con barras verticales y colores para diferenciarlas. En esta cepa EHEC, muchos de los genes tienen más de dos copias, por lo que la amplificación es más abundante que la duplicación. Tanto el localizador de cada una de las copias de estos genes, como la descripción de sus funciones, se recopilaron en la Tabla anexa A2. Un gran número de los genes que se han duplicado en esta cepa EHEC no están bien caracterizados y se desconoce su función, como es el caso de aquellos que se encuentran en la región 3 y 5, y parte de la región 4. En el caso de genes relacionados con fagos, se pueden encontrar en la región 1 (número 32 al 40), la 2 (52 al 65) y en la 3 (92 al 103). Por otra parte, se encuentran duplicados reguladores transcripcionales de la familia Rha (44) y de la familia AraC (50), y, también, genes relacionados con el metabolismo del hierro, como son transportadores del tipo ABC y sideróforos (143 al 147).

La identificación de un número mayor de genes duplicados en estas dos cepas planteó una nueva cuestión: ¿cuál era la distribución de estos genes duplicados entre los diferentes patotipos? De manera homónima, se realizaron los mismos análisis realizados anteriormente con la cepa 042. Los genes duplicados de ambas cepas, englobados en las regiones principales definidas arbitrariamente, se buscaron en las 28 cepas representativas de cada patotipo seleccionadas previamente. En estas cepas se encuentra la cepa comensal MG1655 y las 3 cepas analizadas: 042, CFT073 y O145:H28 RM13516.

En primer lugar, se buscó la distribución de los genes duplicados de la cepa UPEC CFT073. La representación de los resultados se realizó siguiendo el mismo código de colores utilizado en la cepa 042 (Fig. 4.8).

Región 6: 88-94											
	88	89	90	91	92	93	94				
Comensal	0	0	0	0	0	0	0				
E. coli O42	1	0	1	1	1	1	1				
E. coli O104:H4 2011C-3493	0	0	0	0	0	0	0				
E. coli O104:H4 LB226692	0	0	0	0	0	0	0				
E. coli 55989	1	1	1	1	1	1	1				
E. coli O26:H11 11368	0	0	0	0	0	0	0				
E. coli O157:H7 Sakai	0	0	0	0	0	0	0				
E. coli O111:H- 11128	0	0	0	0	0	0	0				
E. coli O145:H28 RM13516	0	0	0	0	0	0	0				
E. coli UMN18	1	1	1	1	1	1	1				
E. coli O139:H28 E24377A	1	1	1	1	1	1	1				
E. coli O103:H2 2011C-3750	1	1	1	1	1	1	1				
E. coli H10407	1	1	1	1	1	1	1				
E. coli O127:H6 E2348-69	0	0	0	0	0	0	0				
E. coli O55:H7 CB9615	1	1	1	1	1	1	1				
E. coli O96:H19	1	1	1	1	1	1	1				
E. coli O143:H26 4608-58	0	0	0	0	0	0	0				
E. coli O28ac:NM 02-3404	0	0	0	0	0	0	0				
E. coli O124:H30 M4163	0	0	0	0	0	0	0				
E. coli 53638	0	0	0	0	0	0	0				
E. coli J11886	0	0	0	0	0	0	0				
E. coli O25b:H4 EC958	0	0	0	0	0	0	0				
E. coli NA114	0	0	0	0	0	0	0				
E. coli 536	2	2	2	2	2	2	2				
E. coli UMN026	1	1	1	1	1	1	1				
E. coli CFT073	2	2	2	2	2	2	2				
E. coli UT189	2	2	2	2	2	2	2				
E. coli IA139	0	0	0	0	0	0	0				

Región 5: 73-86														
	73	74	75	76	77	78	79	80	81	82	83	84	85	86
Comensal	2	0	0	0	0	0	0	0	0	0	0	0	0	0
E. coli O42	2	2	0	0	2	0	0	0	0	0	0	0	0	0
E. coli O104:H4 2011C-3493	2	2	0	0	4	0	0	0	0	0	0	0	0	0
E. coli O104:H4 LB226692	2	2	0	0	3	0	0	0	0	0	0	0	0	0
E. coli 55989	2	2	0	0	2	0	0	0	0	0	0	0	0	0
E. coli O26:H11 11368	2	2	0	0	11	0	0	0	0	0	0	0	0	0
E. coli O157:H7 Sakai	2	1	0	0	10	0	0	0	0	0	0	0	0	0
E. coli O111:H- 11128	1	2	0	0	0	0	0	0	0	0	0	0	0	0
E. coli O145:H28 RM13516	2	2	0	0	10	0	0	0	0	0	0	0	0	0
E. coli UMN18	2	1	0	0	0	0	0	0	0	0	0	0	0	0
E. coli O139:H28 E24377A	2	1	0	0	11	0	0	0	0	0	0	0	0	0
E. coli O103:H2 2011C-3750	2	1	1	0	2	0	0	0	0	0	0	0	0	0
E. coli H10407	2	2	1	0	4	0	0	0	0	0	0	0	0	0
E. coli O127:H6 E2348-69	2	1	0	0	0	0	0	0	0	0	0	0	0	0
E. coli O55:H7 CB9615	2	1	0	0	1	0	0	0	0	0	0	0	0	0
E. coli O96:H19	2	0	0	0	0	0	0	0	0	0	0	0	0	0
E. coli O143:H26 4608-58	2	0	0	0	0	0	0	0	0	0	0	0	0	0
E. coli O28ac:NM 02-3404	1	0	0	0	2	0	0	0	0	0	0	0	0	0
E. coli O124:H30 M4163	2	1	0	0	0	0	0	0	0	0	0	0	0	0
E. coli 53638	2	0	0	0	0	0	0	0	0	0	0	0	0	0
E. coli J11886	2	1	1	1	2	0	0	0	0	0	0	0	0	0
E. coli O25b:H4 EC958	2	1	2	2	1	0	0	0	0	0	0	0	0	0
E. coli NA114	2	2	1	1	2	1	0	0	0	0	0	0	0	0
E. coli 536	2	2	1	1	2	1	1	0	1	1	1	1	1	0
E. coli UMN026	2	1	0	0	0	1	0	1	1	1	1	1	1	1
E. coli CFT073	2	1	1	0	3	1	0	1	1	1	1	1	1	1
E. coli UT189	2	2	2	2	2	2	2	2	2	2	2	2	2	2
E. coli IA139	2	2	1	0	1	0	1	0	1	1	1	1	1	1

Región 4: 56-65											
	56	57	58	59	60	61	62	63	64	65	
Comensal	1	2	0	0	0	0	0	0	0	0	
E. coli K-12 MG1655	0	2	1	1	1	1	2	4	1	3	
E. coli O42	1	1	1	1	0	1	3	3	2	3	
E. coli O104:H4 2011C-3493	1	1	1	1	0	1	3	2	1	1	
E. coli O104:H4 LB226692	1	1	1	1	0	1	3	2	1	1	
E. coli 55989	1	2	1	1	0	1	2	3	2	2	
E. coli O26:H11 11368	1	2	4	5	1	1	4	3	2	1	
E. coli O157:H7 Sakai	0	0	1	1	1	1	2	2	1	1	
E. coli O111:H- 11128	0	0	1	2	1	1	0	0	0	0	
E. coli O145:H28 RM13516	0	1	1	1	0	0	2	1	0	1	
E. coli UMN18	2	1	0	0	0	0	0	0	0	1	
E. coli O139:H28 E24377A	1	1	0	0	0	0	0	0	0	0	
E. coli O103:H2 2011C-3750	1	1	1	1	1	1	0	1	1	0	
E. coli H10407	1	2	2	1	1	2	2	3	2	3	
E. coli O127:H6 E2348-69	0	0	0	0	0	0	1	2	0	0	
E. coli O55:H7 CB9615	1	2	2	3	2	3	5	2	2	2	
E. coli O96:H19	1	4	1	1	0	1	1	3	1	3	
E. coli O143:H26 4608-58	0	3	0	0	0	0	0	2	0	1	
E. coli O28ac:NM 02-3404	0	1	1	1	1	1	1	2	0	1	
E. coli O124:H30 M4163	0	3	0	0	0	0	0	2	0	1	
E. coli 53638	0	3	0	0	0	0	0	2	0	2	
E. coli J11886	0	1	1	1	1	1	1	3	0	3	
E. coli O25b:H4 EC958	0	0	1	1	1	1	1	2	0	2	
E. coli NA114	0	0	1	1	1	1	1	2	0	2	
E. coli 536	2	1	1	1	1	1	1	1	1	1	
E. coli UMN026	1	3	2	2	2	2	2	3	2	3	
E. coli CFT073	2	2	2	2	2	2	2	2	2	2	
E. coli UT189	2	2	1	1	1	1	2	3	1	3	
E. coli IA139	0	2	1	1	1	1	1	1	2	1	

Fig. 4.8. Distribución de los genes duplicados de las 6 regiones principales de la cepa CFT073 en las 28 cepas seleccionadas de varios patotipos. En blanco se muestran los genes que no tienen ninguna copia que cumpla con los parámetros que se analizan. En gris, los genes que presentan solo una copia. En negro, los genes que presentan dos o más copias que cumplen con los parámetros. Los números muestran cuántas copias de cada gen de cada cepa cumplen con los parámetros establecidos.

Los genes duplicados en la cepa CFT073 se encuentran distribuidos en el resto de patotipos de manera desigual. En la región 1, se puede ver que los 6 genes duplicados numerados del 7 al 8 y del 10 al 13 se encuentran distribuidos en todos los patotipos. Este bloque de genes coincide con los genes 55 al 60 de la región 2 de la cepa 042. Como se puede observar en las regiones 2, 3 y 4, los genes 36, 42-44, 53, 63 y 65 se muestran duplicados o amplificados en cepas de todos los patotipos, los cuales son de función desconocida o de origen fágico. Sin embargo, muchos de los genes estudiados parece que han sufrido procesos de duplicación únicamente en la cepa CFT073. En la región 5, los genes 78 al 86 codifican proteínas hipotéticas relacionadas con fimbrias, las cuales solo se encuentran presentes en las cepas uropatógenas, como se puede apreciar. Por otra parte, resulta interesante el patrón que se observa en los genes de la región 6. Pese a que solo se ha duplicado en la cepa CFT073, su presencia se encuentra restringida a las cepas UPEC, ETEC y las cepas 042 y 55989 del patotipo EAEC. Respecto a la cepa comensal MG1655, un porcentaje mínimo de los genes analizados se encuentran presentes en su genoma, y únicamente el 52 y el 73 han sufrido un proceso de duplicación.

De manera paralela a este análisis, se realizó la búsqueda de los genes duplicados de la cepa EHEC O145:H28 RM13516 en las mismas 28 cepas seleccionadas. La distribución de estas duplicaciones en los diferentes patotipos puede visualizarse en la figura 4.9.

Fig. 4.9. Distribución de los genes duplicados de las 6 regiones principales de la cepa O145:H28 RM13516 en las 28 cepas seleccionadas de varios patotipos. En blanco se muestran los genes que no tienen ninguna copia que cumpla con los parámetros que se analizan. En gris, los genes que presentan solo una copia. En negro, los genes que presentan dos o más copias que cumplen con los parámetros. Los números muestran cuántas copias de cada gen de cada cepa cumplen con los parámetros establecidos.

La distribución en los diferentes patotipos de los genes duplicados en la cepa EHEC muestra patrones diferentes, según la región que se analice. De forma semejante a las otras 2 cepas analizadas, la cepa O145:H28 RM13516 también presenta genes duplicados específicos de esta cepa, como por ejemplo los genes que se hallan en la región 5 y 6. Sin embargo, como diferencia a los análisis previos en las cepas 042 y CFT073, se puede observar que algunos de los genes se han duplicado o amplificado únicamente en las cepas enterohemorrágicas. Este patrón, característico de un único patotipo, es visible en los genes englobados en la región 1, 2 y, sobre todo, en la región 4. La función y el origen de muchos de los genes de estas 3 regiones es fágico (Tabla anexa A2), por lo que es posible que la adquisición de este ADN por mecanismos HGT tuviera lugar en un ancestro común. En relación con este patrón, es interesante la presencia de estos genes duplicados fágicos también en la cepa O55:H7, la cual pertenece al patotipo EPEC. Por otra parte, los genes de la región 5 solo se encuentran duplicados en 2 cepas ETEC, aparte de la cepa O145:H28 analizada. Pese a que esta relación podría ser de gran interés, la información disponible sobre estos genes es limitada, por lo que serían necesarios más estudios para poder entender el papel que desempeñan. Finalmente, el bloque de genes numerados del 55 al 60 de la cepa 042, correspondiente a los genes 7-13 de la cepa CFT073, en la cepa O145:H28 RM13516 corresponde a los genes del 71 al 76 de la región 3. Como ya se ha comentado, estos 6 genes se encuentran distribuidos entre todos los patotipos y en un gran número de copias, ya que se puede hablar más de un proceso de

amplificación que de duplicación. Sin embargo, la función que desempeñan es desconocida.

Del conjunto de estos estudios también se desprende el hecho de que existe un número de genes duplicados que, a pesar de la gran variabilidad genómica de los distintos aislamientos de *E. coli*, no pertenecen al *core-genome* pero se encuentran presentes en la mayoría de patotipos.

4.2 El gen *aec69* en la cepa EAEC 042

La búsqueda realizada de genes asociados a la presencia de nuevos reguladores de la familia de proteínas Hha (apartado 4.1.1, Resultados) permitió identificar a dos genes en concreto: *yeeR* y *aec69*. Como se ha comentado anteriormente, estos dos genes se encuentran juntos en el cromosoma de *E. coli*, aguas abajo del gen *flu* (Fig. 1.12). Estudios recientes, llevados a cabo por Moriel y colaboradores³²⁸, resaltan el papel que puede tener en la patogénesis de cepas uropatógenas de *E. coli* el producto del gen *aec69*, denominado en ese trabajo IrmA, terminología que utilizaremos a partir de ahora. La proteína IrmA parece actuar como receptor de interleuquinas, interfiriendo en su función. En este mismo estudio, se indica que los genes *flu*, *yeeR* e *irmA* forman una única unidad transcripcional, donde la expresión de los mismos se ve modulada por la acción del regulador OxyR. Dichos estudios se realizaron en la cepa UPEC CFT073, en la que existen 2 copias de los genes *flu*, *yeeR* e *irmA*. En esa cepa, una de las copias de *irmA* (*c3665*) no presenta actividad, debido a la inserción de varios elementos génicos que rompen la pauta de lectura del operón. En el caso de la cepa EAEC 042, estos tres genes se encuentran duplicados y juntos en dos regiones diferentes del cromosoma (Fig. 1.13). Las copias del gen *irmA* en esta cepa se pueden encontrar por su localizador en la base de datos del NCBI (<https://www.ncbi.nlm.nih.gov/gene>), siendo el mismo EC042_2244 y EC042_4509, respectivamente. En el estudio de las duplicaciones presentes en el cromosoma de la cepa EAEC 042 que hemos llevado a cabo, este duplicado está referenciado por el número 53 (Tabla 4.5) (Fig. 4.4). Las copias del gen *irmA* tienen una longitud de 456 nucleótidos, las cuales codifican proteínas de 151 aminoácidos con un tamaño de 16,9 kDa. En ambos casos, la homología que tienen las dos copias es del 99,34 % (Fig. 4.10). Teniendo en cuenta la asociación encontrada entre estos genes y los reguladores Hha2/Hha3, el siguiente objetivo del trabajo consistió en discernir tanto la expresión de las copias del gen *irmA* en la cepa 042 como su regulación.

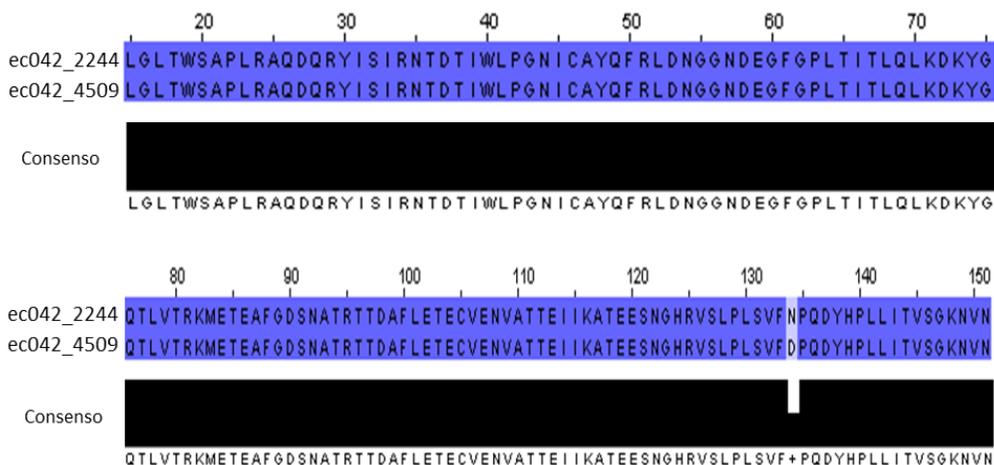


Fig. 4.10. Alineamiento de la secuencia aminoacídica de las dos copias de la proteína IrmA que expresa la cepa 042. Solo hay un único cambio en el aminoácido 134.

4.2.1 Estudio de la expresión de la proteína IrmA en la cepa 042 por *Western Blot*

La primera aproximación para estudiar la expresión de la proteína IrmA codificada por cualquiera de las dos copias del gen fue inmunodetectarla. Para ello, decidimos, en primer lugar, obtener anticuerpos específicos. Se llevó a cabo la adición de una cola de histidinas a la proteína IrmA, siguiendo las instrucciones del kit comercial aLICator LIC Cloning & Expression System (*ThermoScientific*). La proteína fue purificada mediante cromatografía de afinidad haciendo uso de una columna de níquel-agarosa, siguiendo la metodología indicada en el apartado 3.8.5.1 en Materiales y métodos (Fig. 4.11A). Una vez purificada la proteína IrmA, se procedió a inmunizar conejos para obtener un anticuerpo policlonal (apartado 3.8.5.2, Materiales y métodos). Los sueros se comprobaron con la proteína IrmA purificada, tanto los sueros de conejos sin inmunizar (Fig 4.11B) como los de los conejos inmunizados (Fig. 4.11C).

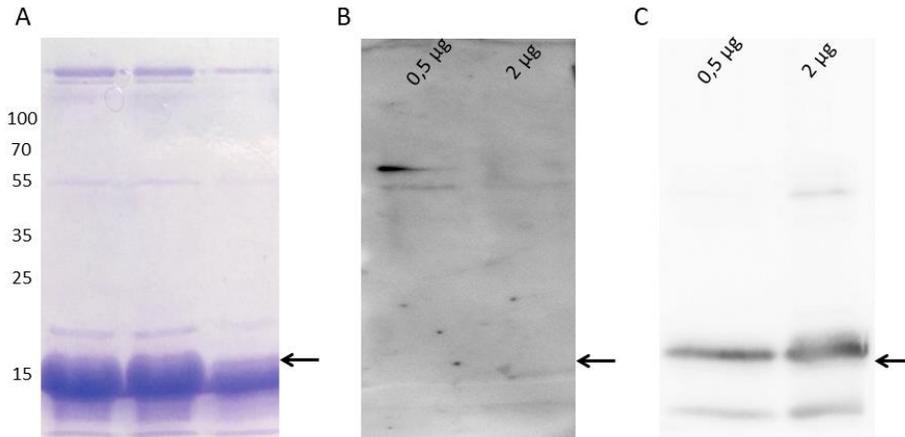


Fig. 4.11. A) Purificación de la proteína IrmA mediante cromatografía de afinidad. Se muestran las tres eluciones realizadas con el tampón A200. B) Inmunodetección de la proteína IrmA purificada mediante el uso de suero de conejo sin inmunizar. Se indican dos concentraciones de proteína IrmA. C) Inmunodetección de la proteína IrmA purificada mediante el uso de suero de conejo inmunizado con la proteína IrmA. Se muestran dos concentraciones diferentes de proteína IrmA purificada. La flecha indica la posición de la proteína IrmA.

Con el fin de detectar la expresión específica de cada una de las dos copias del gen *irmA*, se procedió a obtener mutantes de cada copia para poder estudiarlas por separado, denominándolos *irmA2244* e *irmA4509*. Debido a que este gen se sobreexpresa en un mutante para el regulador OxyR³²⁸, se procedió a construir este mutante en la cepa 042. El control negativo del ensayo se realizó mediante la construcción de un mutante de ambas copias del gen *irmA* en la cepa 042, llamándolo *irmAnull*. La mutación de estos genes se hizo mediante la recombinación homóloga de fragmentos generados por PCR, siguiendo el protocolo descrito por Datsenko y Wanner (apartado 3.6.1, Materiales y métodos)³³⁵.

Una vez conseguidos estos mutantes, se obtuvieron los extractos proteicos totales (proteínas intra y extracelulares), tanto de la cepa salvaje (WT) como de los mutantes de las copias del gen *irmA* y del mutante del gen *oxyR*. Se crecieron

cultivos en medio LB a 37 °C y en agitación hasta una DO_{600nm} de 2, momento en el cual se extrajo 1 mL de cultivo para procesarlo, siguiendo la metodología indicada en el apartado 3.8.1. Se comprobó que la concentración de proteína era equivalente en todos los extractos mediante una tinción con azul de Coomassie, normalizando de esta manera la cantidad de proteína cargada en los geles. La detección de la proteína IrmA se realizó mediante el uso del anticuerpo policlonal anti-IrmA, obteniendo los resultados que se muestran en la figura 4.12.

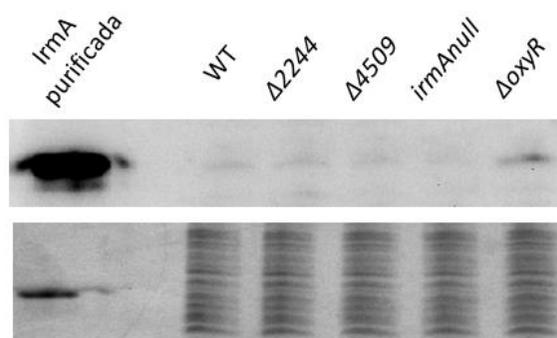


Fig. 4.12. Panel superior. Inmunodetección de la proteína IrmA en extractos totales de las cepas WT, *irmA2244*, *irmA4509*, *irmAnull* y *oxyR* utilizando anticuerpos policlonales anti-IrmA. Como control positivo se utilizó proteína IrmA purificada. Panel inferior. Tinción de Coomassie de los extractos proteicos totales para la inmunodetección, demostrando que los mismos contenían cantidades equivalentes de proteína.

El nivel de expresión que se obtuvo para la proteína IrmA, con independencia del gen que la codificase, fue muy bajo. Como ya se había descrito anteriormente, la expresión del gen aumenta en ausencia del regulador OxyR. Debido a los bajos niveles de la proteína que se detectaron, se decidió estudiar la expresión de la proteína IrmA siguiendo otra metodología, con el fin de poder conocer mejor la expresión de cada copia de manera individual y de mejorar la sensibilidad de su detección.

El estudio de la expresión de las copias *irmA2244* e *irmA4509*, de una manera más precisa, se llevó a cabo mediante la adición de una etiqueta 3XFLAG en el extremo C-terminal de la proteína, mediante la acción de la recombinasa Red (apartado 3.6.2, Materiales y métodos). Este tipo de fusiones permiten la utilización de un mismo anticuerpo para todas las proteínas marcadas, reduciendo los costes que conlleva la obtención de un anticuerpo específico. Las cepas con estas fusiones se nombraron como 042 2244::3X y 042 4509::3X (Tabla 3.1, Materiales y métodos), según la copia que se hubiera marcado en cada caso. Tras la obtención de estas fusiones, se procedió a obtener los extractos proteicos totales, tal y como se hizo en el apartado anterior. Seguidamente, se separaron las proteínas en geles de poliacrilamida y se transfirieron a membranas de PVDF, incubando las mismas con el anticuerpo monoclonal comercial anti-FLAG. Por último, se revelaron las membranas gracias a una reacción de quimioluminiscencia, detectando de esta manera la expresión de las proteínas marcadas con la etiqueta 3XFLAG (Fig. 4.13). Se comprobó la cantidad de proteína de los extractos proteicos mediante tinción con azul de Coomassie.

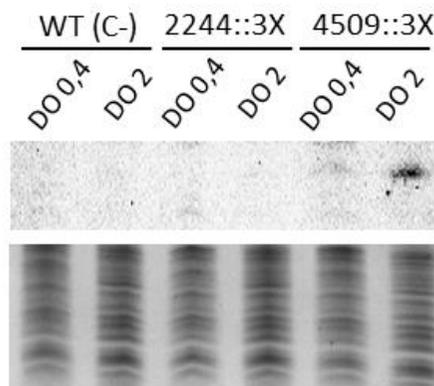


Fig. 4.13. Panel superior. Inmunodetección de las proteínas IrmA2244::3X e IrmA4509::3X con el anticuerpo monoclonal anti-FLAG. Como control negativo se utilizó un extracto proteico de la cepa 042 WT. Panel inferior. Tinción de Coomassie de los extractos totales utilizados para la inmunodetección, demostrando que los mismos contenían cantidades equivalentes de proteína.

La inmunodetección llevada a cabo reveló un patrón de expresión diferencial entre las dos copias del gen *irmA*. Como se puede observar, la expresión de la proteína IrmA2244 es nula o tan baja que no se pudo detectar, a diferencia de la copia IrmA4509. En el caso de la copia IrmA4509, la expresión es mayor en fase estacionaria que en fase exponencial. Los resultados obtenidos en el ensayo anterior con el anticuerpo anti-IrmA (Fig. 4.12) muestran una baja expresión en ambas copias, mientras que en este caso se ve claramente una expresión diferente entre ellas. Es posible que tenga lugar una regulación cruzada entre las dos copias, por lo que en los mutantes *irmA2244* e *irmA4509* (necesarios para poder estudiar cada copia individualmente) esta regulación no tiene lugar y las dos copias se expresan. Gracias al uso del epítipo 3XFLAG se evita este problema, pudiendo estudiar las dos copias en el mismo fondo genético.

La expresión diferencial de los dos genes *irmA* de la cepa 042 puede deberse a un procesamiento post-traducciona l diferente de las correspondientes proteínas. Sin embargo, debido al gran porcentaje de homología existente entre ellas, esta hipótesis parece poco probable. Las diferencias en la expresión pueden tener origen en los niveles de transcripción de cada uno de los genes y, por lo tanto, en la regulación de la expresión de los mismos. En este sentido, el estudio realizado por Wapinski y colaboradores señala que las diferencias entre genes duplicados pueden deberse más a su regulación que a su función³⁰². Es por ello que se decidió estudiar la transcripción de ambas copias, tratando de dilucidar el posible origen de estas diferencias y confirmando los niveles de expresión de las proteínas detectadas (Fig. 4.13).

4.2.2 Estudio de la transcripción de las dos copias del gen *irmA* en la cepa 042 mediante las fusiones cromosómicas *irmA2244::lacZ* e *irmA4509::lacZ*

La transcripción de los genes EC042_2244 y EC042_4509 se estudió mediante un método indirecto, gracias al gen reportero *lacZ*. En primer lugar, siguiendo la metodología descrita en el apartado 3.6.3, se construyeron las fusiones transcripcionales en el cromosoma de la cepa 042, a partir de los mutantes de cada una de las copias (*irmA2244* e *irmA4509*) obtenidos anteriormente (Fig. 4.14). Las cepas obtenidas con las fusiones transcripcionales se denominaron 042 2244::*lacZ* y 042 4509::*lacZ* (Tabla 3.1, Materiales y métodos).

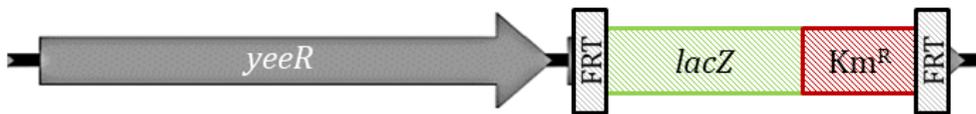


Fig. 4.14. Esquema de la posición de la fusión con el gen reportero *lacZ* para estudiar la transcripción del gen *irmA*.

Tras conseguir estas dos construcciones, se crecieron los cultivos a 37 °C en agitación en medio LB. Se tomaron muestras de los cultivos a DO_{600nm} de 0,4 y 2, correspondientes a fase exponencial e inicio de la estacionaria, respectivamente. Estas muestras se procesaron para llevar a cabo un ensayo de β-galactosidasa, de acuerdo al apartado 3.9.1 (Fig. 4.15).

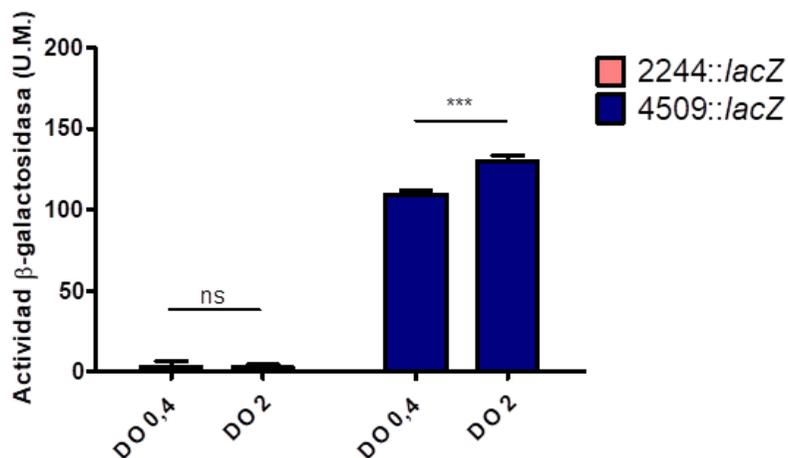


Fig. 4.15. Actividad β -galactosidasa (U. M.) de las cepas 042 2244::lacZ y 042 4509::lacZ. Valores obtenidos del ensayo por triplicado. ns no significativo, ***P < 0,001.

Tal y como es apreciable, la actividad transcripcional en las dos copias del gen *irmA* es muy diferente. La expresión del gen *lacZ* bajo el control del promotor responsable de la transcripción del gen *irmA2244* es muy baja. Por el contrario, existe un buen nivel de expresión del gen *lacZ* bajo el control del promotor responsable de la transcripción del gen *irmA4509*. Estos resultados concuerdan perfectamente con la expresión detectada de las proteínas marcadas con el epítipo 3XFLAG, comentada anteriormente (Fig. 4.13). Por lo tanto, las diferencias que se ven en la expresión de ambas proteínas tienen su origen en la transcripción de los respectivos genes, descartándose que un posible procesamiento post-traduccional sea la razón del nivel diferente de expresión de las proteínas codificadas por ambas copias. Por lo que hace referencia al efecto de la fase de crecimiento sobre la expresión del gen *irmA4509*, se puede observar un ligero aumento de su expresión en fase estacionaria.

4.2.3 Regulación de la expresión de las dos copias del gen *irmA* en la cepa EAEC 042

Tras observar una expresión diferente de cada copia de la proteína IrmA cuando la cepa 042 crece en medio LB a 37 °C, se decidió investigar la regulación de la expresión de ambas copias del gen *irmA*. En primer lugar, decidimos estudiar el efecto de diferentes condiciones ambientales sobre la expresión de ambas copias. Por otra parte, teniendo en cuenta que la presencia del gen *irmA* se correlaciona con la presencia de los reguladores Hha2 y Hha3 (Fig. 4.1) y que en el análisis transcriptómico de la expresión de *irmA* en diferentes fondos genéticos los genes *irmA* aparecen regulados por el sistema Hha/H-NS (Tabla 4.6), se quiso corroborar el papel del sistema H-NS/Hha en la regulación de la expresión de las dos copias del gen *irmA*. En la cepa UPEC CFT073, se describió que este gen forma parte de la misma unidad transcripcional junto a los genes *flu* e *yeeR*³²⁸. El gen *flu* presenta una regulación por variación de fase debida a la competición de las proteínas OxyR y Dam^{324,325}. Es por ello que se quiso determinar si las copias EC042_2244 y EC042_4509 se encuentran reguladas de la misma manera.

4.2.3.1 Efecto de la temperatura y del medio de cultivo sobre la expresión de los genes *irmA2244* e *irmA4509*

El efecto de los factores ambientales sobre la expresión génica ha sido ampliamente estudiado. En este caso, nos focalizamos en el efecto de la temperatura y la disponibilidad de nutrientes sobre la expresión de los genes *irmA2244* e *irmA4509*. Para llevarlo a cabo, se utilizaron las cepas 042 2244::3X y 042 4509::3X para poder examinar, mediante inmunodetección, la expresión de cada proteína por separado en cultivos de la cepa 042 crecidos en condiciones diferentes. En el caso de la temperatura, se crecieron cultivos de

estas dos cepas 37 °C y 25 °C (Fig. 4.16A). Por otra parte, para simular la disponibilidad de concentraciones diferentes de nutrientes, se utilizaron dos medios de cultivo diferentes: un medio rico como LB y un medio mínimo como M9 (Fig. 4.16C). De forma paralela, se utilizaron las cepas 042 2244::*lacZ* y 042 4509::*lacZ* para estudiar el efecto de las condiciones de cultivo sobre los niveles de transcripción de los genes *irmA2244* e *irmA4509*, mediante ensayos de β -galactosidasa (Fig. 4.16B y D). Tanto en el caso de la temperatura como en el de la disponibilidad de nutrientes, se crecieron los cultivos en agitación hasta una DO_{600nm} de 2, ya que esta proteína presenta mayor expresión en una fase estacionaria temprana que en fase exponencial (Fig. 4.13).

Ambos ensayos aportaron información relevante respecto al efecto de estas condiciones en la expresión de la proteína IrmA. En el caso de la temperatura, la proteína IrmA4509 presenta una mayor expresión a 37 °C que a 25 °C en fase estacionaria. Respecto a su expresión en los medios utilizados, la proteína IrmA4509 tiene una expresión ligeramente mayor en medio M9 que en medio LB a 37 °C. Sin embargo, la proteína IrmA2244 no presenta una expresión detectable en ninguna de las condiciones estudiadas. Por lo tanto, su expresión puede verse afectada por otros estímulos y seguir una regulación diferente que la proteína IrmA4509.

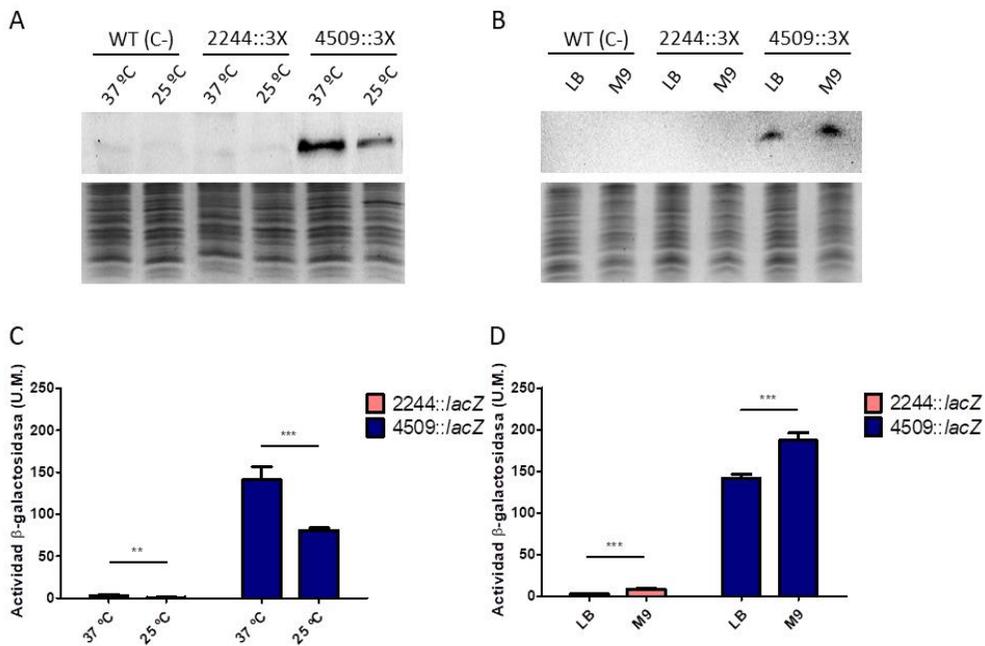


Fig. 16. Efecto de la temperatura (A y C) y de la disponibilidad de nutrientes (B y D) en la expresión de los genes *irmA2244* y *irmA4509*. Como control negativo se utilizó un extracto proteico de la cepa 042 WT. **P < 0,01, ***P < 0,001.

El efecto que ejerce la disponibilidad de nutrientes en el medio de cultivo sobre la expresión de estos genes también se estudió mediante un ensayo de limitación de la fuente de carbono (en nuestro caso, glucosa). Para ello, se utilizó, en primer lugar, medio M9 suplementado con distintas concentraciones de glucosa, para determinar la concentración limitante del crecimiento. Se crecieron cultivos de la cepa 042 en medio M9 y se llevaron a cabo curvas de crecimiento para cada concentración de glucosa (Fig. 4.17).

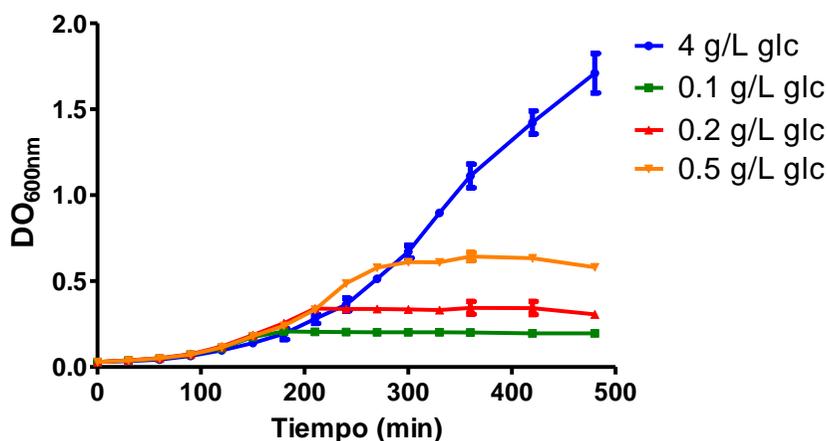


Fig. 4.17. Curvas de crecimiento de la cepa 042 de *E. coli* en medio M9 a 37 °C con concentraciones diferentes de glucosa. Como control se utilizó medio M9 suplementado con 4 g/L de glucosa. Valores obtenidos del ensayo por triplicado.

Observando estos resultados, se decidió utilizar una concentración de glucosa de 0,5 g/L en el medio M9 para analizar cómo la limitación en la fuente de carbono afecta a la expresión de la proteína IrmA en la cepa 042. Para ello, se procedió a crecer las cepas 042 2244::3X y 042 4509::3X en medio M9 con 4 g/L (como control) y 0,5 g/L de glucosa a 37 °C. Una vez que se detuvo el crecimiento, se cogieron muestras tras 1 y 2 horas, procesándolas para llevar a cabo un ensayo de inmunodetección de la proteína IrmA. De forma paralela, se crecieron bajo las mismas condiciones de limitación de la fuente de carbono las cepas 042 2244::lacZ y 042 4509::lacZ, con el fin de analizar indirectamente la transcripción de los genes *irmA2244* e *irmA4509* evaluando la expresión del gen *lacZ* (Fig. 4.18).

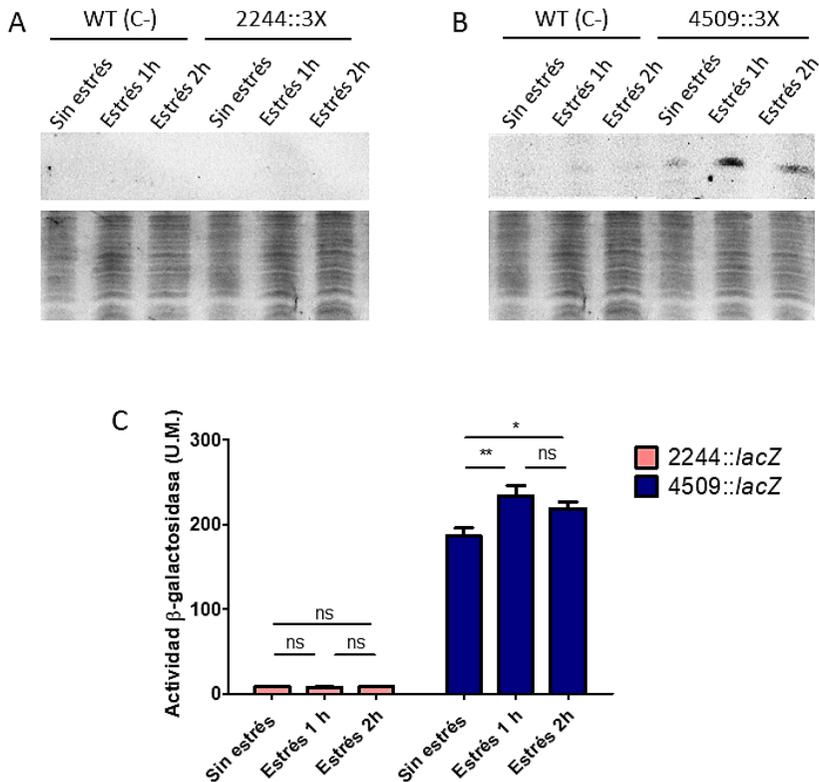


Fig. 4.18. Efecto del estrés por limitación de glucosa sobre la expresión de las proteínas IrmA2244::3XFLAG (A), IrmA4509::3XFLAG (B) y sobre la transcripción de los genes *irmA2244* e *irmA4509* (C). Como control negativo se utilizó un extracto proteico de la cepa 042 WT. ns no significativo, *P < 0,05, **P < 0,01.

Los resultados obtenidos demuestran que, en el caso de la proteína IrmA4509, la expresión aumenta bajo estas condiciones de estrés nutricional. Sin embargo, en el caso de la proteína IrmA2244, la expresión es muy baja o nula en todas las condiciones. Estos resultados se ven confirmados al determinar la transcripción de los genes *irmA2244* e *irmA4509* (Fig. 4.18C).

4.2.3.2 Efecto de la presencia de oxígeno sobre la expresión de los genes *irmA2244* e *irmA4509*

Tras estudiar tanto el efecto de la temperatura como el de la disponibilidad de nutrientes en la expresión de ambas copias del gen *irmA*, se decidió analizar si la presencia de oxígeno influía en la expresión de las mismas. Este análisis se realizó tanto con las cepas 042 2244::3X y 042 4509::3X, para llevar a cabo una inmunodetección de IrmA, como con las cepas 042 2244::lacZ y 042 4509::lacZ, para analizar la transcripción de ambos genes (Fig. 4.19). La condición de anaerobiosis se consiguió añadiendo vaselina a cultivos en medio LB previamente inoculados, los cuales se dejaron crecer durante toda la noche en estático a 37 °C.

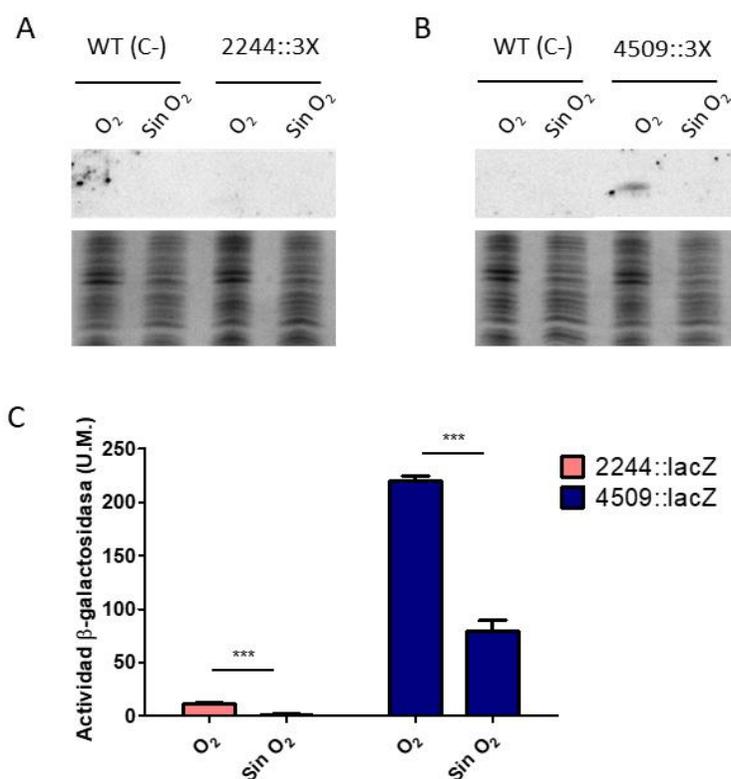


Fig. 4.19. Efecto de la anaerobiosis sobre la expresión de las proteínas IrmA2244::3X (A), IrmA4509::3X (B) y sobre la transcripción de los genes *irmA2244* y *irmA4509* (C). Como control negativo se utilizó un extracto proteico de la cepa 042 WT. ***P < 0,001.

Los resultados obtenidos demuestran que, en ausencia de oxígeno, la expresión del gen *irmA4509* se ve reducida significativamente. En el caso del gen *irmA2244*, la expresión de la proteína mediante *Western blot* no es apreciable.

4.2.3.3 Papel de las proteínas H-NS y Hha en la expresión de los genes *irmA2244* e *irmA4509*

Como se ha comentado en el apartado 4.1.1, el estudio focalizado en la comparación de cepas que codifican nuevos parálogos de Hha frente a cepas que no los tienen permitió asociar estos reguladores con los genes *flu*, *yeeR* e *irmA*. Las proteínas tipo Hha forman complejos con la familia de reguladores H-NS, alterando su especificidad y modulando su actividad^{222,267}. Debido a la correlación entre la presencia del gen *irmA* con parálogos de la familia Hha, y la interacción entre Hha y H-NS, se decidió estudiar la posible función reguladora de este sistema sobre la expresión de las copias *irmA2244* e *irmA4509*.

A fin de estudiar el papel de estos reguladores en la expresión de las dos copias del gen *irmA*, se obtuvieron mutantes sin función para Hha/H-NS en cepas derivadas de la 042. La mutagénesis se llevó a cabo en las cepas 042 2244::3X y 042 4509::3X, obteniendo en cada una los mutantes *hha*, *hhanull* (mutante de *hha* y *hha2*) y *hns* (Tabla 3.1, Materiales y métodos). Una vez obtenidos los mutantes, se procedió a analizar los niveles de expresión de cada copia de la proteína IrmA en los distintos fondos genéticos. Se crecieron cultivos a 37 °C en medio LB con agitación hasta una DO_{600nm} de 2, se obtuvieron los extractos proteicos totales y se llevó a cabo un ensayo de inmunodetección de la proteína IrmA (Fig. 4.20).

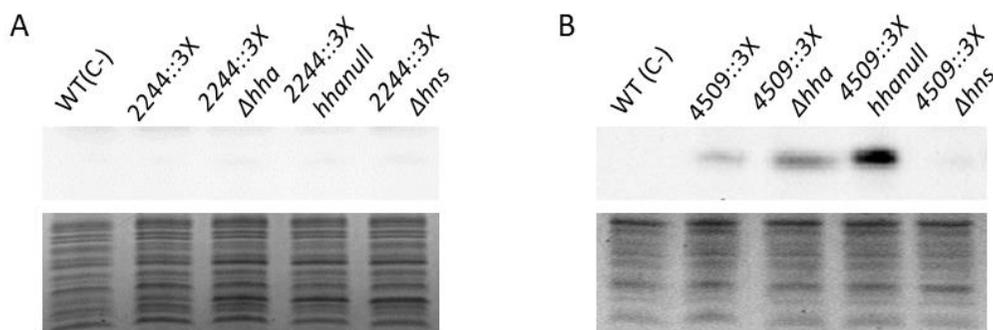


Fig. 4.20. Efecto de la regulación por el sistema H-NS/Hha en la expresión de la proteína IrmA. A) 042 2244::3X y mutantes derivados. B) 042 4509::3X y mutantes derivados. Como control negativo se utilizó un extracto proteico de la cepa 042 WT. Panel superior, inmunodetección realizada con el anticuerpo monoclonal anti-FLAG. Panel inferior, tinción de Coomassie de los extractos totales utilizados para la inmunodetección, demostrando que los mismos contenían cantidades equivalentes de proteína.

La copia IrmA2244 no presenta expresión en ninguno de los fondos genéticos estudiados. En el caso de la copia IrmA4509, la mutación del regulador Hha provoca un aumento en los niveles de expresión, efecto que se ve de manera más significativa en el mutante de Hha y su parálogo Hha2 (*hhanull*). Estudios transcriptómicos realizados en los mutantes *hha*, *hha2* y *hhanull* han puesto de manifiesto que una de las funciones de la proteína Hha2 es hacer de reservorio molecular para Hha, por lo que es en un mutante *hhanull* donde puede comprobarse el papel regulador del sistema Hha en la cepa 042²⁰⁵. Respecto al papel de H-NS en la regulación de ambas copias, en la copia IrmA2244 no se observan diferencias mientras que, en el caso de la copia IrmA4509, los niveles se reducen.

4.2.3.4 Papel de las proteínas OxyR y Dam en la expresión de los genes *irmA2244* e *irmA4509*

Estudios previos, llevados a cabo por Moriel y colaboradores³²⁸, resaltan la regulación que ejerce OxyR sobre el gen *irmA* en la cepa CFT073, ya que se produce una sobreexpresión de la proteína IrmA en un mutante *oxyR*. Debido a estas observaciones, decidimos analizar si, tal y como ha sido descrito para la cepa uropatógena CFT073, la proteína OxyR regula la expresión de alguna de las copias del gen *irmA* en la cepa 042. Teniendo en cuenta el mecanismo de variación de fase llevado a cabo por la competición de OxyR con la metilasa Dam^{303,319}, se comprobó el posible papel de ambas proteínas en la regulación del gen *irmA*.

Para estudiar el papel de OxyR y de Dam, se obtuvieron los correspondientes mutantes mediante recombinación de fragmentos de PCR (apartado 3.6.1, Materiales y métodos). Estos mutantes se realizaron en las cepas 042 2244::3X y 042 4509::3X (Tabla 3.1, Materiales y métodos). Los niveles de expresión se analizaron mediante un ensayo de inmunodetección de muestras de extracto total (Fig. 4.21). Como se realizó anteriormente, los cultivos se crecieron a 37 °C hasta una DO_{600nm} de 2, momento en el cual se obtuvieron los extractos proteicos totales.

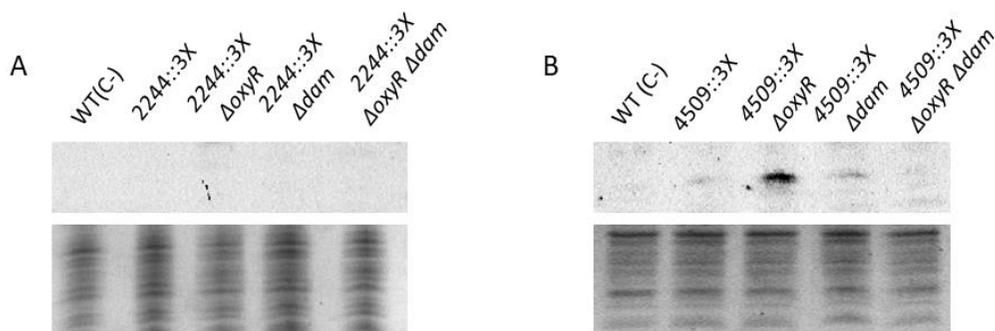


Fig. 4.21. Efecto de OxyR y Dam en la expresión de la proteína IrmA. A) 042 2244::3X y mutantes derivados. B) 042 4509::3X y mutantes derivados. Como control negativo se utilizó un extracto proteico de la cepa 042 WT. Panel superior, inmunodetección realizada con el anticuerpo monoclonal anti-FLAG. Panel inferior, tinción de Coomassie de los extractos totales utilizados para la inmunodetección, demostrando que los mismos contenían cantidades equivalentes de proteína.

La copia IrmA2244 no se expresa tampoco en ninguno de estos fondos genéticos, en las condiciones que se han estudiado. En el caso de la copia IrmA4509, se observa una sobreexpresión en un mutante *oxyR*, hecho que concuerda con las observaciones previas realizadas en la cepa UPEC CFT073³²⁸. Respecto a la metilasa Dam, su mutagénesis no afecta significativamente a los niveles de expresión de la copia IrmA4509. Sin embargo, en el doble mutante *oxyR dam*, no tiene lugar la misma sobreexpresión que se observa en el mutante *oxyR*. Por lo tanto, es posible que la metilación del ADN sea necesaria para la expresión del gen *irmA*.

Tras la obtención de estos resultados, se estudió tanto el efecto de *oxyR* como el de la metilasa *dam* sobre la transcripción. Para ello, se mutaron ambos genes en las fusiones *lacZ* de las dos copias del gen *irmA* (042 2244::*lacZ* y 042 4509::*lacZ*), previamente construidas (Tabla 3.1, Materiales y métodos). Seguidamente, se crecieron los cultivos en medio LB a 37 °C en agitación, recogiendo muestras a DO_{600nm} 2 y procesándolas para llevar a cabo un ensayo de β-galactosidasa (Fig. 4.22).

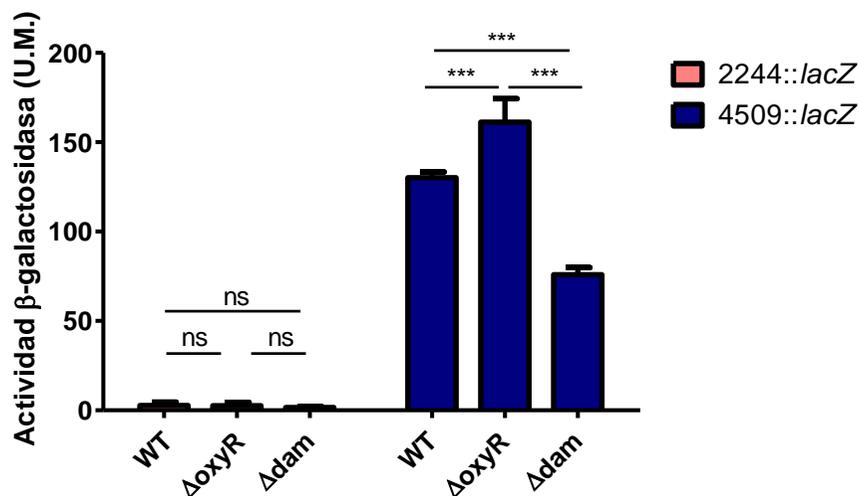


Fig. 4.22. Actividad β -galactosidasa (U.M.) de las cepas 2244::lacZ y 4509::lacZ, tanto en un fondo genético salvaje como en un mutante *oxyR* y en un mutante *dam*. Valores obtenidos del ensayo realizado por triplicado. ns no significativo, *** $P < 0,001$.

La actividad β -galactosidasa que presentan ambas concuerda con los resultados previos, obtenidos mediante la inmunodetección del epítipo 3XFLAG (Fig. 4.21). Por lo tanto, *oxyR* inhibe la expresión de la copia *IrmA4509* en la cepa 042, tal y como ocurre en la cepa uropatógena CFT073. Por otra parte, como se ha comentado anteriormente, la metilación del ADN parece ser necesaria para la expresión del gen *irmA*, ya que en el mutante *dam* la actividad β -galactosidasa es menor que en la cepa salvaje.

Debido a que se produce una sobreexpresión de la proteína *IrmA* en un mutante *oxyR*, se decidió realizar un ensayo de inmunodetección utilizando el anticuerpo anti-*IrmA*. Con la finalidad de poder estudiar el efecto de este regulador sobre la expresión de cada copia por separado, se construyeron los mutantes dobles *irmA2244 oxyR* y *irmA4509 oxyR* (Tabla 3.1, Materiales y métodos). Los cultivos se crecieron hasta una DO_{600nm} de 2 y las muestras de extractos totales fueron igualadas (Fig. 4.23).

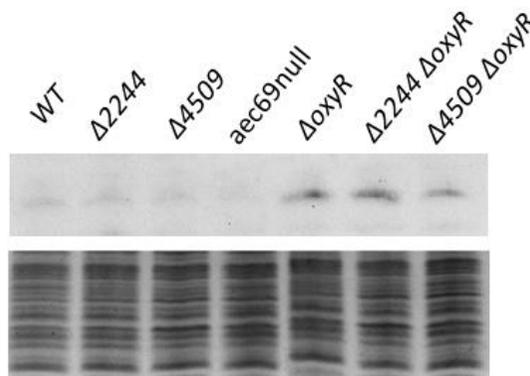


Fig. 4.23. Panel superior. Inmunodetección de la proteína IrmA con el anticuerpo policlonal anti-IrmA. Panel inferior. Tinción de Coomassie de los extractos totales utilizados para la inmunodetección, demostrando que los mismos contenían cantidades equivalentes de proteína.

En los mutantes de cada copia del gen *irmA*, en un fondo genético *oxyR*, se observa una mayor expresión que en un fondo salvaje. La detección con el anticuerpo policlonal anti-IrmA sí que muestra expresión de la copia IrmA2244 en ausencia de *oxyR* (cepa *irmA4509 oxyR*), a diferencia de los resultados obtenidos mediante las fusiones FLAG. Sin embargo, hay que tener en cuenta que son fondos genéticos diferentes. En el caso de las fusiones FLAG, solo se ha adicionado un epítipo en el extremo 3' del gen, mientras que, en este ensayo, se ha eliminado una de las copias completamente para poder estudiar la otra.

4.2.3.5 Regulación cruzada entre las dos copias del gen *irmA*

Con el objetivo de aclarar si hay una relación entre los niveles de expresión de las dos copias del gen *irmA*, se decidió analizar la expresión de cada copia en un fondo genético donde se ha mutado la otra copia. Para llevarlo a cabo, se realizaron fusiones con el epítipo FLAG en los mutantes *irmA2244* e *irmA4509*, previamente conseguidos (Tabla 3.1, Materiales y métodos). Realizándolo a

partir de los mutantes simples nos aseguramos que la copia marcada sería la contraria, facilitando la comprobación de las construcciones. Tras la obtención de estas fusiones, se crecieron los cultivos a 37 °C y se obtuvieron extractos proteicos totales para su análisis mediante inmunodetección (Fig. 4.24).

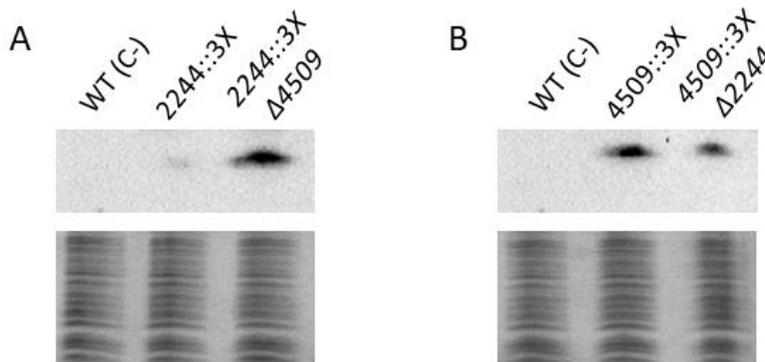


Fig. 4.24. Relación de los niveles de expresión entre las dos copias de la proteína IrmA. A) 042 2244::3X y mutante derivado. B) 042 4509::3X y mutante derivado. Como control negativo se utilizó un extracto proteico de la cepa 042 WT. Panel superior. Inmunodetección realizada con el anticuerpo monoclonal anti-FLAG. Panel inferior. Tinción de Coomassie de los extractos totales utilizados para la inmunodetección, demostrando que los mismos contenían cantidades equivalentes de proteína.

La expresión de la copia IrmA2244 incrementa notablemente en ausencia de la copia IrmA4509, como se puede observar en la parte A de la figura 4.24. Sin embargo, en el caso de la copia IrmA4509 (parte B), no se ve un cambio en la expresión en un mutante para la copia IrmA2244. Los resultados obtenidos marcando con etiquetas 3xFLAG los genes *irmA2244* e *irmA4509* son los mismos que los obtenidos previamente en un fondo genético *oxyR* (Fig. 4.23). Por lo tanto, la expresión de la proteína IrmA4509 es posible que evite que la copia IrmA2244 se exprese. Otra posibilidad es que el promotor del gen *irmA2244* sufra pequeños cambios cuando no se encuentra la copia *irmA4509*, provocando un aumento de su expresión y manteniendo de esta manera los mismos niveles de proteína.

4.2.4 Identificación del promotor de los genes *irmA* de la cepa 042

Los resultados referidos anteriormente ponen de manifiesto que la actividad transcripcional en ambas copias del gen *irmA* en la cepa 042 es muy diferente, siendo esto probablemente la razón de las diferencias en la expresión de las proteínas IrmA2244 e IrmA4509 (Fig. 4.15). Teniendo en cuenta que las secuencias nucleotídicas, tanto en dichos genes como del clúster *agn43 yeeR*, son muy parecidas para ambas copias, se decidió continuar los estudios intentando determinar el promotor de cada copia. Como punto de partida se analizó el promotor del gen *agn43*, el cual se había descrito que forma una única unidad transcripcional junto a los genes *yeeR* e *irmA*³²⁸.

4.2.4.1 Expresión del gen *lacZ* bajo el control del promotor del gen *agn43*

La actividad del promotor del gen *agn43* se estudió mediante el uso del plásmido pUJ8. Este plásmido contiene el gen reportero *lacZ* sin su región promotora, junto a un casete de resistencia a ampicilina para su selección (apartado 3.6.4, Materiales y métodos). Se amplificó la región promotora de cada copia del gen *agn43* (*promagn43_2242* y *promagn43_4511*) (327 pb) y los insertos se digirieron con las dianas de restricción *EcoRI-BamHI* (Fig. 4.25). A continuación, se procedió a su ligación en el plásmido pUJ8, consiguiendo así una fusión transcripcional.

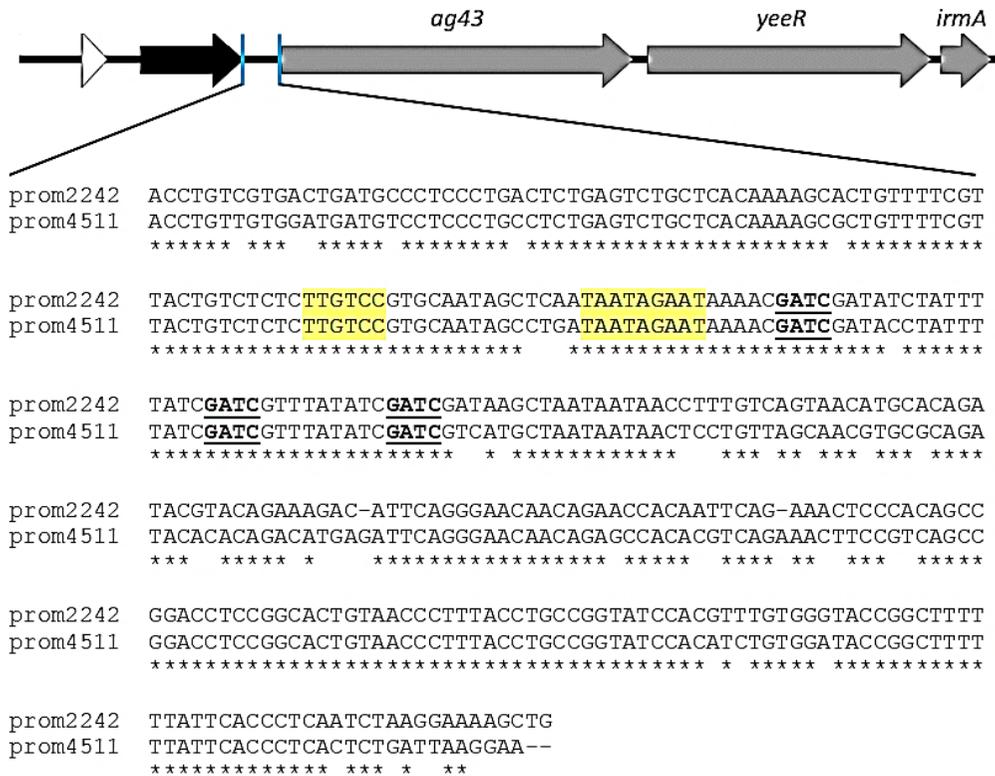


Fig. 4.25. Alineamiento de la región utilizada como promotor en las fusiones transcripcionales de las copias *agn43_2242* y *agn43_4511*. Los sitios de unión de OxyR y Dam (GATC) se encuentran en negrita y subrayados. En amarillo las cajas -10 y -35 del promotor³⁰³.

La región promotora seleccionada para las fusiones transcripcionales presenta algunos cambios entre las dos copias. Sin embargo, tanto los sitios GATC de unión de Dam y OxyR como las cajas -10 y -35 del promotor se conservan idénticas. Una vez obtenidos los clones, se secuenciaron para verificarlos y se transformaron en la cepa 042 WT (Tabla 3.1, Materiales y métodos). Seguidamente, se crecieron cultivos en medio LB en agitación a una temperatura de 37 °C. Cuando los cultivos alcanzaron una DO_{600nm} de 2, se recogieron muestras y se procesaron para valorar la actividad β -galactosidasa

(Fig. 4.26). Debido a la regulación por variación de fase que presenta este gen, se crecieron varios cultivos en paralelo para poder estudiarlo en detalle.

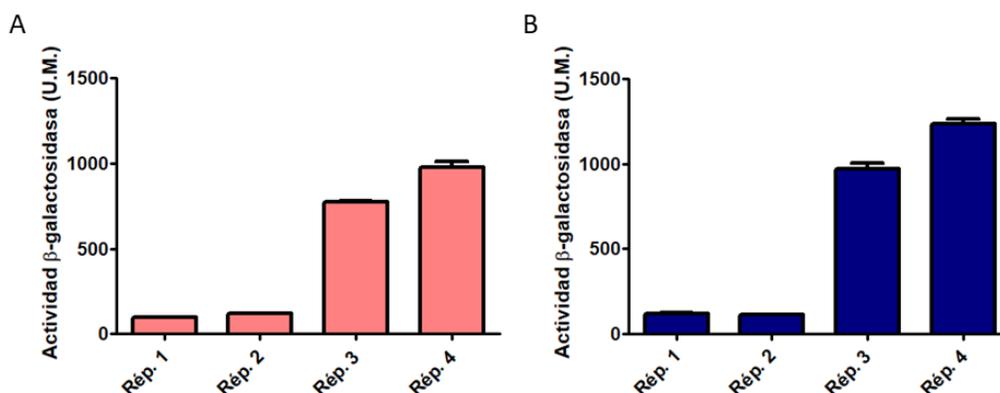


Fig. 4.26. Actividad β -galactosidasa (U.M.) de las fusiones transcripcionales plasmídicas. A) Fusión pUJ8-prom2242. B) Fusión pUJ8-prom4511. Valores obtenidos del ensayo por triplicado.

Como se ha comentado anteriormente, en la regulación por variación de fase del *agn43*, las proteínas OxyR y Dam compiten por su unión con el promotor (sitios GATC), actuando Dam como inductor y OxyR como represor. La unión de la metilasa Dam permite la transcripción del gen (fase ON), mientras que cuando se une OxyR la transcripción se bloquea (fase OFF)³⁰³. Observando los resultados (Fig. 4.26), se puede apreciar en las dos fusiones una fase OFF (rép. 1 y 2) como una fase ON (rép. 3 y 4). Aparentemente, en la fase ON la actividad de la fusión pUJ8-prom4511 es más elevada, en comparación con la fusión pUJ8-prom2242. Sin embargo, las diferencias entre ambas no son significativas, y están muy lejos de las diferencias transcripcionales de los genes *irmA2244* e *irmA4509*, observadas previamente con fusiones cromosómicas con el gen reportero *lacZ*.

Si el gen *agn43* formase una misma unidad transcripcional con los genes *yeeR* e *irmA* (EC042_2243 y EC042_2244), se debería observar también actividad transcripcional en el gen *irmA2244*. En cambio, no se ha apreciado

expresión de proteína (Fig. 4.13) ni actividad transcripcional (Fig. 4.15). Como se ha comentado anteriormente, los estudios realizados previamente en el gen *irmA* han tenido lugar en la cepa UPEC CFT073. De manera semejante, la mayoría de estudios llevados a cabo sobre el gen *agn43* también han sido realizados en esta cepa^{303,305}. Planteamos como hipótesis que la regulación del gen *irmA* en la cepa 042 tenga lugar de una manera diferente o, incluso, dependa de un promotor diferente que en la cepa uropatógena.

4.2.4.2 Determinación del inicio de transcripción del gen *irmA* en la cepa 042 por *Walking RT-PCR*

Se decidió comenzar con la localización del promotor del gen *irmA* en la cepa 042, mediante un experimento de *Walking RT-PCR*. De esta manera, podríamos confirmar que el gen *irmA* no forma una unidad transcripcional con los genes *agn43* e *yeeR* en la cepa 042. Esta técnica se ha llevado a cabo tanto con la cepa WT como con los mutantes simples *irmA2244* e *irmA4509*. En primer lugar, se crecieron los cultivos hasta fase estacionaria y se procesaron para extraer el ARN mensajero (apartado 3.7.1, Materiales y métodos). Este ARN se retrotranscribió, obteniendo el ADN complementario (ADNc) como producto. Paralelamente, se diseñaron oligonucleótidos directos en distintas posiciones entre los genes *irmA* e *yeeR* y, como oligonucleótido reverso, se diseñó *irmA* rv. La posición de los oligonucleótidos diseñados para este experimento se encuentra resumida en la figura 4.27.



Fig. 4.27. Esquema de la posición de los oligonucleótidos diseñados para el experimento de *Walking RT-PCR*. La secuencia de los mismos se encuentra en la Tabla 3.3 y la posición en el cromosoma en la figura anexa A2.

Como se puede ver, la posición de los oligonucleótidos directos está cada vez más alejada respecto al reverso. De esta manera, se “camina” por el cromosoma y se puede observar la longitud del ARN mensajero, gracias a la amplificación del ADN complementario.

Una vez diseñados los oligonucleótidos y obtenido el ADNc, se llevaron a cabo las 3 reacciones de PCR diseñadas. Como control positivo de la amplificación, se utilizó ADN genómico. Por otra parte, el control negativo fue realizado con el ARN extraído, asegurándonos así de que no hubiera contaminaciones con ADN durante la extracción (Fig. 4.28).

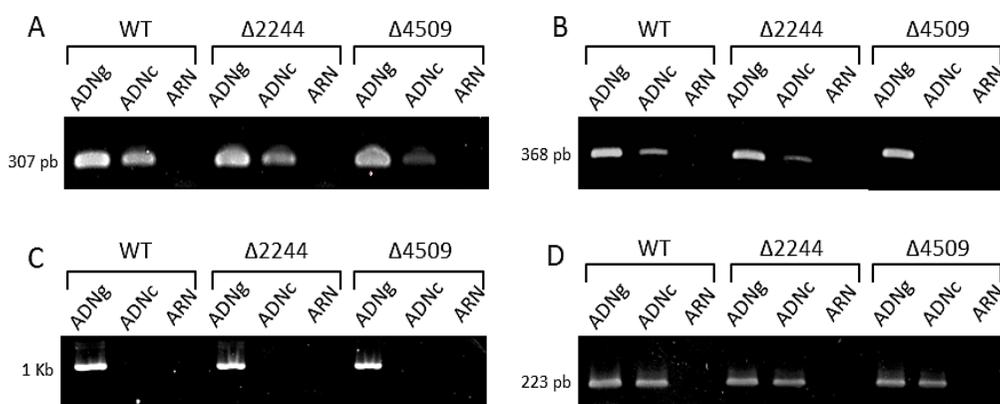


Fig. 4.28. Resultados obtenidos en las reacciones de amplificación por PCR con la técnica de *Walking RT-PCR*. A) Pareja de oligonucleótidos *irmA* fw – *irmA* rv. B) Pareja de oligonucleótidos *int* fw – *irmA* rv. C) Pareja de oligonucleótidos *yeeR* fw – *irmA* rv. D) Control de carga amplificando el gen *gapA* (*gapA* fw – *gapA* rv). ADNg: ADN genómico, ADNc: ADN complementario.

En el mutante *irmA2244* se puede ver los resultados relacionados con la copia *irmA4509*, mientras que en el mutante *irmA4509* se ven los resultados relacionados con la copia *irmA2244*. La figura 4.28A demuestra la transcripción de ambos genes, aunque los niveles de la copia *irmA4509* (mutante *irmA2244*) parecen más elevados que los de la copia *irmA2244* (mutante *irmA4509*). En la figura 4.28B se puede ver cómo hay amplificación de la copia *irmA4509* con los

oligonucleótidos utilizados, sin embargo, no se amplificó ADNc de la copia *irmA2244*. Por otra parte, como se puede observar en la figura 4.28C, no hubo amplificación del ADNc en ninguna de las cepas utilizadas. Por último, la figura 4.28D muestra el control de carga, como es el gen de expresión constitutiva *gapA*. De esta manera, nos aseguramos que la cantidad de molde utilizado en cada reacción fue la misma.

Según estos resultados, la transcripción del gen *irmA* tiene su inicio en la región intergénica entre *yeeR* e *irmA*, ya que en esta región sí que hay amplificación (Fig. 4.28B), mientras que entre *yeeR* y *irmA* no ha habido amplificación del ADNc (Fig. 4.28C). La amplificación del ADNc de cada copia no ha sido idéntica, por lo que los niveles de transcripción de ambas copias no son iguales, tal y como habíamos indicado anteriormente. Por lo tanto, el gen *irmA* en la cepa EAEC 042 no forma una unidad transcripcional junto a los genes *agn43* e *yeeR*, y los niveles de transcripción de cada copia son diferentes en las condiciones estudiadas.

4.2.4.3 Búsqueda del promotor del gen *irmA* mediante análisis bioinformáticos y mediante la determinación del inicio de transcripción por 5' RACE

Tras los resultados obtenidos, se procedió a buscar el inicio de transcripción de las copias *irmA2244* e *irmA4509*. Para llevarlo a cabo, se realizó un ensayo de 5' RACE (apartado 3.7.4, Materiales y métodos) para determinar el extremo 5' del ARN mensajero de estos genes y corroborar los resultados obtenidos mediante *Walking RT-PCR*. Las cepas utilizadas fueron los mutantes simples *irmA2244* e *irmA4509*, consiguiendo de esta manera estudiar cada copia por separado. De manera complementaria, se usó el software en línea BPROM (apartado 3.12.5, Materiales y métodos), el cual está diseñado para la búsqueda de promotores en genomas bacterianos (Fig. 4.29).

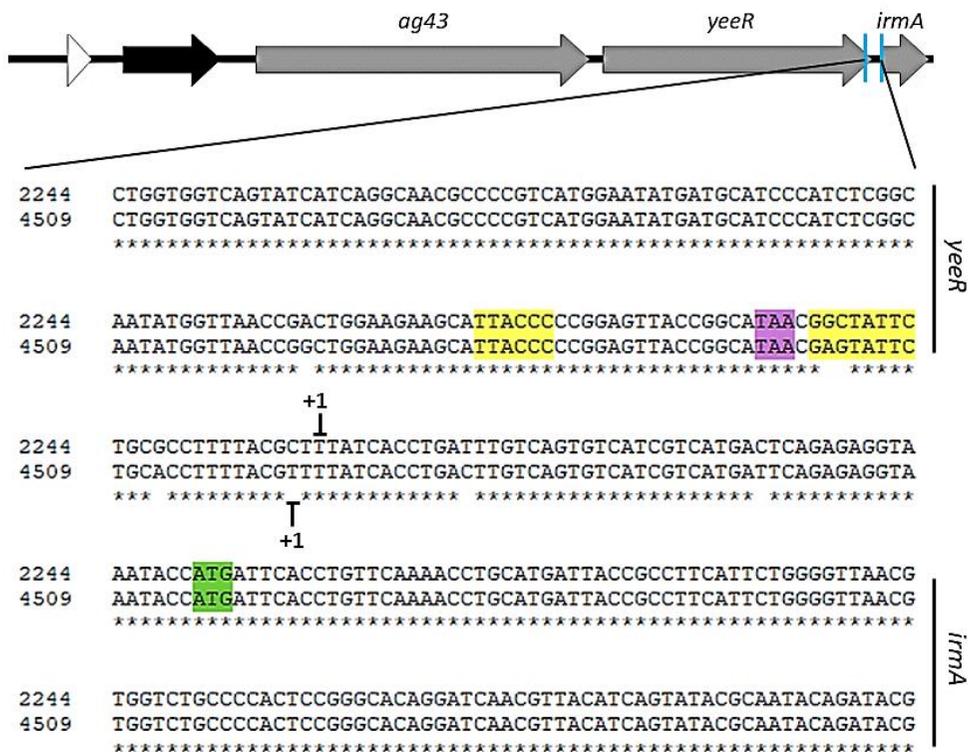


Fig. 4.29. Resultados del ensayo de 5' RACE y el predictor de promotores BPROM. El extremo 5' del ARN mensajero de las copias *irmA*₂₂₄₄ e *irmA*₄₅₀₉ encontrado mediante 5' RACE se indica con +1. En amarillo se encuentran las regiones -10 y -35 obtenidas mediante el predictor BPROM. En morado se indica el codón de parada del gen *yeeR* y en verde el codón de inicio del gen *irmA*.

Los resultados obtenidos, tanto del ensayo de 5' RACE como de la predicción de promotores con BPROM, fueron muy parecidos (Fig. 4.29). El extremo 5' de los ARN mensajeros de las dos copias se encuentra aguas arriba del codón de inicio del gen *irmA* (+1), en la región intergénica entre los genes *yeeR* e *irmA*. Los datos obtenidos mediante el predictor de promotores BPROM apoyan este resultado, ya que coinciden con bastante exactitud. Estos resultados concuerdan con el experimento de *Walking RT-PCR* realizado en el apartado anterior. Por lo tanto, en la cepa 042, los genes *agn43-yeeR-irmA* no forman una única unidad transcripcional.

Como se puede observar, en la secuencia codificante de las dos copias del gen *irmA* no se han producido cambios entre las dos copias, mientras que, en la región intergénica, sí que los hay. Es posible que estos cambios sean críticos para la expresión de los genes, dando lugar a una expresión diferente de las dos copias del gen *irmA*.

4.2.4.4 Actividad del promotor del gen *irmA*

El ensayo de 5' RACE determinó el inicio de transcripción de ambas copias del gen *irmA*. Estas dos copias presentan unos niveles de expresión diferentes, tal y como se puede ver en la figura 4.13. Este patrón diferente también se puede apreciar en los niveles de transcripción (Fig. 4.15). Debido a que la expresión del gen *irmA* no depende del promotor del gen *agn43* en la cepa 042, procedimos a averiguar dónde se encuentra exactamente y la actividad que tiene. Como ya se ha indicado, a diferencia de la secuencia codificante, la región intergénica contiene pequeños cambios entre las dos copias. Es posible que estos cambios tengan un papel fundamental en la transcripción de ambas copias.

Como punto de partida, se utilizó el análisis realizado mediante el predictor de promotores BPPROM y del ensayo de 5' RACE (Fig. 4.29). De manera homónima al estudio del promotor del *agn43* previamente realizado (Fig. 4.26), se construyeron fusiones transcripcionales en el plásmido pUJ8. Los fragmentos amplificados se pueden ver en detalle en la figura 4.30. Estos fragmentos fueron denominados prom2244 y prom4509, los cuales tienen una longitud de 90 pb. Se digirieron con las enzimas *EcoRI-BamHI* y se ligaron en el plásmido. Las colonias transformantes obtenidas se secuenciaron para confirmar la obtención de ambas fusiones.

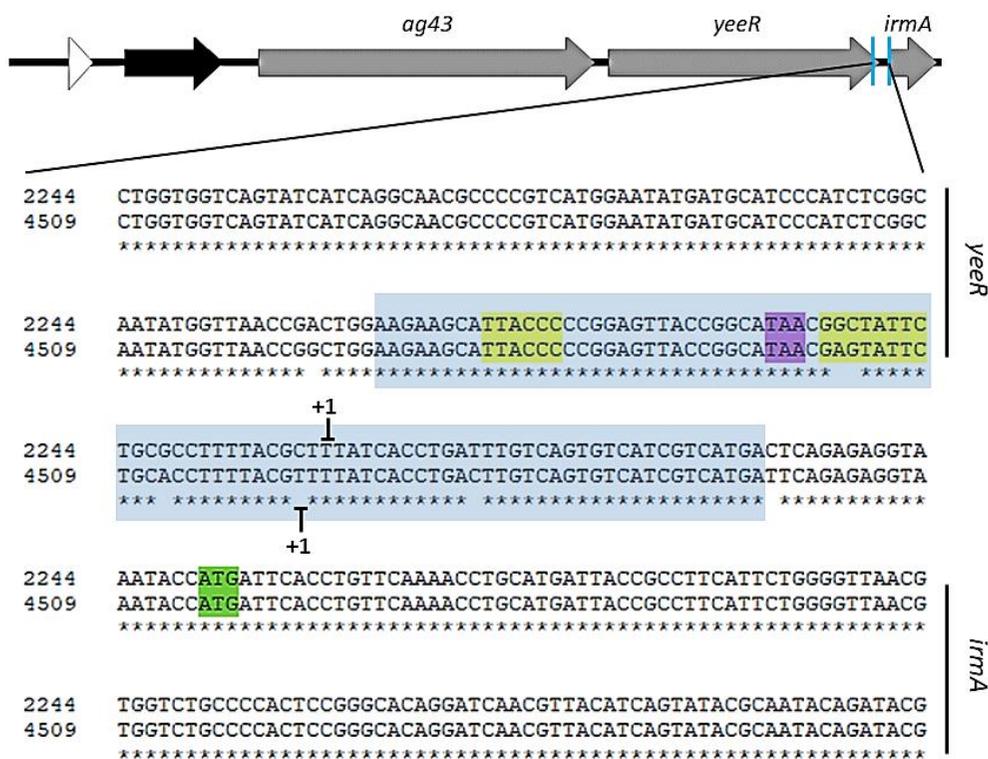


Fig. 4.30. Fragmento utilizado (recuadro azul) para la construcción de las fusiones transcripcionales en el plásmido pUJ8. El extremo 5' del ARN mensajero de las copias *irmA*₂₂₄₄ e *irmA*₄₅₀₉ encontrado mediante 5' RACE se indica con +1. En amarillo se encuentran las regiones -10 y -35 obtenidas mediante el predictor BPROM. En morado se indica el codón de parada del gen *yeeR* y en verde el codón de inicio del gen *irmA*.

Seguidamente, se crecieron cultivos en medio LB a 37 °C hasta fase estacionaria, recogiendo muestras y procesándolas para valorar la actividad β-galactosidasa (Fig. 4.31).

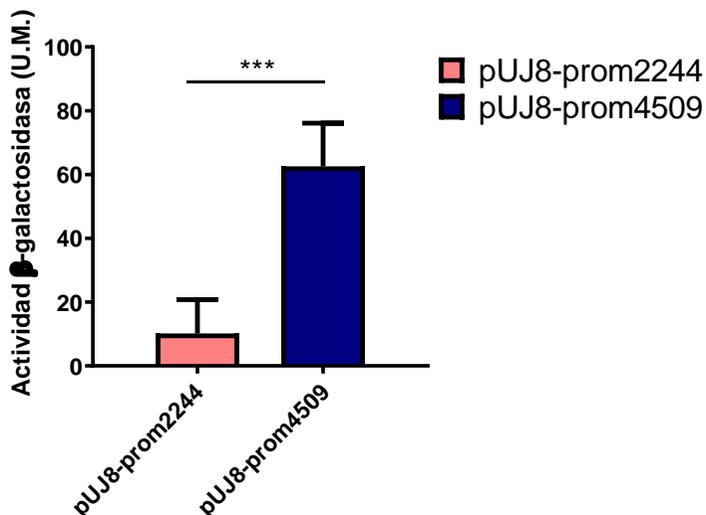


Fig. 4.31 Actividad β -galactosidasa (U. M.) de las fusiones transcripcionales plasmídicas pUJ8-prom2244 y pUJ8-prom4509. Valores obtenidos del ensayo por triplicado. ***P < 0,001

Los resultados muestran actividad en ambas copias, por lo tanto, el promotor sí que se encuentra dentro de la región seleccionada. De la misma forma que en los resultados previos, las dos copias muestran un patrón de expresión diferente, siendo mayor la actividad en la copia *irmA4509*. Consecuentemente, los cambios nucleotídicos existentes entre ambas copias tienen un papel fundamental. Para confirmarlo, se decidió realizar una mutagénesis dirigida mediante la técnica de *quickchange*, usando las fusiones plasmídicas obtenidas en el plásmido pUJ8 (apartado 3.6.5, Materiales y métodos). Se realizaron cuatro cambios puntuales en la región promotora de la copia *irmA4509*, convirtiéndola en la región promotora de la copia *irmA2244* (Fig. 4.32).

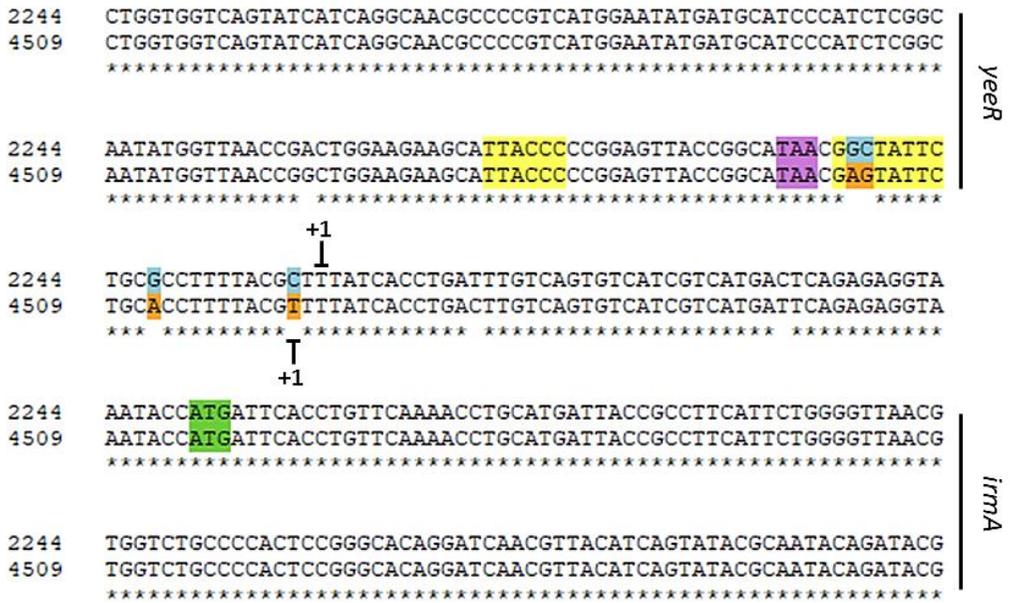


Fig. 4.32. Región promotora de las dos copias del gen *irmA* ligadas en el plásmido pUJ8. En naranja se encuentran los nucleótidos que se han cambiado en la copia *irmA4509* por los de la copia *irmA2244*, en azul. El extremo 5' del ARN mensajero de las copias *irmA2244* e *irmA4509* encontrado mediante 5' RACE se indica con +1. En amarillo se encuentran las regiones -10 y -35 obtenidas mediante el predictor BPROM. En morado se indica el codón de parada del gen *yeeR* y en verde el codón de inicio del gen *irmA*.

Una vez conseguidos estos cuatro cambios (pUJ8-mut*prom4509), se transformó el plásmido en la cepa 042. Se crecieron cultivos en LB a 37 °C hasta fase estacionaria, momento en el que se recogieron muestras y se trataron para realizar un ensayo de β-galactosidasa (Fig. 4.33).

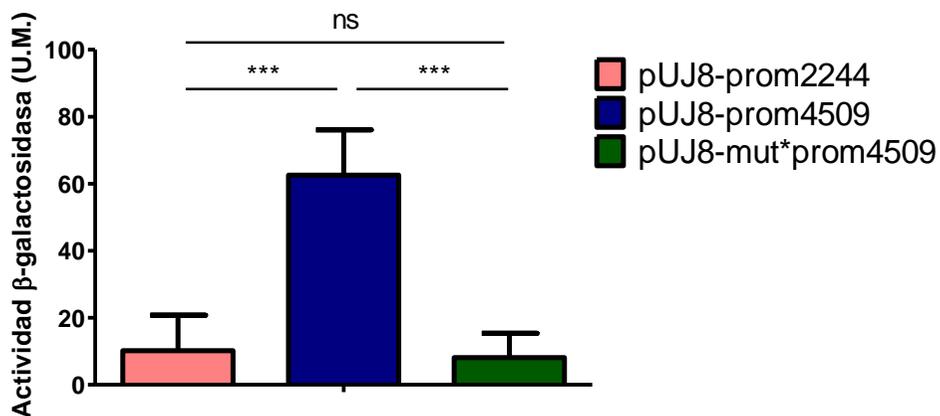


Fig. 4.33. Actividad β -galactosidasa (U.M.) de las fusiones transcripcionales plasmídicas pUJ8-prom2244, pUJ8-prom4509 y la construcción obtenida mediante *quickchange* pUJ8-mut*prom4509. Valores obtenidos del ensayo por triplicado. ns no significativo, *** $P < 0,001$

Los resultados demuestran que, al convertir el promotor de la copia *irmA4509* en el promotor de la copia *irmA2244*, los niveles de expresión cambian. Por lo tanto, estos cambios mínimos que se aprecian en los promotores de las dos copias del gen *irmA* son críticos en los niveles de expresión de ambas.

En el caso de la copia *irmA4509*, se procedió a confirmar la posición del promotor mediante un ensayo de inmunodetección de la proteína. Para ello, en la cepa 042 4509::3X se llevó a cabo la mutagénesis del mismo fragmento utilizado en la construcción del plásmido pUJ8-prom4509. De manera homónima, en la misma cepa, se realizó la mutagénesis del promotor del gen *agn43*, previamente estudiado mediante otra ligación con el plásmido pUJ8 (pUJ8-prom4511). Una vez obtenidas las construcciones, se crecieron cultivos a 37 °C en medio LB y se obtuvieron los extractos proteicos totales para la detección de la proteína IrmA mediante inmunodetección (Fig. 4.34).

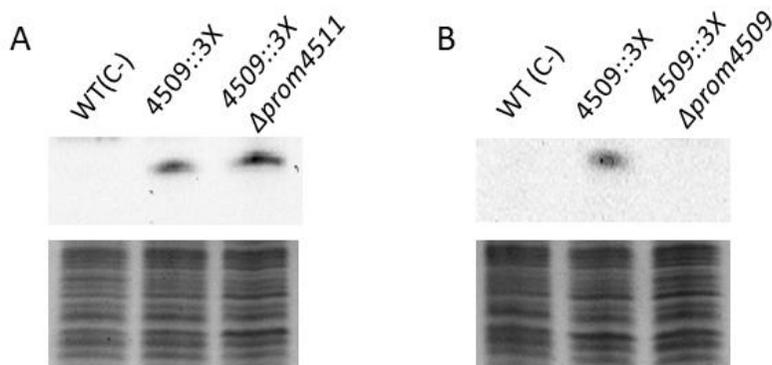


Fig. 4.34. Efecto de la mutagénesis de promotores sobre la expresión de la proteína IrmA4509. A) Mutagénesis del fragmento prom4511. B) Mutagénesis del fragmento prom4509. Como control negativo se utilizó la cepa 042 WT. Panel superior, inmunodetección realizada usando el anticuerpo monoclonal anti-FLAG. Panel inferior, tinción de Coomassie de los extractos totales utilizados para la inmunodetección, demostrando que los mismos contenían cantidades equivalentes de proteína.

La inmunodetección de la proteína IrmA demuestra como la expresión de la proteína IrmA4509 no depende del promotor del gen *agn43*, ya que se mantienen los mismos niveles en ausencia de esta región (Fig. 4.34A). Sin embargo, la mutagénesis del fragmento utilizado en la ligación pUJ8-prom4509 provoca que el gen deje de expresarse (Fig. 4.34B), confirmando la presencia de la región promotora en la zona intergénica entre los genes *yeeR* e *irmA*.

4.2.4.5 Análisis comparativo de la región promotora del gen *irmA* entre la cepa 042 y otras cepas patógenas de *E. coli*

Los ensayos realizados en el capítulo anterior confirman la localización del promotor del gen *irmA* en la cepa 042. Además, como se ha podido observar, cambios nucleotídicos puntuales en la zona donde se sitúa el promotor son cruciales en la expresión de ambas copias del gen *irmA* (Fig. 4.31). Tal y como se ha comentado anteriormente, este gen se encuentra junto a los genes *agn43*

e *yeeR* en una región muy variable, derivada del profago CP4-44. En el caso de la cepa 042, dicha región se encuentra duplicada, pero, en otras cepas, los genes presentes en esta región solo se encuentran en una copia (Fig. 4.6). En el caso concreto de la cepa CFT073, estos genes también se encuentran duplicados, sin embargo, en una de las regiones tuvo lugar un proceso de inserción, el cual rompe la pauta de lectura³²⁸. Debido a la variabilidad que presenta esta región, decidimos realizar, en primer lugar, un alineamiento de la misma entre diferentes cepas patógenas de *E. coli*, incluyendo tanto la cepa 042 como la cepa CFT073 (Fig. 4.35).

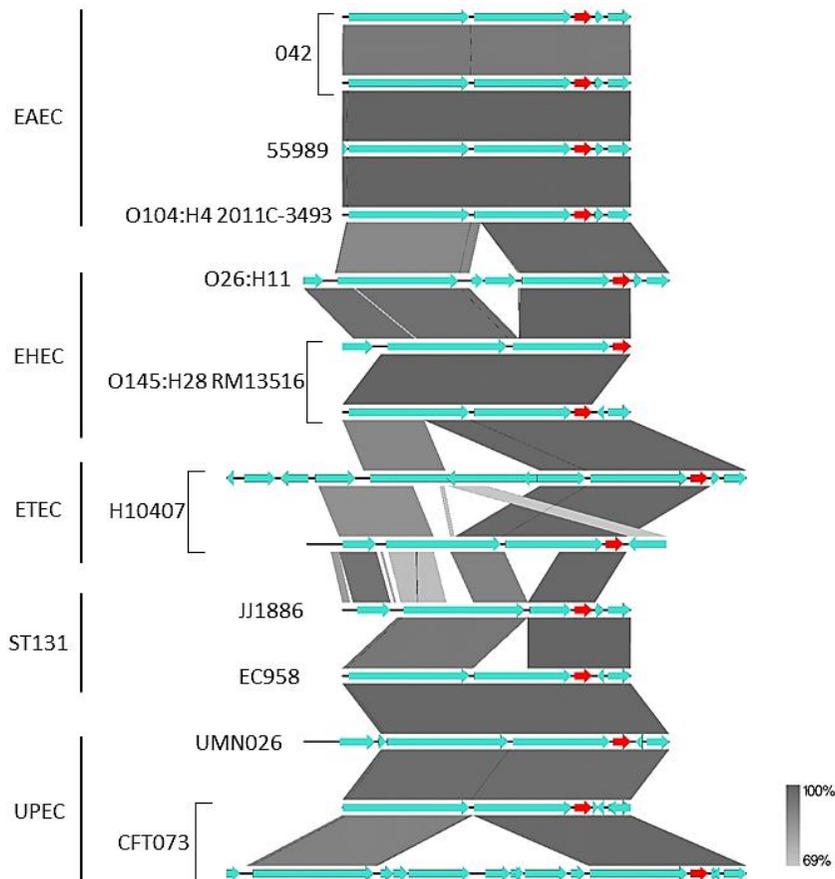


Fig. 4.35. Análisis comparativo de la región *agn43-yeeR-irmA* de 10 cepas patógenas de *E. coli* de diferentes patotipos. A la izquierda se muestra el patotipo al que pertenecen. En rojo se muestra el gen *irmA*. La figura fue realizada con Easyfig³³².

Como se puede observar, entre las 14 regiones que se han comparado se han producido procesos de inserción o deleción en 4 de ellas. Pese a que se producen cambios en esta región, la zona intergénica entre los genes *yeeR* e *irmA*, donde se ha determinado la presencia del promotor del gen *irmA*, mantiene un porcentaje de similitud muy elevado entre todas las cepas estudiadas. Consecuentemente, se procedió a determinar si, de la misma manera que en la cepa 042, se han producido cambios puntuales en esta región en otras cepas. Se realizó un alineamiento de la secuencia nucleotídica, incluyendo tanto la región promotora como la secuencia codificante del gen *irmA* (Fig. anexa A3). Seguidamente, a partir de este alineamiento, se llevó a cabo un análisis filogenético, utilizando el método neighbor-joining³⁵²(Fig. 4.36).

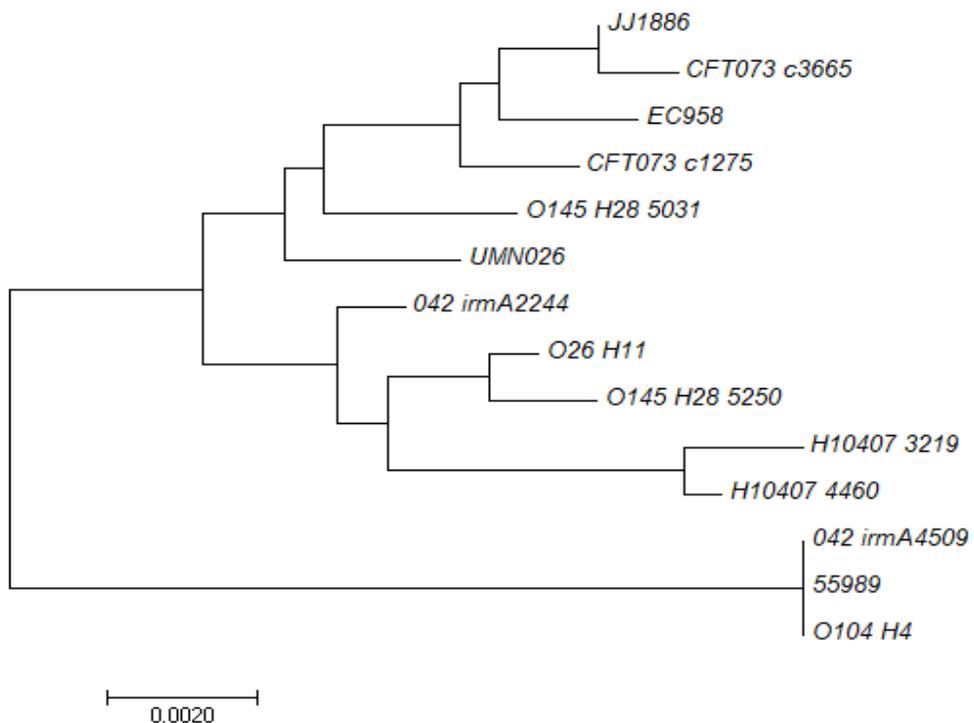


Fig. 4.36. Análisis filogenético realizado a partir de la secuencia promotora y codificante del gen *irmA* de cepas pertenecientes a diferentes patotipos. La barra muestra 0,002 sustituciones por posición. Figura realizada mediante el programa MEGA7³⁵³.

El árbol filogenético muestra dos ramas bien diferenciadas. En el caso de la cepa 042, el gen *irmA2244* es evolutivamente más cercano a la mayoría de secuencias de este gen que se han utilizado en el análisis. En cambio, el gen *irmA4509* se encuentra en la otra rama, siendo la secuencia codificante idéntica a la de las otras cepas del patotipo enteroagregativo analizadas.

4.3 Estudio de la región 55-60, presente en cuatro copias en el genoma de la cepa 042

El análisis realizado con el objetivo de determinar la presencia de duplicaciones distribuidas en los diferentes patotipos de *E. coli*, permitió identificar diferentes genes duplicados en la mayoría de cepas patógenas de *E. coli*. Tomando como modelo de la cepa 042, existen diferentes genes, tanto en la región 1 como en la región 2, que se encuentran duplicados en la mayoría de cepas analizadas (Fig. 4.6). De todos ellos, es destacable la presencia de un grupo de 6 genes situados de forma continua, numerados del 55 al 60 y ubicados en la región 2 (Fig. 4.37).

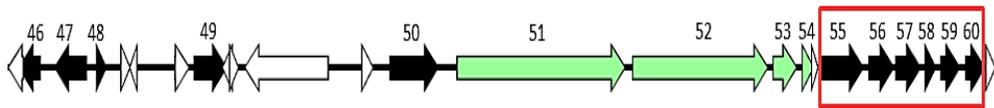


Fig. 37. Detalle de la región 2 que contiene genes duplicados en la cepa 042 de *E. coli*. Los genes duplicados de esta región se muestran en negro y con su número (Tabla 4.5), mientras que los genes que no se han duplicado están en blanco. Los genes en verde corresponden a la región *flu-yeeR-aec69(irmA)-aec70*. Con un recuadro rojo se muestran los 6 genes cuya duplicación se encuentra ampliamente distribuida entre todos los patotipos.

Como se ha podido observar en apartados anteriores, en las cepas CFT073 y O145:H28 RM13516 estos 6 genes también han sufrido procesos GDA. En el caso de la cepa CFT073, se encuentran numerados del 7 al 13 en la región 1 (Tabla anexa A1) y en la cepa O145:H28 RM13516 del 71 al 76 en la región 3 (Tabla anexa A2). La correspondencia de estos 6 genes en las 3 cepas estudiadas y su función se encuentra resumida en la Tabla 4.7.

Tabla 4.7. Relación de los 6 genes entre las cepas EAEC 042, UPEC CFT073 y EHEC O145:H28 RM13516 de *E. coli*. Los números corresponden a los grupos de genes duplicados, numerados en orden empezando desde el origen del mapa cromosómico de cada cepa.

O42	CFT073	O145:H28 RM13516	Descripción
55	13	71	Aec71
56	12	72	Putative antirestriction protein
57	11	73	Putative DNA repair protein
58	10	74	Conserved hypothetical protein
59	8	75	Antitoxin
60	7	76	CbtA toxin

Como se puede ver en la Tabla 4.7, la función de 4 de estos 6 genes se desconoce. Este hecho llama la atención ya que, como se puede observar en los resultados obtenidos previamente en esta Memoria, se encuentran presentes en todos los patotipos y en un número elevado de copias cada uno de ellos. Por ello, se decidió dilucidar el posible papel de estos genes en la virulencia, tomando como referencia la cepa 042 de *E. coli*.

4.3.1 Deleción de las cuatro copias del clúster 55-60 de la cepa 042 y estudio del *fitness* celular

Previo a cualquier análisis, fue necesario delecionar las cuatro copias del grupo de 6 genes amplificado en la cepa 042. Gracias a que estos 6 genes se encuentran juntos en 4 regiones diferentes del cromosoma de la cepa 042, y que el tamaño de cada una de estas regiones es de alrededor de 3 Kb, la deleción de todas las copias se pudo realizar de una manera sencilla. Se trabajó con cada región como un fragmento único, realizando la deleción de toda la región cromosómica mediante fragmentos de PCR 4 veces (apartado 3.6.1, Materiales y métodos) (Fig. 4.38).

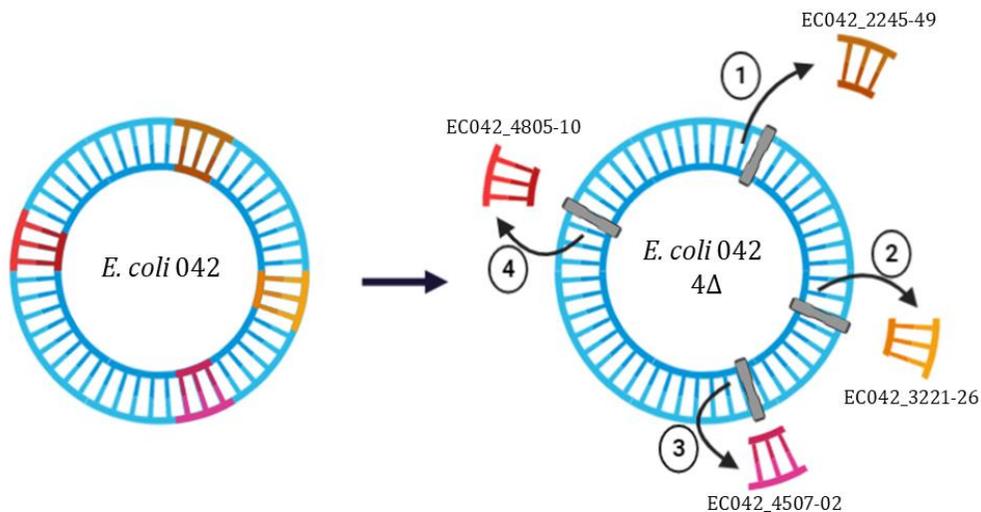


Fig. 4.38. Esquema de la metodología empleada para la delección de las 4 regiones génicas de la cepa 042 mediante la recombinación de fragmentos de PCR. La delección de las 4 regiones se realizó de manera secuencial (números del 1 al 4). En gris se muestran las secuencias FRT resultantes del proceso de delección de cada región.

Una vez realizadas estas delecciones, se obtuvo la cepa mutante para todos los genes, la cual se denominó 042 4Δ (Tabla 3.1, Materiales y métodos). A continuación, se procedió a evaluar el efecto de estas mutagénesis sobre el crecimiento de la cepa 042 mediante curvas de crecimiento en diferentes condiciones (Fig. 4.39).

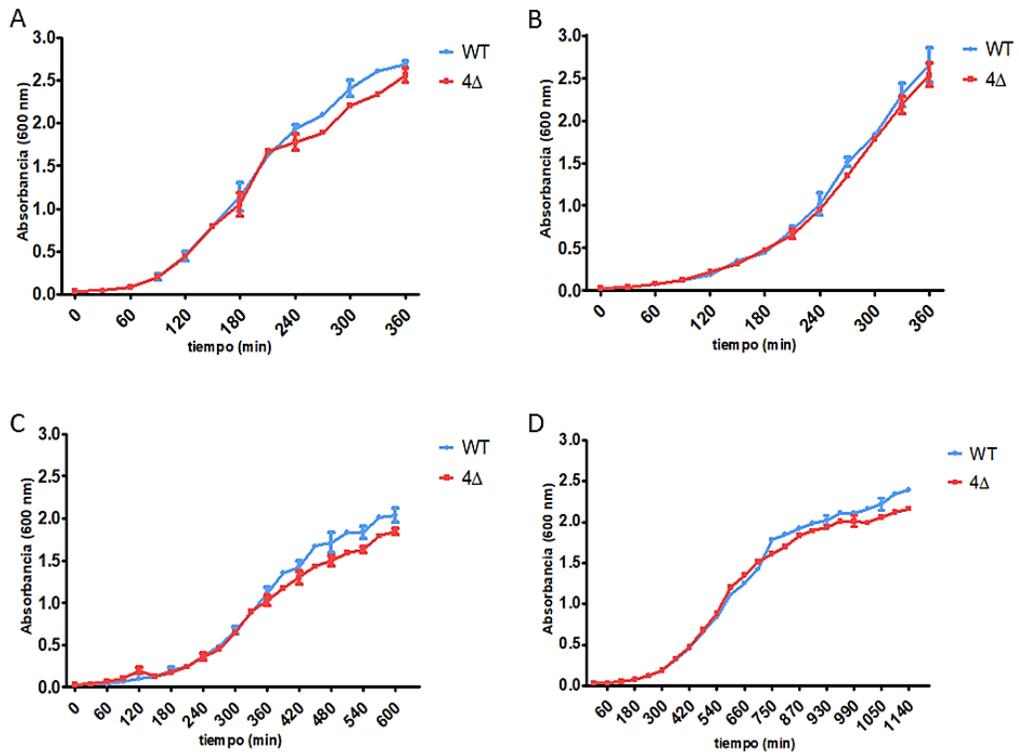


Fig. 4.39. Curvas de crecimientos de la cepa 042 WT y del mutante de las 4 regiones de genes (042 4 Δ) en diferentes condiciones. A) Medio LB a 37 °C. B) Medio LB a 25 °C. C) Medio M9 a 37 °C. D) Medio M9 a 25 °C. Valores obtenidos del ensayo por triplicado.

En las condiciones estudiadas de temperatura y disponibilidad de nutrientes, los resultados obtenidos muestran un efecto mínimo de la ausencia de estos genes en el crecimiento de la cepa 042 de *E. coli*. Por lo tanto, es posible que estos genes no sean vitales para la supervivencia del microorganismo cuando crece en condiciones óptimas, pero sí que tengan un papel importante en la infectividad y/o virulencia de la cepa estudiada.

4.3.2 Estudio de la formación de biofilm en la cepa 042

Como se ha indicado previamente en el apartado de la Introducción, la formación de biofilm es una de las características principales de la patogénesis de cepas enteroagregativas¹⁹⁶. Esta adhesión a superficies está mediada por fimbrias de adherencia agregativa, las cuales en la cepa 042 se encuentran codificadas en el plásmido pAA2. Con el objetivo de determinar si alguno de los genes presentes en la región mutagenizada está involucrado en la adhesión, se procedió a realizar un ensayo de formación de biofilm en diferentes medios de cultivo (Fig. 4.40). En este caso, como control negativo se utilizó la cepa 042 donde el plásmido pAA2 ha sido curado (denominada 042 pAA-, Tabla 3.1 en Materiales y métodos). El protocolo utilizado fue descrito por Sheikh y colaboradores³⁴⁸, tal y como se encuentra detallado en el apartado 3.10, Materiales y métodos.

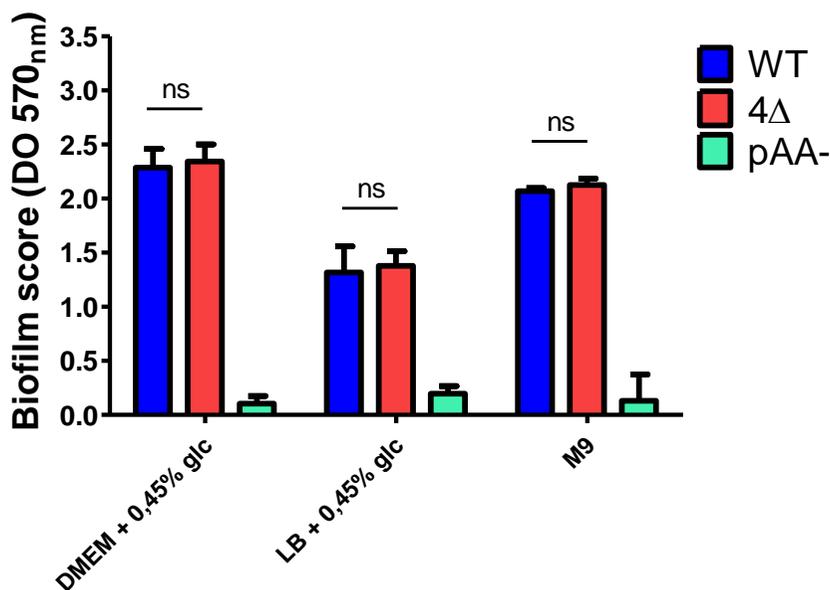


Fig. 4.40. Formación de biofilm en medio DMEM, LB y M9 de las cepas WT, 4Δ y pAA-. Resultados del ensayo por duplicado. ns, no significativo.

Las diferencias entre la cepa salvaje y el mutante de las 4 regiones no son significativas, independientemente de las condiciones analizadas. Por lo tanto, aparentemente, ninguno de estos 6 genes está involucrado en la formación de biofilm de la cepa 042 de *E. coli*.

4.3.3 Ensayo de infectividad usando como organismo modelo *Galleria mellonella*

La patogénesis de las cepas EAEC sigue sin comprenderse en su totalidad, sobre todo debido a la gran heterogeneidad genética que presentan las cepas de este patotipo. En este sentido, la compleja relación entre huésped y patógeno es crítica para la comprensión de la infectividad de ciertas cepas de *E. coli*. Aunque se han propuesto diferentes modelos animales para estudiar el proceso infectivo de este patógeno, ninguno de ellos ha mostrado todas las manifestaciones clínicas de la patogénesis³⁵⁴.

Recientemente, las larvas de *Galleria mellonella* (llamada comúnmente polilla de la cera) se han establecido como un modelo interesante para estudiar la infección de diversos patógenos³⁵⁵⁻³⁵⁷, entre los cuales destaca el patotipo EAEC de *E. coli*³⁵⁸. Este modelo animal presenta muchas ventajas, como, por ejemplo, la rapidez para obtener larvas, la capacidad de realizar experimentos a 37 °C o, más importante, la correlación de los resultados obtenidos entre este modelo y otros modelos animales^{355,359}. En el caso de cepas EAEC, el modelo de *Galleria* resulta muy interesante como herramienta para identificar nuevos marcadores de virulencia y el desarrollo de vacunas³⁵⁸. Además, este sistema no está sujeto a los complejos procedimientos de autorización de experimentos de manipulación de animales como sucede, por ejemplo, con los ratones de laboratorio. Por ello, se decidió utilizar las larvas de *Galleria mellonella* para analizar si la delección de las cuatro copias del grupo de genes 55-60

(EC042_2245-49, EC042_3221-26, EC042_4507-02, EC042_4805-10) altera la virulencia de la cepa 042 (apartado 3.11, Materiales y métodos). Para estos experimentos, se utilizaron tanto la cepa salvaje WT como los mutantes de las copias del gen *irmA* (042 *irmAnull*) y la cepa mutante para todas las copias de la región ampliamente distribuida entre todos los patotipos (042 4 Δ), siguiendo la metodología detallada en el apartado 3.11, Materiales y métodos. Los resultados obtenidos se encuentran en la Figura 4.41.

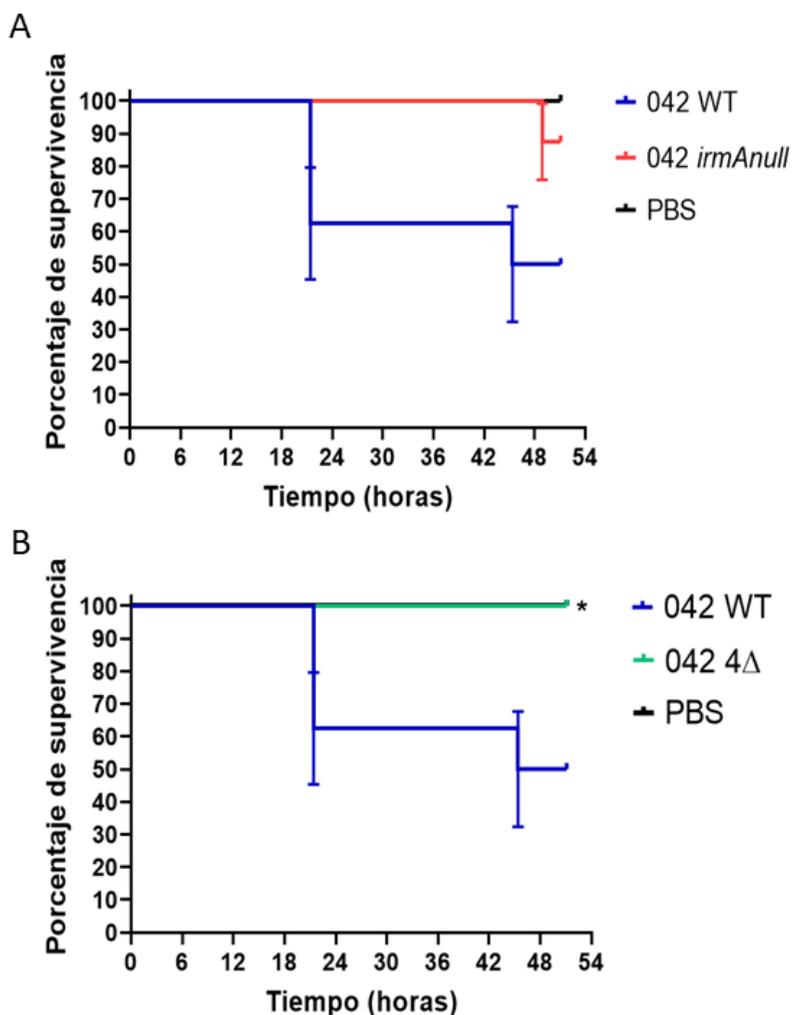


Fig. 4.41. Ensayo de supervivencia en *Galleria mellonella*. A) Comparación entre las cepas WT e *irmAnull*. B) Comparación entre las cepas WT y 4 Δ . Como control negativo se utilizaron larvas inoculadas con PBS. *P < 0,05

Los resultados en el ensayo de supervivencia muestran como, tras 54 horas, la infección de las larvas con la cepa salvaje tiene como consecuencia la reducción de la viabilidad de las mismas en un 50 %. Por otra parte, en el caso de la cepa 042 *irmAnull*, la infección provoca solo una reducción del 15 %. Esta diferencia no llega a ser estadísticamente significativa, en comparación con la cepa salvaje (Fig. 4.41A). Por último, la infección con la cepa 042 4Δ no conlleva una reducción de la viabilidad de las larvas, hecho que, en comparación con la cepa salvaje, sí que es significativo (Fig. 4.41B). Por lo tanto, en el modelo de infección de larvas de *Galleria mellonella*, la mutagénesis de la región ampliamente distribuida provoca una atenuación de la virulencia de la cepa 042. Aunque no es estadísticamente significativa, la mutagénesis de las dos copias del gen *irmA* sugiere una tendencia en el sentido que también afecta a la virulencia en el modelo utilizado.

5. Discusión

El análisis bioinformático del genoma de la cepa EAEC 042 permitió asociar la presencia de copias adicionales del gen *hha* (parálogos *hha2* y *hha3*) con los genes *yeeR* y *aec69* (*irmA*). Estos dos últimos se encuentran en una región genómica que incluye también al gen *flu*, aguas arriba de *yeeR*, y *aec70* *aec71*, aguas abajo de *irmA*. Es ésta una región extremadamente variable en muchas cepas de *E. coli*. Pertenece al profago CP4-44 y, a título de ejemplo, en la cepa comensal MG1655 solamente se encuentra funcional el gen *flu*, encontrándose *yeeR* parcialmente deletado e *irmA*, *aec70* y *aec71* están ausentes. En cambio, en la cepa 042, existen dos copias completas de todos ellos. Pertenecen, por tanto, estos genes al “pool” de genes variables de *E. coli*. Más allá de la relación gen-regulador establecida en este trabajo, al menos para *irmA* y Hha (y Hha2), este estudio nos abrió una nueva perspectiva en el genoma de *E. coli*: la existencia de un número significativo, pero no caracterizado hasta el momento, de genes duplicados en este microorganismo. Por tanto, nos pareció importante obtener un mapa detallado del impacto de las duplicaciones génicas en *E. coli*.

5.1 Distribución de genes duplicados en *E. coli* y su relación entre patotipos

El genoma de partida para estos estudios fue la propia cepa EAEC 042, ampliándose posteriormente el estudio a los genomas de otras dos cepas, pertenecientes a patotipos diferentes. Los parámetros seleccionados, en base a los cuales se han filtrado los resultados, fueron restrictivos. Filtrando los resultados de esta manera, los genes obtenidos son aquellos que han sufrido un proceso de duplicación y no han tenido grandes cambios respecto a su ancestro común. Esto puede ser debido a que se trata de una duplicación reciente o a que el aumento de la dosis génica resulta útil por sí mismo. Pero, también, es posible que haya tenido lugar un proceso de neofuncionalización o

subfuncionalización, donde los cambios necesarios que se han producido hayan sido mínimos.

Es importante resaltar que, de los 80 genes que se encuentran duplicados en el genoma de la cepa 042, un gran porcentaje son de función desconocida (Tabla 4.5), mientras que algunos codifican transposasas (Fig. 4.4, marcados en negro). El hecho de que se agrupen en tres regiones diferentes permite inferir información acerca del origen y del significado de tales duplicaciones. En el caso de la región 1, los genes duplicados provienen de un fago insertado en el genoma, el cual se encuentra distribuido en muchas cepas. Estos datos claramente apoyan que, en muchos casos, la presencia de duplicaciones en los genomas bacterianos se debe a procesos HGT, como se había constatado previamente²⁹¹. Es interesante que un número significativo de genes estén organizados en una misma unidad transcripcional (Fig. 4.5), hecho que sugiere que su regulación puede estar coordinada y su expresión inducirse como respuesta a un estímulo específico. Por lo que hace referencia a la región 2, se trata también de genes ampliamente distribuidos en otros patotipos. El orden de los genes duplicados en dicha región sugiere que el origen de dicha duplicación pudo ser un proceso de inversión. Por lo tanto, las regiones 1 y 2 ejemplifican mecanismos diferentes de generación de duplicación. Los genes de la región 3 representan un significado diferente, ya que corresponden a duplicaciones específicas de un patotipo determinado.

En el caso de las cepas CFT073 y O145:H28 RM13516, ambas poseen un número mayor de genes duplicados que la cepa 042, en concreto, 94 y 154, respectivamente. En el caso de estas cepas, los genes se encuentran en 6 regiones principales del genoma. Es reseñable el número tan elevado de genes duplicados, sobre todo en la cepa EHEC, y el porcentaje tan significativo de los mismos cuya función se desconoce (Tabla anexa A1 y A2). Curiosamente, como norma general en las 3 cepas analizadas, cada una de las copias del gen duplicado se encuentra en cadenas diferentes del ADN (Figs. 4.4 y 4.7).

El análisis comparativo de estas duplicaciones, realizado frente a cepas pertenecientes a diferentes patotipos, aporta información sobre la presencia y distribución de los duplicados. La existencia de un número significativo de duplicaciones distribuidas en todos los patotipos analizados, y ausente en cepas comensales, sugiere que pueden estar implicadas en la patogénesis de las mismas. Los genes 55 al 60 de la región 2 de la cepa 042, encontrados también en las cepas CFT073 y O145:H28 RM13516, están duplicados o amplificados en prácticamente todas las cepas analizadas. Sin embargo, tanto la función de estos genes, como la de muchos otros duplicados, no están definidas. Consecuentemente, el estudio en profundidad de los mismos y la asignación de funciones concretas resulta crítico para discernir cuál es el papel que desempeñan en la biología de estos patógenos.

A partir de este análisis comparativo, también se han encontrado tanto relaciones entre patotipos como duplicaciones restringidas a un patotipo específico, e incluso, a una cepa concreta. Como ejemplo de esta especificidad, se puede ver la región 3 de la cepa 042 (Fig. 4.6) o la región 4 de la cepa O145:H28 RM13516 (Fig. 4.9). La región 4 de la cepa O145:H28 se encuentra conservada entre todas las cepas del patotipo EHEC. Como se ha mencionado anteriormente, el origen de algunos genes de esta región es fágico, pero también se encuentran dos reguladores transcripcionales y genes de función desconocida (Tabla anexa A2). Debido a la especificidad de patotipo que muestra esta región, el estudio de los genes que engloba resultaría interesante para una mejor caracterización del patotipo EHEC. Un ejemplo más de esta especificidad se puede observar en la región 5 de la cepa CFT073 (Fig. 4.8). La mayoría de genes que se encuentran en esta región de la cepa CFT073 codifican proteínas putativas relacionadas con fimbrias. El gran número de tipos de fimbrias que poseen las cepas UPEC está íntimamente relacionada con la patogénesis de este patotipo^{46,47,50}. Por lo tanto, el estudio de la función de estos

genes puede contribuir a entender mejor los mecanismos de virulencia del patotipo UPEC.

Respecto a las relaciones existentes entre patotipos, en el caso de las duplicaciones de la cepa 042 se observa un patrón muy similar entre las cepas del patotipo EAEC y las cepas del patotipo UPEC (Fig. 4.6). Estudios previos han sugerido la relación entre ambos patotipos, donde se han encontrado cepas que codifican marcadores característicos de los dos^{16,192,193} e, incluso, una cepa híbrida EAEC/UPEC fue causa de un brote de infecciones urinarias en Copenhague¹⁹⁴. Por lo tanto, esta hipótesis se ve apoyada por los resultados aquí obtenidos. A diferencia de esta relación que se observa en los patotipos EAEC y UPEC, los genes duplicados de la cepa EHEC O145:H28 RM13516 se encuentran en cepas de los patotipos EPEC y ETEC. Estudios filogenéticos previos resaltan la relación entre el patotipo EHEC y la cepa O55:H7, perteneciente al patotipo EPEC. Estos estudios sugieren que esta cepa adquirió los factores de virulencia característicos del patotipo EHEC mediante transferencia horizontal y recombinación, como por ejemplo la toxina Shiga⁹³⁻⁹⁵. Observando los resultados presentados en la figura 4.9, es evidente que esta cepa guarda una íntima relación con el patotipo EHEC. Muchos de los genes duplicados presentes en las cepas EHEC también lo están en la cepa EPEC O55:H7, por lo que es posible que algunos de estos procesos de duplicación se originaran en esta cepa y se han conservado. Por otra parte, respecto a la relación con el patotipo ETEC, la característica principal que permite diferenciar ambos patotipos se encuentra en las toxinas que codifican, es decir, la toxina Stx en el caso de las cepas EHEC y las toxinas ST o LT en el de las ETEC^{1,2,17,25}. Estos dos patotipos están bien diferenciados, tanto por su filogenia como por la sintomatología asociada, sin embargo, se han reportado casos debidos a híbridos recientemente. Estos híbridos EHEC/ETEC contienen los genes que codifican ambos tipos de toxinas características, mencionadas anteriormente. La combinación de múltiples factores de virulencia provoca que

estos híbridos exhiban propiedades asociadas a múltiples patotipos, hecho que demuestra la gran plasticidad y variabilidad que muestran los genomas de *E. coli*. En cuanto a su sintomatología, los híbridos EHEC/EPEC son capaces de provocar tanto diarreas acuosas como HUS, por lo que suponen una emergencia pública sanitaria emergente^{360,361}. A partir del análisis realizado, esta relación queda puesta de manifiesto, sobre todo, en la región 5 de genes duplicados de la cepa O145:H28 RM13516 (Fig. 4.9). Como se ha comentado anteriormente, la información disponible sobre los genes duplicados englobados en esta región 5 es escasa. Debido a la emergencia que representan las cepas híbridas que han surgido en los últimos años, un estudio en profundidad sobre estos genes de función desconocida resulta necesario.

Queda patente que, realizando un análisis de genes duplicados, es posible saber fácilmente cuáles son aquellos que se encuentran distribuidos en mayor medida entre los diferentes patotipos de *E. coli*. Por lo tanto, un estudio más amplio de genes duplicados en los diferentes patotipos puede aportar mucha más información sobre su hipotética función. Tanto estos genes como sus productos pueden ser candidatos para el diseño de nuevas vacunas. Concretamente, en el caso de las cepas EPEC, no existe una vacuna eficaz en humanos contra este patógeno³⁶². Es posible que alguno de los productos codificados por genes duplicados sea antigénico, por lo que pueden ser candidatos para el desarrollo de nuevas vacunas de *E. coli*.

5.2 Papel del sistema H-NS/Hha en la regulación de genes duplicados

En cuanto a la regulación de los genes duplicados en la cepa 042, se ha realizado un estudio focalizado en la expresión de estos genes, tanto en la cepa salvaje como en mutantes *hns* y *hhanull*. Teniendo en cuenta los niveles de

expresión en la cepa salvaje y en los mutantes *hhanull* y *hns*, puede afirmarse que, bajo ciertas condiciones (crecimiento a 37 °C en medio LB), el sistema H-NS/Hha inhibe la expresión de un número significativo de dichos genes (Tabla 4.6). Considerando el origen fágico de los genes de la región 1, el sistema de regulación H-NS/Hha estaría llevando a cabo un silenciamiento xenogénico^{236,237}, posibilitando la incorporación de nuevas funciones sin afectar al *fitness* celular. Sin embargo, en las otras dos regiones, los datos resaltan una nueva función de este sistema, silenciando los genes que se encuentran duplicados en la cepa 042. Es posible que la duplicación de genes regulados por este sistema H-NS/Hha sea la causa del origen de nuevas copias de estos reguladores, con la finalidad de mantener estable toda la red reguladora. La derepresión de los genes regulados por el sistema H-NS/Hha puede que tenga lugar si las condiciones del entorno cambian. Por lo tanto, la duplicación génica puede resultar ventajosa si las dos copias muestran patrones de expresión diferentes o si responden a estímulos diferentes.

5.3 Expresión y regulación del gen *irmA*

A diferencia de los resultados obtenidos previamente en las cepas UPEC CFT073 y EC958³²⁸, la transcripción del gen *irmA4509* ocurre de manera independiente a la del gen *agn43* en la cepa 042, tal y como hemos podido comprobar tanto con los ensayos de *Walking RT-PCR* (Fig. 4.28) como con los ensayos de 5' RACE (Fig. 4.29). Además, en ausencia de la región promotora del gen *agn43*, la proteína IrmA4509 continúa expresándose (Fig. 4.34). Esta transcripción se encuentra bien regulada por diversos factores. De igual manera que ocurre en las cepas UPEC, la transcripción del gen *irmA4509* está reprimida por la acción de OxyR, ya que la delección de este regulador conlleva una sobreexpresión de la proteína IrmA4509. Dicha sobreexpresión puede observarse en la inmunodetección de la proteína IrmA con el anticuerpo

policlonal anti-IrmA (Fig. 4.12). Sin embargo, este anticuerpo presentaba una baja sensibilidad y no permitía el estudio de cada copia del gen *irmA* de la cepa 042 por separado. El marcaje de la proteína IrmA con el epítipo 3XFLAG mejoró la sensibilidad de la inmunodetección, pudiendo así detectar esta proteína, tanto en la cepa 042 salvaje como en otros fondos genéticos.

La expresión de la copia *irmA4509* se ve influenciada por factores ambientales, como son la temperatura de crecimiento o la disponibilidad de nutrientes (Fig. 4.16). Una elevada temperatura y una disponibilidad limitada de nutrientes son condiciones que pueden encontrarse en el intestino, las cuales coinciden con condiciones donde se produce un aumento de la expresión de la proteína IrmA4509. Respecto a la disponibilidad de oxígeno, la expresión de la proteína IrmA4509 se reduce en anaerobiosis. Las cepas enteroagregativas son capaces de unirse tanto a enterocitos del íleon como del colon, pero las concentraciones de oxígeno son diferentes en estos dos órganos. Es posible que la expresión de IrmA4509 tenga lugar cuando las bacterias se unen al íleon, donde la concentración de oxígeno es mayor, mientras que, cuando se unen al colon, la expresión se inhibe al haber una concentración de oxígeno menor. Por otra parte, hay que tener en cuenta que el oxígeno se distribuye en el intestino de forma radial, donde las zonas cercanas al tejido están más oxigenadas y en el lumen menos³⁶³. Por lo tanto, los factores que influyen en su expresión muestran un patrón relacionado con los procesos infecciosos de la cepa 042^{195,196}.

Respecto a la regulación que ejerce el sistema H-NS/Hha sobre el gen *irmA4509*, tiene lugar un patrón poco frecuente. Se puede observar como la expresión se encuentra reprimida por la proteína Hha, pero esta regulación no tiene lugar por la proteína H-NS (Fig. 4.20). La represión del gen *irmA4509* mediada por la proteína Hha es consistente con el análisis bioinformático realizado, ya que se correlaciona la presencia de nuevos parálogos de la proteína Hha (Hha2 y Hha3) con los genes *yeeR* e *irmA* (Fig. 4.1). Sin embargo,

como se indica, no ocurre de la misma manera con la proteína H-NS. Como regla general, la proteína Hha interacciona con H-NS, modulando su actividad y alterando su especificidad^{262,270,273}. La regulación génica mediada únicamente por la acción de la familia de proteínas Hha se ha descrito previamente, pero el mecanismo por el cual tiene lugar continúa sin esclarecerse²⁷⁴. Como ejemplo de esta disparidad en el patrón de regulación se encuentra la regulación en la formación de biofilm o la excisión de profagos, donde Hha y H-NS son antagónicos^{364,365}. Es posible que este patrón esté relacionado con la regulación de profagos, debido a que, como se ha indicado previamente, el gen *irmA* forma parte del profago CP4-44. En este sentido, serían necesarios otros ensayos para confirmarlo.

Por otra parte, la expresión de la copia *irmA2244* se encuentra silenciada en todas las condiciones que se han estudiado en la cepa 042. La razón de este silenciamiento se encuentra probablemente en la región promotora, donde han tenido lugar sustituciones nucleotídicas puntuales. La conversión de la región promotora de la copia *irmA4509* en la región de la copia *irmA2244* conlleva el silenciamiento de la misma, demostrando este resultado que los cambios nucleotídicos entre ambas copias son críticos para la expresión del gen *irmA* (Fig. 4.33). Cabe señalar que, en mutantes de la copia *irmA4509*, se ha podido detectar la expresión de la copia *irmA2244* (Fig. 4.23 y 4.24). La secuenciación del promotor de *irmA2244* en este mutante no muestra diferencias respecto a la cepa salvaje. Por lo tanto, el mecanismo subyacente a la inducción de *irmA2244* cuando cesa la expresión de *irmA4509* queda por dilucidar, pero esta dependencia de la expresión de uno de los alelos duplicados por la represión del otro es consistente con uno de los patrones de expresión entre genes duplicados. La presencia de un segundo alelo del gen *irmA* en la cepa 042 puede facilitar la reducción del ruido y la expresión apropiada de la proteína IrmA cuando se requiera en el proceso de infección, consiguiendo de esta manera una mayor robustez genética³⁶⁶.

El gen *irmA* se localiza junto a los genes *agn43-yeeR*, formando un clúster de genes ampliamente distribuido entre los diferentes patotipos de *E. coli* que presenta una elevada variabilidad genética. Dentro de un mismo patotipo, esta región se puede encontrar tanto en una única copia como duplicada e, incluso, elementos de inserción pueden romper la pauta de lectura de una de las copias, como es el caso de la cepa UPEC CFT073. Estos procesos de inserción son frecuentes en esta región, sobre todo en la zona entre los genes *agn43* e *yeeR* (Fig. 4.35). Sorprendentemente, los genes *agn43-yeeR-irmA* también pueden estar ausentes en algunas cepas patógenas de *E. coli* (Fig. 4.6). La presencia o ausencia de elementos génicos específicos relacionados con la virulencia, que también pueden estar presentes en una o más copias exhibiendo patrones de expresión diferenciales, aboga porque sus productos génicos correspondientes se expresen solo en condiciones de infección muy específicas, lo que a su vez habla de la evolución sofisticada de *E. coli*. En el caso del gen *irmA*, las diferencias en la secuencia promotora y codificante son mínimas, pero su análisis nos permite diferenciar perfectamente dos alelos, como son *irmA2244* e *irmA4509* en la cepa 042 (Fig. 4.36).

5.4 Papel de la región de genes duplicados ampliamente distribuida entre los patotipos de *E. coli*

El análisis de genes duplicados y su distribución entre los patotipos de *E. coli* mostró un conjunto de genes presente en la mayoría de patotipos y en múltiples copias, los cuales en la cepa 042 son los enumerados del 55 al 60 (Tabla 4.7). Entre estos genes, la pareja 59-60 forma el sistema toxina-antitoxina *yeeU-yeeV*, donde la toxina *yeeV* (*cbtA*) está involucrada en la inhibición del crecimiento y la muerte celular en condiciones de estrés^{367,368}. Sin embargo, es reseñable la falta de información sobre el resto de genes de esta región. Debido al elevado número de copias de estos genes y su distribución en

todos los patotipos estudiados, se procedió a caracterizar la función de esta región de genes en la cepa 042.

La mutagénesis de todas las copias de esta región no afecta de manera significativa al *fitness* celular de esta cepa de *E. coli*, en las condiciones estudiadas (Fig. 4.39). Este hecho concuerda con las observaciones previas realizadas por Arun y colaboradores, donde indican que los genes involucrados en mecanismos que ayudan a la adaptación a las condiciones del medio son más propensos a sufrir algún proceso de duplicación, mientras que aquellos implicados en funciones esenciales muestran una mayor conservación y un menor grado de duplicación³⁰¹. En consecuencia, se procedió a analizar el papel de esta región dentro de la patogénesis de la cepa 042, donde la adaptación a las condiciones del medio juega un papel fundamental.

Como hemos comentado anteriormente, la formación de biofilm es un paso clave dentro de la patogénesis de las cepas del patotipo EAEC^{172,196}. Por ello, se analizó la posible implicación de estos genes en este proceso de adhesión, pero no se encontró relación debido a que no hubo diferencias entre los resultados obtenidos de la cepa salvaje y la cepa mutante 4 Δ (Fig 4.40). Por otra parte, sí que se obtuvieron resultados significativos en el ensayo de infección del organismo *Galleria mellonella* (Fig. 4.41). La mutagénesis tanto de las copias del gen *irmA* como de las copias de esta región conlleva una reducción de la virulencia de la cepa 042, mostrando la relación que tienen estos genes duplicados con el proceso infectivo.

El estudio de dichos genes de manera más detallada resultaría interesante, con el objetivo de dilucidar cuál o cuáles de ellos están involucrados en el proceso de patogénesis y qué patrón de expresión sigue cada uno. Es probable que el patrón de expresión de cada copia responda a estímulos diferentes o que actúen como *backup* celular, aportando una mayor robustez genética³⁶⁶ y siendo, de esta manera, más eficientes en el proceso de infección. Por otra

parte, alguno de sus productos génicos puede resultar antigénico, lo que podría abrir nuevas perspectivas para la mejora de vacunas contra infecciones de *E. coli* ya que, a diferencia de otros muchos antígenos vacunales que se utilizan contra este microorganismo, estos productos génicos están ampliamente distribuidos en todas las cepas virulentas. Asimismo, puede ser factible utilizar cepas como el mutante cuádruple obtenido, donde se han delecionado todas y cada una de las copias de estos genes, como cepas atenuadas de la virulencia para el diseño de nuevas vacunas vivas.

6. Conclusiones

Las conclusiones que se pueden extraer de este trabajo son las siguientes:

- La presencia de los nuevos parálogos descritos *hha2* y *hha3* de la familia de proteínas Hha en cepas de *E. coli* está relacionada significativamente con la presencia de los genes *yeeR* e *irmA*.
- La posición de los genes *hha2/3* en el cromosoma se encuentra junto al conjunto de genes *flu-yeeR-irmA*, a los que están asociados.
- El sistema H-NS/Hha muestra una nueva función silenciando los genes en los que han tenido lugar procesos de duplicación.
- La duplicación de genes regulados por el sistema H-NS/Hha puede ser la causa del origen de nuevos parálogos de estos reguladores, manteniendo de esta forma la estabilidad de la red reguladora.
- El análisis comparativo de los patrones de duplicación permite establecer tanto relaciones entre patotipos como duplicaciones restringidas a un patotipo específico, e incluso, a una cepa concreta.
- La existencia de un número significativo de genes duplicados o amplificados en la mayoría de patotipos muestra que estos genes pueden desempeñar un papel en la virulencia de *E. coli*.
- El gen *irmA* en la cepa 042 se transcribe de forma independiente al gen *agn43*, a diferencia de lo previamente descrito en la cepa UPEC CFT073.
- El gen *irmA4509* se encuentra regulado siguiendo un patrón poco frecuente, donde la transcripción está reprimida tanto por la proteína OxyR como por Hha y su parálogo Hha2, pero no por H-NS.

- La transcripción del gen *irmA4509* responde a factores ambientales, aumentando en condiciones de temperatura, disponibilidad de nutrientes y de oxígeno semejantes a las de infección.
- La expresión del gen *irmA2244* de la cepa 042 está silenciada en todas las condiciones que se han estudiado, excepto cuando cesa la expresión del gen *irmA4509*.
- Cambios nucleotídicos puntuales en la región promotora de las dos copias del gen *irmA* son cruciales para la expresión de dichos genes.
- Los genes que se encuentran en la región amplificada en la cepa 042 y distribuida en todos los patotipos están involucrados en procesos infectivos.

7. Bibliografía

the 1990s, the number of people with a disability in the United States has increased by 50% (U.S. Census Bureau, 1997). The number of people with a disability in the United States is expected to increase to 100 million by the year 2020 (U.S. Census Bureau, 1997).

As the number of people with a disability increases, the need for accessible information and services also increases. The National Center for Accessible Information (NCAI) has estimated that the total number of people with a disability who are unable to access information is 100 million (NCAI, 1997). The NCAI has also estimated that the total number of people with a disability who are unable to access services is 100 million (NCAI, 1997).

The National Center for Accessible Information (NCAI) has estimated that the total number of people with a disability who are unable to access information is 100 million (NCAI, 1997). The NCAI has also estimated that the total number of people with a disability who are unable to access services is 100 million (NCAI, 1997).

The National Center for Accessible Information (NCAI) has estimated that the total number of people with a disability who are unable to access information is 100 million (NCAI, 1997). The NCAI has also estimated that the total number of people with a disability who are unable to access services is 100 million (NCAI, 1997).

The National Center for Accessible Information (NCAI) has estimated that the total number of people with a disability who are unable to access information is 100 million (NCAI, 1997). The NCAI has also estimated that the total number of people with a disability who are unable to access services is 100 million (NCAI, 1997).

The National Center for Accessible Information (NCAI) has estimated that the total number of people with a disability who are unable to access information is 100 million (NCAI, 1997). The NCAI has also estimated that the total number of people with a disability who are unable to access services is 100 million (NCAI, 1997).

The National Center for Accessible Information (NCAI) has estimated that the total number of people with a disability who are unable to access information is 100 million (NCAI, 1997). The NCAI has also estimated that the total number of people with a disability who are unable to access services is 100 million (NCAI, 1997).

The National Center for Accessible Information (NCAI) has estimated that the total number of people with a disability who are unable to access information is 100 million (NCAI, 1997). The NCAI has also estimated that the total number of people with a disability who are unable to access services is 100 million (NCAI, 1997).

The National Center for Accessible Information (NCAI) has estimated that the total number of people with a disability who are unable to access information is 100 million (NCAI, 1997). The NCAI has also estimated that the total number of people with a disability who are unable to access services is 100 million (NCAI, 1997).

- (1) Kaper, J. B.; Nataro, J. P.; Mobley, H. L. T. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **2004**, *2* (2), 123–140. <https://doi.org/10.1038/nrmicro818>.
- (2) Gomes, T. A. T.; Elias, W. P.; Scaletsky, I. C. A.; Guth, B. E. C.; Rodrigues, J. F.; Piazza, R. M. F.; Ferreira, L. C. S.; Martinez, M. B. Diarrheagenic *Escherichia coli*. *Brazilian J. Microbiol.* **2016**, *47*, 3–30. <https://doi.org/10.1016/j.bjm.2016.10.015>.
- (3) Troeger, C.; Forouzanfar, M.; Rao, P. C.; Khalil, I.; Brown, A.; Reiner, R. C.; Fullman, N.; Thompson, R. L.; Abajobir, A.; Ahmed, M.; Alemayohu, M. A.; Alvis-Guzman, N.; Amare, A. T.; Antonio, C. A.; Asayesh, H.; Avokpaho, E.; Awasthi, A.; Bacha, U.; Barac, A.; Betsue, B. D.; Beyene, A. S.; Boneya, D. J.; Malta, D. C.; Dandona, L.; Dandona, R.; Dubey, M.; Eshrati, B.; Fitchett, J. R. A.; Gebrehiwot, T. T.; Hailu, G. B.; Horino, M.; Hotez, P. J.; Jibat, T.; Jonas, J. B.; Kasaeian, A.; Kisosoon, N.; Kotloff, K.; Koyanagi, A.; Kumar, G. A.; Rai, R. K.; Lal, A.; El Razek, H. M. A.; Mengistie, M. A.; Moe, C.; Patton, G.; Platts-Mills, J. A.; Qorbani, M.; Ram, U.; Roba, H. S.; Sanabria, J.; Sartorius, B.; Sawhney, M.; Shigematsu, M.; Sreeramareddy, C.; Swaminathan, S.; Tedla, B. A.; Jagiellonian, R. T. M.; Ukwaja, K.; Werdecker, A.; Widdowson, M. A.; Yonemoto, N.; El Sayed Zaki, M.; Lim, S. S.; Naghavi, M.; Vos, T.; Hay, S. I.; Murray, C. J. L.; Mokdad, A. H. Estimates of Global, Regional, and National Morbidity, Mortality, and Aetiologies of Diarrhoeal Diseases: A Systematic Analysis for the Global Burden of Disease Study 2015. *Lancet Infect. Dis.* **2017**, *17* (9), 909–948. [https://doi.org/10.1016/S1473-3099\(17\)30276-1](https://doi.org/10.1016/S1473-3099(17)30276-1).
- (4) Croxen, M. A.; Law, R. J.; Scholz, R.; Keeney, K. M.; Wlodarska, M.; Finlay, B. B. Recent Advances in Understanding Enteric Pathogenic *Escherichia coli*. *Clin. Microbiol. Rev.* **2013**, *26* (4), 822–880. <https://doi.org/10.1128/CMR.00022-13>.
- (5) Dobrindt, U.; Chowdary, M. G.; Krumbholz, G.; Hacker, J. Genome Dynamics and Its Impact on Evolution of *Escherichia coli*. *Med. Microbiol. Immunol.* **2010**, *199* (3), 145–154. <https://doi.org/10.1007/s00430-010-0161-2>.
- (6) Wirth, T.; Falush, D.; Lan, R.; Colles, F.; Mensa, P.; Wieler, L. H.; Karch, H.; Reeves, P. R.; Maiden, M. C. J.; Ochman, H.; Achtman, M. Sex and Virulence in *Escherichia coli*: An Evolutionary Perspective. *Mol. Microbiol.* **2006**, *60* (5), 1136–1151. <https://doi.org/10.1111/j.1365-2958.2006.05172.x>.
- (7) Sokurenko, E. V.; Hasty, D. L.; Dykhuizen, D. E. Pathoadaptive Mutations: Gene Loss and Variation in Bacterial Pathogens. *Trends Microbiol.* **1999**, *7* (5), 191–195. [https://doi.org/10.1016/S0966-842X\(99\)01493-6](https://doi.org/10.1016/S0966-842X(99)01493-6).
- (8) Blakely, G. W. *Mechanisms of Horizontal Gene Transfer and DNA Recombination*; Elsevier Ltd, 2014; Vol. 1–3. <https://doi.org/10.1016/B978-0-12-397169-2.00015-9>.

Bibliografía

- (9) Johnston, C.; Martin, B.; Fichant, G.; Polard, P.; Claverys, J. P. Bacterial Transformation: Distribution, Shared Mechanisms and Divergent Control. *Nat. Rev. Microbiol.* **2014**, *12* (3), 181–196. <https://doi.org/10.1038/nrmicro3199>.
- (10) Chiang, Y. N.; Penadés, J. R.; Chen, J. Genetic Transduction by Phages and Chromosomal Islands: The New and Noncanonical. *PLoS Pathog.* **2019**, *15* (8), 1–7. <https://doi.org/10.1371/journal.ppat.1007878>.
- (11) Cabezón, E.; Ripoll-Rozada, J.; Peña, A.; de la Cruz, F.; Arechaga, I. Towards an Integrated Model of Bacterial Conjugation. *FEMS Microbiol. Rev.* **2015**, *39* (1), 81–95. <https://doi.org/10.1111/1574-6976.12085>.
- (12) Sokurenko, E. Pathoadaptive Mutations in Uropathogenic *Escherichia coli*. *Microbiol. Spectr.* **2016**, *4* (2), 1–21. <https://doi.org/10.1128/microbiolspec.uti-0020-2015>.
- (13) Bliven, K. A.; Maurelli, A. T. Antivirulence Genes: Insights into Pathogen Evolution through Gene Loss. *Infect. Immun.* **2012**, *80* (12), 4061–4070. <https://doi.org/10.1128/IAI.00740-12>.
- (14) Suzart, S.; Gomes, T. A. T.; Guth, B. E. C. Characterization of Serotypes and Outer Membrane Protein Profiles in Enteroaggregative *Escherichia coli* Strains. *Microbiol. Immunol.* **1999**, *43* (3), 201–205. <https://doi.org/10.1111/j.1348-0421.1999.tb02394.x>.
- (15) Wallace-Gadsden, F.; Johnson, J. R.; Wain, J.; Okeke, I. N. Enteroaggregative *Escherichia coli* Related to Uropathogenic Clonal Group A. *Emerg. Infect. Dis.* **2007**, *13* (5), 757–760. <https://doi.org/10.3201/eid1305.061057>.
- (16) Nunes, K. O.; Santos, A. C. P.; Bando, S. Y.; Silva, R. M.; Gomes, T. A. T.; Elias, W. P. Enteroaggregative *Escherichia coli* with Uropathogenic Characteristics Are Present in Feces of Diarrheic and Healthy Children. *Pathog. Dis.* **2017**, *75* (8), 2010–2013. <https://doi.org/10.1093/femspd/ftx106>.
- (17) Nataro, J. P.; Kaper, J. B. Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* **1998**, *11* (1), 142–201. <https://doi.org/10.1128/cmr.11.1.142>.
- (18) Russo, T. A.; Johnson, J. R. Proposal for a New Inclusive Designation for Extraintestinal Pathogenic Isolates of *Escherichia coli*: ExPEC. *J. Infect. Dis.* **2000**, *181* (5), 1753–1754. <https://doi.org/10.1086/315418>.
- (19) Rasko, D. A.; Rosovitz, M. J.; Myers, G. S. A.; Mongodin, E. F.; Fricke, W. F.; Gajer, P.; Crabtree, J.; Sebaihia, M.; Thomson, N. R.; Chaudhuri, R.; Henderson, I. R.; Sperandio, V.; Ravel, J. The Pangenome Structure of *Escherichia coli*: Comparative Genomic Analysis of *E. coli* Commensal and Pathogenic Isolates. *J. Bacteriol.* **2008**, *190* (20), 6881–6893.

<https://doi.org/10.1128/JB.00619-08>.

- (20) Touchon, M.; Hoede, C.; Tenailon, O.; Barbe, V.; Baeriswyl, S.; Bidet, P.; Bingen, E.; Bonacorsi, S.; Bouchier, C.; Bouvet, O.; Calteau, A.; Chiappello, H.; Clermont, O.; Cruveiller, S.; Danchin, A.; Diard, M.; Dossat, C.; El Karoui, M.; Frapy, E.; Garry, L.; Ghigo, J. M.; Gilles, A. M.; Johnson, J.; Le Bouguéne, C.; Lescat, M.; Mangenot, S.; Martinez-Jéhanne, V.; Matic, I.; Nassif, X.; Oztas, S.; Petit, M. A.; Pichon, C.; Rouy, Z.; Ruf, C. Saint; Schneider, D.; Turret, J.; Vacherie, B.; Vallenet, D.; Médigue, C.; Rocha, E. P. C.; Denamur, E. Organised Genome Dynamics in the *Escherichia coli* Species Results in Highly Diverse Adaptive Paths. *PLoS Genet.* **2009**, *5* (1). <https://doi.org/10.1371/journal.pgen.1000344>.
- (21) Van Elsas, J. D.; Semenov, A. V.; Costa, R.; Trevors, J. T. Survival of *Escherichia coli* in the Environment: Fundamental and Public Health Aspects. *ISME J.* **2011**, *5* (2), 173–183. <https://doi.org/10.1038/ismej.2010.80>.
- (22) Mira, A.; Martín-Cuadrado, A. B.; D'Auria, G.; Rodríguez-Valera, F. The Bacterial Pan-Genome: A New Paradigm in Microbiology. *Int. Microbiol.* **2010**, *13* (2), 45–57. <https://doi.org/10.2436/20.1501.01.110>.
- (23) Foxman, B. Epidemiology of Urinary Tract Infections: Incidence, Morbidity, and Economic Costs. *Am. J. Med.* **2002**, *113* (1), 5–13. [https://doi.org/10.1016/S0002-9343\(02\)01054-9](https://doi.org/10.1016/S0002-9343(02)01054-9).
- (24) Hagan, E. C.; Donnenberg, M. S.; Mobley, H. L. T. Uropathogenic *Escherichia coli*. *EcoSal Plus* **2009**, *3* (2), 275–304. <https://doi.org/10.1128/ecosalplus.8.6.1.3>.
- (25) Croxen, M. A.; Finlay, B. B. Molecular Mechanisms of *Escherichia coli* Pathogenicity. *Nat. Rev. Microbiol.* **2010**, *8* (1), 26–38. <https://doi.org/10.1038/nrmicro2265>.
- (26) Wiles, T. J.; Kulesus, R. R.; Mulvey, M. A. Origins and Virulence Mechanisms of Uropathogenic *Escherichia coli*. *Exp. Mol. Pathol.* **2008**, *85* (1), 11–19. <https://doi.org/10.1016/j.yexmp.2008.03.007>.
- (27) Russo, T. A.; Stapleton, A.; Wenderoth, S.; Hooton, T. M.; Stamm, W. E. Chromosomal Restriction Fragment Length Polymorphism Analysis of *Escherichia coli* Strains Causing Recurrent Urinary Tract Infections in Young Women. *J. Infect. Dis.* **1995**, *172* (2), 440–445. <https://doi.org/10.1093/infdis/172.2.440>.
- (28) Foxman, B. The Epidemiology of Urinary Tract Infection. *Nat. Rev. Urol.* **2010**, *7* (12), 653–660. <https://doi.org/10.1038/nrurol.2010.190>.
- (29) Manges, A. R.; Johnson, J. R.; Foxman, B.; O'Bryan, T. T.; Fullerton, K. E.; Riley, L. W. Widespread Distribution of Urinary Tract Infections Caused

- by a Multidrug-Resistant *Escherichia coli* Clonal Group. *N. Engl. J. Med.* **2001**, 345 (14), 1007–1013. <https://doi.org/10.1056/NEJMoa011265>.
- (30) Foxman, B.; Manning, S. D.; Tallman, P.; Bauer, R.; Zhang, L.; Koopman, J. S.; Gillespie, B.; Sobel, J. D.; Marrs, C. F. Uropathogenic *Escherichia coli* Are More Likely than Commensal *E. coli* to Be Shared between Heterosexual Sex Partners. *Am. J. Epidemiol.* **2002**, 156 (12), 1133–1140. <https://doi.org/10.1093/aje/kwf159>.
- (31) Johnson, J. R.; Delavari, P. Concurrent Fecal Colonization with Extraintestinal Pathogenic *Escherichia coli* in a Homosexual Man with Recurrent Urinary Tract Infection and in His Male Sex Partner. *Clin. Infect. Dis.* **2002**, 35 (6), e65–e68. <https://doi.org/10.1086/342301>.
- (32) Moore, E. E.; Hawes, S. E.; Scholes, D.; Boyko, E. J.; Hughes, J. P.; Fihn, S. D. Sexual Intercourse and Risk of Symptomatic Urinary Tract Infection in Post-Menopausal Women. *J. Gen. Intern. Med.* **2008**, 23 (5), 595–599. <https://doi.org/10.1007/s11606-008-0535-y>.
- (33) Brzuszkiewicz, E.; Brüggemann, H.; Liesegang, H.; Emmerth, M.; Ölschläger, T.; Nagy, G.; Albermann, K.; Wagner, C.; Buchrieser, C.; Emody, L.; Gottschalk, G.; Hacker, J.; Dobrindt, U. How to Become a Uropathogen: Comparative Genomic Analysis of Extraintestinal Pathogenic *Escherichia coli* Strains. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, 103 (34), 12879–12884. <https://doi.org/10.1073/pnas.0603038103>.
- (34) Chen, S. L.; Hung, C. S.; Xu, J.; Reigstad, C. S.; Magrini, V.; Sabo, A.; Blasiar, D.; Bieri, T.; Meyer, R. R.; Ozersky, P.; Armstrong, J. R.; Fulton, R. S.; Latreille, J. P.; Spieth, J.; Hooton, T. M.; Mardis, E. R.; Hultgren, S. J.; Gordon, J. I. Identification of Genes Subject to Positive Selection in Uropathogenic Strains of *Escherichia coli*: A Comparative Genomics Approach. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, 103 (15), 5977–5982. <https://doi.org/10.1073/pnas.0600938103>.
- (35) Welch, R. A.; Burland, V.; Plunkett, G.; Redford, P.; Roesch, P.; Rasko, D.; Buckles, E. L.; Liou, S. R.; Boutin, A.; Hackett, J.; Stroud, D.; Mayhew, G. F.; Rose, D. J.; Zhou, S.; Schwartz, D. C.; Perna, N. T.; Mobley, H. L. T.; Sonnenberg, M. S.; Blattner, F. R. Extensive Mosaic Structure Revealed by the Complete Genome Sequence of Uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99 (26), 17020–17024. <https://doi.org/10.1073/pnas.252529799>.
- (36) Hacker J, Blum-Oehler G, Mühldorfer I, T. H. Pathogenicity Islands of Virulent Bacteria: Structure, Function and Impact on Microbial Evolution. *Mol. Microbiol.* **1997**, 23 (6 PG-1089–97), 1089–1097.
- (37) Schmidt, H.; Hensel, M. Pathogenicity Islands in Bacterial Pathogenesis. *Clin. Microbiol. Rev.* **2004**, 17 (1), 14–56. <https://doi.org/10.1128/CMR.17.1.14>.

- (38) Flores-Mireles, A. L.; Walker, J. N.; Caparon, M.; Hultgren, S. J. Urinary Tract Infections: Epidemiology, Mechanisms of Infection and Treatment Options. *Nat. Rev. Microbiol.* **2015**, *13* (5), 269–284. <https://doi.org/10.1038/nrmicro3432>.
- (39) Bacheller, C. D.; Bernstein, J. M. Urinary Tract Infections. *Med. Clin. North Am.* **1997**, *81* (3), 719–730. [https://doi.org/10.1016/S0025-7125\(05\)70542-3](https://doi.org/10.1016/S0025-7125(05)70542-3).
- (40) Wright, K. J.; Seed, P. C.; Hultgren, S. J. Development of Intracellular Bacterial Communities of Uropathogenic *Escherichia coli* Depends on Type 1 Pili. *Cell. Microbiol.* **2007**, *9* (9), 2230–2241. <https://doi.org/10.1111/j.1462-5822.2007.00952.x>.
- (41) Martinez, J. J. Type 1 Pilus-Mediated Bacterial Invasion of Bladder Epithelial Cells. *EMBO J.* **2000**, *19* (12), 2803–2812. <https://doi.org/10.1093/emboj/19.12.2803>.
- (42) Min, G.; Stolz, M.; Zhou, G.; Liang, F.; Sebbel, P.; Stoffler, D.; Glockshuber, R.; Sun, T. T.; Aebi, U.; Kong, X. P. Localization of Uroplakin Ia, the Urothelial Receptor for Bacterial Adhesin FimH, on the Six Inner Domains of the 16 Nm Urothelial Plaque Particle. *J. Mol. Biol.* **2002**, *317* (5), 697–706. <https://doi.org/10.1006/jmbi.2002.5442>.
- (43) Lane, M. C.; Alteri, C. J.; Smith, S. N.; Mobley, H. L. T. Expression of Flagella Is Coincident with Uropathogenic *Escherichia coli* Ascension to the Upper Urinary Tract. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (42), 16669–16674. <https://doi.org/10.1073/pnas.0607898104>.
- (44) Bhakdi, S.; Mackman, N.; Menestrina, G.; Gray, L.; Hugo, F.; Seeger, W.; Holland, I. B. The Hemolysin of *Escherichia coli*. *Eur. J. Epidemiol.* **1988**, *4* (2), 135–143. <https://doi.org/10.1007/BF00144740>.
- (45) Marrs, C. F.; Zhang, L.; Foxman, B. *Escherichia coli* Mediated Urinary Tract Infections: Are There Distinct Uropathogenic *E. coli* (UPEC) Pathotypes? *FEMS Microbiol. Lett.* **2005**, *252* (2), 183–190. <https://doi.org/10.1016/j.femsle.2005.08.028>.
- (46) Miyazaki, J.; Ba-Thein, W.; Kumao, T.; Obata Yasuoka, M.; Akaza, H.; Hayashi, H. Type 1, P and S Fimbriae, and Afimbrial Adhesin I Are Not Essential for Uropathogenic *Escherichia coli* to Adhere to and Invade Bladder Epithelial Cells. *FEMS Immunol. Med. Microbiol.* **2002**, *33* (1), 23–26. [https://doi.org/10.1016/S0928-8244\(01\)00311-X](https://doi.org/10.1016/S0928-8244(01)00311-X).
- (47) Snyder, J. A.; Haugen, B. J.; Lockatell, C. V.; Maroncle, N.; Hagan, E. C.; Johnson, D. E.; Welch, R. A.; Mobley, H. L. T. Coordinate Expression of Fimbriae in Uropathogenic *Escherichia coli*. *Infect. Immun.* **2005**, *73* (11), 7588–7596. <https://doi.org/10.1128/IAI.73.11.7588-7596.2005>.

- (48) Snyder, J. A.; Lloyd, A. L.; Lockett, C. V.; Johnson, D. E.; Mobley, H. L. T. Role of Phase Variation of Type 1 Fimbriae in a Uropathogenic *Escherichia coli* Cystitis Isolate during Urinary Tract Infection. *Infect. Immun.* **2006**, *74* (2), 1387–1393. <https://doi.org/10.1128/IAI.74.2.1387-1393.2006>.
- (49) Holden, N.; Totsika, M.; Dixon, L.; Catherwood, K.; Gally, D. L. Regulation of P-Fimbrial Phase Variation Frequencies in *Escherichia coli* CFT073. *Infect. Immun.* **2007**, *75* (7), 3325–3334. <https://doi.org/10.1128/IAI.01989-06>.
- (50) Holden, N. J.; Totsika, M.; Mahler, E.; Roe, A. J.; Catherwood, K.; Lindner, K.; Dobrindt, U.; Gally, D. L. Demonstration of Regulatory Cross-Talk between P Fimbriae and Type 1 Fimbriae in Uropathogenic *Escherichia coli*. *Microbiology* **2006**, *152* (4), 1143–1153. <https://doi.org/10.1099/mic.0.28677-0>.
- (51) Holden, N. J.; Gally, D. L. Switches, Cross-Talk and Memory in *Escherichia coli* Adherence. *J. Med. Microbiol.* **2004**, *53* (7), 585–593. <https://doi.org/10.1099/jmm.0.05491-0>.
- (52) Lane, M. C.; Mobley, H. L. T. Role of P-Fimbrial-Mediated Adherence in Pyelonephritis and Persistence of Uropathogenic *Escherichia coli* (UPEC) in the Mammalian Kidney. *Kidney Int.* **2007**, *72* (1), 19–25. <https://doi.org/10.1038/sj.ki.5002230>.
- (53) Leffler, H.; Svanborg-Eden, C. Glycolipid Receptors for Uropathogenic *Escherichia coli* on Human Erythrocytes and Uroepithelial Cells. *Infect. Immun.* **1981**, *34* (3), 920–929.
- (54) Leffler, H.; N, C. S. E. D. I. Chemical Identification of a Glycosphingolipid Receptor for *Escherichia coli* Attaching to Human Urinary Tract Epithelial Cells and Agglutinating Human Erythrocytes. **1980**, *8*, 127–134.
- (55) Rice, J. C.; Peng, T.; Spence, J. S.; Wang, H. Q.; Goldblum, R. M.; Corthésy, B.; Nowicki, B. J. Pyelonephritic *Escherichia coli* Expressing P Fimbriae Decrease Immune Response of the Mouse Kidney. *J. Am. Soc. Nephrol.* **2005**, *16* (12), 3583–3591. <https://doi.org/10.1681/ASN.2005030243>.
- (56) Bray, J. Isolation of Antigenically Homogeneous Strains Of *Bact. Coli* Neapolitanum from Summer Diarrhoea of Infants. *J. Pathol. Bacteriol.* **1945**, *57* (2), 239–247. <https://doi.org/10.1002/path.1700570210>.
- (57) Deborah Chen, H.; Frankel, G. Enteropathogenic *Escherichia coli*: Unravelling Pathogenesis. *FEMS Microbiol. Rev.* **2005**, *29* (1), 83–98. <https://doi.org/10.1016/j.femsre.2004.07.002>.
- (58) Cravioto, A.; Reyes, R. E.; Ortega, R.; Fernández, G.; Hernández, R.; López, D. Prospective Study of Diarrhoeal Disease in a Cohort of Rural Mexican

- Children: Incidence and Isolated Pathogens during the First Two Years of Life. *Epidemiol. Infect.* **1988**, *101* (1), 123–134. <https://doi.org/10.1017/S0950268800029289>.
- (59) Tardelli Gomes, T. A.; Rassi, V.; MacDonald, K. L.; Silva Ramos, S. R. T.; Trabulsi, L. R.; Vieira, M. A. M.; Guth, B. E. C.; Candeias, J. A. N.; Ivey, C.; Toledo, M. R. F.; Blake, P. A. Enteropathogens Associated with Acute Diarrheal Disease in Urban Infants in Sao Paulo, Brazil. *J. Infect. Dis.* **1991**, *164* (2), 331–337. <https://doi.org/10.1093/infdis/164.2.331>.
- (60) Ansaruzzaman, M.; Albert, M. J.; Nahar, S.; Byun, R.; Katouli, M.; Kühn, I.; Möllby, R. Clonal Groups of Enteropathogenic *Escherichia coli* Isolated in Case-Control Studies of Diarrhoea in Bangladesh. *J. Med. Microbiol.* **2000**, *49* (2), 177–185. <https://doi.org/10.1099/0022-1317-49-2-177>.
- (61) Robins-browne, R. O. Y. M.; Levine, M. M.; Rowe, B.; Ernest, M. Failure to Detect Conventional Enterotoxins in Classical Enteropathogenic (Serotyped) *Escherichia coli* Strains of Proven Pathogenicity. **1982**, *38* (2), 798–801.
- (62) Uribe, F.; Navarro, A.; Roca, J. Risk of diarrhea during the first year of life. **1990**, *131* (5).
- (63) Andrade, J. R. C.; Da Veiga, V. F.; De Santa Rosa, M. R.; Suassuna, I. An Endocytic Process in HEp-2 Cells Induced by Enteropathogenic *Escherichia coli*. *J. Med. Microbiol.* **1989**, *28* (1), 49–57. <https://doi.org/10.1099/00222615-28-1-49>.
- (64) Moon, H. W.; Whipp, S. C.; Argenzio, R. A.; Levine, M. M.; Giannella, R. A. Attaching and Effacing Activities of Rabbit and Human Enteropathogenic *Escherichia coli* in Pig and Rabbit Intestines. *Infect. Immun.* **1983**, *41* (3), 1340–1351.
- (65) Baldwin, T. O. M. J.; Lee-delaunay, M. B.; Knutson, S.; Williams, P. H. Calcium-Calmodulin Dependence of Actin Accretion and Lethality in Cultured HEp-2 Cells Infected with Enteropathogenic *Escherichia coli*. **1993**, *61* (2), 760–763.
- (66) Baldwin, T. J.; Ward, W.; Aitken, A.; Knutton, S.; Williams, P. H. Elevation of Intracellular Free Calcium Levels in HEp-2 Cells Infected with Enteropathogenic *Escherichia coli*. *Infect. Immun.* **1991**, *59* (5), 1599–1604.
- (67) Dytoc, M.; Fedorko, L.; Sherman, P. M. Signal Transduction in Human Epithelial Cells Infected with Attaching and Effacing *Escherichia coli* in Vitro. *Gastroenterology* **1994**, *106* (5), 1150–1161. [https://doi.org/10.1016/0016-5085\(94\)90004-3](https://doi.org/10.1016/0016-5085(94)90004-3).
- (68) Jerse, A. E.; Yu, J.; Tall, B. D.; Kaper, J. B. A Genetic Locus of

- Enteropathogenic *Escherichia coli* Necessary for the Production of Attaching and Effacing Lesions on Tissue Culture Cells. *Proc. Natl. Acad. Sci.* **1990**, *87* (20), 7839–7843. <https://doi.org/10.1073/pnas.87.20.7839>.
- (69) Sanger, J. M.; Chang, R.; Ashton, F.; Kaper, J. B.; Sanger, J. W. Novel Form of Actin-Based Motility Transports Bacteria on the Surfaces of Infected Cells. *Cell Motil. Cytoskeleton* **1996**, *34* (4), 279–287. [https://doi.org/10.1002/\(SICI\)1097-0169\(1996\)34:4<279::AID-CM3>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0169(1996)34:4<279::AID-CM3>3.0.CO;2-3).
- (70) Swimm, A.; Bommarius, B.; Li, Y.; Cheng, D.; Reeves, P.; Sherman, M.; Veach, D.; Bornmann, W.; Kalman, D. Enteropathogenic *Escherichia coli* Use Redundant Tyrosine Kinases to Form Actin Pedestals. *Mol. Biol. Cell* **2004**, *15* (8), 3520–3529. <https://doi.org/10.1091/mbc.e04-02-0093>.
- (71) Swimm, A. I.; Kalman, D. Cytosolic Extract Induces Tir Translocation and Pedestals in EPEC-Infected Red Blood Cells. *PLoS Pathog.* **2008**, *4* (1), 0038–0054. <https://doi.org/10.1371/journal.ppat.0040004>.
- (72) Jarvis, K. G.; Girón, J. A.; Jerse, A. E.; Mcdaniel, T. K.; Donnenberg, M. S.; Kaper, J. B. Enteropathogenic *Escherichia coli* Contains a Putative Type III Secretion System Necessary for the Export of Proteins Involved in Attaching and Effacing Lesion Formation. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92* (17), 7996–8000. <https://doi.org/10.1073/pnas.92.17.7996>.
- (73) Rosenshine, I.; Ruschkowski, S.; Stein, M.; Reinscheid, D. J.; Mills, S. D.; Finlay, B. B. A Pathogenic Bacterium Triggers Epithelial Signals to Form a Functional Bacterial Receptor That Mediates Actin Pseudopod Formation. *EMBO J.* **1996**, *15* (11), 2613–2624. <https://doi.org/10.1002/j.1460-2075.1996.tb00621.x>.
- (74) Kenny, B.; DeVinney, R.; Stein, M.; Reinscheid, D. J.; Frey, E. A.; Finlay, B. B. Enteropathogenic *E. coli* (EPEC) Transfers Its Receptor for Intimate Adherence into Mammalian Cells. *Cell* **1997**, *91* (4), 511–520. [https://doi.org/10.1016/S0092-8674\(00\)80437-7](https://doi.org/10.1016/S0092-8674(00)80437-7).
- (75) Mcdaniel, T. K.; Jarvis, K. G.; Donnenberg, M. S.; Kaper, J. B. A Genetic Locus of Enterocyte Effacement Conserved among Diverse Enterobacterial Pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92* (5), 1664–1668. <https://doi.org/10.1073/pnas.92.5.1664>.
- (76) Elliott, E.; Li, Z.; Bell, C.; Stiel, D.; Buret, A.; Wallace, J.; Brzuszczyk, I.; O'Loughlin, E. Modulation of Host Response to *Escherichia coli* O157:H7 Infection by Anti-CD18 Antibody in Rabbits. *Gastroenterology* **1994**, *106* (6), 1554–1561. [https://doi.org/10.1016/0016-5085\(94\)90410-3](https://doi.org/10.1016/0016-5085(94)90410-3).
- (77) Yu, J.; Kaper, J. B. Cloning and Characterization of the Eae Gene of Enterohaemorrhagic *Escherichia coli* O157:H7. *Mol. Microbiol.* **1992**, *6*

- (3), 411–417. <https://doi.org/10.1111/j.1365-2958.1992.tb01484.x>.
- (78) Brady, M. J.; Radhakrishnan, P.; Liu, H.; Magoun, L.; Murphy, K. C.; Mukherjee, J.; Donohue-Rolfe, A.; Tzipori, S.; Leong, J. M. Enhanced Actin Pedestal Formation by Enterohemorrhagic *Escherichia coli* O157:H7 Adapted to the Mammalian Host. *Front. Microbiol.* **2011**, *2* (NOV), 1–10. <https://doi.org/10.3389/fmicb.2011.00226>.
- (79) Rumer, L.; Jores, J.; Kirsch, P.; Cavnignac, Y.; Zehmke, K.; Wieler, L. H. Dissemination of PheU- and PheV-Located Genomic Islands among Enteropathogenic (EPEC) and Enterohemorrhagic (EXEC) *E. coli* and Their Possible Role in the Horizontal Transfer of the Locus of Enterocyte Effacement (LEE). *Int. J. Med. Microbiol.* **2003**, *292* (7–8), 463–475. <https://doi.org/10.1078/1438-4221-00229>.
- (80) Riley, L. W.; Remis, R. S.; Helgerson, S. D.; McGee, H. B.; Wells, J. G.; Davis, B. R.; Hebert, R. J.; Olcott, E. S.; Johnson, L. M.; Hargrett, N. T.; Blake, P. A.; Cohen, M. L. Hemorrhagic Colitis Associated with a Rare *Escherichia coli* Serotype. *N. Engl. J. Med.* **1983**, *308* (12), 681–685. <https://doi.org/10.1056/NEJM198303243081203>.
- (81) Karmali, M.; Petric, M.; Steele, B.; Lim, C. Sporadic cases of haemolytic-uremic syndrome associated with fecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. *Lancet* **1983**, *321* (8325), 619–620. [https://doi.org/10.1016/S0140-6736\(83\)91795-6](https://doi.org/10.1016/S0140-6736(83)91795-6).
- (82) Watanabe, Y.; Ozasa, K.; Mermin, J. H.; Griffin, P. M.; Masuda, K.; Imashuku, S.; Sawada, T. Factory Outbreak of *Escherichia coli* O157:H7 Infection in Japan. *Emerg. Infect. Dis.* **1999**, *5* (3), 424–428. <https://doi.org/10.3201/eid0503.990313>.
- (83) Caro, I.; García-Armesto, M. R. Occurrence of Shiga Toxin-Producing *Escherichia coli* in a Spanish Raw Ewe's Milk Cheese. *Int. J. Food Microbiol.* **2007**, *116* (3), 410–413. <https://doi.org/10.1016/j.ijfoodmicro.2007.02.015>.
- (84) Gerber, A.; Karch, H.; Allerberger, F.; Verweyen, H. M.; Zimmerhackl, L. B. Clinical Course and the Role of Shiga Toxin-Producing *Escherichia coli* Infection in the Hemolytic-Uremic Syndrome in Pediatric Patients, 1997–2000, in Germany and Austria: A Prospective Study. *J. Infect. Dis.* **2002**, *186* (4), 493–500. <https://doi.org/10.1086/341940>.
- (85) Mead, P. S.; Slutsker, L.; Dietz, V.; McCaig, L. F.; Bresee, J. S.; Shapiro, C.; Griffin, P. M.; Tauxe, R. V. Food-Related Illness and Death in the United States. *J. Environ. Health* **2000**, *62* (7), 9–18. <https://doi.org/10.3201/eid0506.990624>.
- (86) Pruijboom-Brees, I. M.; Morgan, T. W.; Ackermann, M. R.; Nystrom, E. D.; Samuel, J. E.; Cornick, N. A.; Moon, H. W. Cattle Lack Vascular Receptors

- for *Escherichia coli* O157:H7 Shiga Toxins. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, 97 (19), 10325–10329. <https://doi.org/10.1073/pnas.190329997>.
- (87) Keene, W. E. An Outbreak of *Escherichia coli* O157:H7 Infections Traced to Jerky Made From Deer Meat. *JAMA J. Am. Med. Assoc.* **1997**, 277 (15), 1229. <https://doi.org/10.1001/jama.1997.03540390059036>.
- (88) Kudva, I. T.; Hatfield, P. G.; Hovde, C. J. *Escherichia coli* O157:H7 in Microbial Flora of Sheep. *J. Clin. Microbiol.* **1996**, 34 (2), 431–433.
- (89) Karch, H.; Tarr, P. I.; Bielaszewska, M. Enterohaemorrhagic *Escherichia coli* in Human Medicine. *Int. J. Med. Microbiol.* **2005**, 295 (6–7), 405–418. <https://doi.org/10.1016/j.ijmm.2005.06.009>.
- (90) Tozzi, A. E.; Caprioli, A.; Minelli, F.; Gianviti, A.; De Petris, L.; Edefonti, A.; Montini, G.; Ferretti, A.; De Palo, T.; Gaido, M.; Rizzoni, G.; Bettinelli, A.; Capasso, G.; Caringella, A.; Coppo, R.; Lama, G.; Li Volti, S.; Maffei, S.; Maringhini, S.; Miglietti, N.; Pecoraro, C.; Pela, I.; Pennesi, M.; Penza, R.; Peratoner, L.; Perfumo, F.; Ratsche, I.; Salvaggio, E.; Setzu, C.; Zacchello, G. Shiga Toxin-Producing *Escherichia coli* Infections Associated with Hemolytic Uremic Syndrome, Italy, 1988-2000. *Emerg. Infect. Dis.* **2003**, 9 (1), 106–108. <https://doi.org/10.3201/eid0901.020266>.
- (91) Eklund, M.; Scheutz, F.; Siitonen, A. Clinical Isolates of Non-O157 Shiga Toxin-Producing *Escherichia coli*: Serotypes, Virulence Characteristics, and Molecular Profiles of Strains of the Same Serotype. *J. Clin. Microbiol.* **2001**, 39 (8), 2829–2834. <https://doi.org/10.1128/JCM.39.8.2829-2834.2001>.
- (92) Elliott, E. J.; Robins-Browne, R. M.; O'Loughlin, E. V.; Bennett-Wood, V.; Bourke, J.; Henning, P.; Hogg, G. G.; Knight, J.; Powell, H.; Redmond, D. Nationwide Study of Haemolytic Uraemic Syndrome: Clinical, Microbiological, and Epidemiological Features. *Arch. Dis. Child.* **2001**, 85 (2), 125–131. <https://doi.org/10.1136/adc.85.2.125>.
- (93) Whittam, T. S.; Wolfe, M. L.; Wachsmuth, I. K.; Orskov, F.; Orskov, I.; Wilson, R. A. Clonal Relationships among *Escherichia coli* Strains That Cause Hemorrhagic Colitis and Infantile Diarrhea. *Infect. Immun.* **1993**, 61 (5), 1619–1629.
- (94) Zhou, Z.; Li, X.; Liu, B.; Beutin, L.; Xu, J.; Ren, Y.; Feng, L.; Lan, R.; Reeves, P. R.; Wang, L. Derivation of *Escherichia coli* O157:H7 from Its O55:H7 Precursor. *PLoS One* **2010**, 5 (1). <https://doi.org/10.1371/journal.pone.0008700>.
- (95) Kyle, J. L.; Cummings, C. A.; Parker, C. T.; Quiñones, B.; Vatta, P.; Newton, E.; Huynh, S.; Swimley, M.; Degoricija, L.; Barker, M.; Fontanoz, S.; Nguyen, K.; Patel, R.; Fang, R.; Tebbs, R.; Petrauskene, O.; Furtado, M.; Mandrell, R. E. *Escherichia coli* Serotype O55:H7 Diversity Supports Parallel

- Acquisition of Bacteriophage at Shiga Toxin Phage Insertion Sites during Evolution of the O157:H7 Lineage. *J. Bacteriol.* **2012**, *194* (8), 1885–1896. <https://doi.org/10.1128/JB.00120-12>.
- (96) LeClerc, J. E.; Li, B.; Payne, W. L.; Cebula, T. A. High Mutation Frequencies Among *Escherichia coli* and Salmonella Pathogens. *Science (80-.)*. **1996**, *274* (5290), 1208–1211. <https://doi.org/10.1126/science.274.5290.1208>.
- (97) Bielaszewska, M.; Mellmann, A.; Zhang, W.; Köck, R.; Fruth, A.; Bauwens, A.; Peters, G.; Karch, H. Characterisation of the *Escherichia coli* Strain Associated with an Outbreak of Haemolytic Uraemic Syndrome in Germany, 2011: A Microbiological Study. *Lancet Infect. Dis.* **2011**, *11* (9), 671–676. [https://doi.org/10.1016/S1473-3099\(11\)70165-7](https://doi.org/10.1016/S1473-3099(11)70165-7).
- (98) Bell, B. P. A Multistate Outbreak of *Escherichia coli* O157:H7—Associated Bloody Diarrhea and Hemolytic Uremic Syndrome From Hamburgers. *JAMA* **1994**, *272* (17), 1349. <https://doi.org/10.1001/jama.1994.03520170059036>.
- (99) Holme, R. Drinking Water Contamination in Walkerton, Ontario: Positive Resolutions from a Tragic Event. *Water Sci. Technol.* **2003**, *47* (3), 1–6. <https://doi.org/10.2166/wst.2003.0144>.
- (100) Lingwood, C. A.; Law, H.; Richardson, S.; Petric, M.; Brunton, J. L.; De Grandis, S.; Karmali, M. Glycolipid Binding of Purified and Recombinant *Escherichia coli* Produced Verotoxin in Vitro. *J. Biol. Chem.* **1987**, *262* (18), 8834–8839.
- (101) Lindberg, A. A.; Brown, J. E.; Strömberg, N.; Westling-Ryd, M.; Schultz, J. E.; Karlsson, K. A. Identification of the Carbohydrate Receptor for Shiga Toxin Produced by *Shigella Dysenteriae* Type 1. *J. Biol. Chem.* **1987**, *262* (4), 1779–1785.
- (102) Waddell, T.; Head, S.; Petric, M.; Cohen, A.; Lingwood, C. Globotriosyl Ceramide Is Specifically Recognized by the *Escherichia coli* Verocytotoxin 2. *Biochem. Biophys. Res. Commun.* **1988**, *152* (2), 674–679. [https://doi.org/10.1016/S0006-291X\(88\)80091-3](https://doi.org/10.1016/S0006-291X(88)80091-3).
- (103) Proulx, F.; Seidman, E. G.; Karpman, D. Pathogenesis of Shiga Toxin-Associated Hemolytic Uremic Syndrome. *Pediatr. Res.* **2001**, *50* (2), 163–171. <https://doi.org/10.1203/00006450-200108000-00002>.
- (104) Obrig, T. G. Shiga Toxin Mode of Action in *E. coli* O157:H7 Disease. *Front. Biosci.* **1997**, *2* (4), A219. <https://doi.org/10.2741/A219>.
- (105) Taylor, F. B.; Tesh, V. L.; DeBault, L.; Li, A.; Chang, A. C. K.; Kosanke, S. D.; Pysker, T. J.; Siegler, R. L. Characterization of the Baboon Responses to Shiga-like Toxin. Descriptive Study of a New Primate Model of Toxic

- Responses to Stx-1. *Am. J. Pathol.* **1999**, *154* (4), 1285–1299. [https://doi.org/10.1016/S0002-9440\(10\)65380-1](https://doi.org/10.1016/S0002-9440(10)65380-1).
- (106) Melton-Celsa, A. R. Shiga Toxin (Stx) Classification, Structure, and Function. *Microbiol. Spectr.* **2014**, *2* (4), 1–21. <https://doi.org/10.1128/microbiolspec.ehec-0024-2013>.
- (107) Hughes, A. K.; Ergonal, Z.; Stricklett, P. K.; Kohan, D. E. Molecular Basis for High Renal Cell Sensitivity to the Cytotoxic Effects of Shigatoxin-1: Upregulation of Globotriaosylceramide Expression. *J. Am. Soc. Nephrol.* **2002**, *13* (9), 2239–2245. <https://doi.org/10.1097/01.ASN.0000027873.85792.52>.
- (108) Weiss, S. M.; Ladwein, M.; Schmidt, D.; Ehinger, J.; Lommel, S.; Städing, K.; Beutling, U.; Disanza, A.; Frank, R.; Jänsch, L.; Scita, G.; Gunzer, F.; Rottner, K.; Stradal, T. E. B. IRSp53 Links the Enterohemorrhagic *E. coli* Effectors Tir and EspFU for Actin Pedestal Formation. *Cell Host Microbe* **2009**, *5* (3), 244–258. <https://doi.org/10.1016/j.chom.2009.02.003>.
- (109) Vingadassalom, D.; Kazlauskas, A.; Skehan, B.; Cheng, H. C.; Magoun, L.; Robbins, D.; Rosen, M. K.; Saksela, K.; Leong, J. M. Insulin Receptor Tyrosine Kinase Substrate Links the *E. coli* O157:H7 Actin Assembly Effectors Tir and EspF U during Pedestal Formation. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (16), 6754–6759. <https://doi.org/10.1073/pnas.0809131106>.
- (110) Campellone, K. G.; Robbins, D.; Leong, J. M. EspFU Is a Translocated EHEC Effector That Interacts with Tir and N-WASP and Promotes Nck-Independent Actin Assembly. *Dev. Cell* **2004**, *7* (2), 217–228. <https://doi.org/10.1016/j.devcel.2004.07.004>.
- (111) Garmendia, J.; Phillips, A. D.; Carlier, M. F.; Chong, Y.; Schüller, S.; Marches, O.; Dahan, S.; Oswald, E.; Shaw, R. K.; Knutton, S.; Frankel, G. TccP Is an Enterohaemorrhagic *Escherichia coli* O157:H7 Type III Effector Protein That Couples Tir to the Actin-Cytoskeleton. *Cell. Microbiol.* **2004**, *6* (12), 1167–1183. <https://doi.org/10.1111/j.1462-5822.2004.00459.x>.
- (112) DeVinney, R.; Stein, M.; Reinscheid, D.; Abe, A.; Ruschkowski, S.; Brett Finlay, B. Enterohemorrhagic *Escherichia coli* O157:H7 Produces Tir, Which Is Translocated to the Host Cell Membrane but Is Not Tyrosine Phosphorylated. *Infect. Immun.* **1999**, *67* (5), 2389–2398.
- (113) Kotloff, K. L.; Nataro, J. P.; Blackwelder, W. C.; Nasrin, D.; Farag, T. H.; Panchalingam, S.; Wu, Y.; Sow, S. O.; Sur, D.; Breiman, R. F.; Faruque, A. S. G.; Zaidi, A. K. M.; Saha, D.; Alonso, P. L.; Tamboura, B.; Sanogo, D.; Onwuchekwa, U.; Manna, B.; Ramamurthy, T.; Kanungo, S.; Ochieng, J. B.; Omore, R.; Oundo, J. O.; Hossain, A.; Das, S. K.; Ahmed, S.; Qureshi, S.; Quadri, F.; Adegbola, R. A.; Antonio, M.; Hossain, M. J.; Akinsola, A.; Mandomando, I.; Nhampossa, T.; Acácio, S.; Biswas, K.; O'Reilly, C. E.;

- Mintz, E. D.; Berkeley, L. Y.; Muhsen, K.; Sommerfelt, H.; Robins-Browne, R. M.; Levine, M. M. Burden and Aetiology of Diarrhoeal Disease in Infants and Young Children in Developing Countries (the Global Enteric Multicenter Study, GEMS): A Prospective, Case-Control Study. *Lancet* **2013**, *382* (9888), 209–222. [https://doi.org/10.1016/S0140-6736\(13\)60844-2](https://doi.org/10.1016/S0140-6736(13)60844-2).
- (114) Fleckenstein, J. M.; Kuhlmann, F. M. Enterotoxigenic *Escherichia coli* Infections. *Curr. Infect. Dis. Rep.* **2019**, *21* (3). <https://doi.org/10.1007/s11908-019-0665-x>.
- (115) Khalil, I. A.; Troeger, C.; Blacker, B. F.; Rao, P. C.; Brown, A.; Atherly, D. E.; Brewer, T. G.; Engmann, C. M.; Houpt, E. R.; Kang, G.; Kotloff, K. L.; Levine, M. M.; Luby, S. P.; MacLennan, C. A.; Pan, W. K.; Pavlinac, P. B.; Platts-Mills, J. A.; Qadri, F.; Riddle, M. S.; Ryan, E. T.; Shoultz, D. A.; Steele, A. D.; Walson, J. L.; Sanders, J. W.; Mokdad, A. H.; Murray, C. J. L.; Hay, S. I.; Reiner, R. C. Morbidity and Mortality Due to Shigella and Enterotoxigenic *Escherichia coli* Diarrhoea: The Global Burden of Disease Study 1990–2016. *Lancet Infect. Dis.* **2018**, *18* (11), 1229–1240. [https://doi.org/10.1016/S1473-3099\(18\)30475-4](https://doi.org/10.1016/S1473-3099(18)30475-4).
- (116) Riddle, M. S.; Sanders, J. W.; Putnam, S. D.; Tribble, D. R. Incidence, Etiology, and Impact of Diarrhea among Long-Term Travelers (US Military and Similar Populations): A Systematic Review. *Am. J. Trop. Med. Hyg.* **2006**, *74* (5), 891–900. <https://doi.org/10.4269/ajtmh.2006.74.891>.
- (117) Hameed, J. M.; McCaffrey, R. L.; McCoy, A.; Brannock, T.; Martin, G. J.; Scouten, W. T.; Brooks, K.; Putnam, S. D.; Riddle, M. S. Incidence, Etiology and Risk Factors for Travelers' Diarrhea during a Hospital Ship-Based Military Humanitarian Mission: Continuing Promise 2011. *PLoS One* **2016**, *11* (5), 1–13. <https://doi.org/10.1371/journal.pone.0154830>.
- (118) Gupta, S. K.; Keck, J.; Ram, P. K.; Crump, J. A.; Miller, M. A.; Mintz, E. D. Part III. Analysis of Data Gaps Pertaining to Enterotoxigenic *Escherichia coli* Infections in Low and Medium Human Development Index Countries, 1984-2005. *Epidemiol. Infect.* **2008**, *136* (6), 721–738. <https://doi.org/10.1017/S095026880700934X>.
- (119) Qadri, F.; Svennerholm, A.-M.; Faruque, A. S. G.; Sack, R. B. Enterotoxigenic *Escherichia coli* in Developing Countries: Epidemiology, Microbiology, Clinical Features, Treatment, and Prevention. *Clin. Microbiol. Rev.* **2005**, *18* (3), 465–483. <https://doi.org/10.1128/CMR.18.3.465-483.2005>.
- (120) Fairbrother, J. M.; Nadeau, É.; Gyles, C. L. *Escherichia coli* in Postweaning Diarrhea in Pigs: An Update on Bacterial Types, Pathogenesis, and Prevention Strategies. *Anim. Heal. Res. Rev.* **2005**, *6* (1), 17–39. <https://doi.org/10.1079/ahr2005105>.

Bibliografía

- (121) Nagy, B.; Fekete, P. Z. Enterotoxigenic *Escherichia coli* in Veterinary Medicine. *Int. J. Med. Microbiol.* **2005**, *295* (6–7), 443–454. <https://doi.org/10.1016/j.ijmm.2005.07.003>.
- (122) Dubreuil, J. D.; Isaacson, R. E.; Schifferli, D. M. Animal Enterotoxigenic *Escherichia coli*. *EcoSal Plus* **2016**, *7* (1), 1–26. <https://doi.org/10.1128/ecosalplus.ESP-0006-2016>.
- (123) Wang, H.; Zhong, Z.; Luo, Y.; Cox, E.; Devriendt, B. Heat-Stable Enterotoxins of Enterotoxigenic *Escherichia coli* and Their Impact on Host Immunity. *Toxins (Basel)*. **2019**, *11* (1), 1–12. <https://doi.org/10.3390/toxins11010024>.
- (124) Mudrak, B.; Kuehn, M. J. Heat-Labile Enterotoxin: Beyond GM1 Binding. *Toxins (Basel)*. **2010**, *2* (6), 1445–1470. <https://doi.org/10.3390/toxins2061445>.
- (125) Chen, T.; Kocinsky, H. S.; Cha, B.; Murtazina, R.; Yang, J.; Tse, C. M.; Singh, V.; Cole, R.; Aronson, P. S.; De Jonge, H.; Sarker, R.; Donowitz, M. Cyclic GMP Kinase II (CGKII) Inhibits NHE3 by Altering Its Trafficking and Phosphorylating NHE3 at Three Required Sites: Identification of a Multifunctional Phosphorylation Site. *J. Biol. Chem.* **2015**, *290* (4), 1952–1965. <https://doi.org/10.1074/jbc.M114.590174>.
- (126) Sears, C. L.; Kaper, J. B. Enteric Bacterial Toxins: Mechanisms of Action and Linkage to Intestinal Secretion. *Microbiol. Rev.* **1996**, *60* (1), 167–215.
- (127) Turner, S. M.; Scott-Tucker, A.; Cooper, L. M.; Henderson, I. R. Weapons of Mass Destruction: Virulence Factors of the Global Killer Enterotoxigenic *Escherichia coli*. *FEMS Microbiol. Lett.* **2006**, *263* (1), 10–20. <https://doi.org/10.1111/j.1574-6968.2006.00401.x>.
- (128) Fleckenstein, J. M.; Hardwidge, P. R.; Munson, G. P.; Rasko, D. A.; Sommerfelt, H.; Steinsland, H. Molecular Mechanisms of Enterotoxigenic *Escherichia coli* Infection. *Microbes Infect.* **2010**, *12* (2), 89–98. <https://doi.org/10.1016/j.micinf.2009.10.002>.
- (129) Gaastra, W.; Svennerholm, A. M. Colonization Factors of Human Enterotoxigenic *Escherichia coli* (ETEC). *Trends Microbiol.* **1996**, *4* (11), 444–452. [https://doi.org/10.1016/0966-842X\(96\)10068-8](https://doi.org/10.1016/0966-842X(96)10068-8).
- (130) Torres, A. G.; Zhou, X.; Kaper, J. B. Adherence of Diarrheagenic *Escherichia coli* Strains to Epithelial Cells. *Microbiology* **2005**, *73* (1), 18–29. <https://doi.org/10.1128/IAI.73.1.18>.
- (131) Evans, D. G.; Silver, R. P.; Evans, D. J.; Chase, D. G.; Gorbach, S. L. Plasmid Controlled Colonization Factor Associated with Virulence in *Escherichia coli* Enterotoxigenic for Humans. *Infect. Immun.* **1975**, *12* (3), 656–667.

- (132) Von Mentzer, A.; Connor, T. R.; Wieler, L. H.; Semmler, T.; Iguchi, A.; Thomson, N. R.; Rasko, D. A.; Joffe, E.; Corander, J.; Pickard, D.; Wiklund, G.; Svennerholm, A. M.; Sjöling, Å.; Dougan, G. Identification of Enterotoxigenic *Escherichia coli* (ETEC) Clades with Long-Term Global Distribution. *Nat. Genet.* **2014**, *46* (12), 1321–1326. <https://doi.org/10.1038/ng.3145>.
- (133) Wolf, M. K. Occurrence, Distribution, and Associations of O and H Serogroups, Colonization Factor Antigens, and Toxins of Enterotoxigenic *Escherichia coli*. *Clin. Microbiol. Rev.* **1997**, *10* (4), 569–584. <https://doi.org/10.1128/cmr.10.4.569>.
- (134) Pasqua, M.; Michelacci, V.; Di Martino, M. L.; Tozzoli, R.; Grossi, M.; Colonna, B.; Morabito, S.; Prosseda, G. The Intriguing Evolutionary Journey of Enteroinvasive *E. coli* (EIEC) toward Pathogenicity. *Front. Microbiol.* **2017**, *8* (DEC), 1–12. <https://doi.org/10.3389/fmicb.2017.02390>.
- (135) Vieira, N.; Bates, S. J.; Solberg, O. D.; Ponce, K.; Cevallos, W.; Trueba, G.; Riley, L. E. E.; Joseph, N. S. Isolated in a Remote Region of Northern Coastal. *Am J Trop Med Hyg* **2007**, *76* (3), 528–533.
- (136) Beutin, L.; Gleier, K.; Kontny, I.; Echeverria, P.; Scheutz, F. Origin and Characteristics of Enteroinvasive Strains of *Escherichia coli* (EIEC) Isolated in Germany. *Epidemiol. Infect.* **1997**, *118* (3), 199–205. <https://doi.org/10.1017/S0950268897007413>.
- (137) Svenungsson, B.; Lagergren, A.; Ekwall, E.; Evengard, B.; Hedlund, K. O.; Karnell, A.; Lofdahl, S.; Svensson, L.; Weintraub, A. Enteropathogens in Adult Patients with Diarrhea and Healthy Control Subjects: A 1-Year Prospective Study in a Swedish Clinic for Infectious Diseases. *Clin. Infect. Dis.* **2000**, *30* (5), 770–778. <https://doi.org/10.1086/313770>.
- (138) Michelacci, V.; Prosseda, G.; Maugliani, A.; Tozzoli, R.; Sanchez, S.; Herrera-León, S.; Dallman, T.; Jenkins, C.; Caprioli, A.; Morabito, S. Characterization of an Emergent Clone of Enteroinvasive *Escherichia coli* Circulating in Europe. *Clin. Microbiol. Infect.* **2016**, *22* (3), 287.e11–287.e19. <https://doi.org/10.1016/j.cmi.2015.10.025>.
- (139) Newitt, S.; MacGregor, V.; Robbins, V.; Bayliss, L.; Anne Chattaway, M.; Dallman, T.; Ready, D.; Aird, H.; Puleston, R.; Hawker, J. Two Linked Enteroinvasive *Escherichia coli* Outbreaks, Nottingham, UK, June 2014. *Emerg. Infect. Dis.* **2016**, *22* (7), 1178–1184. <https://doi.org/10.3201/eid2207.152080>.
- (140) Escher, M.; Scavia, G.; Morabito, S.; Tozzoli, R.; Maugliani, A.; Cantoni, S.; Fracchia, S.; Bettati, A.; Casa, R.; Gesu, G. P.; Torresani, E.; Caprioli, A. A Severe Foodborne Outbreak of Diarrhoea Linked to a Canteen in Italy Caused by Enteroinvasive *Escherichia coli*, an Uncommon Agent.

- Epidemiol. Infect.* **2014**, *142* (12), 2559–2566.
<https://doi.org/10.1017/S0950268814000181>.
- (141) Lan, R.; Reeves, P. R. *Escherichia coli* in Disguise: Molecular Origins of Shigella. *Microbes Infect.* **2002**, *4* (11), 1125–1132.
[https://doi.org/10.1016/S1286-4579\(02\)01637-4](https://doi.org/10.1016/S1286-4579(02)01637-4).
- (142) Pavlovic, M.; Luze, A.; Konrad, R.; Berger, A.; Sing, A.; Busch, U.; Huber, I. Development of a Duplex Real-Time PCR for Differentiation between *E. coli* and *Shigella* Spp. *J. Appl. Microbiol.* **2011**, *110* (5), 1245–1251.
<https://doi.org/10.1111/j.1365-2672.2011.04973.x>.
- (143) Van Den Beld, M. J. C.; Reubsæet, F. A. G. Differentiation between *Shigella*, Enteroinvasive *Escherichia coli* (EIEC) and Noninvasive *Escherichia coli*. *Eur. J. Clin. Microbiol. Infect. Dis.* **2012**, *31* (6), 899–904.
<https://doi.org/10.1007/s10096-011-1395-7>.
- (144) Martinez, M. B.; Whittan, T. S.; McGraw, E. A.; Rodrigues, J.; Trabulsi, L. R. Clonal Relationship among Invasive and Non-Invasive Strains of Enteroinvasive *Escherichia coli* Serogroups. *FEMS Microbiol. Lett.* **1999**, *172* (2), 145–151. [https://doi.org/10.1016/S0378-1097\(99\)00031-2](https://doi.org/10.1016/S0378-1097(99)00031-2).
- (145) Lan, R.; Alles, M. C.; Bonohoe, K.; Martinez, M. B.; Reeves, P. R. Molecular Evolutionary Relationships of Enteroinvasive *Escherichia coli* and *Shigella* Spp. *Infect. Immun.* **2004**, *72* (9), 5080–5088.
<https://doi.org/10.1128/IAI.72.9.5080-5088.2004>.
- (146) DuPont, H. L.; Formal, S. B.; Hornick, R. B.; Snyder, M. J.; Libonati, J. P.; Sheahan, D. G.; LaBrec, E. H.; Kalas, J. P. Pathogenesis of *Escherichia coli* Diarrhea. *N. Engl. J. Med.* **1971**, *285* (1), 1–9.
<https://doi.org/10.1056/NEJM197107012850101>.
- (147) Ramos Moreno, A. C.; Gonçalves Ferreira, L.; Baquerizo Martinez, M. Enteroinvasive *Escherichia coli* vs. *Shigella* Flexneri: How Different Patterns of Gene Expression Affect Virulence. *FEMS Microbiol. Lett.* **2009**, *301* (2), 156–163. <https://doi.org/10.1111/j.1574-6968.2009.01815.x>.
- (148) Harris, J. R.; Wachsmuth, I. K.; Davis, B. R.; Cohen, M. L. High-Molecular-Weight Plasmid Correlates with *Escherichia coli* Enteroinvasiveness. *Infect. Immun.* **1982**, *37* (3), 1295–1298.
- (149) Sansonetti, P. J.; Kopecko, D. J.; Formal, S. B. Involvement of a Plasmid in the Invasive Ability of *Shigella* Flexneri. *Infect. Immun.* **1982**, *35* (3), 852–860.
- (150) Lan, R.; Lumb, B.; Ryan, D.; Reeves, P. R. Molecular Evolution of Large Virulence Plasmid in *Shigella* Clones and Enteroinvasive *Escherichia coli*. *Infect. Immun.* **2001**, *69* (10), 6303–6309.
<https://doi.org/10.1128/IAI.69.10.6303-6309.2001>.

- (151) Sansonetti, P. J.; Phalipon, A. M Cells as Ports of Entry for Enteroinvasive Pathogens: Mechanisms of Interaction, Consequences for the Disease Process. *Semin. Immunol.* **1999**, *11* (3), 193–203. <https://doi.org/10.1006/smim.1999.0175>.
- (152) Ambrosi, C.; Pompili, M.; Scribano, D.; Limongi, D.; Petrucca, A.; Cannavacciuolo, S.; Schippa, S.; Zagaglia, C.; Grossi, M.; Nicoletti, M. The *Shigella* Flexneri OspB Effector: An Early Immunomodulator. *Int. J. Med. Microbiol.* **2015**, *305* (1), 75–84. <https://doi.org/10.1016/j.ijmm.2014.11.004>.
- (153) Mattock, E.; Blocker, A. J. How Do the Virulence Factors of *Shigella* Work Together to Cause Disease? *Front. Cell. Infect. Microbiol.* **2017**, *7* (MAR), 1–24. <https://doi.org/10.3389/fcimb.2017.00064>.
- (154) Schroeder, G. N.; Hilbi, H. Molecular Pathogenesis of *Shigella* Spp.: Controlling Host Cell Signaling, Invasion, and Death by Type III Secretion. *Clin. Microbiol. Rev.* **2008**, *21* (1), 134–156. <https://doi.org/10.1128/CMR.00032-07>.
- (155) Taylor, D. N.; Echeverria, P.; Sethabuter, O.; Pitarangsi, C.; Leksomboon, U.; Blacklow, N. R.; Rowe, B.; Gross, R.; Cross, J. Clinical and Microbiologic Features of *Shigella* and Enteroinvasive *Escherichia coli* Infections Detected by DNA Hybridization. *J. Clin. Microbiol.* **1988**, *26* (7), 1362–1366.
- (156) O'Brien, A. D.; Gentry, M. K.; Thompson, M. R.; Doctor, B. P.; Gemski, P.; Formal, S. B. Shigellosis and *Escherichia coli* Diarrhea: Relative Importance of Invasive and Toxigenic Mechanisms. *Am. J. Clin. Nutr.* **1979**, *32* (1), 229–233. <https://doi.org/10.1093/ajcn/32.1.229>.
- (157) Snelling, A. M.; MacFarlane-Smith, L. R.; Fletcher, J. N.; Okeke, I. N. The Commonly-Used DNA Probe for Diffusely-Adherent *Escherichia coli* Cross-React with a Subset of Enteroaggregative *E. coli*. *BMC Microbiol.* **2009**, *9*, 1–10. <https://doi.org/10.1186/1471-2180-9-269>.
- (158) Scaletsky, I. C. A.; Silva, M. L. M.; Trabulsi, L. R. Distinctive Patterns of Adherence of Enteropathogenic *Escherichia coli* to HeLa Cells. *Infect. Immun.* **1984**, *45* (2), 534–536.
- (159) Lopes, L. M.; Fabbriotti, S. H.; Ferreira, A. J. P.; Kato, M. A. M. F.; Michalski, J.; Scaletsky, I. C. A. Heterogeneity among Strains of Diffusely Adherent *Escherichia coli* Isolated in Brazil. *J. Clin. Microbiol.* **2005**, *43* (4), 1968–1972. <https://doi.org/10.1128/JCM.43.4.1968-1972.2005>.
- (160) Gunzburg, S. T.; Chang, B. J.; Gunzburg, S. T.; Elliott, S. J.; Burke, V. Diffuse and Enteroaggregative Patterns of Adherence of Enteric *Escherichia coli* Isolated from Aboriginal Children from the Kimberley Region of Western Australia. *J. Infect. Dis.* **1993**, *167* (3), 755–758.

<https://doi.org/10.1093/infdis/167.3.755>.

- (161) Girón, J. A.; Jones, T.; Miilán-Velasco, F.; Castro-Muñoz, E.; Zárate, L.; Fry, J.; Frankel, G.; Moseley, S. L.; Baudry, B.; Kaper, J. B.; Schoolnik, G. K.; Riley, L. W. Diffuse-Adhering *Escherichia coli* (DAEC) as a Putative Cause of Diarrhea in Mayan Children in Mexico. *J. Infect. Dis.* **1991**, *163* (3), 507–513. <https://doi.org/10.1093/infdis/163.3.507>.
- (162) Knutton, S.; Shaw, R.; Phillips, A. D.; Smith, H. R.; Willshaw, G. A.; Watson, P.; Price, E. Phenotypic and Genetic Analysis of Diarrhea-Associated *Escherichia coli* Isolated From Children in the United Kingdom. *J. Pediatr. Gastroenterol. Nutr.* **2001**, *33* (1), 32–40. <https://doi.org/10.1097/00005176-200107000-00006>.
- (163) Servin, A. L. Pathogenesis of Human Diffusely Adhering *Escherichia coli* Expressing Afa/Dr Adhesins (Afa/Dr DAEC): Current Insights and Future Challenges. *Clin. Microbiol. Rev.* **2014**, *27* (4), 823–869. <https://doi.org/10.1128/CMR.00036-14>.
- (164) Le Bouguéneq, C.; Servin, A. L. Diffusely Adherent *Escherichia coli* Strains Expressing Afa/Dr Adhesins (Afa/Dr DAEC): Hitherto Unrecognized Pathogens. *FEMS Microbiol. Lett.* **2006**, *256* (2), 185–194. <https://doi.org/10.1111/j.1574-6968.2006.00144.x>.
- (165) Bernet-Camard, M. F.; Coconnier, M. H.; Hudault, S.; Servin, A. L. Pathogenicity of the Diffusely Adhering Strain *Escherichia coli* C1845: F1845 Adhesin-Decay Accelerating Factor Interaction, Brush Border Microvillus Injury, and Actin Disassembly in Cultured Human Intestinal Epithelial Cells. *Infect. Immun.* **1996**, *64* (6), 1918–1928.
- (166) Servin, A. L. Pathogenesis of Afa/Dr Diffusely Adhering *Escherichia coli*. *Clin. Microbiol. Rev.* **2005**, *18* (2), 264–292. <https://doi.org/10.1128/CMR.18.2.264>.
- (167) Guignot, J.; Chaplais, C.; Coconnier-Polter, M. H.; Servin, A. L. The Secreted Autotransporter Toxin, Sat, Functions as a Virulence Factor in Afa/Dr Diffusely Adhering *Escherichia coli* by Promoting Lesions in Tight Junction of Polarized Epithelial Cells. *Cell. Microbiol.* **2007**, *9* (1), 204–221. <https://doi.org/10.1111/j.1462-5822.2006.00782.x>.
- (168) Nataro, J. P.; Kaper, J. B.; Robins-Browne, R.; Prado, V.; Vial, P.; Levine, M. M. Patterns of Adherence of Diarrheagenic *Escherichia coli* to HEp-2 Cells. *Pediatr. Infect. Dis. J.* **1987**, *6* (9), 829–831. <https://doi.org/10.1097/00006454-198709000-00008>.
- (169) Nataro, J. P.; Yikang, D.; Cookson, S.; Cravioto, A.; Savarino, S. J.; Guers, L. D.; Levine, M. M.; Tacket, C. O. Heterogeneity of Enteroaggregative *Escherichia coli* Virulence Demonstrated. *J. Infect. Dis.* **1995**, *171* (2), 465–468. <https://doi.org/10.1093/infdis/171.2.465>.

- (170) Asea, A.; Kaur, P.; Chakraborti, A. Enteroaggregative *Escherichia coli*: An Emerging Enteric Food Borne Pathogen. *Interdiscip. Perspect. Infect. Dis.* **2010**, *2010*. <https://doi.org/10.1155/2010/254159>.
- (171) Huang, D. B.; Mohamed, J. A.; Nataro, J. P.; DuPont, H. L.; Jiang, Z. D.; Okhuysen, P. C. Virulence Characteristics and the Molecular Epidemiology of Enteroaggregative *Escherichia coli* Isolates from Travellers to Developing Countries. *J. Med. Microbiol.* **2007**, *56* (10), 1386–1392. <https://doi.org/10.1099/jmm.0.47161-0>.
- (172) Harrington, S. M.; Dudley, E. G.; Nataro, J. P. Pathogenesis of Enteroaggregative *Escherichia coli* Infection. *FEMS Microbiol. Lett.* **2006**, *254* (1), 12–18. <https://doi.org/10.1111/j.1574-6968.2005.00005.x>.
- (173) Wilson, A.; Evans, J.; Chart, H.; Cheasty, T.; Wheeler, J. G.; Tompkins, D.; Smith, H. R. Characterisation of Strains of Enteroaggregative *Escherichia coli* Isolated during the Infectious Intestinal Disease Study in England. *Eur. J. Epidemiol.* **2001**, *17* (12), 1125–1130. <https://doi.org/10.1023/A:1021224915322>.
- (174) Okeke, I. N.; Lamikanra, A.; Czczulin, J.; Dubovsky, F.; Kaper, J. B.; Nataro, J. P. Heterogeneous Virulence of Enteroaggregative *Escherichia coli* Strains Isolated from Children in Southwest Nigeria. *J. Infect. Dis.* **2000**, *181* (1), 252–260. <https://doi.org/10.1086/315204>.
- (175) Nataro, J. P.; Mai, V.; Johnson, J.; Blackwelder, W. C.; Heimer, R.; Tirrell, S.; Edberg, S. C.; Braden, C. R.; Morris, J. G.; Hirshon, J. M. Diarrheagenic *Escherichia coli* Infection in Baltimore, Maryland, and New Haven, Connecticut. *Clin. Infect. Dis.* **2006**, *43* (4), 402–407. <https://doi.org/10.1086/505867>.
- (176) Cohen, M. B.; Nataro, J. P.; Bernstein, D. I.; Hawkins, J.; Roberts, N.; Staat, M. A. Prevalence of Diarrheagenic *Escherichia coli* in Acute Childhood Enteritis: A Prospective Controlled Study. *J. Pediatr.* **2005**, *146* (1), 54–61. <https://doi.org/10.1016/j.jpeds.2004.08.059>.
- (177) Vial, P. A.; Robins-Browne, R.; Lior, H.; Prado, V.; Kaper, J. B.; Nataro, J. P.; Maneval, D.; Elsayed, A. E. D.; Levine, M. M. Characterization of Enteroadherent-Aggregative *Escherichia coli*, a Putative Agent of Diarrheal Disease. *J. Infect. Dis.* **1988**, *158* (1), 70–79. <https://doi.org/10.1093/infdis/158.1.70>.
- (178) Jenkins, C.; Van Ijperen, C.; Dudley, E. G.; Chart, H.; Willshaw, G. A.; Cheasty, T.; Smith, H. R.; Nataro, J. P. Use of a Microarray to Assess the Distribution of Plasmid and Chromosomal Virulence Genes in Strains of Enteroaggregative *Escherichia coli*. *FEMS Microbiol. Lett.* **2005**, *253* (1), 119–124. <https://doi.org/10.1016/j.femsle.2005.09.040>.
- (179) Gallegos, M. trinidad; Michán, C.; Ramos, J. L. The XylS/AraC Family of

- Regulators. *Nucleic Acids Res.* **1993**, *21* (4), 807–810. <https://doi.org/10.1093/nar/21.4.807>.
- (180) Nataro, J. P.; Yikang, D.; Yingkang, D.; Walker, K. AggR, a Transcriptional Activator of Aggregative Adherence Fimbria I Expression in Enteroaggregative *Escherichia coli*. *J. Bacteriol.* **1994**, *176* (15), 4691–4699. <https://doi.org/10.1128/jb.176.15.4691-4699.1994>.
- (181) Morin, N.; Santiago, A. E.; Ernst, R. K.; Guillot, S. J.; Nataro, J. P. Characterization of the AggR Regulon in Enteroaggregative *Escherichia coli*. *Infect. Immun.* **2013**, *81* (1), 122–132. <https://doi.org/10.1128/IAI.00676-12>.
- (182) Okeke, I. N.; Wallace-Gadsden, F.; Simons, H. R.; Matthews, N.; Labar, A. S.; Hwang, J.; Wain, J. Multi-Locus Sequence Typing of Enteroaggregative *Escherichia coli* Isolates from Nigerian Children Uncovers Multiple Lineages. *PLoS One* **2010**, *5* (11). <https://doi.org/10.1371/journal.pone.0014093>.
- (183) Frank C, Werber D, Cramer JP, Askar M, Faber M, an der Heiden M, Bernard H, F. A.; Prager R, Spode A, Wadl M, Zoufaly A, Jordan S, Kemper MJ, Follin P, Müller L, King LA, Rosner B, Buchholz U, Stark K, K. G. H. I. T. Epidemic Profile of Shiga-Toxin-Producing *Escherichia coli* O104:H4 Outbreak in Germany. *N. Engl. J. Med.* **2011**, *365* (19), 1771–1780. <https://doi.org/10.1056/NEJMoa1106483>.
- (184) Mayer, C. L.; Leibowitz, C. S.; Kurosawa, S.; Stearns-Kurosawa, D. J. Shiga Toxins and the Pathophysiology of Hemolytic Uremic Syndrome in Humans and Animals. *Toxins (Basel)*. **2012**, *4* (11), 1261–1287. <https://doi.org/10.3390/toxins4111261>.
- (185) Scheutz, F.; Nielsen, E. M.; Frimodt-Møller, J.; Boisen, N.; Morabito, S.; Tozzoli, R.; Nataro, J. P.; Caprioli, A. Characteristics of the Enteroaggregative Shiga Toxin/Verotoxin-Producing *Escherichia coli* O104:H4 Strain Causing the Outbreak of Haemolytic Uraemic Syndrome in Germany, May to June 2011. *Eurosurveillance* **2011**, *16* (24), 1–6. <https://doi.org/10.2807/ese.16.24.19889-en>.
- (186) Morabito, S.; Karch, H.; Mariani-Kurkdjian, P.; Schmidt, H.; Minelli, F.; Bingen, E.; Caprioli, A. Enteroaggregative, Shiga Toxin-Producing *Escherichia coli* O111:H2 Associated with an Outbreak of Hemolytic-Uremic Syndrome. *J. Clin. Microbiol.* **1998**, *36* (3), 840–842.
- (187) King, L. A.; Nogareda, F.; Weill, F. X.; Mariani-Kurkdjian, P.; Loukiadis, E.; Gault, G.; Jourdan-Dasilva, N.; Bingen, E.; MacÉ, M.; Thevenot, D.; Ong, N.; Castor, C.; Noël, H.; Van Cauteren, D.; Charron, M.; Vaillant, V.; Aldabe, B.; Goulet, V.; Delmas, G.; Couturier, E.; Le Strat, Y.; Combe, C.; Delmas, Y.; Terrier, F.; Vendrely, B.; Rolland, P.; De Valk, H. Outbreak of Shiga Toxin-Producing *Escherichia coli* O104:H4 Associated with Organic Fenugreek

- Sprouts, France, June 2011. *Clin. Infect. Dis.* **2012**, *54* (11), 1588–1594. <https://doi.org/10.1093/cid/cis255>.
- (188) Iyoda, S.; Tamura, K.; Itoh, K.; Izumiya, H.; Ueno, N.; Nagata, K.; Togo, M.; Terajima, J.; Watanabe, H. Inducible Stx2 Phages Are Lysogenized in the Enteroaggregative and Other Phenotypic *Escherichia coli* O86:HNM Isolated from Patients. *FEMS Microbiol. Lett.* **2000**, *191* (1), 7–10. [https://doi.org/10.1016/S0378-1097\(00\)00360-8](https://doi.org/10.1016/S0378-1097(00)00360-8).
- (189) Scavia, G.; Morabito, S.; Tozzoli, R.; Michelacci, V.; Marziano, M. L.; Minelli, F.; Ferreri, C.; Paglialonga, F.; Edefonti, A.; Caprioli, A. Similarity of Shiga Toxin–Producing *Escherichia coli* O104:H4 Strains from Italy and Germany. *Emerg. Infect. Dis.* **2011**, *17* (10), 1957–1958. <https://doi.org/10.3201/eid1710.111072>.
- (190) Dallman, T.; Smith, G. P.; O'Brien, B.; Chattaway, M. A.; Finlay, D.; Grant, K. A.; Jenkins, C. Characterization of a Verocytotoxin-Producing Enteroaggregative *Escherichia coli* Serogroup O111:H21 Strain Associated with a Household Outbreak in Northern Ireland. *J. Clin. Microbiol.* **2012**, *50* (12), 4116–4119. <https://doi.org/10.1128/JCM.02047-12>.
- (191) Tozzoli, R.; Grande, L.; Michelacci, V.; Ranieri, P.; Maugliani, A.; Caprioli, A.; Morabito, S. Shiga Toxin-Converting Phages and the Emergence of New Pathogenic *Escherichia coli*: A World in Motion. *Front. Cell. Infect. Microbiol.* **2014**, *4* (JUN), 1–8. <https://doi.org/10.3389/fcimb.2014.00080>.
- (192) Abe, C. M.; Salvador, F. A.; Falsetti, I. N.; Vieira, M. A. M.; Blanco, J.; Blanco, J. E.; Blanco, M.; MacHado, A. M. O.; Elias, W. P.; Hernandez, R. T.; Gomes, T. A. T. Uropathogenic *Escherichia coli* (UPEC) Strains May Carry Virulence Properties of Diarrhoeagenic *E. coli*. *FEMS Immunol. Med. Microbiol.* **2008**, *52* (3), 397–406. <https://doi.org/10.1111/j.1574-695X.2008.00388.x>.
- (193) Lara, F. B. M.; Nery, D. R.; de Oliveira, P. M.; Araujo, M. L.; Carvalho, F. R. Q.; Messias-Silva, L. C. F.; Ferreira, L. B.; Faria-Junior, C.; Pereira, A. L. Virulence Markers and Phylogenetic Analysis of *Escherichia coli* Strains with Hybrid EAEC/UPEC Genotypes Recovered from Sporadic Cases of Extraintestinal Infections. *Front. Microbiol.* **2017**, *8* (FEB), 1–11. <https://doi.org/10.3389/fmicb.2017.00146>.
- (194) Olesen, B.; Scheutz, F.; Andersen, R. L.; Menard, M.; Boisen, N.; Johnston, B.; Hansen, D. S.; Kroghfelt, K. A.; Nataro, J. P.; Johnson, J. R. Enteroaggregative *Escherichia coli* O78:H10, the Cause of an Outbreak of Urinary Tract Infection. *J. Clin. Microbiol.* **2012**, *50* (11), 3703–3711. <https://doi.org/10.1128/JCM.01909-12>.

- (195) Jenkins, C. Enteroaggregative *Escherichia coli*. In *Current Topics in Microbiology and Immunology*; 2018; Vol. 340, pp 27–50. https://doi.org/10.1007/82_2018_105.
- (196) Jensen, B. H.; Olsen, K. E. P.; Struve, C.; Krogfelt, K. A.; Petersen, A. M. Epidemiology and Clinical Manifestations of Enteroaggregative *Escherichia coli*. *Clin. Microbiol. Rev.* **2014**, 27 (3), 614–630. <https://doi.org/10.1128/CMR.00112-13>.
- (197) Chaudhuri, R. R.; Sebaihia, M.; Hobman, J. L.; Webber, M. A.; Leyton, D. L.; Goldberg, M. D.; Cunningham, A. F.; Scott-Tucker, A.; Ferguson, P. R.; Thomas, C. M.; Frankel, G.; Tang, C. M.; Dudley, E. G.; Roberts, I. S.; Rasko, D. A.; Pallen, M. J.; Parkhill, J.; Nataro, J. P.; Thomson, N. R.; Henderson, I. R. Complete Genome Sequence and Comparative Metabolic Profiling of the Prototypical Enteroaggregative *Escherichia coli* Strain 042. *PLoS One* **2010**, 5 (1). <https://doi.org/10.1371/journal.pone.0008801>.
- (198) Liebert, C. A.; Hall, R. M.; Summers, A. O. Transposon Tn21, Flagship of the Floating Genome. *Microbiol. Mol. Biol. Rev.* **1999**, 63 (3), 507–522.
- (199) Yamamoto, T.; Echeverria, P.; Yokota, T. Drug Resistance and Adherence to Human Intestines of Enteroaggregative *Escherichia coli*. *J. Infect. Dis.* **1992**, 165 (4), 744–749. <https://doi.org/10.1093/infdis/165.4.744>.
- (200) Raju, B.; Ballal, M. Multidrug Resistant Enteroaggregative *Escherichia coli* Diarrhoea in Rural Southern Indian Population. *Scand. J. Infect. Dis.* **2009**, 41 (2), 105–108. <https://doi.org/10.1080/00365540802641856>.
- (201) Sang, W. K.; Oundo, J. O.; Mwituria, J. K.; Waiyaki, P. G.; Yoh, M.; Iida, T.; Honda, T. Multidrug-Resistant Enteroaggregative *Escherichia coli* Associated with Persistent Diarrhea in Kenyan Children. *Emerg. Infect. Dis.* **1997**, 3 (3), 373–374. <https://doi.org/10.3201/eid0303.970317>.
- (202) Mendez Arancibia, E.; Pitart, C.; Ruiz, J.; Marco, F.; Gascón, J.; Vila, J. Evolution of Antimicrobial Resistance in Enteroaggregative *Escherichia coli* and Enterotoxigenic *Escherichia coli* Causing Traveller's Diarrhoea. *J. Antimicrob. Chemother.* **2009**, 64 (2), 343–347. <https://doi.org/10.1093/jac/dkp178>.
- (203) Hunt, S.; Green, J.; Artymiuk, P. J. Hemolysin E (HlyE, ClyA, SheA) and Related Toxins. In *Advances in experimental medicine and biology*; 2010; Vol. 677, pp 116–126. https://doi.org/10.1007/978-1-4419-6327-7_10.
- (204) Prieto, A.; Bernabeu, M.; Aznar, S.; Ruiz-Cruz, S.; Bravo, A.; Queiroz, M. H.; Juárez, A. Evolution of Bacterial Global Modulators: Role of a Novel H-NS Parologue in the Enteroaggregative *Escherichia coli* Strain 042. *mSystems* **2018**, 3 (3), 1–14. <https://doi.org/10.1128/msystems.00220-17>.

- (205) Prieto, A.; Urcola, I.; Blanco, J.; Dahbi, G.; Muniesa, M.; Quirós, P.; Falgenhauer, L.; Chakraborty, T.; Hüttener, M.; Juárez, A. Tracking Bacterial Virulence: Global Modulators as Indicators. *Sci. Rep.* **2016**, *6* (1), 25973. <https://doi.org/10.1038/srep25973>.
- (206) Valkenburg, J. A. C.; Woldringh, C. L. Phase Separation between Nucleoid and Cytoplasm in *Escherichia coli* as Defined by Immersive Refractometry. *J. Bacteriol.* **1984**, *160* (3), 1151–1157.
- (207) Qin, L.; Erkelens, A. M.; Ben Bdira, F.; Dame, R. T. The Architects of Bacterial DNA Bridges: A Structurally and Functionally Conserved Family of Proteins. *Open Biol.* **2019**, *9* (12), 190223. <https://doi.org/10.1098/rsob.190223>.
- (208) Dillon, S. C.; Dorman, C. J. Bacterial Nucleoid-Associated Proteins, Nucleoid Structure and Gene Expression. *Nat. Rev. Microbiol.* **2010**, *8* (3), 185–195. <https://doi.org/10.1038/nrmicro2261>.
- (209) Stavans, J.; Oppenheim, A. DNA-Protein Interactions and Bacterial Chromosome Architecture. *Phys. Biol.* **2006**, *3* (4). <https://doi.org/10.1088/1478-3975/3/4/R01>.
- (210) De Vries, R. DNA Condensation in Bacteria: Interplay between Macromolecular Crowding and Nucleoid Proteins. *Enferm. Infec. Microbiol. Clin.* **2010**, *28* (SUPPL. 3), 1715–1721. <https://doi.org/10.1016/j.biochi.2010.06.024>.
- (211) Zhou, H.-X.; Rivas, G.; Minton, A. P. Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences. *Annu. Rev. Biophys.* **2008**, *37* (1), 375–397. <https://doi.org/10.1146/annurev.biophys.37.032807.125817>.
- (212) Pelletier, J.; Halvorsen, K.; Ha, B. Y.; Paparcone, R.; Sandler, S. J.; Woldringh, C. L.; Wong, W. P.; Jun, S. Physical Manipulation of the *Escherichia coli* Chromosome Reveals Its Soft Nature. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (40). <https://doi.org/10.1073/pnas.1208689109>.
- (213) Cunha, S.; Woldringh, C. L.; Odijk, T. Polymer-Mediated Compaction and Internal Dynamics of Isolated *Escherichia coli* Nucleoids. *J. Struct. Biol.* **2001**, *136* (1), 53–66. <https://doi.org/10.1006/jsbi.2001.4420>.
- (214) Dorman, C. J. DNA Supercoiling and Transcription in Bacteria: A Two-Way Street. *BMC Mol. and Cell Bio.* **2019**, *6*, 1–9 <https://doi.org/10.1186/s12860-019-0211-6>.
- (215) Dame, R. T. The Role of Nucleoid-Associated Proteins in the Organization and Compaction of Bacterial Chromatin. *Mol. Microbiol.* **2005**, *56* (4), 858–870. <https://doi.org/10.1111/j.1365-2958.2005.04598.x>.

- (216) Travers, A.; Muskhelishvili, G. Bacterial Chromatin. *Curr. Opin. Genet. Dev.* **2005**, *15* (5 SPEC. ISS.), 507–514. <https://doi.org/10.1016/j.gde.2005.08.006>.
- (217) McLeod, S. M.; Johnson, R. C. Control of Transcription by Nucleoid Proteins. *Curr. Opin. Microbiol.* **2001**, *4* (2), 152–159. [https://doi.org/10.1016/S1369-5274\(00\)00181-8](https://doi.org/10.1016/S1369-5274(00)00181-8).
- (218) Bertin, P.; Hommais, F.; Krin, E.; Soutourina, O.; Tendeng, C.; Derzelle, S.; Danchin, A. H-NS and H-NS-like Proteins in Gram-Negative Bacteria and Their Multiple Role in the Regulation of Bacterial Metabolism. *Biochimie* **2001**, *83* (2), 235–241. [https://doi.org/10.1016/S0300-9084\(01\)01247-0](https://doi.org/10.1016/S0300-9084(01)01247-0).
- (219) Jacquet, M.; Cukier-Kahn, R.; Pla, J.; Gros, F. A Thermostable Protein Factor Acting on in Vitro DNA Transcription. *Biochem. Biophys. Res. Commun.* **1971**, *45* (6), 1597–1607. [https://doi.org/10.1016/0006-291X\(71\)90204-X](https://doi.org/10.1016/0006-291X(71)90204-X).
- (220) Varshavsky, A. J.; Nedospasov, S. A.; Bakayev, V. V.; Bakayeva, T. G.; Georgiev, G. P. Histone-like Proteins in the Purified *Escherichia coli* Deoxyribonucleoprotein. *Nucleic Acids Res.* **1977**, *4* (8), 2725–2746. <https://doi.org/10.1093/nar/4.8.2725>.
- (221) Dorman, C. J. H-NS: A Universal Regulator for a Dynamic Genome. *Nat. Rev. Microbiol.* **2004**, *2* (5), 391–400. <https://doi.org/10.1038/nrmicro883>.
- (222) Baños, R. C.; Vivero, A.; Aznar, S.; García, J.; Pons, M.; Madrid, C.; Juárez, A. Differential Regulation of Horizontally Acquired and Core Genome Genes by the Bacterial Modulator H-NS. *PLoS Genet.* **2009**, *5* (6). <https://doi.org/10.1371/journal.pgen.1000513>.
- (223) Schröder, O.; Wagner, R. The Bacterial Regulatory Protein H-NS A Versatile Modulator of Nucleic Acid Structures. *Biol. Chem.* **2002**, *383* (6), 1836–1845. <https://doi.org/10.1515/BC.2002.101>.
- (224) Hommais, F.; Krin, E.; Laurent-Winter, C.; Soutourina, O.; Malpertuy, A.; Le Caer, J. P.; Danchin, A.; Bertin, P. Large-Scale Monitoring of Pleiotropic Regulation of Gene Expression by the Prokaryotic Nucleoid-Associated Protein, H-NS. *Mol. Microbiol.* **2001**, *40* (1), 20–36. <https://doi.org/10.1046/j.1365-2958.2001.02358.x>.
- (225) Srinivasan, R.; Scolari, V. F.; Lagomarsino, M. C.; Seshasayee, A. S. N. The Genome-Scale Interplay amongst Xenogene Silencing, Stress Response and Chromosome Architecture in *Escherichia coli*. *Nucleic Acids Res.* **2015**, *43* (1), 295–308. <https://doi.org/10.1093/nar/gku1229>.

- (226) Tendeng, C.; Bertin, P. N. H-NS in Gram-Negative Bacteria: A Family of Multifaceted Proteins. *Trends Microbiol.* **2003**, *11* (11), 511–518. <https://doi.org/10.1016/j.tim.2003.09.005>.
- (227) Göransson, M.; Sondén, B.; Nilsson, P.; Dagberg, B.; Foreman, K.; Emanuelsson, K.; Uhlin, B. E. Transcriptional Silencing and Thermoregulation of Gene Expression in *Escherichia coli*. *Nature* **1990**, *344* (6267), 682–685. <https://doi.org/10.1006/jmbi.1994.1404>.
- (228) Ono, S.; Goldberg, M. D.; Olsson, T.; Esposito, D.; Hinton, J. C. D.; Ladbury, J. E. H-NS Is a Part of a Thermally Controlled Mechanism for Bacterial Gene Regulation. *Biochem. J.* **2005**, *391* (2), 203–213. <https://doi.org/10.1042/BJ20050453>.
- (229) White-Ziegler, C. A.; Davis, T. R. Genome-Wide Identification of H-NS-Controlled, Temperature-Regulated Genes in *Escherichia coli* K-12. *J. Bacteriol.* **2009**, *191* (3), 1106–1110. <https://doi.org/10.1128/JB.00599-08>.
- (230) Yamada, H.; Muramatsu, S.; Mizuno, T. An *Escherichia coli* Protein That Preferentially Binds to Sharply Curved DNA. *J. Biochem.* **1990**, *108* (3), 420–425. <https://doi.org/10.1093/oxfordjournals.jbchem.a123216>.
- (231) Queiroz, M. H.; Madrid, C.; Paytubi, S.; Balsalobre, C.; Juárez, A. Integration Host Factor Alleviates H-NS Silencing of the *Salmonella* Enterica Serovar Typhimurium Master Regulator of SPI1, HilA. *Microbiology* **2011**, *157* (9), 2504–2514. <https://doi.org/10.1099/mic.0.049197-0>.
- (232) van Elsas, J. D.; Semenov, A. V.; Costa, R.; Trevors, J. T. Survival of *Escherichia coli* in the Environment: Fundamental and Public Health Aspects. *ISME J.* **2011**, *5* (2), 173–183. <https://doi.org/10.1038/ismej.2010.80>.
- (233) Smith, J. Superinfection Drives Virulence Evolution in Experimental Populations of Bacteria and Plasmids. *Evolution (N. Y.)* **2011**, *65* (3), 831–841. <https://doi.org/10.1111/j.1558-5646.2010.01178.x>.
- (234) Bennett, P. M. Plasmid Encoded Antibiotic Resistance: Acquisition and Transfer of Antibiotic Resistance Genes in Bacteria. *Br. J. Pharmacol.* **2008**, *153* (SUPPL. 1), 347–357. <https://doi.org/10.1038/sj.bjp.0707607>.
- (235) Svava, F.; Rankin, D. J. The Evolution of Plasmid-Carried Antibiotic Resistance. *BMC Evol. Biol.* **2011**, *11* (1), 130. <https://doi.org/10.1186/1471-2148-11-130>.
- (236) Navarre, W. W.; McClelland, M.; Libby, S. J.; Fang, F. C. Silencing of Xenogeneic DNA by H-NS - Facilitation of Lateral Gene Transfer in Bacteria by a Defense System That Recognizes Foreign DNA. *Genes Dev.* **2007**, *21* (12), 1456–1471. <https://doi.org/10.1101/gad.1543107>.

- (237) Ali, S. S.; Xia, B.; Liu, J.; Navarre, W. W. Silencing of Foreign DNA in Bacteria. *Curr. Opin. Microbiol.* **2012**, *15* (2), 175–181. <https://doi.org/10.1016/j.mib.2011.12.014>.
- (238) Forsman, K.; Sonden, B.; Goransson, M.; Uhlin, B. E. Antirepression Function in *Escherichia coli* for the CAMP-CAMP Receptor Protein Transcriptional Activator. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (20), 9880–9884. <https://doi.org/10.1073/pnas.89.20.9880>.
- (239) Bustamante, V. H.; Santana, F. J.; Calva, E.; Puente, J. L. Transcriptional Regulation of Type III Secretion Genes in Enteropathogenic *Escherichia coli*: Ler Antagonizes H-NS-Dependent Repression. *Mol. Microbiol.* **2001**, *39* (3), 664–678. <https://doi.org/10.1046/j.1365-2958.2001.02209.x>.
- (240) Ueguchi, C.; Suzuki, T.; Yoshida, T.; Tanaka, K. I.; Mizuno, T. Systematic Mutational Analysis Revealing the Functional Domain Organization of *Escherichia coli* Nucleoid Protein H-NS. *J. Mol. Biol.* **1996**, *263* (2), 149–162. <https://doi.org/10.1006/jmbi.1996.0566>.
- (241) Bloch, V.; Yang, Y.; Margeat, E.; Chavanieu, A.; Augé, M. T.; Robert, B.; Arold, S.; Rimsky, S.; Kochoyan, M. The H-NS Dimerization Domain Defines a New Fold Contributing to DNA Recognition. *Nat. Struct. Biol.* **2003**, *10* (3), 212–218. <https://doi.org/10.1038/nsb904>.
- (242) Shindo, H.; Ohnuki, A.; Ginba, H.; Katoh, E.; Ueguchi, C.; Mizuno, T.; Yamazaki, T. Identification of the DNA Binding Surface of H-NS Protein from *Escherichia coli* by Heteronuclear NMR Spectroscopy. *FEBS Lett.* **1999**, *455* (1–2), 63–69. [https://doi.org/10.1016/S0014-5793\(99\)00862-5](https://doi.org/10.1016/S0014-5793(99)00862-5).
- (243) Badaut, C.; Williams, R.; Arluison, V.; Bouffartigues, E.; Robert, B.; Buc, H.; Rimsky, S. The Degree of Oligomerization of the H-NS Nucleoid Structuring Protein Is Related to Specific Binding to DNA. *J. Biol. Chem.* **2002**, *277* (44), 41657–41666. <https://doi.org/10.1074/jbc.M206037200>.
- (244) Dame, R. T.; Wyman, C.; Wurm, R.; Wagner, R.; Goosen, N. Structural Basis for H-NS-Mediated Trapping of RNA Polymerase in the Open Initiation Complex at the RrnB P1. *J. Biol. Chem.* **2002**, *277* (3), 2146–2150. <https://doi.org/10.1074/jbc.C100603200>.
- (245) Dorman, C. J.; Hinton, J. C. D.; Free, A. Domain Organization and Oligomerization among H-NS-like Nucleoid-Associated Proteins in Bacteria. *Trends Microbiol.* **1999**, *7* (3), 124–128. [https://doi.org/10.1016/S0966-842X\(99\)01455-9](https://doi.org/10.1016/S0966-842X(99)01455-9).
- (246) Müller, C. M.; Dobrindt, U.; Nagy, G.; Emödy, L.; Uhlin, B. E.; Hacker, J. Role of Histone-like Proteins H-NS and StpA in Expression of Virulence Determinants of Uropathogenic *Escherichia coli*. *J. Bacteriol.* **2006**, *188*

- (15), 5428–5438. <https://doi.org/10.1128/JB.01956-05>.
- (247) Zhang, A.; Derbyshire, V.; Galloway Salvo, J. L.; Belfort, M. *Escherichia coli* Protein StpA Stimulates Self-Splicing by Promoting RNA Assembly in Vitro. *Rna*. 1995, pp 783–793.
- (248) Dorman, C. J. H-NS-like Nucleoid-Associated Proteins, Mobile Genetic Elements and Horizontal Gene Transfer in Bacteria. *Plasmid* **2014**, *75*, 1–11. <https://doi.org/10.1016/j.plasmid.2014.06.004>.
- (249) Mayer, O.; Rajkowitsch, L.; Lorenz, C.; Konrat, R.; Schroeder, R. RNA Chaperone Activity and RNA-Binding Properties of the *E. coli* Protein StpA. *Nucleic Acids Res.* **2007**, *35* (4), 1257–1269. <https://doi.org/10.1093/nar/gkl1143>.
- (250) Paytubi, S.; Madrid, C.; Forn, N.; Nieto, J. M.; Balsalobre, C.; Uhlin, B. E.; Juárez, A. YdgT, the Hha Parologue in *Escherichia coli*, Forms Heteromeric Complexes with H-NS and StpA. *Mol. Microbiol.* **2004**, *54* (1), 251–263. <https://doi.org/10.1111/j.1365-2958.2004.04268.x>.
- (251) Johansson, J.; Eriksson, S.; Sondén, B.; Sun Nyunt Wai; Uhlin, B. E. Heteromeric Interactions among Nucleoid-Associated Bacterial Proteins: Localization of StpA-Stabilizing Regions in H-NS of *Escherichia coli*. *J. Bacteriol.* **2001**, *183* (7), 2343–2347. <https://doi.org/10.1128/JB.183.7.2343-2347.2001>.
- (252) Deighan, P.; Free, A.; Dorman, C. J. A Role for the *Escherichia coli* H-NS-like Protein StpA in OmpF Porin Expression through Modulation of MicF RNA Stability. *Mol. Microbiol.* **2000**, *38* (1), 126–139. <https://doi.org/10.1046/j.1365-2958.2000.02120.x>.
- (253) Lim, C. J.; Whang, Y. R.; Kenney, L. J.; Yan, J. Gene Silencing H-NS Parologue StpA Forms a Rigid Protein Filament along DNA That Blocks DNA Accessibility. *Nucleic Acids Res.* **2012**, *40* (8), 3316–3328. <https://doi.org/10.1093/nar/gkr1247>.
- (254) Free, A.; Dorman, C. J. The *Escherichia coli* StpA Gene Is Transiently Expressed during Growth in Rich Medium and Is Induced in Minimal Medium and by Stress Conditions. **1997**, *179* (3), 909–918.
- (255) Sondén, B.; Uhlin, B. E. Coordinated and Differential Expression of Histone-like Proteins in *Escherichia coli*: Regulation and Function of the H-NS Analog StpA. *EMBO J.* **1996**, *15* (18), 4970–4980. <https://doi.org/10.1002/j.1460-2075.1996.tb00877.x>.
- (256) Free, A.; Dorman, C. J. The *Escherichia coli* StpA Gene Is Transiently Expressed during Growth in Rich Medium and Is Induced in Minimal Medium and by Stress Conditions. *J. Bacteriol.* **1997**, *179* (3), 909–918. <https://doi.org/10.1128/jb.179.3.909-918.1997>.

Bibliografía

- (257) Johansson, J.; Uhlin, B. E. Differential Protease-Mediated Turnover of H-NS and StpA Revealed by a Mutation Altering Protein Stability and Stationary-Phase Survival of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (19), 10776–10781. <https://doi.org/10.1073/pnas.96.19.10776>.
- (258) Müller, C. M.; Schneider, G.; Dobrindt, U.; Emödy, L.; Hacker, J.; Uhlin, B. E. Differential Effects and Interactions of Endogenous and Horizontally Acquired H-NS-like Proteins in Pathogenic *Escherichia coli*. *Mol. Microbiol.* **2010**, *75* (2), 280–293. <https://doi.org/10.1111/j.1365-2958.2009.06995.x>.
- (259) Baños, R. C.; Aznar, S.; Madrid, C.; Juárez, A. Differential Functional Properties of Chromosomal- and Plasmid-Encoded H-NS Proteins. *Res. Microbiol.* **2011**, *162* (4), 382–385. <https://doi.org/10.1016/j.resmic.2011.02.003>.
- (260) Doyle, M.; Fookes, M.; Ivens, A.; Mangan, M. W.; Wain, J.; Dorman, C. J. An H-NS-like Stealth Protein Aids Horizontal DNA Transmission in Bacteria. **2011**, *251* (2007), 10–12. <https://doi.org/10.1126/science.1137550>.
- (261) Deighan, P.; Beloin, C.; Dorman, C. J. Three-Way Interactions among the Sfh, StpA and H-NS Nucleoid-Structuring Proteins of *Shigella Flexneri* 2a Strain 2457T. *Mol. Microbiol.* **2003**, *48* (5), 1401–1416. <https://doi.org/10.1046/j.1365-2958.2003.03515.x>.
- (262) Madrid, C.; García, J.; Pons, M.; Juárez, A. Molecular Evolution of the H-NS Protein: Interaction with Hha-like Proteins Is Restricted to Enterobacteriaceae. *J. Bacteriol.* **2007**, *189* (1), 265–268. <https://doi.org/10.1128/JB.01124-06>.
- (263) Nieto, J. M.; Carmona, M.; Bolland, S.; Jubete, Y.; Cruz, F.; Juárez, A. The Hha Gene Modulates Haemolysin Expression in *Escherichia coli*. *Mol. Microbiol.* **1991**, *5* (5), 1285–1293. <https://doi.org/10.1111/j.1365-2958.1991.tb01902.x>.
- (264) Cornells, G. R.; Sluiters, C.; Delor, I.; Geib, D.; Kaniga, K.; Rouvroit, C. L.; Sory, M.-P.; Vanooteghem, J.-C.; Michiels, T. YmoA, a *Yersinia Enterocolitica* Chromosomal Gene Modulating the Expression of Virulence Functions. *Mol. Microbiol.* **1991**, *5* (5), 1023–1034. <https://doi.org/10.1111/j.1365-2958.1991.tb01875.x>.
- (265) Carmona, M.; Balsalobre, C.; Muñoa, F.; Mouriño, M.; Jubete, Y.; Cruz, F.; Juárez, A. *Escherichia coli* Hha Mutants, DNA Supercoiling and Expression of the Haemolysin Genes from the Recombinant Plasmid PANN202-312. *Mol. Microbiol.* **1993**, *9* (5), 1011–1018. <https://doi.org/10.1111/j.1365-2958.1993.tb01230.x>.
- (266) Mikulskis, A. V.; Cornetis, G. R. A New Class of Proteins Regulating Gene

- Expression in Enterobacteria. *Mol. Microbiol.* **1994**, *11* (1), 77–86. <https://doi.org/10.1111/j.1365-2958.1994.tb00291.x>.
- (267) Madrid, C.; Nieto, J. M.; Juárez, A. Role of the Hha/YmoA Family of Proteins in the Thermoregulation of the Expression of Virulence Factors. *Int. J. Med. Microbiol.* **2001**, *291* (6–7), 425–432. <https://doi.org/10.1078/1438-4221-00149>.
- (268) Nieto, J. M.; Madrid, C.; Prenafeta, A.; Miquelay, E.; Balsalobre, C.; Carrascal, M.; Juárez, A. Expression of the Hemolysin Operon in *Escherichia coli* Is Modulated by a Nucleoid-Protein Complex That Includes the Proteins Hha and H-NS. *Mol. Gen. Genet.* **2000**, *263* (2), 349–358. <https://doi.org/10.1007/s004380051178>.
- (269) Aznar, S.; Paytubi, S.; Juárez, A. The Hha Protein Facilitates Incorporation of Horizontally Acquired DNA in Enteric Bacteria. *Microbiol. (United Kingdom)* **2013**, *159* (PART3), 545–554. <https://doi.org/10.1099/mic.0.062448-0>.
- (270) Nieto, J. M.; Madrid, C.; Miquelay, E.; Parra, J. L.; Rodríguez, S.; Juárez, A. Evidence for Direct Protein-Protein Interaction between Members of the Enterobacterial Hha/YmoA and H-NS Families of Proteins. *J. Bacteriol.* **2002**, *184* (3), 629–635. <https://doi.org/10.1128/JB.184.3.629-635.2002>.
- (271) Paytubi, S.; García, J.; Juárez, A. Bacterial Hha-like Proteins Facilitate Incorporation of Horizontally Transferred DNA. *Cent. Eur. J. Biol.* **2011**, *6* (6), 879–886. <https://doi.org/10.2478/s11535-011-0071-3>.
- (272) Ali, S. S.; Whitney, J. C.; Stevenson, J.; Robinson, H.; Howell, P. L.; Navarre, W. W. Structural Insights into the Regulation of Foreign Genes in Salmonella by the Hha/H-NS Complex. *J. Biol. Chem.* **2013**, *288* (19), 13356–13369. <https://doi.org/10.1074/jbc.M113.455378>.
- (273) Madrid, C.; Balsalobre, C.; García, J.; Juárez, A. The Novel Hha/YmoA Family of Nucleoid-Associated Proteins: Use of Structural Mimicry to Modulate the Activity of the H-NS Family of Proteins. *Mol. Microbiol.* **2007**, *63* (1), 7–14. <https://doi.org/10.1111/j.1365-2958.2006.05497.x>.
- (274) Solórzano, C.; Srikumar, S.; Canals, R.; Juárez, A.; Paytubi, S.; Madrid, C. Hha Has a Defined Regulatory Role That Is Not Dependent upon H-NS or StpA. *Front. Microbiol.* **2015**, *6* (JUL), 1–14. <https://doi.org/10.3389/fmicb.2015.00773>.
- (275) F, T.; Wilson, R. L.; Boddicker, J. D. Hha Is a Negative Modulator of Transcription Of. *Microbiology* **2001**, *183* (22), 6620–6629. <https://doi.org/10.1128/JB.183.22.6620>.

Bibliografía

- (276) Olekhovich, I. N.; Kadner, R. J. Crucial Roles of Both Flanking Sequences in Silencing of the HilA Promoter in *Salmonella* Enterica. *J. Mol. Biol.* **2006**, *357* (2), 373–386. <https://doi.org/10.1016/j.jmb.2006.01.007>.
- (277) Takeda, T.; Yun, C.-S.; Shintani, M.; Yamane, H.; Nojiri, H. Distribution of Genes Encoding Nucleoid-Associated Protein Homologs in Plasmids. *Int. J. Evol. Biol.* **2011**, *2011*, 1–30. <https://doi.org/10.4061/2011/685015>.
- (278) Kopotsa, K.; Osei Sekyere, J.; Mbelle, N. M. Plasmid Evolution in Carbapenemase-Producing Enterobacteriaceae: A Review. *Ann. N. Y. Acad. Sci.* **2019**, *1457* (1), 61–91. <https://doi.org/10.1111/nyas.14223>.
- (279) Rozwandowicz, M.; Brouwer, M. S. M.; Fischer, J.; Wagenaar, J. A.; Gonzalez-Zorn, B.; Guerra, B.; Mevius, D. J.; Hordijk, J. Plasmids Carrying Antimicrobial Resistance Genes in Enterobacteriaceae. *J. Antimicrob. Chemother.* **2018**, *73* (5), 1121–1137. <https://doi.org/10.1093/jac/dkx488>.
- (280) Shintani, M.; Suzuki-Minakuchi, C.; Nojiri, H. Nucleoid-Associated Proteins Encoded on Plasmids: Occurrence and Mode of Function. *Plasmid* **2015**, *80*, 32–44. <https://doi.org/10.1016/j.plasmid.2015.04.008>.
- (281) Sonnenfield, J. M.; Burns, C. M.; Higgins, C. F.; Hinton, J. C. D. The Nucleoid-Associated Protein StpA Binds Curved DNA, Has a Greater DNA-Binding Affinity than H-NS and Is Present in Significant Levels in Hns Mutants. *Biochimie* **2001**, *83* (2), 243–249. [https://doi.org/10.1016/S0300-9084\(01\)01232-9](https://doi.org/10.1016/S0300-9084(01)01232-9).
- (282) Qin, X.; Hu, F.; Wu, S.; Ye, X.; Zhu, D.; Zhang, Y.; Wang, M. Comparison of Adhesin Genes and Antimicrobial Susceptibilities between Uropathogenic and Intestinal Commensal *Escherichia coli* Strains. *PLoS One* **2013**, *8* (4), 1–7. <https://doi.org/10.1371/journal.pone.0061169>.
- (283) Andersson, D. I.; Hughes, D. Gene Amplification and Adaptive Evolution in Bacteria. *Annu. Rev. Genet.* **2009**, *43* (1), 167–195. <https://doi.org/10.1146/annurev-genet-102108-134805>.
- (284) Serres, M. H.; Kerr, A. R. W.; McCormack, T. J.; Riley, M. Evolution by Leaps: Gene Duplication in Bacteria. *Biol. Direct* **2009**, *4*, 1–17. <https://doi.org/10.1186/1745-6150-4-46>.
- (285) Zhang, J. Evolution by Gene Duplication: An Update. *Trends Ecol. Evol.* **2003**, *18* (6), 292–298. [https://doi.org/10.1016/S0169-5347\(03\)00033-8](https://doi.org/10.1016/S0169-5347(03)00033-8).
- (286) He, X.; Zhang, J. Gene Complexity and Gene Duplicability. *Curr. Biol.* **2005**, *15* (11), 1016–1021. <https://doi.org/10.1016/j.cub.2005.04.035>.

- (287) Conant, G. C.; Wolfe, K. H. Turning a Hobby into a Job: How Duplicated Genes Find New Functions. *Nat. Rev. Genet.* **2008**, *9* (12), 938–950. <https://doi.org/10.1038/nrg2482>.
- (288) Voordeckers, K.; Verstrepen, K. J. Experimental Evolution of the Model Eukaryote *Saccharomyces Cerevisiae* Yields Insight into the Molecular Mechanisms Underlying Adaptation. *Curr. Opin. Microbiol.* **2015**, *28* (July), 1–9. <https://doi.org/10.1016/j.mib.2015.06.018>.
- (289) Force, A.; Lynch, M.; Pickett, F. B.; Amores, A.; Yan, Y. L.; Postlethwait, J. Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* **1999**, *151* (4), 1531–1545.
- (290) AL, H. The Evolution of Functionally Novel Proteins after Gene Duplication. *Proc. R. Soc. London. Ser. B Biol. Sci.* **1994**, *256* (1346), 119–124. <https://doi.org/10.1098/rspb.1994.0058>.
- (291) Elliott, K. T.; Cuff, L. E.; Neidle, E. L. Copy Number Change: Evolving Views on Gene Amplification. *Future Microbiol.* **2013**, *8* (7), 887–899. <https://doi.org/10.2217/fmb.13.53>.
- (292) Kondrashov, F. A.; Rogozin, I. B.; Wolf, Y. I.; Koonin, E. V. Selection in the Evolution of Gene Duplications. *Genome Biol.* **2002**, *3* (2), 1–9. <https://doi.org/10.1186/gb-2002-3-2-research0008>.
- (293) Kondrashov, F. A. Gene Duplication as a Mechanism of Genomic Adaptation to a Changing Environment. *Proc. R. Soc. B Biol. Sci.* **2012**, *279* (1749), 5048–5057. <https://doi.org/10.1098/rspb.2012.1108>.
- (294) Hooper, S. D.; Berg, O. G. Duplication Is More Common among Laterally Transferred Genes than among Indigenous Genes. *Genome Biol.* **2003**, *4* (8), R48.
- (295) Gevers, D.; Vandepoele, K.; Simillion, C.; Van de Peer, Y. Gene Duplication and Biased Functional Retention of Paralogs in Bacterial Genomes. *Trends Microbiol.* **2004**, *12* (4), 148–154. <https://doi.org/10.1016/j.tim.2004.02.007>.
- (296) Song, S.; Berg, O. G.; Roth, J. R.; Andersson, D. I. Contribution of Gene Amplification to Evolution of Increased Antibiotic Resistance in *Salmonella Typhimurium*. *Genetics* **2009**, *182* (4), 1183–1195. <https://doi.org/10.1534/genetics.109.103028>.
- (297) Sandegren, L.; Andersson, D. I. Bacterial Gene Amplification: Implications for the Evolution of Antibiotic Resistance. *Nat. Rev. Microbiol.* **2009**, *7* (8), 578–588. <https://doi.org/10.1038/nrmicro2174>.
- (298) Reams, A. B.; Kofoid, E.; Savageau, M.; Roth, J. R. Duplication Frequency in a Population of *Salmonella Enterica* Rapidly Approaches Steady State

- with or without Recombination. *Genetics* **2010**, *184* (4), 1077–1094. <https://doi.org/10.1534/genetics.109.111963>.
- (299) Kugelberg, E.; Kofoid, E.; Reams, A. B.; Andersson, D. I.; Roth, J. R. Multiple Pathways of Selected Gene Amplification during Adaptive Mutation. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (46), 17319–17324. <https://doi.org/10.1073/pnas.0608309103>.
- (300) Sun, S.; Ke, R.; Hughes, D.; Nilsson, M.; Andersson, D. I. Genome-Wide Detection of Spontaneous Chromosomal Rearrangements in Bacteria. *PLoS One* **2012**, *7* (8). <https://doi.org/10.1371/journal.pone.0042639>.
- (301) Arun, P. V. P. S.; Miryala, S. K.; Chattopadhyay, S.; Thiyyagura, K.; Bawa, P.; Bhattacharjee, M.; Yellaboina, S. Identification and Functional Analysis of Essential, Conserved, Housekeeping and Duplicated Genes. *FEBS Lett.* **2016**, *590* (10), 1428–1437. <https://doi.org/10.1002/1873-3468.12192>.
- (302) Wapinski, I.; Pfeffer, A.; Friedman, N.; Regev, A. Natural History and Evolutionary Principles of Gene Duplication in Fungi. *Nature* **2007**, *449* (7158), 54–61. <https://doi.org/10.1038/nature06107>.
- (303) van der Woude, M. W.; Henderson, I. R. Regulation and Function of Ag43 (Flu). *Annu. Rev. Microbiol.* **2008**, *62* (1), 153–169. <https://doi.org/10.1146/annurev.micro.62.081307.162938>.
- (304) Chauhan, A.; Sakamoto, C.; Ghigo, J. M.; Beloin, C. Did I Pick the Right Colony? Pitfalls in the Study of Regulation of the Phase Variable Antigen 43 Adhesin. *PLoS One* **2013**, *8* (9). <https://doi.org/10.1371/journal.pone.0073568>.
- (305) Ulett, G. C.; Valle, J.; Beloin, C.; Sherlock, O.; Ghigo, J. M.; Schembri, M. A. Functional Analysis of Antigen 43 in Uropathogenic *Escherichia coli* Reveals a Role in Long-Term Persistence in the Urinary Tract. *Infect. Immun.* **2007**, *75* (7), 3233–3244. <https://doi.org/10.1128/IAI.01952-06>.
- (306) Korea, C. G.; Badouraly, R.; Prevost, M. C.; Ghigo, J. M.; Beloin, C. *Escherichia coli* K-12 Possesses Multiple Cryptic but Functional Chaperone-Usher Fimbriae with Distinct Surface Specificities. *Environ. Microbiol.* **2010**, *12* (7), 1957–1977. <https://doi.org/10.1111/j.1462-2920.2010.02202.x>.
- (307) Roux, A.; Beloin, C.; Ghigo, J. M. Combined Inactivation and Expression Strategy to Study Gene Function under Physiological Conditions: Application to Identification of New *Escherichia coli* Adhesins. *J. Bacteriol.* **2005**, *187* (3), 1001–1013. <https://doi.org/10.1128/JB.187.3.1001-1013.2005>.
- (308) Nuccio, S.-P.; Baumler, A. J. Evolution of the Chaperone/Usher Assembly

- Pathway: Fimbrial Classification Goes Greek. *Microbiol. Mol. Biol. Rev.* **2007**, *71* (4), 551–575. <https://doi.org/10.1128/mnbr.00014-07>.
- (309) Wurpel, D. J.; Beatson, S. A.; Totsika, M.; Petty, N. K.; Schembri, M. A. Chaperone-Usher Fimbriae of *Escherichia coli*. *PLoS One* **2013**, *8* (1). <https://doi.org/10.1371/journal.pone.0052835>.
- (310) Wells, T. J.; Totsika, M.; Schembri, M. A. Autotransporters of *Escherichia coli*: A Sequence-Based Characterization. *Microbiology* **2010**, *156* (8), 2459–2469. <https://doi.org/10.1099/mic.0.039024-0>.
- (311) Leo, J. C.; Grin, I.; Linke, D. Type V Secretion: Mechanism(S) of Autotransport through the Bacterial Outer Membrane. *Philos. Trans. R. Soc. B Biol. Sci.* **2012**, *367* (1592), 1088–1101. <https://doi.org/10.1098/rstb.2011.0208>.
- (312) Klemm, P.; Vejborg, R. M.; Sherlock, O. Self-Associating Autotransporters, SAATs: Functional and Structural Similarities. *Int. J. Med. Microbiol.* **2006**, *296* (4–5), 187–195. <https://doi.org/10.1016/j.ijmm.2005.10.002>.
- (313) Sherlock, O.; Schembri, M. a; Reisner, A.; Klemm, P. Novel Roles for the AIDA Adhesin from Diarrheagenic. *J. Bacteriol.* **2004**, *186* (23), 8058–8065. <https://doi.org/10.1128/JB.186.23.8058>.
- (314) Kjærgaard, K.; Schembri, M. A.; Ramos, C.; Molin, S.; Klemm, P. Antigen 43 Facilitates Formation of Multispecies Biofilms. *Environ. Microbiol.* **2000**, *2* (6), 695–702. <https://doi.org/10.1046/j.1462-2920.2000.00152.x>.
- (315) Diderichsen, B. Flu, A Metastable Gene Controlling Surface Properties of *Escherichia coli*. *J. Bacteriol.* **1980**, *141* (2), 858–867.
- (316) Restieri, C.; Garriss, G.; Locas, M. C.; Dozois, C. M. Autotransporter-Encoding Sequences Are Phylogenetically Distributed among *Escherichia coli* Clinical Isolates and Reference Strains. *Appl. Environ. Microbiol.* **2007**, *73* (5), 1553–1562. <https://doi.org/10.1128/AEM.01542-06>.
- (317) Dautin, N.; Bernstein, H. D. Protein Secretion in Gram-Negative Bacteria via the Autotransporter Pathway. *Annu. Rev. Microbiol.* **2007**, *61* (1), 89–112. <https://doi.org/10.1146/annurev.micro.61.080706.093233>.
- (318) Selkrig, J.; Mosbahi, K.; Webb, C. T.; Belousoff, M. J.; Perry, A. J.; Wells, T. J.; Morris, F.; Leyton, D. L.; Totsika, M.; Phan, M. D.; Celik, N.; Kelly, M.; Oates, C.; Hartland, E. L.; Robins-Browne, R. M.; Ramarathinam, S. H.; Purcell, A. W.; Schembri, M. A.; Strugnell, R. A.; Henderson, I. R.; Walker, D.; Lithgow, T. Discovery of an Archetypal Protein Transport System in Bacterial Outer Membranes. *Nat. Struct. Mol. Biol.* **2012**, *19* (5), 506–510. <https://doi.org/10.1038/nsmb.2261>.
- (319) Roche, A. J.; McFadden, J. P.; Owen, P. Antigen 43, the Major Phase-

- Variable Protein of the *Escherichia coli* Outer Membrane, Can Exist as a Family of Proteins Encoded by Multiple Alleles. *Microbiology* **2001**, *147* (1), 161–169. <https://doi.org/10.1099/00221287-147-1-161>.
- (320) Klemm, P.; Hjerrild, L.; Gjermansen, M.; Schembri, M. A. Structure-Function Analysis of the Self-Recognizing Antigen 43 Autotransporter Protein from *Escherichia coli*. *Mol. Microbiol.* **2004**, *51* (1), 283–296. <https://doi.org/10.1046/j.1365-2958.2003.03833.x>.
- (321) Ageorges, V.; Schiavone, M.; Jubelin, G.; Caccia, N.; Ruiz, P.; Chafsey, I.; Bailly, X.; Dague, E.; Leroy, S.; Paxman, J.; Heras, B.; Chaucheyras-Durand, F.; Rossiter, A. E.; Henderson, I. R.; Desvaux, M. Differential Homotypic and Heterotypic Interactions of Antigen 43 (Ag43) Variants in Autotransporter-Mediated Bacterial Autoaggregation. *Sci. Rep.* **2019**, *9* (1), 11100. <https://doi.org/10.1038/s41598-019-47608-4>.
- (322) Anderson, G. G.; Palermo, J. J.; Schilling, J. D.; Roth, R.; Heuser, J.; Hultgren, S. J. Intracellular Bacterial Biofilm-like Pods in Urinary Tract Infections. *Science* (80-.). **2003**, *301* (5629), 105–107. <https://doi.org/10.1126/science.1084550>.
- (323) Van Der Woude, M. W.; Bäumlér, A. J. Phase and Antigenic Variation in Bacteria. *Clin. Microbiol. Rev.* **2004**, *17* (3), 581–611. <https://doi.org/10.1128/CMR.17.3.581-611.2004>.
- (324) Henderson, I. R.; Meehan, M.; Owen, P. A Novel Regulatory Mechanism for a Novel Phase-Variable Outer Membrane Protein of *Escherichia coli*; 1997; pp 349–355. https://doi.org/10.1007/978-1-4899-1828-4_56.
- (325) Henderson, I. R.; Owen, P. The Major Phase-Variable Outer Membrane Protein of *Escherichia coli* Structurally Resembles the Immunoglobulin A1 Protease Class of Exported Protein and Is Regulated by a Novel Mechanism Involving Dam and OxyR. *J. Bacteriol.* **1999**, *181* (7), 2132–2141.
- (326) Waldron, D. E.; Owen, P.; Dorman, C. J. Competitive Interaction of the OxyR DNA-Binding Protein and the Dam Methylase at the Antigen 43 Gene Regulatory Region in *Escherichia coli*. *Mol. Microbiol.* **2002**, *44* (2), 509–520. <https://doi.org/10.1046/j.1365-2958.2002.02905.x>.
- (327) Haagmans, W.; Van Der Woude, M. Phase Variation of Ag43 in *Escherichia coli*: Dam-Dependent Methylation Abrogates OxyR Binding and OxyR-Mediated Repression of Transcription. *Mol. Microbiol.* **2000**, *35* (4), 877–887. <https://doi.org/10.1046/j.1365-2958.2000.01762.x>.
- (328) Moriel, D. G.; Heras, B.; Paxman, J. J.; Lo, A. W.; Tan, L.; Sullivan, M. J.; Dando, S. J.; Beatson, S. A.; Ulett, G. C.; Schembri, M. A. Molecular and Structural Characterization of a Novel *Escherichia coli* Interleukin Receptor Mimic Protein. *MBio* **2016**, *7* (2), 1–11.

<https://doi.org/10.1128/mBio.02046-15>.

- (329) Moriel, D. G.; Bertoldi, I.; Spagnuolo, A.; Marchi, S.; Rosini, R.; Nesta, B.; Pastorello, I.; Mariani Corea, V. A.; Torricelli, G.; Cartocci, E.; Savino, S.; Scarselli, M.; Dobrindt, U.; Hacker, J.; Tettelin, H.; Tallon, L. J.; Sullivan, S.; Wieler, L. H.; Ewers, C.; Pickard, D.; Dougan, G.; Fontana, M. R.; Rappuoli, R.; Pizza, M.; Serino, L. Identification of Protective and Broadly Conserved Vaccine Antigens from the Genome of Extraintestinal Pathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (20), 9072–9077. <https://doi.org/10.1073/pnas.0915077107>.
- (330) Moriel, D. G.; Scarselli, M.; Serino, L.; Mora, M.; Rappuoli, R.; Masignani, V. Genome-Based Vaccine Development: A Short Cut for the Future. *Hum. Vaccin.* **2008**, *4* (3), 184–188. <https://doi.org/10.4161/hv.4.3.6313>.
- (331) Akdis, M.; Burgler, S.; Cramer, R.; Eiwegger, T.; Fujita, H.; Gomez, E.; Klunker, S.; Meyer, N.; O'Mahony, L.; Palomares, O.; Rhyner, C.; Quaked, N.; Schaffartzik, A.; Van De Veen, W.; Zeller, S.; Zimmermann, M.; Akdis, C. A. Interleukins, from 1 to 37, and Interferon- γ : Receptors, Functions, and Roles in Diseases. *J. Allergy Clin. Immunol.* **2011**, *127* (3), 701–721.e70. <https://doi.org/10.1016/j.jaci.2010.11.050>.
- (332) Sullivan, M. J.; Petty, N. K.; Beatson, S. A. Easyfig: A Genome Comparison Visualizer. *Bioinformatics* **2011**, *27* (7), 1009–1010. <https://doi.org/10.1093/bioinformatics/btr039>.
- (333) Taylor, R. G.; Walker, D. C.; McInnes, R. R. *E. coli* Host Strains. *Nucleic Acids Res.* **1993**, *21* (7), 1677–1678.
- (334) Studier, F. W.; Moffatt, B. A. Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-Level Expression of Cloned Genes. *J. Mol. Biol.* **1986**, *189* (1), 113–130. [https://doi.org/10.1016/0022-2836\(86\)90385-2](https://doi.org/10.1016/0022-2836(86)90385-2).
- (335) Datsenko, K. A.; Wanner, B. L. One-Step Inactivation of Chromosomal Genes in *Escherichia coli* K-12 Using PCR Products. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (12), 6640–6645. <https://doi.org/10.1073/pnas.120163297>.
- (336) Cherepanov, P. P.; Wackernagel, W. Gene Disruption in *Escherichia coli*: TcR and KmR Cassettes with the Option of FLP-Catalyzed Excision of the Antibiotic-Resistance Determinant. *Gene* **1995**, *158* (1), 9–14. [https://doi.org/10.1016/0378-1119\(95\)00193-A](https://doi.org/10.1016/0378-1119(95)00193-A).
- (337) Uzzau, S.; Figueroa-Bossi, N.; Rubino, S.; Bossi, L. Epitope Tagging of Chromosomal Genes in *Salmonella*. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (26), 15264–15269. <https://doi.org/10.1073/pnas.261348198>.
- (338) Ellermeier, C. D.; Janakiraman, A.; Slauch, J. M. Construction of Targeted Single Copy Lac Fusions Using λ Red and FLP-Mediated Site-Specific

- Recombination in Bacteria. **2002**, *290*, 153–161.
- (339) De Lorenzo, V.; Herrero, M.; Jakubzik, U.; Timmis, K. N. Mini-Tn5 Transposon Derivatives for Insertion Mutagenesis, Promoter Probing, and Chromosomal Insertion of Cloned DNA in Gram-Negative Eubacteria. *J. Bacteriol.* **1990**, *172* (11), 6568–6572. <https://doi.org/10.1128/jb.172.11.6568-6572.1990>.
- (340) Hanahan, D.; Jessee, J.; Bloom, F. R. Plasmid Transformation of *Escherichia coli* and Other Bacteria. *Methods Enzymol.* **1991**, *204* (C), 63–113. [https://doi.org/10.1016/0076-6879\(91\)04006-A](https://doi.org/10.1016/0076-6879(91)04006-A).
- (341) Cohen, S. N.; Chang, A. C.; Hsu, L. Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of *Escherichia coli* by R-Factor DNA. *Proc. Natl. Acad. Sci. U. S. A.* **1972**, *69* (8), 2110–2114. <https://doi.org/10.1073/pnas.69.8.2110>.
- (342) Dower, W. J.; Miller, J. F.; Ragsdale, C. W. High Efficiency Transformation of *E. coli* by High Voltage Electroporation. *Nucleic Acids Res.* **1988**, *16* (13), 6127–6145. <https://doi.org/10.1093/nar/16.13.6127>.
- (343) Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H. Construction of *Escherichia coli* K-12 in-Frame, Single-Gene Knockout Mutants: The Keio Collection. *Mol. Syst. Biol.* **2006**, *2*. <https://doi.org/10.1038/msb4100050>.
- (344) Ellermeier, C. D.; Janakiraman, A.; Slauch, J. M. Construction of Targeted Single Copy Lac Fusions Using λ Red and FLP-Mediated Site-Specific Recombination in Bacteria. *Gene* **2002**, *290* (1–2), 153–161. [https://doi.org/10.1016/S0378-1119\(02\)00551-6](https://doi.org/10.1016/S0378-1119(02)00551-6).
- (345) Vogel, J.; Wagner, E. G. H. Approaches to Identify Novel Non-Messenger RNAs in Bacteria and to Investigate Their Biological Functions: RNA Mining. *Handb. RNA Biochem.* **2008**, 595–613. <https://doi.org/10.1002/9783527619504.ch36>.
- (346) Schägger, H.; von Jagow, G. Tricine-Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis for the Separation of Proteins in the Range from 1 to 100 KDa. *Anal. Biochem.* **1987**, *166* (2), 368–379. [https://doi.org/10.1016/0003-2697\(87\)90587-2](https://doi.org/10.1016/0003-2697(87)90587-2).
- (347) Miller, J. A Short Course in Bacterial Genetics – A Laboratory Manual and Handbook for *Escherichia coli* and Related Bacteria. *J. Basic Microbiol.* **1993**, *33* (4), 278–278. <https://doi.org/10.1002/jobm.3620330412>.
- (348) Sheikh, J.; Hicks, S.; Dall’Agnol, M.; Phillips, A. D.; Nataro, J. P. Roles for Fis and YafK in Biofilm Formation by Enteroaggregative *Escherichia coli*. *Mol. Microbiol.* **2001**, *41* (5), 983–997. <https://doi.org/10.1046/j.1365->

2958.2001.02512.x.

- (349) Moya-Andérico, L.; Admella, J.; Fernandes, R.; Torrents, E. Monitoring Gene Expression during a *Galleria mellonella* Bacterial Infection. *Microorganisms* **2020**, *8* (11), 1–14. <https://doi.org/10.3390/microorganisms8111798>.
- (350) Vallenet, D.; Engelen, S.; Mornico, D.; Cruveiller, S.; Fleury, L.; Lajus, A.; Rouy, Z.; Roche, D.; Salvignol, G.; Scarpelli, C.; Médigue, C. MicroScope: A Platform for Microbial Genome Annotation and Comparative Genomics. *Database* **2009**, *2009*, 1–12. <https://doi.org/10.1093/database/bap021>.
- (351) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- (352) Saitou, N.; Nei, M. The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* **1987**, *4* (4), 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>.
- (353) Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33* (7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>.
- (354) Philipson, C. W.; Bassaganya-Riera, J.; Hontecillas, R. Animal Models of Enterotoxigenic *Escherichia coli* Infection. *Gut Microbes* **2013**, *4* (4), 281–291. <https://doi.org/10.4161/gmic.24826>.
- (355) Insua, J. L.; Llobet, E.; Moranta, D.; Pérez-Gutiérrez, C.; Tomás, A.; Garmendia, J.; Bengoechea, J. A. Modeling *Klebsiella pneumoniae* Pathogenesis by Infection of the Wax Moth *Galleria mellonella*. *Infect. Immun.* **2013**, *81* (10), 3552–3565. <https://doi.org/10.1128/IAI.00391-13>.
- (356) Mukherjee, K.; Altincicek, B.; Hain, T.; Domann, E.; Vilcinskas, A.; Chakraborty, T. *Galleria mellonella* as a Model System for Studying *Listeria* Pathogenesis. *Appl. Environ. Microbiol.* **2010**, *76* (1), 310–317. <https://doi.org/10.1128/AEM.01301-09>.
- (357) Miyata, S.; Casey, M.; Frank, D. W.; Ausubel, F. M.; Drenkard, E. Use of the *Galleria mellonella* Caterpillar as a Model Host to Study the Role of the Type III Secretion System in *Pseudomonas aeruginosa* Pathogenesis. *Infect. Immun.* **2003**, *71* (5), 2404–2413. <https://doi.org/10.1128/IAI.71.5.2404-2413.2003>.
- (358) Jønsson, R.; Struve, C.; Jenssen, H.; Krogfelt, K. A. The Wax Moth *Galleria mellonella* as a Novel Model System to Study Enterotoxigenic *Escherichia coli* Pathogenesis. *Virulence* **2017**, *8* (8), 1894–1899. <https://doi.org/10.1080/21505594.2016.1256537>.

- (359) Jander, G.; Rahme, L. G.; Ausubel, F. M. Positive Correlation between Virulence of *Pseudomonas aeruginosa* Mutants in Mice and Insects. *J. Bacteriol.* **2000**, *182* (13), 3843–3845. <https://doi.org/10.1128/JB.182.13.3843-3845.2000>.
- (360) Bai, X.; Zhang, J.; Ambikan, A.; Jernberg, C.; Ehricht, R.; Scheutz, F.; Xiong, Y.; Matussek, A. Molecular Characterization and Comparative Genomics of Clinical Hybrid Shiga Toxin-Producing and Enterotoxigenic *Escherichia coli* (STEC/ETEC) Strains in Sweden. *Sci. Rep.* **2019**, *9* (1), 5619. <https://doi.org/10.1038/s41598-019-42122-z>.
- (361) Nyholm, O.; Halkilahti, J.; Wiklund, G.; Okeke, U.; Paulin, L.; Auvinen, P.; Haukka, K.; Siitonen, A. Comparative Genomics and Characterization of Hybrid Shigatoxigenic and Enterotoxigenic *Escherichia coli* (STEC/ETEC) Strains. *PLoS One* **2015**, *10* (8), 1–17. <https://doi.org/10.1371/journal.pone.0135936>.
- (362) Zhang, W.; Sack, D. A. Current Progress in Developing Subunit Vaccines against Enterotoxigenic *Escherichia coli*-Associated Diarrhea. *Clin. Vaccine Immunol.* **2015**, *22* (9), 983–991. <https://doi.org/10.1128/CVI.00224-15>.
- (363) Albenberg, L.; Esipova, T. V.; Judge, C. P.; Bittinger, K.; Chen, J.; Laughlin, A.; Grunberg, S.; Baldassano, R. N.; Lewis, J. D.; Li, H.; Thom, S. R.; Bushman, F. D.; Vinogradov, S. A.; Wu, G. D. Correlation between Intraluminal Oxygen Gradient and Radial Partitioning of Intestinal Microbiota. *Gastroenterology* **2014**, *147* (5), 1055–1063.e8. <https://doi.org/10.1053/j.gastro.2014.07.020>.
- (364) Zhang, X.; Kim, Y.; Wood, T. K.; Garcı, R. Protein Translation and Cell Death : The Role of Rare TRNAs in Biofilm Formation and in Activating Dormant Phage Killer Genes. **2008**, *3* (6). <https://doi.org/https://doi.org/10.1371/journal.pone.0002394>.
- (365) Hong, S. H.; Wang, X.; Wood, T. K. Controlling Biofilm Formation , Prophage Excision and Cell Death by Rewiring Global Regulator H-NS of *Escherichia coli*. **2010**, *3*, 344–356. <https://doi.org/10.1111/j.1751-7915.2010.00164.x>.
- (366) Keane, O. M.; Toft, C.; Carretero-Paulet, L.; Jones, G. W.; Fares, M. A. Preservation of Genetic and Regulatory Robustness in Ancient Gene Duplicates of *Saccharomyces cerevisiae*. *Genome Res.* **2014**, *24* (11), 1830–1841. <https://doi.org/10.1101/gr.176792.114>.
- (367) Brown, J. M.; Shaw, K. J. A Novel Family of *Escherichia coli* Toxin-Antitoxin Gene Pairs. *J. Bacteriol.* **2003**, *185* (22), 6600–6608. <https://doi.org/10.1128/JB.185.22.6600-6608.2003>.

- (368) Masuda, H.; Tan, Q.; Awano, N.; Wu, K.; Inouye, M. YeeU Enhances the Bundling of Cytoskeletal Polymers of MreB and FtsZ , Antagonizing the CbtA (YeeV) Toxicity in *Escherichia coli*. **2012**, *84* (May), 979–989. <https://doi.org/10.1111/j.1365-2958.2012.08068.x>.

8. Anexos

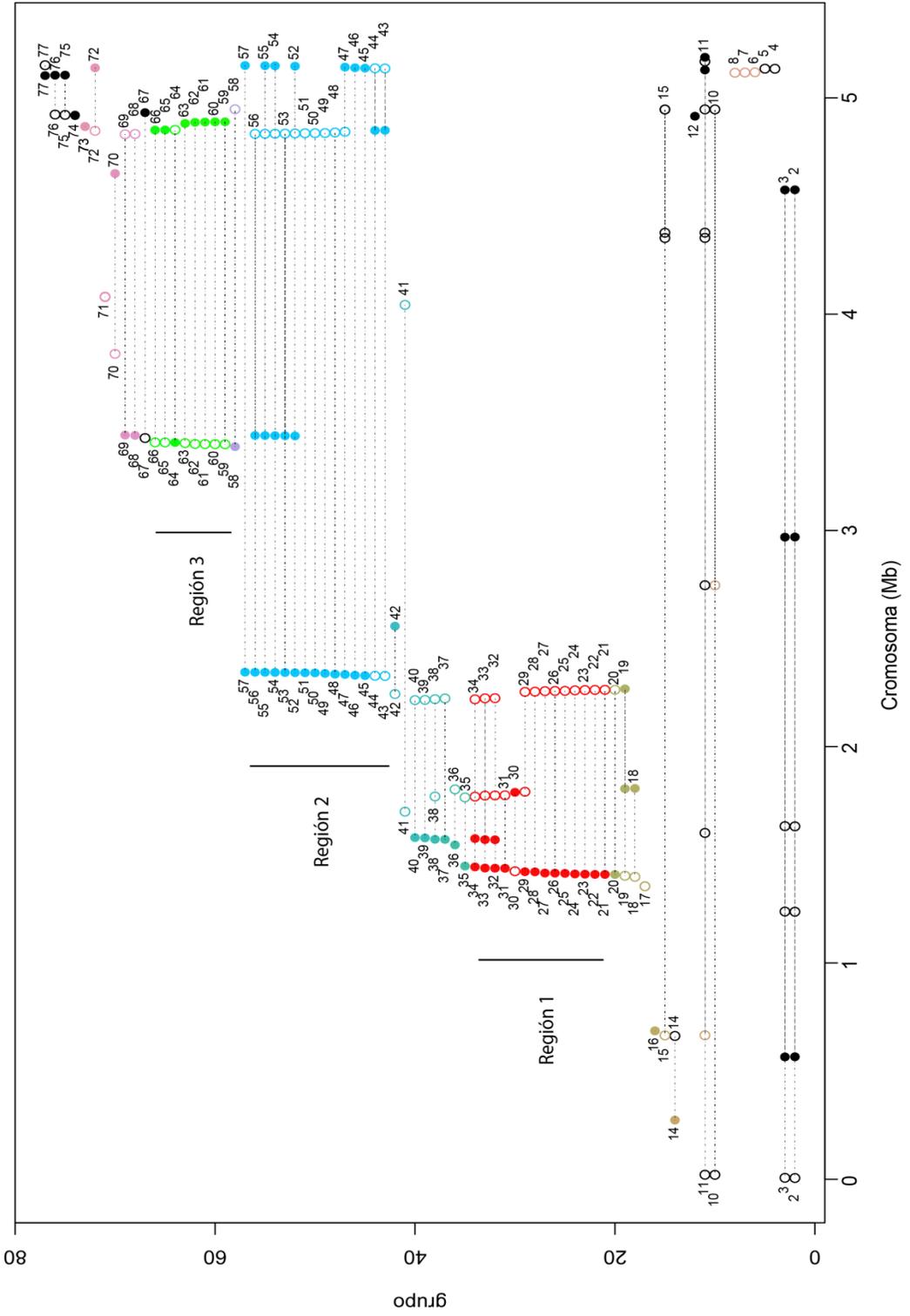


Fig. A1. Genes duplicados en el genoma de la cepa EAEC 042. El eje de las X representa el mapa lineal del cromosoma. Cada grupo de círculos unidos con una línea horizontal discontinua corresponde a un gen duplicado o amplificado. Los círculos indican la posición que ocupa cada duplicado en el genoma. Los círculos enteros corresponden a los genes que se encuentran en la cadena positiva, mientras que los círculos vacíos corresponden a los de la cadena negativa. Los números corresponden a los grupos de genes duplicados, numerados en orden empezando desde el origen del mapa cromosómico. En negro se muestran las transposasas. Los colores (rojo, azul y verde) y las barras verticales definen las tres regiones principales que contienen genes duplicados. Análisis realizado mediante BLASTn.

Tabla A1. Localizador de cada una de las copias de los genes duplicados o amplificados (Locus tag) en la cepa CFT073 y descripción de su función. A) Genes de la región 1: 5-25. B) Genes de la región 2: 32-44. C) Genes de la región 3: 45-55. D) Genes de la región 4: 56-65. E) Genes de la región 5: 73-86. F) Genes de la región 6: 88-94.

A	Grupo	Locus tag 1	Locus tag 2	Locus tag 3	Locus tag 4	Locus tag 5	Locus tag 6	Descripción
Región 1: 5-25	5	C_RS01210	C_RS17180	C_RS17295				IS66 family insertion sequence hypothetical protein
	6	C_RS01265	C_RS06025	C_RS07130	C_RS11895	C_RS17135	C_RS24345	ATPase AAA
	7	C_RS01270	C_RS06050	C_RS17440	C_RS21565	C_RS24365		Toxin
	8	C_RS01275	C_RS12005	C_RS17435	C_RS21560	C_RS24370		type IV toxin_antitoxin system YeeU family antitoxin
	9	C_RS01280	C_RS06035	C_RS28815				hypothetical protein
	10	C_RS01285	C_RS06030	C_RS12000	C_RS17425	C_RS21555	C_RS24375	DUF987 domain containing protein
	11	C_RS01290	C_RS06015	C_RS11995	C_RS17420	C_RS21550	C_RS24380	hypothetical protein
	12	C_RS01295	C_RS06010	C_RS11990	C_RS17415	C_RS24385		antirestriction protein
	13	C_RS01305	C_RS06005	C_RS11985	C_RS17410	C_RS21540	C_RS24390	DUF945 domain containing protein
	14	C_RS01335	C_RS11970					phospholipase
	15	C_RS01340	C_RS11965					chemotaxis protein
	16	C_RS01380	C_RS11950					TonB dependent siderophore receptor
	17	C_RS01385	C_RS11945					ABC transporter substrate binding protein
	18	C_RS01390	C_RS11940					iron ABC transporter permease
	19	C_RS01395	C_RS11935					ABC transporter ATP binding protein
	20	C_RS01400	C_RS11930					hypothetical protein
	21	C_RS26150	C_RS27170					hypothetical protein
	22	C_RS01440	C_RS11850	C_RS17315				hypothetical protein
	23	C_RS01505	C_RS26585					hypothetical protein
	24	C_RS01585	C_RS05670					hypothetical protein
	25	C_RS01595	C_RS05680					hypothetical protein

B	Grupo	Locus tag 1	Locus tag 2	Locus tag 3	Locus tag 4	Locus tag 5	Descripción
Región 2: 32-44	32	C_RS26480	C_RS27670				hypothetical protein
	33	C_RS05805	C_RS17055	C_RS24530			Major pilus subunit operon regulatory protein
	34	C_RS05855	C_RS16995				transcriptional regulator
	35	C_RS05950	C_RS21295				transposase
	36	C_RS05980	C_RS17355				50S ribosome_binding GTPase
	37	C_RS05985	C_RS17360				autotransporter domain_containing protein
	38	C_RS05990	C_RS17395				hypothetical protein
	39	C_RS05995	C_RS17400				hypothetical protein
	40	C_RS06020	C_RS07135	C_RS11890	C_RS17130	C_RS24340	IS21 family transposase ISEc10
	41	C_RS06055	C_RS17445				hypothetical protein
	42	C_RS06060	C_RS17450	C_RS24355			DUF957 domain_containing protein
	43	C_RS06070	C_RS17455	C_RS21580			DUF4942 domain_containing protein
	44	C_RS06735	C_RS15130				lysozyme

C	Grupo	Locus tag 1	Locus tag 2	Locus tag 3	Descripción
Región 3: 45-55	45	C_RS06780	C_RS07355		phage terminase large subunit family protein
	46	C_RS06785	C_RS07360		head_tail joining protein
	47	C_RS06790	C_RS07365		phage portal protein
	48	C_RS06795	C_RS07370		S49 family peptidase
	49	C_RS06800	C_RS07375		head decoration protein
	50	C_RS06805	C_RS07380		minor capsid protein E
	51	C_RS06815	C_RS07390		tail attachment protein
	52	C_RS06845	C_RS07420	C_RS15020	phage tail tape measure protein
	53	C_RS06860	C_RS07435	C_RS15005	phage tail protein
	54	C_RS06880	C_RS14990		Ail/Lom family protein
	55	C_RS06895	C_RS07460	C_RS14965	hypothetical protein

D	Grupo	Locus tag 1	Locus tag 2	Descripción
Región 4: 56-65	56	C_RS07310	C_RS11125	phosphoporin PhoE
	57	C_RS07315	C_RS15140	lysis protein S
	58	C_RS07395	C_RS15045	tail protein
	59	C_RS07400	C_RS15040	tail protein
	60	C_RS07410	C_RS15030	phage minor tail protein G
	61	C_RS07415	C_RS15025	phage tail assembly protein T
	62	C_RS07425	C_RS15015	tail protein
	63	C_RS07430	C_RS15010	phage minor tail protein L
	64	C_RS07440	C_RS15000	tail assembly protein
	65	C_RS07445	C_RS14995	host specificity protein J

E	Grupo	Locus tag 1	Locus tag 2	Descripción
Región 5: 73-86	73	C_RS16485	C_RS24305	lysine_tRNA ligase
	74	C_RS16885	C_RS24645	DUF4102 domain_containing protein
	75	C_RS16890	C_RS21210	hypothetical protein
	76	C_RS27505	C_RS27835	hypothetical protein
	77	C_RS16905	C_RS16980	IS66 family insertion sequence hypothetical protein
	78	C_RS17000	C_RS24480	fimbrial protein
	79	C_RS17005	C_RS24485	fimbrial protein
	80	C_RS17010	C_RS24490	fimbrial protein
	81	C_RS17015	C_RS24495	type 1 fimbrial protein
	82	C_RS17020	C_RS24500	protein papJ
	83	C_RS17025	C_RS24505	molecular chaperone
	84	C_RS17030	C_RS24510	PapC/FimD family outer membrane usher protein
	85	C_RS17035	C_RS24515	type 1 fimbrial protein
	86	C_RS17045	C_RS24525	major pilus subunit operon regulatory protein PapB

F	Grupo	Locus tag 1	Locus tag 2	Descripción
Region 6: 88-94	88	C_RS21495	C_RS24540	AlpA family transcriptional regulator
	89	C_RS21500	C_RS24545	hypothetical protein
	90	C_RS21505	C_RS24445	inovirus Gp2 family protein
	91	C_RS21515	C_RS24410	GTPase
	92	C_RS21525	C_RS24400	WYL domain_containing protein
	93	C_RS21530	C_RS24395	hypothetical protein
	94	C_RS21535	C_RS28070	DUF905 domain_containing protein

Tabla A2. Localizador de cada una de las copias de los genes duplicados o amplificados (Locus tag) en la cepa O145:H28 RM13516 y descripción de su función. A) Genes de la región 1: 13-40. B) Genes de la región 2: 41-65. C) Genes de la región 3: 67-79. D) Genes de la región 4: 80-103. E) Genes de la región 5: 129-137. F) Genes de la región 6: 139-147.

A	Grupo	Locus tag 1	Locus tag 2	Locus tag 3	Locus tag 4	Descripción
Región 1: 13-40	13	ECRM13516_RS03805	ECRM13516_RS16985			Replication protein P
	14	ECRM13516_RS03810	ECRM13516_RS16980			protein ren
	15	ECRM13516_RS03815	ECRM13516_RS16975			recombination protein NinB
	16	ECRM13516_RS03820	ECRM13516_RS16970			Phage N-6-adenine methyltransferase
	17	ECRM13516_RS03825	ECRM13516_RS30035			NinE family protein
	18	ECRM13516_RS03835	ECRM13516_RS16955			Endonuclease
	19	ECRM13516_RS03840	ECRM13516_RS16950			protein ninG
	20	ECRM13516_RS03845	ECRM13516_RS16945			serine/threonine protein phosphatase
	21	ECRM13516_RS03875	ECRM13516_RS16935			hypothetical protein
	22	ECRM13516_RS03885	ECRM13516_RS07050	ECRM13516_RS14160	ECRM13516_RS16925	DUF1737 domain_containing protein
	23	ECRM13516_RS03890	ECRM13516_RS07060	ECRM13516_RS14155	ECRM13516_RS16920	Holin
	24	ECRM13516_RS03895	ECRM13516_RS09815	ECRM13516_RS14150	ECRM13516_RS16915	DUF1327 domain_containing protein
	25	ECRM13516_RS03900	ECRM13516_RS07075	ECRM13516_RS14145	ECRM13516_RS16910	Lysozyme
	26	ECRM13516_RS03905	ECRM13516_RS05270	ECRM13516_RS12555		hypothetical protein
	27	ECRM13516_RS03910	ECRM13516_RS05275	ECRM13516_RS09800	ECRM13516_RS11655	lysis protein
	28	ECRM13516_RS03915	ECRM13516_RS29045	ECRM13516_RS29455	ECRM13516_RS29675	hypothetical protein
	29	ECRM13516_RS03920	ECRM13516_RS05280	ECRM13516_RS09795	ECRM13516_RS12545	PerC family transcriptional regulator
	30	ECRM13516_RS03925	ECRM13516_RS29050	ECRM13516_RS29450	ECRM13516_RS29670	DUF3950 domain_containing protein
	31	ECRM13516_RS03935	ECRM13516_RS05285	ECRM13516_RS12540		Terminase
	32	ECRM13516_RS03940	ECRM13516_RS05290	ECRM13516_RS12535		phage terminase large subunit family protein
	33	ECRM13516_RS03945	ECRM13516_RS05295	ECRM13516_RS12530		head_tail joining protein
	34	ECRM13516_RS03950	ECRM13516_RS05300	ECRM13516_RS12525		phage portal protein
	35	ECRM13516_RS03955	ECRM13516_RS05305	ECRM13516_RS12520		S49 family peptidase
	36	ECRM13516_RS03960	ECRM13516_RS05310	ECRM13516_RS12515		head decoration protein
	37	ECRM13516_RS03965	ECRM13516_RS05315	ECRM13516_RS12510		minor capsid protein E
	38	ECRM13516_RS03980	ECRM13516_RS09685	ECRM13516_RS12440	ECRM13516_RS14005	hypothetical protein
	39	ECRM13516_RS03985	ECRM13516_RS09680	ECRM13516_RS12435	ECRM13516_RS14000	tail fiber protein
	40	ECRM13516_RS03990	ECRM13516_RS09675	ECRM13516_RS12430	ECRM13516_RS13995	phage tail protein

B	Grupo	Locus tag 1	Locus tag 2	Locus tag 3	Locus tag 4	Descripción
Región 2: 41-65	41	ECRM13516_RS04700	ECRM13516_RS10020			dimethyl sulfoxide reductase subunit B
	42	ECRM13516_RS05115	ECRM13516_RS06915	ECRM13516_RS07975		Exonuclease
	43	ECRM13516_RS05135	ECRM13516_RS09930	ECRM13516_RS12695		DUF1391 domain_containing protein
	44	ECRM13516_RS05160	ECRM13516_RS12680			Rha family transcriptional regulator
	45	ECRM13516_RS05165	ECRM13516_RS12675			hypothetical protein
	46	ECRM13516_RS05190	ECRM13516_RS09865	ECRM13516_RS12635		Hok/Gef family protein
	47	ECRM13516_RS31170	ECRM13516_RS31245	ECRM13516_RS31350		hypothetical protein
	48	ECRM13516_RS05200	ECRM13516_RS07025	ECRM13516_RS09855	ECRM13516_RS12620	DUF968 domain_containing protein
	49	ECRM13516_RS05205	ECRM13516_RS07030	ECRM13516_RS09850	ECRM13516_RS12615	endodeoxyribonuclease
	50	ECRM13516_RS05225	ECRM13516_RS12600			transcriptional regulator AraC
	51	ECRM13516_RS05320	ECRM13516_RS12505			hypothetical protein
	52	ECRM13516_RS05325	ECRM13516_RS12500			tail attachment protein
	53	ECRM13516_RS05330	ECRM13516_RS12495			tail protein
	54	ECRM13516_RS05335	ECRM13516_RS09735	ECRM13516_RS12490		tail protein
	55	ECRM13516_RS05340	ECRM13516_RS09730	ECRM13516_RS12485		tail protein
	56	ECRM13516_RS05345	ECRM13516_RS12480			phage minor tail protein G
	57	ECRM13516_RS05350	ECRM13516_RS09720	ECRM13516_RS12475		phage tail assembly protein T
	58	ECRM13516_RS05355	ECRM13516_RS12470			phage tail tape measure protein
	59	ECRM13516_RS05360	ECRM13516_RS09710	ECRM13516_RS12465		tail protein
	60	ECRM13516_RS05365	ECRM13516_RS09705	ECRM13516_RS12460	ECRM13516_RS14025	phage minor tail protein L
	61	ECRM13516_RS05370	ECRM13516_RS09700	ECRM13516_RS12455	ECRM13516_RS14020	phage tail protein
	62	ECRM13516_RS05375	ECRM13516_RS09695	ECRM13516_RS12450	ECRM13516_RS14015	tail assembly protein
	63	ECRM13516_RS05380	ECRM13516_RS29570			hypothetical protein
	64	ECRM13516_RS05385	ECRM13516_RS11550			superoxide dismutase
	65	ECRM13516_RS05390	ECRM13516_RS09690	ECRM13516_RS12445	ECRM13516_RS14010	host specificity protein J

C	Grupo	Locus tag 1	Locus tag 2	Locus tag 3	Locus tag 4	Descripción
Región 3: 67-79	67	ECRM13516_RS06345	ECRM13516_RS25305	ECRM13516_RS26370		50S ribosome_binding GTPase
	68	ECRM13516_RS06355	ECRM13516_RS13030			hypothetical protein
	69	ECRM13516_RS06360	ECRM13516_RS13035			chemotaxis protein
	70	ECRM13516_RS06365	ECRM13516_RS13040			phospholipase
	71	ECRM13516_RS06375	ECRM13516_RS13055	ECRM13516_RS25280	ECRM13516_RS26395	DUF945 domain_containing protein
	72	ECRM13516_RS06380	ECRM13516_RS13060	ECRM13516_RS25270	ECRM13516_RS26400	antirestriction protein
	73	ECRM13516_RS06385	ECRM13516_RS13065	ECRM13516_RS25265	ECRM13516_RS26405	hypothetical protein
	74	ECRM13516_RS14095	ECRM13516_RS13070	ECRM13516_RS25260	ECRM13516_RS26410	DUF987 domain_containing protein
	75	ECRM13516_RS06400	ECRM13516_RS13075	ECRM13516_RS25255	ECRM13516_RS26420	type IV toxin_antitoxin system YeeU family antitoxin
	76	ECRM13516_RS06405	ECRM13516_RS25250	ECRM13516_RS26425		Toxin
	77	ECRM13516_RS06410	ECRM13516_RS26430			hypothetical protein
	78	ECRM13516_RS06415	ECRM13516_RS25240	ECRM13516_RS26435		DUF957 domain_containing protein
	79	ECRM13516_RS06420	ECRM13516_RS25235			DUF4942 domain_containing protein

D	Grupo	Locus tag 1	Locus tag 2	Locus tag 3	Locus tag 4	Descripción
Región 4: 80-103	80	ECRM13516_RS06950	ECRM13516_RS07995			hypothetical protein
	81	ECRM13516_RS06960	ECRM13516_RS08005			transcriptional regulator
	82	ECRM13516_RS06965	ECRM13516_RS08010			transcriptional regulator
	83	ECRM13516_RS06970	ECRM13516_RS08015			hypothetical protein
	84	ECRM13516_RS31235	ECRM13516_RS31280			replication protein
	85	ECRM13516_RS06990	ECRM13516_RS09880			hypothetical protein
	86	ECRM13516_RS06995	ECRM13516_RS09875			hypothetical protein
	87	ECRM13516_RS31240	ECRM13516_RS31355			hypothetical protein
	88	ECRM13516_RS07045	ECRM13516_RS12580	ECRM13516_RS14170		membrane protein
	89	ECRM13516_RS07115	ECRM13516_RS11645	ECRM13516_RS14105	ECRM13516_RS16885	HNH endonuclease
	90	ECRM13516_RS07120	ECRM13516_RS11640	ECRM13516_RS14100	ECRM13516_RS16880	terminase small subunit
	91	ECRM13516_RS07125	ECRM13516_RS11635	ECRM13516_RS14095	ECRM13516_RS16875	terminase large subunit
	92	ECRM13516_RS07130	ECRM13516_RS11630	ECRM13516_RS14090	ECRM13516_RS16870	phage major capsid protein
	93	ECRM13516_RS07135	ECRM13516_RS11625	ECRM13516_RS14085	ECRM13516_RS16865	hypothetical protein
	94	ECRM13516_RS07145	ECRM13516_RS11610	ECRM13516_RS14070	ECRM13516_RS16850	phage gp6_like head_tail connector protein
	95	ECRM13516_RS07150	ECRM13516_RS11605	ECRM13516_RS14065	ECRM13516_RS16845	head_tail adaptor protein
	96	ECRM13516_RS07155	ECRM13516_RS11600	ECRM13516_RS14060	ECRM13516_RS16840	hypothetical protein
	97	ECRM13516_RS07980	ECRM13516_RS09950	ECRM13516_RS12715		cell division inhibitor
	98	ECRM13516_RS08140	ECRM13516_RS11595	ECRM13516_RS14055	ECRM13516_RS16835	DUF3168 domain_containing protein
	99	ECRM13516_RS08145	ECRM13516_RS11590	ECRM13516_RS14050	ECRM13516_RS16830	tail protein
	100	ECRM13516_RS08150	ECRM13516_RS11585	ECRM13516_RS14045	ECRM13516_RS16825	tail assembly chaperon
	101	ECRM13516_RS08155	ECRM13516_RS11580	ECRM13516_RS14040	ECRM13516_RS16820	phage tail protein
	102	ECRM13516_RS08160	ECRM13516_RS11575	ECRM13516_RS14035	ECRM13516_RS16815	phage tail tape measure protein
	103	ECRM13516_RS08165	ECRM13516_RS11570	ECRM13516_RS14030	ECRM13516_RS16810	tail protein

E	Grupo	Locus tag 1	Locus tag 2	Descripción
Región 5: 129-137	129	ECRM13516_RS14240	ECRM13516_RS26985	hypothetical protein
	130	ECRM13516_RS14245	ECRM13516_RS26980	RusA family crossover junction endodeoxyribonuclease
	131	ECRM13516_RS14250	ECRM13516_RS26975	LexA family transcriptional regulator
	132	ECRM13516_RS14265	ECRM13516_RS26965	helix_turn_helix domain_containing protein
	133	ECRM13516_RS29885	ECRM13516_RS30820	DUF4222 domain_containing protein
	134	ECRM13516_RS14295	ECRM13516_RS26915	hypothetical protein
	135	ECRM13516_RS14300	ECRM13516_RS26910	DUF2303 domain_containing protein
	136	ECRM13516_RS14305	ECRM13516_RS26905	HD family hydrolase
	137	ECRM13516_RS14310	ECRM13516_RS26900	hypothetical protein

F	Grupo	Locus tag 1	Locus tag 2	Descripción
Región 6: 139-147	139	ECRM13516_RS18210	ECRM13516_RS25205	lysine_tRNA ligase
	140	ECRM13516_RS18645	ECRM13516_RS26195	DUF4102 domain_containing protein
	141	ECRM13516_RS18665	ECRM13516_RS25425	RNA polymerase sigma factor FecI
	142	ECRM13516_RS18670	ECRM13516_RS25430	protein FecR
	143	ECRM13516_RS18675	ECRM13516_RS25435	TonB_dependent siderophore receptor
	144	ECRM13516_RS18680	ECRM13516_RS25440	Fe(3+)_dicitrate ABC transporter substrate_binding protein FecB
	145	ECRM13516_RS18685	ECRM13516_RS25445	Fe3+ dicitrate ABC transporter permease FecC
	146	ECRM13516_RS18690	ECRM13516_RS25450	Fe3+ dicitrate ABC transporter permease FecD
	147	ECRM13516_RS18695	ECRM13516_RS25455	Fe(3+) dicitrate transport ATP_binding protein FecE

yeeR fw →

```

2244 TTCAGAAGGGGACCGGTGCAGGAAATATCAGTGCCCTGAACAAAGTAATGAAGGGTACGC
4509 TTCAGAAGGGGACCGGTACAGGAAATATCAGTGCCCTGAACAAAGTGATGAAGGGTACGC
*****

2244 TGGTGACATCTCTGGCACTGGTAGCTGTCAACAACCGGCCCTGACATGATCAAAATGTTGC
4509 TGGTGACATCTCTGGCACTGGTAGCTGTCAACAACCGGCCCTGACATGATCAAAATGTTGC
*****

2244 GGGGACGGATCTCCGGTGCAGTTCATCAGGAATCTTGCCGTGGCATCTTCCGGTGTGG
4509 GGGGACGGATCTCCGGTGCAGTTCATCAGGAATCTTGCCGTGGCATCTTCCGGTGTGG
*****

2244 CAGGTGGAGCTGTCCGGTGCAGTGGCGGGCGGATACTGTTTCAGTCCACTGGGACCATTTG
4509 CAGGTGGAGCTGTCCGGTGCAGTGGCGGGCGGATACTGTTTCAGTCCACTGGGACCATTTG
*****

2244 GTGCAC TGACAGGGCGTGTGGTTGGCGGTGTCTGGGGGAATGATGTCCTCCGCTGTAT
4509 GTGCAC TGACAGGGCGTGTGGTTGGCGGTGTCTGGGGGAATGATGTCCTCCGCTGTAT
*****

2244 CAGGAAAAAATGCCGGAGCGCTGGTTGAAGAAGATCGCGTCAAAATCTGGCAATGATTC
4509 CAGGAAAAAATGCCGGAGCGCTGGTTGAAGAAGATCGCGTCAAAATCTGGCAATGATTC
*****

2244 AGGAGCAGGTGACATGGCTTGCCGGCAGTTTCTGCTGACCGGACATGAGATTGAAAATC
4509 AGGAACAGGTGACATGGCTTGCCGGCAGTTTCTGCTGACCGGACATGAAAACAGAAAATC
**** *

2244 TGAACGAGAATCTGGCCCGTGTATCGATCAGAATGCTCTGGAGATCATTTTCGCCGCCG
4509 TGAACGCGAATCTGGCCCGTGTATCGATCAGAATGCTCTGGAGATCATTTTCGCCGCCG
*****

2244 GTATACAACAACGGGCCGCGACCAATATGTTAATCAAACCACTGGTGGTCAGTATCATCA
4509 GTATACAACAACGGGCCGCGACCAATATGTTAATCAAACCACTGGTGGTCAGTATCATCA
*****

2244 GGCAACGCCCCGTCATGGAATATGATGCATCCCATCTCGGCAATATGGTTAACCGACTGG
4509 GGCAACGCCCCGTCATGGAATATGATGCATCCCATCTCGGCAATATGGTTAACCGACTGG
*****

2244 AAGAAGCATTACCCCGGAGTTACCGGCATAACGGCTATTCTGCGCCTTTTACGCTTTAT
4509 AAGAAGCATTACCCCGGAGTTACCGGCATAACGGCTATTCTGCGCCTTTTACGCTTTAT
*****
    
```

yeeR

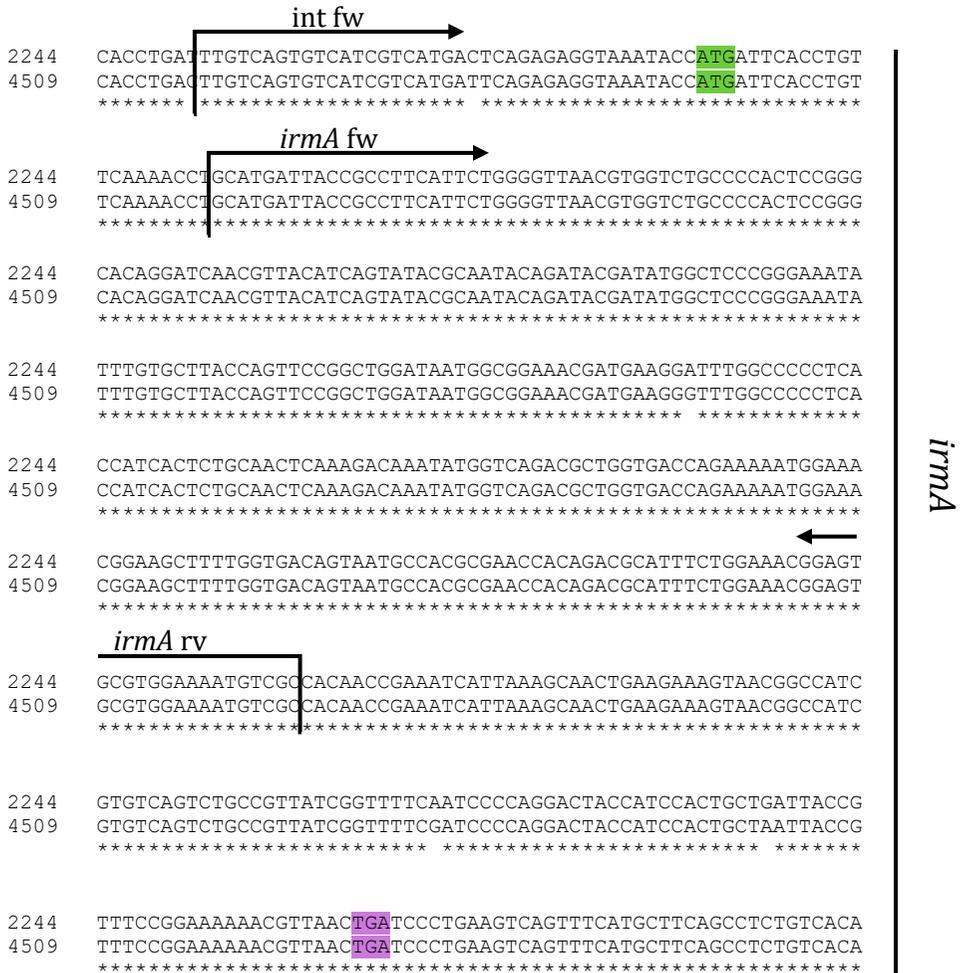


Fig. A2. Posición de los oligonucleótidos utilizados en el experimento de *Walking RT-PCR*. En morado se indica el codón de parada del gen *yeeR* y en verde el codón de inicio del gen *irmA*.

Anexos

042_4509 CATCCCATCTCGGCAATATGGTTAACCGGCTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
55989 CATCCCATCTCGGCAATATGGTTAACCGGCTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
O104:H4 CATCCCATCTCGGCAATATGGTTAACCGGCTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
H10407_3219 CATCCCATCTCGGCAATATGGTTAACCGACTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
H10407_4460 CATCCCATCTCGGCAATATGGTTAACCGACTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
O145:H28_5031 CATCCCATCTCGGCAATATGGTTAACCGACTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
CFT073_c1275 CATCCCATCTCGGCAATATGGTTAACCGACTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
EC958 CATCCCATCTCGGCAATATGGTTAACCGACTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
JJ1886 CATCCCATCTCGGCAATATGGTTAACCGACTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
CFT073_c3665 CATCCCATCTCGGCAATATGGTTAACCGACTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
UMN026 CATCCCATCTCGGCAATATGGTTAACCGACTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
O145:H28_5250 CATCCCATCTCGGCAATATGGTTAACCGACTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
042_2244 CATCCCATCTCGGCAATATGGTTAACCGACTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG
O26:H11 CATCCCATCTCGGCAATATGGTTAACCGACTGGAAGAAGCA**TTACCC**CCGGAGTTACCGG

042_4509 CAT**TAAC****GAGTATTC**GCACCTTTTACGTTTTATCACCTGACTTGTGTCAGTGTATCGTCAT
55989 CAT**TAAC****GAGTATTC**GCACCTTTTACGTTTTATCACCTGACTTGTGTCAGTGTATCGTCAT
O104:H4 CAT**TAAC****GAGTATTC**GCACCTTTTACGTTTTATCACCTGACTTGTGTCAGTGTATCGTCAT
H10407_3219 CAT**TAAC****GGGTATTC**GCGCCTTTTACGTTTTATCACCTGATTTGCCAGTGTATCGTCAT
H10407_4460 CAT**TAAC****GGGTATTC**GCGCCTTTTACGTTTTATCACCTGATTTGCCAGTGTATCGTCAT
O145:H28_5031 CAT**TAAC****GGGTATTC**GCGCCTTTTACGTTTTATCACCTGATTTGCCAGTGTATCGTCAT
CFT073_c1275 CAT**TAAC****GGGTATTC**GCGCCTTTTACGTTTTATCACCTGATTTGCCAGTGTATCGTCAT
EC958 CAT**TAAC****GGGTATTC**GCGCCTTTTACGTTTTATCACCTGATTTGCCAGTGTATCGTCAT
JJ1886 CAT**TAAC****GGGTATTC**GCGCCTTTTACGTTTTATCACCTGATTTGCCAGTGTATCGTCAT
CFT073_c3665 CAT**TAAC****GGGTATTC**GCGCCTTTTACGTTTTATCACCTGATTTGCCAGTGTATCGTCAT
UMN026 CAT**TAAC****GGGTATTC**GCGCCTTTTACGTTTTATCACCTGATTTGCCAGTGTATCGTCAT
O145:H28_5250 CAT**TAAC****GGGTATTC**GCGCCTTTTACGTTTTATCACCTGATTTGCCAGTGTATCGTCAT
042_2244 CAT**TAAC****GGGTATTC**GCGCCTTTTACGTTTTATCACCTGATTTGCCAGTGTATCGTCAT
O26:H11 CAT**TAAC****GGGTATTC**GCGCCTTTTACGTTTTATCACCTGATTTGCCAGTGTATCGTCAT

042_4509 GATT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
55989 GATT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
O104:H4 GATT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
H10407_3219 GACT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
H10407_4460 GACT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
O145:H28_5031 GACT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
CFT073_c1275 GACT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
EC958 GACT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
JJ1886 GACT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
CFT073_c3665 GACT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
UMN026 GACT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
O145:H28_5250 GACT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
042_2244 GACT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
O26:H11 GACT**CAGAGAGG**TAATACC**ATG**ATTACCTGTTCAAACCTGCATGATTACCGCCTTCA
** *****

042_4509 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
55989 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
O104:H4 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
H10407_3219 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
H10407_4460 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
O145:H28_5031 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
CFT073_c1275 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
EC958 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
JJ1886 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
CFT073_c3665 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
UMN026 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
O145:H28_5250 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
042_2244 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC
O26:H11 TTCTGGGGT**TAA**CGTGGTCTGCCCCACTCCGGGCACAGGATCAACGTTACATCAGTATAC

```

042_4509      GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
55989        GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
0104:H4      GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
H10407_3219 GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
H10407_4460 GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
0145:H28_5031 GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
CFT073_c1275 GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
EC958       GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
JJ1886      GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
CFT073_c3665 GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
UMN026      GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
0145:H28_5250 GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
042_2244    GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
026:H11     GCAATACAGATACGATATGGCTCCCGGAAAAATTTGTGCTTACCAGTTCGGCTGGATA
*****

042_4509      ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
55989        ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
0104:H4      ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
H10407_3219 ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
H10407_4460 ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
0145:H28_5031 ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
CFT073_c1275 ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
EC958       ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
JJ1886      ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
CFT073_c3665 ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
UMN026      ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
0145:H28_5250 ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
042_2244    ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
026:H11     ATGGCGGAAACGATGAAGGGTTTGGCCCCCTCACCATCACTCTGCAACTCAAAGACAAAT
*****

042_4509      ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
55989        ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
0104:H4      ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
H10407_3219 ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
H10407_4460 ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
0145:H28_5031 ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
CFT073_c1275 ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
EC958       ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
JJ1886      ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
CFT073_c3665 ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
UMN026      ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
0145:H28_5250 ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
042_2244    ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
026:H11     ATGGTCAGACGCTGGTGACCAGAAAAATGAAACCGGAAGCTTTTGGTGACAGTAATGCCA
***

042_4509      CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
55989        CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
0104:H4      CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
H10407_3219 CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
H10407_4460 CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
0145:H28_5031 CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
CFT073_c1275 CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
EC958       CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
JJ1886      CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
CFT073_c3665 CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
UMN026      CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
0145:H28_5250 CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
042_2244    CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
026:H11     CGCGAACCACAGACGCATTTCTGGAACCGGAGTGCCTGGAAAAATGTCGCCACAACCGAAA
*****

```

```

042_4509      TCATTAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCG
55989        TCATTAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCG
O104:H4       TCATTAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCG
H10407_3219  TCATCAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCA
H10407_4460  TCATTAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCG
O145:H28_5031 TCATTAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCG
CFT073_c1275 TCATTAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCG
EC958        TCATTAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCG
JJ1886       TCATTAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCG
CFT073c3665 TCATTAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCG
UMN026       TCATTAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCG
O145:H28_5250 TCATCAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCA
042_2244     TCATTAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCA
O26:H11      TCATCAAGCAACTGAAGAAAGTAACGGCCATCGTGTCAAGTCTGCCGTTATCGGTTTTCA
*****

042_4509      ATCCCCAGGACTACCATCCACTGCTAATTACCGTTTCCGGAAAAAACGTTAAC TGA
55989        ATCCCCAGGACTACCATCCACTGCTAATTACCGTTTCCGGAAAAAACGTTAAC TGA
O104:H4       ATCCCCAGGACTACCATCCACTGCTAATTACCGTTTCCGGAAAAAACGTTAAC TGA
H10407_3219  ATCCCCAGGACTACCATCCACTGCTGATTACCGTTTCCGGAAAAAACGTTAAC TGA
H10407_4460  ATCCCCAGGACTACCATCCACTGCTGATTACCGTTTCCGGAAAAAACGTTAAC TGA
O145:H28_5031 ATCCCCAGGACTACCATCCACTGCTGATTACCGTTTCCGGAAAAAACGTTAAC TGA
CFT073_c1275 ATCCCCAGGACTACCATCCACTGCTGATTACCGTTTCCGGAAAAAACGTTAAC TGA
EC958        ATCCCCAGGACTACCATCCACTGCTGATTACCGTTTCCGGAAAAAACGTTAAC TGA
JJ1886       ATCCCCAGGACTACCATCCACTGCTGATTACCGTTTCCGGAAAAAACGTTAAC TGA
CFT073c3665 ATCCCCAGGACTACCATCCACTGCTGATTACCGTTTCCGGAAAAAACGTTAAC TGA
UMN026       ATCCCCAGGACTACCATCCACTGCTGATTACCGTTTCCGGAAAAAACGTTAAC TGA
O145:H28_5250 ATCCCCAGGACTACCATCCACTGCTGATTACCGTTTCCGGAAAAAACGTTAAC TGA
043_2244     ATCCCCAGGACTACCATCCACTGCTGATTACCGTTTCCGGAAAAAACGTTAAC TGA
O26:H11      ATCCCCAGGACTACCATCCACTGCTGATTACCGTTTCCGGAAAAAACGTTAAC TGA
*****

```

Fig. A3. Alineamiento de la secuencia nucleotídica de la región promotora y codificante del gen *irmA* en diferentes cepas de *E. coli*. En morado se indica el codón de parada, en verde el codón de inicio y en amarillo la región -10 y -35 predichas mediante el programa BPROM.

9. Publicaciones

the *Journal of Applied Behavior Analysis* (1974), and the *Journal of Experimental Psychology: Applied* (1995).

There are a number of reasons why the *Journal of Applied Behavior Analysis* is the most widely cited journal in the field of behavior analysis.

First, the journal has a long history of publishing high-quality research. The journal was founded in 1968 and has since published numerous landmark studies in the field of behavior analysis.

Second, the journal has a broad scope of coverage. It publishes research in a wide range of areas, including clinical behavior analysis, experimental behavior analysis, and applied behavior analysis.

Third, the journal has a high impact factor. This is a measure of the journal's influence in the field, and it is one of the highest in the field of behavior analysis.

Finally, the journal has a high level of readability. The articles are written in a clear and concise style, making them accessible to a wide range of researchers and practitioners.

Overall, the *Journal of Applied Behavior Analysis* is the most widely cited journal in the field of behavior analysis because of its long history of publishing high-quality research, its broad scope of coverage, its high impact factor, and its high level of readability.

References
 Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Journal of Applied Behavior Analysis (1974)
 Journal of Experimental Psychology: Applied (1995)

Durante el desarrollo de la tesis doctoral se ha publicado el siguiente artículo, correspondiente a los resultados mostrados en el primer capítulo de esta memoria:

- Capítulo 1 (apartado 4.1): Bernabeu M, Sánchez-Herrero JF, Huedo P, Prieto A, Hüttener M, Rozas J and Juárez A. Gene duplications in the *E. coli* genome: common themes among pathotypes. *BMC Genomics.* 2019;20(1):313. doi:10.1186/s12864-019-5683-4

Bernabeu et al. *BMC Genomics* (2019) 20:313
<https://doi.org/10.1186/s12864-019-5683-4>

BMC Genomics

RESEARCH ARTICLE

Open Access

Gene duplications in the *E. coli* genome: common themes among pathotypes



Manuel Bernabeu¹, José Francisco Sánchez-Herrero^{1,2}, Pol Huedo³, Alejandro Prieto¹, Mário Hüttener¹, Julio Rozas^{1,2} and Antonio Juárez^{1,4*}

Abstract

Background: Gene duplication underlies a significant proportion of gene functional diversity and genome complexity in both eukaryotes and prokaryotes. Although several reports in the literature described the duplication of specific genes in *E. coli*, a detailed analysis of the extent of gene duplications in this microorganism is needed.

Results: The genomes of the *E. coli* enteroaggregative strain 042 and other pathogenic strains contain duplications of the gene that codes for the global regulator Hha. To determine whether the presence of additional copies of the *hha* gene correlates with the presence of other genes, we performed a comparative genomic analysis between *E. coli* strains with and without *hha* duplications. The results showed that strains harboring additional copies of the *hha* gene also encode the *yeeR-immA (oac69)* gene cluster, which, in turn, is also duplicated in strain 042 and several other strains. The identification of these duplications prompted us to obtain a global map of gene duplications, first in strain 042 and later in other *E. coli* genomes.

Duplications in the genomes of the enteroaggregative strain 042, the uropathogenic strain CFT073 and the enterohemorrhagic strain O145H28 have been identified by a BLASTp protein similarity search. This algorithm was also used to evaluate the distribution of the identified duplicates among the genomes of a set of 28 representative *E. coli* strains. Despite the high genomic diversity of *E. coli* strains, we identified several duplicates in the genomes of almost all studied pathogenic strains. Most duplicated genes have no known function. Transcriptomic analysis also showed that most of these duplications are regulated by the H-NS/Hha proteins.

Conclusions: Several duplicated genes are widely distributed among pathogenic *E. coli* strains. In addition, some duplicated genes are present only in specific pathotypes, and others are strain specific. This gene duplication analysis shows novel relationships between *E. coli* pathotypes and suggests that newly identified genes that are duplicated in a high percentage of pathogenic *E. coli* isolates may play a role in virulence. Our study also shows a relationship between the duplication of genes encoding regulators and genes encoding their targets.

Keywords: Pathotypes, Gene duplication, *Escherichia coli* 042, H-NS, Hha

Por otra parte, se ha logrado publicar otros artículos derivados de la colaboración con la actividad realizada por nuestro grupo de investigación:

- Prieto A, Bernabeu M, Falgenhauer L, Chakraborty T, Hüttener M, Juárez A. The regulated expression of the third H-NS paralogue H-NS2 influences the fitness of the enteroaggregative *Escherichia coli* strain 042. Sci Rep. 2020 Oct 22; 10(1):18131. doi: 10.1038/s41598-020-75204-4.
- Sánchez-Herrero JF, Bernabeu M, Prieto A, Hüttener M, Juárez A. Gene duplications in the genomes of Staphilococci and Enterococci. Front. Mol. Biosci 2020 doi: 10.3389/fmolb.2020.00160. eCollection 2020.
- Hüttener M, Prieto A, Aznar S, Bernabeu M, Glaría E, Valledor AF, Paytubi S, Merino S, Tomás J, Juárez A. Expression of a novel class of bacterial Ig-like proteins is required for IncHI plasmid conjugation. PLoS Genet.2019 Sep 17;15(9):e1008399. doi: 0.1371/journal.pgen.1008399. eCollection 2019 Sep.
- Prieto A, Bernabeu M, Aznar S, Ruiz-Cruz S, Bravo A, Queiroz MH, Juárez A. Evolution of bacterial global modulators: role of a novel H-NS paralogue in the enteroaggregative *Escherichia coli* strain 042. mSystems. 2018 Mar 20;3(3):e00220-17. doi: 10.1128/mSystems.00220-17. eCollection May-Jun 2018.
- Queiroz MH, Prieto A, Espelt J, Bernabeu M, Juárez A. Stringent response and AggR-dependent virulence regulation in the enteroaggregative *Escherichia coli* strain 042. Front Microbiol doi: 10.3389/fmicb.2018.00717.

