



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

ESTUDIO DE LOS GENES ESENCIALES EN BACTERIAS, CON ÉNFASIS EN
Escherichia coli y *Pseudomonas aeruginosa*.

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

M. en C. ENRIQUE MARTÍNEZ CARRANZA

DRA. GLORIA SOBERÓN CHÁVEZ
TUTOR PRINCIPAL

[Instituto de Investigaciones Biomédicas](#)

MIEMBROS DEL COMITÉ TUTOR:
DR. LUIS DAVID ALCARAZ PERAZA

[Facultad de Ciencias](#)

DR. LUIS SERVÍN GONZÁLEZ
[Instituto de Investigaciones Biomédicas](#)

Ciudad Universitaria, CDMX, Junio de 2019.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

AGRADECIMIENTOS

A mi familia, porque siempre e incondicionalmente me apoyaron. A mis papás, gracias por abrirme camino, siempre estaré agradecido. A mi esposa Cecilia (espero algún día poder devolverte todo lo que me has dado, paciencia, amor, tiempo, apoyo, en fin, muchas gracias), mi hija Diana (discúlpame por no haber estado contigo tantos días y noches, por no haberte leído y por no jugar contigo durante mucho tiempo por estar lejos), a quien ya casi llega (aquí te esperamos, contentos y nerviosos), a mis hermanas Lizeth y Erika. Gracias, sé que esto no fue fácil para nadie, gracias por estar ahí.

Agradezco a la Doctora Gloria Soberón Chávez, una de las personas más humanitarias y solidarias que he conocido, gracias por recibirme en su laboratorio, gracias por su interés en la formación de sus estudiantes, gracias por la oportunidad.

Agradezco a la Universidad Nacional Autónoma de México por recibirme entre sus aulas y laboratorios, aquí conocí excelentes investigadores, profesores y pude hacer amistades invaluable.

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo para realizar mis estudios de doctorado CVU 483310.

Muchas gracias al “Programa de Apoyo a los Estudios de Posgrado” (PAEP) por los apoyos que me otorgó para la asistencia a Congresos Nacionales e Internacionales.

Agradezco a los Doctores Luis David Alcaraz Peraza y Luis Servín González por la formación que me brindaron a lo largo del doctorado, su disponibilidad y sus consejos.

Agradezco al Dr. Miguel Cócotl Yáñez por su amistad, por su ayuda en el laboratorio y por sus enseñanzas.

Gracias a mis compañeros del laboratorio por hacer placentera y alegre la estancia en el laboratorio; Gabriel, Miguel, Selene, Martín, Fernando, René y Uriel, muchas gracias por su amistad y enseñanzas.

Puedo nombrar muchos profesores que han dejado huella en mi formación a lo largo de mi carrera, pero agradezco enormemente a los Doctores: Gloria Soberón, Luis David Alcaraz, Antonio Lazcano, Miguel Cócotl, Valeria Souza, Carlos Cortés Penagos, Armida Sánchez Gallegos, Rosalio Mercado, Jesús Campos.

ABREVIATURAS

DGNO: Desplazamiento de genes no ortólogos

µg: microgramos

µl: microlitros

ml: mililitros

mM: milimolar

THG: Transferencia horizontal de genes

DNA: Ácido desoxirribonucleico

tRNA: Ácido ribonucleico de transferencia

RNA: Ácido ribonucleico

ORF: Marco de lectura abierto (Open Reading Frame)

COGs: Grupos de genes ortólogos (Clusters of Orthologous Genes)

PCR: Reacción en cadena de la polimerasa

NCBI: National Center for Biotechnology Information

RESUMEN

El estudio de los genes esenciales en bacterias es fundamental para comprender los principios básicos de la vida y tiene implicaciones en varias ramas de las ciencias biológicas, básicas y aplicadas, como son la evolución de las bacterias, la biología sintética, la medicina, etc. Desde que se ha incrementado la disponibilidad de genomas bacterianos completos, se ha visto aumentada también la inquietud por conocer cuáles son los genes fundamentales para el desarrollo de la vida. Con el desarrollo de técnicas de secuenciación y desarrollo de métodos para analizar las secuencias, también ha florecido el desarrollo de técnicas que permiten el análisis de los genomas y de esta forma conocer los genes esenciales. En los últimos años se han realizado análisis para determinar los genes esenciales en una amplia variedad de especies bacterianas con varios enfoques utilizados. Dos de las especies bacterianas mejor caracterizadas a la fecha son *Escherichia coli* y *Pseudomonas aeruginosa*, en las que varios trabajos han propuesto los genes esenciales en variadas condiciones de cultivo. Se ha reportado que los genes esenciales son variables entre especies bacterianas, pero también entre cepas de una misma especie, lo que representaría una significativa incongruencia evolutiva; la variabilidad de los genes esenciales puede explicarse bajo el fenómeno del desplazamiento de genes no ortólogos, en el que secuencias no homólogas reemplazan las funciones de genes específicos, sin embargo este fenómeno contrasta con el actual modelo de evolución y filogenética bacteriana. También se ha explicado que la variación de los genes esenciales puede deberse a errores en las secuencias genómicas. En este trabajo tratamos sin éxito de encontrar ejemplos de desplazamiento de genes no ortólogos en *E. coli*, analizando la conservación de los genes esenciales en *E. coli* en 63 genomas completos y se evaluó la esencialidad del gen *bcsB*, que es un gen esencial condicional dependiente de una efectiva producción de celulosa. Se analizó la variabilidad reportada de los genes esenciales entre las cepas de *P. aeruginosa* PAO1 y PA14, observándose que los genes esenciales están muy conservados en estas cepas. También se analizó y se corroboró la conservación de los genes esenciales generales de *P. aeruginosa* en la bacteria cercanamente relacionada *Azotobacter vinelandii*, lo que tiene implicaciones en la evolución de esta especie bacteriana.

ABSTRACT

The study of essential genes in bacteria is fundamental to understand the basic principles of life and has implications in several areas of the biological, basic and applied sciences, such as the evolution of bacteria, synthetic biology, medicine, etc. Since the availability of complete bacterial genomes has increased, there has also been an increase in the concern to know which genes are fundamental for the development of life. With the development of techniques to improve DNA sequencing and bioinformatics analysis, there has been an explosion of bacterial genome sequenced and thus, the search of essential genes has considerably increased. In recent years, analyzes have been carried out to determine the essential genes in a wide variety of bacterial species using several approaches. Two of the best-characterized bacterial species to date are *Escherichia coli* and *Pseudomonas aeruginosa*, in which several papers have defined the essential genes in various culture conditions. It has been reported that the essential genes are variable between bacterial species and also between strains of the same species, which would represent a significant evolutionary incongruence. However the variability of essential genes has been explained by the phenomenon of non-orthologous genes displacement, in which non-homologous sequences replace the functions of specific genes, However, this phenomenon contrasts with the current model of evolution and bacterial phylogenetics. It has also been explained that the variation of the essential genes can be due to errors in the sequencing of genomes. In this work we tried to find examples of non-orthologous gene displacement in *E. coli* without success, analyzing the conservation of 303 essential genes *E. coli* in 63 complete genomes deposited in NCBI. I also report that the essentiality of the *bcsB* gene is dependent of a functional cellulose production and is thus a conditional essential gene (it is only essential when cellulose is produce to translocate it out of the cell). The reported variability of the essential genes between *P. aeruginosa* PAO1 and PA14 type strains was analyzed, observing that the essential genes are highly conserved in these strains. The conservation of the general essential genes of *P. aeruginosa* in *Azotobacter vinelandii* was also analyzed and corroborated. We discuss the implications of the high degree of conservation essential genes in the evolution of *A. vinelandii*.

Contenido

AGRADECIMIENTOS.....	i
ABREVIATURAS	iii
RESUMEN.....	iv
ABSTRACT.....	v
ÍNDICE DE FIGURAS.....	viii
ÍNDICE DE TABLAS	ix
I.- INTRODUCCIÓN.....	1
I.I. ESTUDIO DE LOS GENOMAS BACTERIANOS.....	1
I.I.I. El genoma bacteriano.....	1
I.I.II. Avances en el estudio de genomas bacterianos completos	2
I.I.III. Variabilidad genética de los genomas bacterianos.....	4
I.I.IV. Genes esenciales en bacterias	10
I.I.V. Métodos para la identificación de genes esenciales	12
I.I.V.I Mutagénesis por inserción de transposones.....	14
I.I.V.II Silenciamiento de genes con RNA antisentido.....	16
I.I.V.III Mutagénesis específica de genes	16
I.I.VI. Análisis de los genes esenciales en E. coli	17
I.I.VII. Análisis de los genes esenciales en P. aeruginosa.....	20
II.- JUSTIFICACIÓN.....	23
III.- HIPÓTESIS.....	24
IV.- OBJETIVOS	24
IV.I.- OBJETIVO GENERAL	24

IV.II.- OBJETIVOS ESPECÍFICOS	25
V.- MATERIALES Y MÉTODOS	25
V.I.- Estrategia de búsqueda de los genes esenciales de E. coli K-12 en otros 63 genomas de cepas de E. coli.....	25
V.II.- Material biológico y medios de cultivo	28
V.III.- Técnicas de manipulación de DNA	29
V.IV.- Construcción de plásmidos.....	30
V.V.- Selección de cepas mutantes en bcsB y bcsA en E. coli BW25113 y CFT073.....	31
V.VI.- Evaluación de la producción de celulosa.....	32
V.VII.- Análisis de los ortólogos de los genes esenciales de P. aeruginosa en las cepas PAO1, PA14 y en A. vinelandii.....	32
VI. RESULTADOS	34
VI.I.- Estudio de los genes esenciales de E. coli	34
VI.II.- Estudio de la variabilidad de los genes esenciales en las cepas PAO1 y PA14 de P. aeruginosa.....	47
VII.- DISCUSIÓN	55
VIII.- CONCLUSIONES.....	66
IX.- ANEXOS	66
X.- PUBLICACIONES	76
XI.- REFERENCIAS.....	77

ÍNDICE DE FIGURAS

Figura 1.- Representación esquemática de los genes que componen el Pangenoma, genoma central o “core” y el genoma accesorio.....	6
Figura 2.- Genes esenciales propuestos en <i>E. coli</i> con diferentes estrategias y en diferentes estudios.....	20
Figura 3.- Línea de tiempo que representa los análisis sobre genes esenciales en <i>P. aeruginosa</i> , cepas PAO1 y PA14.	23
Figura 4.- Representación esquemática (mapa de colores) de la presencia de los 303 genes esenciales propuestos por Baba et al., en 63 genomas de diferentes cepas de <i>E. coli</i>	35
Figura 5.- Representación sinténica de los genes <i>secA</i> (A) e <i>infC</i> (B) en las cepas UM146 y P12b, respectivamente, comparado con la cepa referencia K-12 en cada caso.	38
Figura 6.- Representación sinténica del gen <i>dxr</i> en la cepa CFT073 en comparación con la cepa de referencia K-12 (A) y alineamiento de las secuencias del gen <i>dxr</i> de las cepas CFT073 y cepa K-12 (B), se encierran en rojo los huecos producto de errores de secuenciación.....	40
Figura 7.- Representación sinténica de la región del gen <i>alsK</i> en tres cepas representativas de las tres composiciones observadas en los 63 genomas.	40
Figura 8.- Representación sinténica de la delección del operón <i>bcs</i> (involucrado en la síntesis de celulosa) en la cepa HS que abarca aproximadamente 12.5 Kb, en comparación con la cepa de referencia CFT073, que parece tener el operón completo y funcional.....	42
Figura 9.- Identificación de la delección del operón <i>bcs</i> en la cepa HS que abarca aproximadamente 12.5 Kb, en comparación con la cepa de referencia CFT073.	44

Figura 10.- Evaluación de la producción de celulosa por las cepas BW25113, HS y CFT073.	46
Figura 11.- Genes esenciales de <i>P. aeruginosa</i> PAO1 y PA14 reportados en diferentes estudios que se comparten.....	47
Figura 12.- Alineamiento de los genomas de <i>P. aeruginosa</i> PAO1 y <i>A. vinelandii</i> DJ con el programa MAUVE.....	53
Figura 13.- Comparación sinténica de los loci de los genes esenciales generales propuestos por Lee (2015) de <i>P. aeruginosa</i> en <i>A. vinelandii</i>	54
Figura 14.- Esquema de los componentes del sistema de síntesis funcional de celulosa en <i>E. coli</i>	61

ÍNDICE DE TABLAS

Tabla 1.- Número de genes esenciales identificados en diferentes especies bacterianas por diferentes métodos. Modificada de Juhas, M. et al., 2012.....	13
Tabla 2.- Cepas y números de acceso de los 63 genomas de <i>E. coli</i> utilizados en la búsqueda de los 303 genes esenciales propuestos por Baba et al., 2006....	26
Tabla 3.- Oligonucleótidos utilizados en este trabajo.....	29
Tabla 4.- Genes esenciales definidos por Baba et al., que están ausentes en el genoma de alguna o varias cepas analizadas.....	36
Tabla 5.- Relación del número de candidatas al hacer la delección de los genes <i>bcsA</i> y <i>bcsB</i> (A) y número de transconjugantes al transformar con los plásmidos pUCP- <i>bcsA</i> y pUCP- <i>bcsAB</i> (B) para evaluar la esencialidad de <i>bcsB</i> en las cepas CFT073 y BW25113. La expresión de los genes se realiza en presencia de IPTG 0.5mM.....	45
Tabla 6.- Conjunto de genes esenciales reportados en <i>P. aeruginosa</i> PA14 (Skurnik et al., 2013), ausentes en el genoma de <i>P. aeruginosa</i> PAO1.....	49

Tabla 7.- Genes esenciales en <i>P. aeruginosa</i> PAO1 (Turner et al., 2015), ausentes en el genoma de <i>P. aeruginosa</i> PA14.....	50
Tabla 8.- Genes esenciales generales en el genoma de <i>P. aeruginosa</i> PAO1 propuestos por Lee (2015), ausentes en el genoma de <i>P. aeruginosa</i> PA14.....	51
Tabla 9.- Genes esenciales generales propuestos por Lee (2015), ausentes en el genoma de <i>A. vinelandii</i>	54
Tabla 10.- Conjunto de 210 genes esenciales reportados por Skurnik et al., 2013, como ausentes en el genoma de <i>P. aeruginosa</i> PAO1.....	66
Tabla 11.- Conjunto de 352 genes esenciales generales reportados en <i>P. aeruginosa</i> PAO1 (Lee et al., 2015).....	68
Tabla 12.- Conjunto de genes esenciales en <i>P. aeruginosa</i> PAO1 para su desarrollo en medio con esputo de fibrosis quística (Turner et al., 2015).....	71
Tabla 13.- Plásmidos utilizados en este trabajo.....	74

I.- INTRODUCCIÓN

La presente tesis se trata del estudio de los genes esenciales en bacterias, que son los genes fundamentales para la viabilidad celular, específicamente en dos de las especies bacterianas mejor caracterizadas y de las que más se conoce hasta el momento, *Escherichia coli* y *Pseudomonas aeruginosa*. La importancia de la comprensión de los genes esenciales en bacterias es enorme, pues impacta en un gran número de áreas de la biología, así como en ciencias de la salud. En este trabajo se realizaron estudios experimentales y bioinformáticos para tratar de comprender mejor lo que se conoce sobre los genes esenciales de estas dos especies, y se pretende contribuir con conocimiento sobre la esencialidad de los genes y la variabilidad de los mismos entre cepas de estas especies.

I.I. ESTUDIO DE LOS GENOMAS BACTERIANOS

I.I.I. El genoma bacteriano

Las bacterias y arqueas poseen un nucleoide, que es un complejo supramolecular donde se localizan el material genético y moléculas asociadas a él, los detalles de la estructura del nucleoide han sido apenas dilucidados gracias a novedosas técnicas bioquímicas y de microscopía electrónica. En la mayoría de los casos el genoma bacteriano está compuesto por un cromosoma circular y está organizado por un conjunto de proteínas y por mecanismos físicos que contribuyen en su compactación (Dame et al., 2012), El tamaño de los genomas procarióticos va de alrededor de 50 Kbp a más de 13 Mbp (Schneiker et al., 2007; Tatusova et al., 2015; Touchon and Rocha, 2016), los genomas más pequeños corresponden a genomas arqueanos y bacterias endosimbiontas abligadas, en las que todo su repertorio genético no es suficiente para cubrir los requerimientos mínimos de la

vida autónoma en bacterias de vida libre como *Escherichia coli*, además se ha observado que son muy compactos comparados con los genomas de eucariotes, ya que generalmente presentan una densidad de genes del 87%, con un intervalo de 85-90% (McCutcheon and Moran, 2012; Mira et al., 2001), lo que indica que hay una correlación directa entre el número de genes codificantes y el tamaño del genoma.

El cromosoma bacteriano de *E. coli* está organizado a una escala estructural de nanómetros y contiene en promedio 4,288 genes codificantes de proteínas y un tamaño de 4.6 Mbp (Dorman, 2013). Se ha propuesto que la arquitectura de los genomas bacterianos está moldeada por diversos procesos, por ejemplo, la variación en los genomas bacterianos se debe a la integración y a la pérdida de genes accesorios donde a lo largo de las generaciones, las secuencias no funcionales son eliminadas del genoma. Se conoce además, que los genomas bacterianos son de menor tamaño en bacterias que parasitan a otras células, siendo así los endosimbiontes obligados las bacterias con genomas más pequeños (Bobay and Ochman, 2017; McCutcheon and Moran, 2012; Mira et al., 2001).

I.I.II. Avances en el estudio de genomas bacterianos completos

En bacteriología, la era genómica comenzó en el año de 1995, cuando el primer genoma bacteriano fue secuenciado utilizando la secuenciación convencional de Sanger. Desde que el primer genoma bacteriano fue secuenciado (Fleischmann et al., 1995) en *Haemophilus influenzae* se ha incrementado en gran medida la capacidad de secuenciación de los genomas microbianos debido a mejoras en las técnicas de secuenciación y su gran reducción en costos, también el desarrollo de nuevos programas bioinformáticos que han permitido acelerar el ensamblado y análisis de más y más genomas, ha sido una parte fundamental para el progreso

de la genómica (Loman et al., 2012). La notable mejora en este sentido, ha permitido un avance significativo en el conocimiento de la biología bacteriana y en el entendimiento de los genomas bacterianos en los últimos años (Land et al., 2015; Loman and Pallen, 2015). El número de genomas bacterianos secuenciados ha crecido enormemente en los últimos diez años, por ejemplo, en el año 2014 (Land et al., 2015) se reportaban más de 30,000 genomas disponibles en la base de datos del NCBI y actualmente se reportan 198,443, de los cuales 13,878 son genomas completos y 2,309 genomas ensamblados a nivel de cromosoma, adicionalmente se encuentran 81,487 en nivel scaffold (andamio) y 100,769 contigs.

La secuenciación de genomas bacterianos y la disponibilidad de los mismos ha aportado tanta información que ahora está disponible para hacer taxonomía molecular y realizar estudios filogenéticos con el análisis de genomas completos, que poco a poco sustituyen al análisis de secuencias nucleotídicas más cortas, como son el RNA ribosomal 16S. La secuenciación de genomas completos, en conjunto con herramientas bioinformáticas y filogenéticas permiten llevar a cabo análisis, clasificación de los genomas y una comprensión de las relaciones genéticas entre especies y cepas bacterianas a gran escala y con tal magnitud, que supera enormemente los resultados que los análisis con 16S RNAr como marcadores sólo aproximaban (Land et al., 2015).

El avance en la secuenciación de genomas bacterianos completos cada vez más rápido y a menor costo ha permitido también el avance de la metagenómica. Cada vez más proyectos de metagenomas son reportados y éstos son cada vez más ambiciosos, lo que ha permitido conocer una gran diversidad microbiana debido a la cantidad de genomas bacterianos secuenciados en diversos ambientes; de igual manera, el aumento de estudios metagenómicos en los últimos años ha crecido a pasos gigantescos (Land et al., 2015).

Desde que se tiene la posibilidad de secuenciación de genomas bacterianos completos en menos tiempo, con mayor calidad y en una gran diversidad de especies bacterianas, se ha tenido también la posibilidad de analizar detalladamente los genomas de diferentes especies así como de varias cepas de la misma especie mediante genómica comparativa. Hoy en día se tiene en las bases de datos de genomas completos una gran cantidad de genomas disponibles (Land et al., 2015; Loman et al., 2012), lo que hizo posible el nacimiento y desarrollo de la genómica comparativa.

1.1.III. Variabilidad genética de los genomas bacterianos

Gracias al análisis comparativo de los genomas bacterianos de distintas especies y entre cepas de una misma especie, se sabe que los genomas bacterianos son mosaicos de información genética compuestos por fragmentos de diversos orígenes, es decir, su origen es polifilético ya que comparten una fracción de los genes que componen el genoma con otras especies cercanas, e incluso con otras cepas de la misma especie (Fraser et al., 2009; Mira et al., 2004; Zhaxybayeva et al., 2009).

Contrario a lo que podría pensarse, la variación de genes entre bacterias cercanamente relacionadas e incluso entre miembros de una especie bacteriana puede llegar a ser considerable. Hay algunas especies que muestran muy poca variación de genes entre sus miembros y la variación observada entre estas puede observarse principalmente en elementos de DNA móviles y en polimorfismos de un único nucleótido (SNPs, por sus siglas en inglés), tal es el caso de *Mycobacterium tuberculosis* (Land et al., 2015). En muchas otras especies bacterianas donde la variación entre sus miembros es más importante a lo largo del cromosoma, estas observaciones han hecho posible considerar como un hecho que la diversidad genética en el mundo bacteriano es mucho mayor de lo que se esperaba, entonces no solamente el tamaño del genoma bacteriano es

variable entre especies y entre los miembros de una misma especie, sino que también el contenido de genes varía considerablemente. Se ha propuesto que las variaciones en el tamaño del genoma afecta funciones celulares de diferentes maneras, por ejemplo, los repertorios de genes asociados con funciones “housekeeping” como traducción muestran poca variación, en contraste con los repertorios de genes para otras funciones que son mucho más variables, como funciones de transporte, comunicación y funciones regulatorias (Touchon and Rocha, 2016). Un ejemplo de esto es la variabilidad reportada en *E. coli*, en la que cualquiera de sus cepas tiene aproximadamente sólo un 10% del número total de genes presentes en el pangenoma de esta especie (Land et al., 2015), de tal manera que al menos un tercio del genoma de *E. coli* es específico de cepa.

La alta variabilidad genética en las bacterias ha sido también correlacionada con la alta incidencia de transferencia horizontal de genes (THG) entre los genomas bacterianos (Bobay and Ochman, 2017; Dagan et al., 2008; Lan and Reeves, 1996; Soucy et al., 2015), además se ha relacionado directamente el tamaño del genoma bacteriano con la frecuencia de THG en los mismos, a mayor tamaño del genoma mayor frecuencia de THG (Cordero and Hogeweg, 2009) y codifican una mayor cantidad de elementos transposables (Touchon and Rocha, 2007).

La alta cantidad de anomalías en la filogenética bacteriana, debida a los fenómenos anteriormente mencionados ha originado algunos conceptos que intentan dar sentido a la variabilidad genética de las especies bacterianas; estos conceptos son el pangenoma (Medini et al., 2005; Tettelin et al., 2005, 2008), que representa toda la diversidad y/o repertorio genético de una especie bacteriana y que está conformado por el genoma “core” o genoma “central” que representan el conjunto de genes compartidos por todos los individuos que pertenecen a una especie bacteriana (y que se ha propuesto contiene a los genes esenciales de esa especie) y el genoma flexible, que contiene a los genes que se encuentran presentes sólo en uno o algunos miembros de la especie en cuestión, mas no en todos (Figura 1). La mayoría de los genes que pertenecen al genoma accesorio

son adquiridos por THG y se considera que codifican para funciones adaptativas y que por lo tanto no representan funciones o genes esenciales para la biología de ninguna especie bacteriana (Cordero and Hogeweg, 2009; Darmon and Leach, 2014; Lawrence, 1999; Soucy et al., 2015).

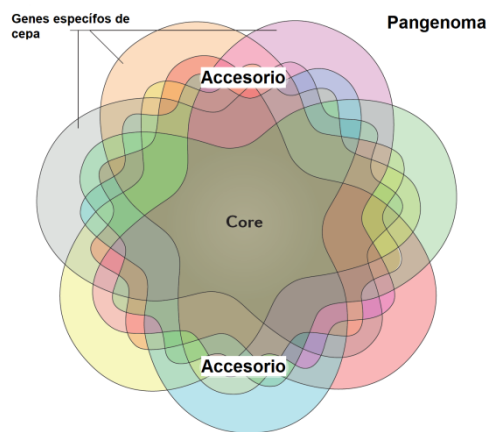


Figura 1 | Representación esquemática de los genes que componen el Pangenoma, genoma central o “core” y el genoma accesorio. Los colores representan los genes los diferentes genes, compartidos o no, por los miembros de la especie, en gris se representa el genoma “Core”. Modificada de Soucy et al., 2015.

Los genes compartidos entre todos los miembros de una especie, genoma “core”, han sido considerados como los genes que representan la historia evolutiva y es por esto que han sido los genes utilizados como marcadores moleculares para estudiar la filogenia bacteriana y construcción de árboles filogenéticos (Ciccarelli et al., 2006; Hug et al., 2016; Zeigler, 2003) ya que se pensaba que son genes no transferibles horizontalmente. Sin embargo, recientemente se han reportado genes del “core” que muestran inconsistencias filogenéticas en algunas bacterias, tal es el caso del 16S RNAr, que ha sido transferido horizontalmente en el género

Enterobacter (Sato and Miyazaki, 2017), tRNA sintasas (Soucy et al., 2015), algunos ejemplos de la familia del regulador transcripcional CRP en diferentes especies bacterianas y genes regulados por estos reguladores, que también se ha encontrado pueden ser transferidos horizontalmente (Soberón-Chávez et al., 2017). Estos datos confirman que la transferencia horizontal de genes no es exclusiva de genes accesorios y que genes conservados pueden estar sujetos a este sistema de transferencia de genes en bacterias.

También se han reportado genomas bacterianos con orígenes polifiléticos e inconsistencias con el modelo de especies bacterianas como un grupo de cepas “genómicamente coherentes” que comparten muchos rasgos en común (Rosselló-Mora and Amann, 2001); tal es el caso de los Thermotogales (Zhaxybayeva et al., 2009) y de la bacteria *Azotobacter vinelandii* bacteria en la que varios de sus genes, que codifican para funciones esenciales y que la definen biológicamente, han sido heredados por THG. Algunos ejemplos de estos genes son *vnfY*, *arsDCBA* y *anfA*, que codifican para proteínas relacionadas a la fijación de nitrógeno, síntesis de alquilresorcinoles y como proteínas flagelares, respectivamente. Estos genes forman parte del genoma accesorio de las Pseudomonadaceas, que es el grupo filogenético que alberga a las especies de *Pseudomonas* y *Azotobacter* (González-Casanova et al., 2014).

Se ha reportado que la variabilidad genética que ha sido documentada en los genomas de distintas especies bacterianas y entre los miembros de una misma especie incluye también a los genes esenciales (Bergmiller et al., 2012; Ciccarelli et al., 2006; Juhas et al., 2011; Skurnik et al., 2013). Varias técnicas se han utilizado para caracterizar los genes esenciales en distintas especies bacterianas (Deng et al., 2012), la genómica comparativa ha sido la metodología mediante la cual se han hecho las primeras búsquedas de los genes que se encuentran conservados en varias especies de bacterias e incluso de arqueas (Charlebois and Doolittle, 2004). Sin embargo, los resultados finales arrojaron un número muy bajo de genes conservados (35) cuando se analizaron 147 genomas de procariontes,

de los cuales 17 fueron arqueanos, lo que ha generado conclusiones que apoyan una muy baja conservación de los genes esenciales en procariontes.

Además de este estudio, un análisis posterior que comparó 1000 secuencias genómicas, reportó que no hubo ninguna secuencia genética común a todos los genomas, además reportaron que sólo encontraron 4 genes comunes entre los genomas de bacterias, dos de los cuales codifican proteínas (el factor de elongación de la traducción EF-Tu y la proteína ribosomal S12) y dos de ellos para RNA's (RNAr 16S y 23S) (Lagesen et al., 2010). Debido a la poca conservación de genes encontrada en los análisis de los genomas, es que se ha propuesto que el genoma "core suave" para los genomas estudiados sea un concepto un poco menos rígido, donde se relaja un poco la exigencia de ubicuidad, con la finalidad de poder tener un conjunto de genes comunes a la mayoría de los genomas que se estudian (Charlebois and Doolittle, 2004). Sin embargo se ha propuesto ya anteriormente que la calidad de las secuencias genómicas a la hora de realizar análisis de genómica comparativa puede afectar seriamente los resultados obtenidos y subestimar la conservación de los genes (Charlebois and Doolittle, 2004). También se ha reportado que la variabilidad de los genes puede ser resultado de la mala calidad de las secuencias genómicas y de anotación (Cook and Ussery, 2013).

Además de los estudios que reportan una muy baja conservación de genes en los genomas bacterianos incluso entre cepas de una misma especie, como se menciona en el texto, existen algunos reportes de alta variabilidad de los genes esenciales (Ibberson et al., 2017; Juhas et al., 2011; Skurnik et al., 2013; Turner et al., 2015), que al ser esenciales debieran estar conservados y por definición deben estar en el genoma "core". Sin embargo, la baja conservación de los genes esenciales en varias especies bacterianas ha conducido a la creación del concepto de desplazamiento de genes no ortólogos (DGNO), que permite explicar la variabilidad de los genes esenciales en los casos en que una secuencia genética es reemplazada por un gen funcional y equivalente pero sin homología

en la secuencia de DNA y que además difiere en su origen desde el punto de vista evolutivo (Charlebois and Doolittle, 2004).

La importancia del fenómeno del DGNO para la evolución de las bacterias puede explicarse y es congruente con los postulados del modelo probabilístico de genética de poblaciones del “banco de semillas” propuesto para explicar el origen de una enorme poza génica a la que pueden acceder los genomas bacterianos además de la alta frecuencia de THG que se observa en procariontes. El modelo probabilístico postula que los organismos que forman estructuras como esporas o quistes y que les permiten a las células sobrevivir por largos periodos de tiempo, que son mucho mayores que los tiempos de generación de esos organismos, dan lugar a una poza génica que logra evadir a las fuerzas evolutivas como es la selección natural y también evitan la extinción. Es claro que en bacterias se presentan fenómenos de esporulación y enquistamiento (Hutchison et al., 2014), por lo tanto pueden ser consideradas como formas durmientes, además de que los bacteriófagos contienen fragmentos genómicos de bacterias (Lang et al., 2012; Penadés et al., 2015) y pueden mantenerlos a salvo de las fuerzas evolutivas por periodos de tiempo muchas veces mayor que los tiempos de generación de las bacterias a las que infectan (Koskella and Brockhurst, 2014; Lang et al., 2012; Penadés et al., 2015; Shoemaker and Lennon, 2018).

Además de definir las diversas propiedades de los genomas bacterianos y de tratar de entender su evolución, su funcionamiento y su estructura, también se han tratado de definir los genes esenciales de cada especie a través de métodos experimentales (Luo et al., 2014). La búsqueda de bacterias con genomas reducidos ha sido también un campo de gran interés que ha intentado determinar el conjunto de genes mínimo necesario para la vida bacteriana con aplicaciones en diversas áreas de la biología, como son la farmacología, la biotecnología, la biología sintética, entre otras. El aislamiento de células bacterianas con genomas reducidos ha sido el primer enfoque utilizado para tratar de determinar los genes esenciales, uno de los ejemplos de bacterias con genomas de pequeño tamaño

son las especies del género *Mycoplasma* (Hutchison et al., 1999), entre ellas se encuentra *M. mycoides*, que aunque naturalmente contiene 828 ORFs, por biología sintética se ha logrado reducir su genoma a menos de 400 genes. Sin embargo, el genoma mínimo de *M. mycoides* no representa al genoma mínimo necesario en bacterias para desarrollarse naturalmente, ya que varios de los genes esenciales en *M. mycoides* son exclusivos de esta célula desarrollada sintéticamente (Hutchison et al., 2016; Juhas et al., 2011, 2012).

Con el objetivo de encontrar los genes mínimos necesarios para la vida que sean representativos de las bacterias, se han tratado de determinar los genes esenciales en una variedad de géneros y especies bacterianas y analizar la variabilidad de estos genes entre las distintas especies.

La única manera de definir que un gen sea esencial es mediante métodos experimentales, de modo que no se encuentran mutantes en ese gen. Una vez que se define de esta manera, se puede buscar la secuencia del gen en los genomas bacterianos secuenciados y disponibles, esta es la manera en que se generan las bases de datos y se construyen los bancos.

I.I.IV. Genes esenciales en bacterias

Los genes en general pueden ser clasificados como esenciales o no esenciales, dependiendo de su efecto biológico en el organismo (Jordan et al., 2002), de tal manera que un gen es considerado esencial si su delección lleva a un fenotipo letal y se considera que un gen no es esencial si aún con la delección del gen, el fenotipo del organismo es viable. Así pues, los genes esenciales son aquellos que juegan un papel decisivo en la sobrevivencia y desarrollo de un organismo bajo condiciones generales (Koonin, 2000) o condiciones donde la mayoría de los nutrientes requeridos son provistos en el medio de crecimiento y a una temperatura que permita el mejor desarrollo del organismo (Juhas et al., 2014; Peng et al., 2017; Peng and Gao, 2014).

La definición de un gen esencial es dependiente del contexto ambiental, ya que sólo cuando las condiciones ambientales en que se desarrolla el microorganismo son claramente definidas se puede clasificar a los genes como esenciales o no esenciales (Gil et al., 2004; Zhang et al., 2015).

La determinación de los genes esenciales puede acercarnos a conocer los requerimientos básicos para el desarrollo de la vida, además contribuye en gran medida al entendimiento fundamental de los procesos biológicos necesarios en la célula que puede ser aprovechado por diversas áreas, como la biología sintética o la farmacología. En el último caso, el conocimiento y comprensión de los genes esenciales en bacterias, puede abrir un enorme potencial de nuevos blancos terapéuticos para combatir bacterias patógenas (Juhas et al., 2011, 2012).

Los marcos de lectura abiertos (ORFs, por sus siglas en inglés) pueden ser clasificados en grupos de genes ortólogos (COGs) que pueden pertenecer a diferentes categorías funcionales; para proteínas que presentan dominios múltiples puede ocurrir que sean clasificadas en uno o más COGs. De esta manera, los genes pueden ser clasificados por su probable función en diferentes categorías. Los genes esenciales en bacterias generalmente se clasifican en grupos con funciones especializadas y con un alto grado de importancia para la vida de la célula (Juhas et al., 2011, 2012, 2014). De manera general, los genes esenciales participan en funciones de almacenamiento y procesamiento de la información genética (replicación, transcripción y traducción), procesos celulares como división celular, desarrollo de la envoltura celular, motilidad celular y secreción, transporte de iones y metabolismo, transducción de señales o en el metabolismo en general, que puede involucrar a la producción y conversión de energía, metabolismo de carbohidratos, ácidos nucleicos, ácidos grasos, entre otros. Es importante señalar, que en los diversos estudios de genes esenciales en bacterias, una gran cantidad de los genes asignados como esenciales en varias especies, no tienen una función asignada o son muy pobremente caracterizados (Baba et al., 2006; Gerdes et al., 2003; Goodall et al., 2018; Lee et al., 2015;

Skurnik et al., 2013). En *E. coli*, la gran mayoría de los genes esenciales pertenecen a COGs involucrados en división celular, metabolismo de lípidos, traducción, transcripción y biogénesis de la envoltura celular (Baba et al., 2006).

I.I.V. Métodos para la identificación de genes esenciales

Si un gen es esencial o no, como se ha mencionado, está basado en la posibilidad de aislar mutantes, ser eliminado mediante delección de la secuencia o si el gen puede o no ser silenciado (Ish-Am et al., 2015; Rusmini et al., 2014). Un ejemplo de las estrategias utilizadas son la mutagénesis por transposones, que fue realizada por primera vez en *Mycoplasma genitalium* (Hutchison et al., 1999), que es una de las bacterias que se pueden cultivar in vitro con uno de los genomas más pequeños. Otro ejemplo es la mutación por delección de genes únicos (Baba et al., 2006; Kobayashi et al., 2003), con técnicas moleculares que permitían la delección de un gen en específico sin afectar por polaridad a los genes vecinos (Datsenko and Wanner, 2000) y el silenciamiento o inhibición por RNA antisentido (Ji et al., 2001).

Con el avance de técnicas de biología molecular y la tecnología de secuenciación mejorando día a día se ha desarrollado la genómica comparativa, que permite proponer la esencialidad de los genes cada vez en más y más especies bacterianas usando la secuencia de los genes que se han definido como esenciales en una especie (Peng et al., 2017).

La combinación de dos estrategias de análisis, la mutagénesis por transposones y la secuenciación de alto rendimiento ha sido utilizada también ampliamente en el análisis de los genes esenciales en una gran variedad de especies bacterianas (Chao et al., 2016; Judson and Mekalanos, 2000), con lo que actualmente ya se cuenta con bases de datos de los genes esenciales, definidos por distintas

técnicas de inactivación, en muchas especies bacterianas (Tabla 1) lo que hace posible que las bases de datos de genes esenciales puedan ser comparadas (Luo et al., 2014).

Tabla 1 | Número de genes esenciales identificados en diferentes especies bacterianas por diferentes métodos. Modificada de Juhas, M. et al., 2012.

Organismo	Número de genes esenciales codificantes de proteínas	Método de identificación	Ref.
<i>Haemophilus influenzae</i>	642	Mutagénesis por saturación con transposones	(Mekalanos et al., 2002)
<i>Acinetobacter baylyi</i>	499	Mutagénesis específica	(de Berardinis et al., 2008)
<i>Mycoplasma pulmonis</i>	461	Mutagénesis por saturación con transposones	(French et al., 2008)
<i>Mycoplasma genitalium</i>	381	Mutagénesis por saturación con transposones	(Glass et al., 2006)
<i>Salmonella enterica Typhi</i>	356	Mutagénesis por saturación con transposones	(Langridge et al., 2009)
<i>Pseudomonas aeruginosa</i>	335	Mutagénesis por saturación con transposones	(Liberati et al., 2006)
<i>Helicobacter pylori</i>	323	Mutagénesis por saturación con transposones	(Salama et al., 2004)
<i>Escherichia coli</i>	303	Mutagénesis específica	(Baba et al., 2006)
<i>Staphylococcus aureus</i>	302	Inhibición por RNA antisentido	(Ji et al., 2001)
<i>Bacillus subtilis</i>	271	Mutagénesis específica	(Kobayashi et al., 2003)

I.I.V.I Mutagénesis por inserción de transposones.

Los transposones han sido la base para el desarrollo de técnicas que han aportado enormemente para definir los genes esenciales en el genoma de cualquier especie bacteriana, en este momento hay varios enfoques derivados disponibles.

Los transposones son segmentos de DNA que pueden moverse o transposicionarse de una localización en el genoma a otra, los más simples segmentos de DNA flanqueados por secuencias frecuentemente repetidas e invertidas que son reconocidas por una proteína denominada transposasa, misma que le permite insertarse en una nueva secuencia de DNA. Las localizaciones en el genoma donde un transposón puede insertarse dependen de las secuencias en que la transposasa reconoce y escinde, sin embargo, se busca usar transposones con una baja selectividad de la secuencia en la que se insertan, con la finalidad de obtener la mayor cantidad posible de transposones insertos en el genoma. De este modo, la mutagénesis por transposición resulta en la interrupción de la región del genoma donde el transposón se ha insertado. De tal manera que si el transposón se inserta en un gen o un ORF predicho y la cepa bacteriana resultante es capaz de sobrevivir y formar una colonia, por definición ese es un gen no esencial para la viabilidad de esa especie en las condiciones en que se hicieron los ensayos mencionados (Chao et al., 2016; Judson and Mekalanos, 2000). Sin embargo, se debe tener mucho cuidado con la posición en el marco de lectura donde se ha insertado el transposón, pues se puede tener aún la posibilidad de tener una proteína funcional si el transposón se ha insertado en uno de los extremos 5' o 3' del marco de lectura. Se han diseñado también transposones que tienen promotores que permiten la expresión de los genes corriente abajo y así se evita la posibilidad de que el gen esencial sea uno que esté codificado por abajo del sitio en el que se insertó el transposón, pero su expresión quede inactivada por un fenómeno de polaridad.

Hay dos enfoques con transposones comúnmente utilizados para definir genes esenciales; en los enfoques “negativos” se identifican los genes o regiones genómicas que no son esenciales, infiriéndose así que las regiones restantes son esenciales. El otro enfoque, al que se ha denominado “positivo” consiste en identificar los genes que son esenciales generando fenotipos letales para la célula, para lograr esto, se reemplaza el promotor natural del gen con un promotor inducible, de tal manera que si el gen es esencial, la bacteria dependerá del inductor para la sobrevivencia de la misma. Cada uno de los métodos tiene ventajas y desventajas, como en el enfoque negativo, donde las regiones o genes sin inserciones pueden aún ser no esenciales, lo que lleva obligatoriamente a la demostración experimental de la no esencialidad de dicho gen o región. En el enfoque positivo puede ocurrir que un promotor inducible no provea suficiente expresión como para superar el defecto creado al eliminar el promotor natural o la expresión basal del promotor inducible puede ser suficientemente alta para lograr identificar genes esenciales en los que se requiere una cantidad mínima de expresión (Hutchison et al., 1999; Judson and Mekalanos, 2000).

Existen ahora métodos mejor elaborados gracias a la mejora en la calidad y abaratamiento de la secuenciación, estos métodos se denominan inserción de transposones y secuenciación (TIS, por sus siglas en inglés), que es el producto más reciente de los análisis genómicos basados en transposones y de las que hay algunas variables como son: Tn-Seq (inserción de transposones y secuenciación en el sitio), INSeq (inserción y secuenciación), HITS (inserción de alto rendimiento y rastreo por secuenciación profunda) y TraDIS (inserción dirigida por transposones y secuenciación de sitio. Sin embargo todas estas técnicas utilizan los mismos fundamentos básicos (Chao et al., 2016). Con estas estrategias recientemente se han caracterizado los genes esenciales de muchas especies bacterianas y debido a la saturación de la mutagénesis por transposones son los que mejores resultados genera, pues se minimizan los falsos positivos (Goodall et al., 2018; Lee et al., 2015; Luo et al., 2014; Peng et al., 2017).

I.I.V.II Silenciamiento de genes con RNA antisentido

Otro de los enfoques experimentales que se han utilizado para definir los genes esenciales en bacterias es el uso de RNAs antisentido al impedir la expresión de los genes (Ji et al., 2001; Peng et al., 2017). La inhibición de los genes por RNAs antisentido reduce el nivel de expresión de los genes blanco a través de la unión por RNA de doble hebra. También hay varios trabajos que emplean esta técnica en varias especies bacterianas (Ji et al., 2001; Rusmini et al., 2014). Sin embargo esta técnica también tiene limitaciones, entre las que se encuentran: el uso de RNAs antisentido está limitado a los genes para los que se puede obtener una expresión adecuada del RNA inhibitorio en ese organismo (Gil et al., 2004). Recientemente se ha utilizado este enfoque para analizar detalladamente genes esenciales que se encuentran dentro de operones (Goh et al., 2015), de esta manera se intentó silenciar genes esenciales que se encuentran dentro de un mismo operón y que parecen no estar regulándose entre ellos y de esta manera evaluar la esencialidad de los mismos. La presente metodología tiene el inconveniente de que un gen aparentemente silenciado con RNAs antisentido podría siempre expresarse en mínimas cantidades, lo que podría arrojar resultados falsos negativos (Peng et al., 2017).

I.I.V.III Mutagénesis específica de genes

Un tercer enfoque experimental utilizado para la identificación de los genes esenciales es la inactivación sistemática de cada gen individual presente en el genoma. Los genes que no pueden ser inactivados con esta estrategia son considerados esenciales. Una desventaja de esta técnica es que se requiere de

una anotación detallada del genoma del organismo a analizar. El enfoque de mutagénesis específica también se ha utilizado en varias especies bacterianas, entre ellas la interrupción a escala del genoma de los genes de *Bacillus subtilis*, que se llevó a cabo inactivando cada uno de los genes (Kobayashi et al., 2003). Una eficiente técnica que permitió fácilmente realizar la inactivación cromosomal de cada uno de los genes en el genoma *E. coli* fue reportada en Japón y utiliza un plásmido de bajo número de copias que expresa la recombinasa lambda Red, además de genes de resistencia a antibióticos flanqueados por regiones homólogas adyacentes al gen objetivo de la delección obtenidos a partir de reacciones en cadena de la polimerasa (PCR), (Datsenko and Wanner, 2000). Con el desarrollo de esta técnica se logró obtener la delección completa para la mayor parte de los genes de una cepa derivada de *E. coli* K-12 (BW25113). De un total de 4288 genes, se obtuvieron mutantes en 3985, conjunto de mutantes a la que se denominó colección de Keio (Baba et al., 2006). Con este método (Baba et al., 2006; Datsenko and Wanner, 2000) se tenía la capacidad de eliminar el cassette de resistencia a kanamicina que reemplazaba al gen cromosomal seleccionado y obtener una delección de toda la región en fase de lectura que impide tener efectos polares y generándose una cicatriz con ausencia de marco de lectura en esa localización, el cassette de resistencia a antibióticos permite seleccionar fácilmente a las clonas con la delección en el gen específico. Esta estrategia impedía tener interrupciones incompletas de los genes, como en el caso de la saturación con transposones de esa época, que era mucho menos saturada que la de ahora.

I.I.VI. Análisis de los genes esenciales en E. coli

E. coli es una enterobacteria Gram negativa que pertenece a las gammaproteobacterias, habitualmente encontrada en el tracto intestinal de

animales y humanos, algunas cepas tienen implicaciones patológicas en humanos. Es una bacteria que ha sido muy utilizada en estudios de bioquímica, fisiología, genética, biología molecular y biotecnología por un gran número de grupos de investigación; gran parte de los mecanismos moleculares que se conocen son resultado de la investigación que se ha realizado con ella (Gopal and Kumar, 2013; Jang et al., 2017). Por la facilidad con la que es posible manipular genéticamente a esta bacteria y por muchas otras características intrínsecas a ella se la considera un organismo modelo. Desde que se conoce el genoma de *E. coli* se han reportado una gran variedad de trabajos que aportan información genómica en cada vez más cepas de la especie (Gerdes et al., 2003; Martínez-Carranza et al., 2018a; Moriel et al., 2016; Rasko et al., 2008; Rouli et al., 2015).

En *E. coli*, se ha tratado de determinar el conjunto de genes esenciales de la especie por varias de las estrategias mencionadas anteriormente; en el primer análisis se evaluó la esencialidad de 3,746 genes codificantes de proteínas de un total de 4,291 marcos de lectura (ORFs), se propusieron 620 genes como esenciales y 3,126 genes fueron considerados dispensables en las condiciones de cultivo utilizadas: crecimiento aéreo en medio rico basado en triptona (Gerdes et al., 2003). Un estudio de los genes esenciales de *E. coli* BW25113 basado en deleciones específicas en cada marco de lectura arrojó un número significativamente menor, que consistía en 303 genes esenciales en total (Baba et al., 2006), que se correlacionaba con la especificidad de cada una de las deleciones en esta estrategia de análisis, siendo más específico en los genes que se mutaban. Además en este estudio se realizaron deleciones en todos los marcos de lectura y no sólo en una parte del cromosoma en donde se insertaron los transposones, como en el estudio anterior. En otro estudio donde se realizaron deleciones de gran magnitud en el cromosoma, se identificaron 35 genes como no esenciales, mismos que en el estudio anterior habían sido considerados esenciales, por lo tanto ellos proponen un total de 268 genes como esenciales (Kato and Hashimoto, 2007) y atribuyen las diferencias a las cepas utilizadas en

cada estudio y a las condiciones de cultivo. Además suponen que la diferencia en la estrategia de análisis también contribuye a las diferencias en los resultados, por ejemplo, en este último estudio, encontraron que genes que pertenecen a sistemas de toxina-antitoxina como *yefM* y *chpR* y que anteriormente fueron considerados como esenciales por Baba et al., no son esenciales en su estudio, ya que logran ser deletados en conjunto con otros genes que pertenecen a sistemas toxina-antitoxina, por lo que los clasifican como no esenciales en su estudio.

Otro estudio que emplea RNAs antisentido para inhibir la expresión de los genes logra identificar 152 genes esenciales (Meng et al., 2012), el número significativamente menor que los estudios anteriores puede deberse a que con esta estrategia de análisis aún puede haber expresión de los genes inhibidos cuando es comparado con las estrategias de mutagénesis. En un análisis por saturación con transposones en *E. coli* K-12 cepa BW25113, se determinaron 358 genes como esenciales (Goodall et al., 2018), de los cuales 248 se comparten con los genes esenciales propuestos por Baba et al., en este estudio se obtuvieron además 47 genes esenciales propios de este estudio (Figura 1). Como puede observarse, de los conjuntos de genes esenciales propuestos en los diferentes trabajos con diferentes enfoques, sólo se comparten 164 genes.

muy valiosa para la comprensión del genoma y de sus características (Klockgether et al., 2011; Mosquera-Rendón et al., 2016; Rouli et al., 2015; Valot et al., 2015). Ahora se tiene una gran colección de cepas secuenciadas y esta información está disponible en las bases de datos (www.pseudomonas.com). La gran cantidad de información genómica ha llevado a realizar análisis de genómica comparativa para tratar de establecer el conjunto de genes comunes a todos los genomas disponibles, así como describir los genes esenciales (Luo et al., 2015). Varios trabajos que se han enfocado en determinar los genes esenciales de *P. aeruginosa* a través de diferentes enfoques experimentales y con diferentes condiciones metabólicas han concluido que los genes esenciales son muy variables entre las dos cepas mejor estudiadas de esta especie, la cepa PAO1 y la cepa PA14 (Jacobs et al., 2003; Lee et al., 2015; Liberati et al., 2006; Rusmini et al., 2014; Skurnik et al., 2013; Turner et al., 2015). El primer genoma de *P. aeruginosa* disponible fue el de la cepa PAO1, así también el primer análisis de los genes esenciales de esta especie fue realizado en el genoma de esta cepa, donde se estimó un conjunto de entre 300 y 400 genes esenciales basándose en el análisis de mutagénesis por inserción de transposones a lo largo del cromosoma, en las condiciones de cultivo establecidas (medio rico y 37°C). En este estudio no se recuperaron inserciones de transposones en un total de 678 marcos de lectura, sin embargo en este estudio no tuvo una alta saturación y se cree que algunos genes pudieron no haber sido mutagenizados (Jacobs et al., 2003). En un estudio posterior de mutagénesis por transposones no redundante en la cepa PA14 se consideró un conjunto de 335 genes esenciales en medio rico (Liberati et al., 2006). Posteriormente se analizó nuevamente la posibilidad de obtener los genes esenciales de la cepa PA14, cambiando las condiciones de cultivo; en esta ocasión se propusieron 636 genes esenciales en ambientes sistémicos y de la mucosa en un modelo de infección tanto *in vivo* como *in vitro*. De este conjunto de 636 genes esenciales, se reportó a 210 genes esenciales como ausentes en el genoma de la cepa PAO1, este reporte respaldaba la alta variabilidad de los genes esenciales entre miembros de una misma especie bacteriana (Skurnik et al.,

2013). Un estudio adicional en la cepa PAO1 realizado con mutagénesis por transposones en un medio que emula las condiciones de infección en pulmón de pacientes con fibrosis quística, identificó a 336 genes como esenciales en esta cepa y en estas condiciones de cultivo además analizó los genes esenciales en las mismas condiciones de cultivo en la cepa PA14 y reportó que 434 genes son esenciales, 20 de estos 434 genes fueron reportados sin ortólogo en la cepa PAO1. Esto lo que apoya también la alta variabilidad de los genes esenciales entre cepas de la misma especie, sin embargo, en esta ocasión en menor proporción al total de genes esenciales reportados (Turner et al., 2015). Los trabajos antes mencionados han identificado genes esenciales en una condición de cultivo específica, sin embargo un estudio que utiliza mutagénesis por transposones en la cepa PAO1 utilizó un total de 6 condiciones de cultivo para determinar cuáles son los genes esenciales específicos de cada condición de cultivo y los genes esenciales generales a todas las condiciones de cultivo probadas, utilizando medios mínimos, medios ricos y medios de cultivo que emulan las condiciones de infección en pulmón (esputo de paciente con fibrosis quística); ellos reportaron 352 genes esenciales generales y 199 genes esenciales específicos de las condiciones de cultivo (Lee et al., 2015).

En el caso de *P. aeruginosa* es difícil explicar la variabilidad de los genes esenciales debido al grado de conservación de los genomas que presentan entre las cepas que han sido secuenciadas (Klockgether et al., 2011; Valot et al., 2015). Sin embargo, de manera general se ha reportado que la variabilidad de los genes esenciales puede estar dada por el fenómeno de DGNO que se ha mencionado anteriormente (Acevedo-Rocha et al., 2013; Juhas et al., 2014; Koonin et al., 1997). Sin embargo también se ha propuesto que la variabilidad observada puede deberse a errores de secuenciación y de anotación de los genomas (Charlebois and Doolittle, 2004; Juhas et al., 2014).

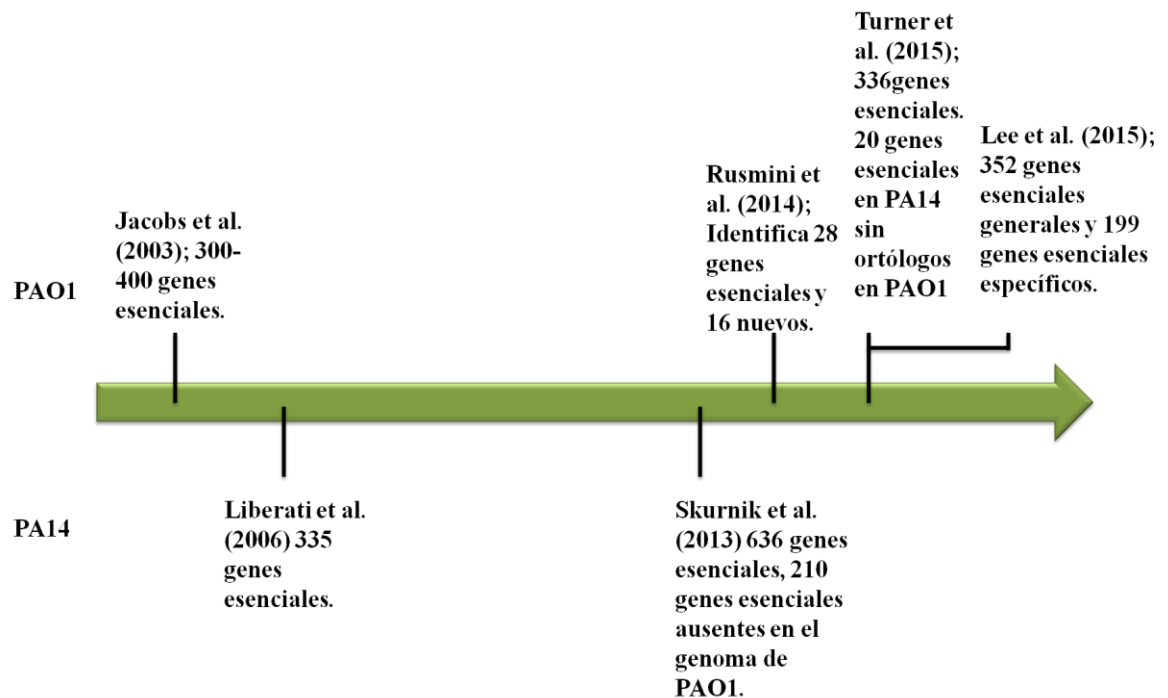


Figura 3 | Línea de tiempo que representa los análisis sobre genes esenciales en *P. aeruginosa*, cepas PAO1 y PA14.

II.- JUSTIFICACIÓN

Se ha reportado una alta variabilidad de los genes esenciales entre cepas de una misma especie bacteriana o la ausencia de genes esenciales (desplazamiento de genes no ortólogos, DGNO). Se han documentado varios casos de genomas como el de *Azotobacter vinelandii* donde los genes que definen a la especie tienen orígenes polifiléticos.

Por esta razón nos propusimos analizar los genes esenciales de *Escherichia coli* y *Pseudomonas aeruginosa*, dos especies bacterianas que cuentan con varios análisis de sus genes esenciales con la finalidad de encontrar ejemplos de «non-orthologous gene displacement» y estudiar su variabilidad.

Adicionalmente comparamos el repertorio de genes esenciales de *P. aeruginosa* con el de *Azotobacter vinelandii* para sustentar la hipótesis de que *A. vinelandii* tiene un genoma con un esqueleto que comparte con *Pseudomonas*.

III.- HIPÓTESIS

- En el genoma de *Escherichia coli* hay genes esenciales para su biología que no fueron heredados por ancestría clásica y se podrían encontrar ejemplos de desplazamiento de genes no ortólogos (non-orthologous gene displacement).
- Los genes esenciales de *Pseudomonas aeruginosa* no son tan variables como ha sido reportado.
- *Azotobacter vinelandii* tiene un esqueleto genómico que comparte con *P. aeruginosa*, al que se han adicionado genes por THG y son fundamentales para su biología.

IV.- OBJETIVOS

IV.I.- OBJETIVO GENERAL

Estudiar la variabilidad de los genes esenciales en bacterias haciendo énfasis en *E. coli* y *P. aeruginosa*.

IV.II.- OBJETIVOS ESPECÍFICOS

1. Analizar la variabilidad de los genes esenciales (Baba y col., 2006) de *E. coli* para buscar ausencia de estos en genomas de distintas cepas.
2. Buscar los genes que realizan la misma función esencial con secuencia distinta (DGNO) y evaluar la esencialidad de los genes ausentes.
3. Analizar la variabilidad de los genes esenciales de *P. aeruginosa* PAO1 y PA14.
4. Identificar los genes esenciales de *P. aeruginosa* en *A. vinelandii* para evaluar su conservación en ambos genomas.

V.- MATERIALES Y MÉTODOS

V.I.- Estrategia de búsqueda de los genes esenciales de E. coli K-12 en otros 63 genomas de cepas de E. coli

Los genes esenciales propuestos por Baba et al., fueron utilizados para obtener secuencias de proteínas del genoma de *E. coli* K-12 y fueron utilizadas como secuencias a buscar, de esta manera, realizando alineamientos locales contra los genomas completos de cada una de las 63 cepas de *E. coli* utilizadas en este estudio. Estos 63 genomas son los que presentan una mejor calidad de secuencia de los genes presentes en la base de datos de NCBI. Los alineamientos locales fueron realizados con TBLASTN (Altschul et al., 1997) con un valor de corte (e-

value) de $1e^{-5}$. Las secuencias encontradas en los genomas se consideraron homólogas si tenían al menos un 70% de la longitud de la secuencia problema y al menos un 30% de identidad de aminoácidos. El perfil de abundancia de genes en los 63 genomas fue graficado en paquete estadístico R con la función heatmap.2 del paquete ggplots.

El análisis detallado de los genes que parecían ausentes en al menos un genoma de las cepas de *E. coli* fue realizado buscando en los locus correspondiente al genoma de referencia, en este caso la cepa K-12, se analizó entonces detalladamente si los genes estaban presentes o no en los genomas, todos los datos fueron obtenidos de las bases de datos del NCBI. Los genes esenciales ausentes en sólo un genoma fueron buscados sintéticamente en los genomas de *E. coli* utilizando el programa Artemis (Carver et al., 2012) y confirmado con MAUVE (Darling et al., 2004). Algunos de los genes que parecían ausentes luego del análisis con alineamientos locales y presentes al realizar la búsqueda mediante análisis sintético, fueron amplificados mediante PCR (cuando contábamos con la cepa de la que analizábamos el genoma, como es el caso de las cepas HS y CFT073), purificados y secuenciados, las secuencias resultantes fueron comparadas con BLASTN con *E. coli* K-12 y CFT073 como genomas de referencia. Las cepas y números de acceso de cada genoma de *E. coli* utilizado en el análisis se encuentra en la Tabla 2

Tabla 2 | Cepas y números de acceso de los 63 genomas de *E. coli* utilizados en la búsqueda de los 303 genes esenciales propuestos por Baba et al., 2006.

Cepa	Número de acceso
Escherichia_coli_042_uid161985	NC_017626, NC_017627
Escherichia_coli_536_uid58531	NC_008253
Escherichia_coli_55989_uid59383	NC_011748
Escherichia_coli_ABU_83972_uid161975	NC_017629, NC_017631
Escherichia_coli_APEC_O1_uid58623	NC_008563, NC_009837, NC_009838
Escherichia_coli_APEC_O78_uid187277	NC_020163
Escherichia_coli_ATCC_8739_uid58783	NC_010468

Escherichia_coli_BL21_DE3_uid161947	NC_012971
Escherichia_coli_BL21_DE3_uid161949	NC_012892
Escherichia_coli_BL21_Gold_DE3_pLysS_AG_uid59245	NC_012947
Escherichia_coli_B_REL606_uid58803	NC_012967
Escherichia_coli_BW2952_uid59391	NC_012759
Escherichia_coli_CFT073_uid57915	NC_004431
Escherichia_coli_clone_D_i14_uid162049	NC_017652
Escherichia_coli_clone_D_i2_uid162047	NC_017651
Escherichia_coli_DH1_uid161951	NC_017625
Escherichia_coli_DH1_uid162051	NC_017638
Escherichia_coli_E24377A_uid58395	NC_009786, NC_009788, NC_009790, NC_009801
Escherichia_coli_ED1a_uid59379	NC_011745, NC_009787, NC_009789, NC_009791
Escherichia_coli_ETEC_H10407_uid161993	NC_017633, NC_017721, NC_017722, NC_017723, NC_017724
Escherichia_coli_HS_uid58393	NC_009800
Escherichia_coli_IAI1_uid59377	NC_011741
Escherichia_coli_IAI39_uid59381	NC_011750
Escherichia_coli_IHE3034_uid162007	NC_017628
Escherichia_coli_JJ1886_uid226103	NC_022648, NC_022650, NC_022661
Escherichia_coli_K_12_substr_DH10B_uid58979	NC_010473, NC_022649, NC_022651, NC_022662
Escherichia_coli_K_12_substr_MDS42_uid193705	NC_020518
Escherichia_coli_K_12_substr_MG1655_uid57779	NC_000913
Escherichia_coli_K_12_substr_W3110_uid161931	NC_007779
Escherichia_coli_KO11FL_uid162099	NC_017660, NC_017661
Escherichia_coli_KO11FL_uid52593	NC_016902, NC_016903, NC_016904
Escherichia_coli_LF82_uid161965	NC_011993
Escherichia_coli_LY180_uid219461	NC_022364
Escherichia_coli_NA114_uid162139	NC_017644
Escherichia_coli_O103_H2_12009_uid41013	NC_013353, NC_013354
Escherichia_coli_O104_H4_2009EL_2050_uid175905	NC_018650, NC_018651, NC_018652, NC_018654
Escherichia_coli_O104_H4_2009EL_2071_uid176128	NC_018661, NC_018662, NC_018663
Escherichia_coli_O104_H4_2011C_3493_uid176127	NC_018658, NC_018659, NC_018660, NC_018666
Escherichia_coli_O111_H_11128_uid41023	NC_013364, NC_013366, NC_013368, NC_013365, NC_013367, NC_013370
Escherichia_coli_O127_H6_E2348_69_uid59343	NC_011601, NC_011602, NC_011603
Escherichia_coli_O157_H7_EC4115_uid59091	NC_011350, NC_011351, NC_011353
Escherichia_coli_O157_H7_EDL933_uid57831	NC_002655, NC_007414
Escherichia_coli_O157_H7_TW14359_uid59235	NC_013008, NC_013010
Escherichia_coli_O157_H7_uid57781	NC_002127, NC_002128, NC_002695
Escherichia_coli_O26_H11_11368_uid41021	NC_013361, NC_013362, NC_013363, NC_013369, NC_014543
Escherichia_coli_O55_H7_CB9615_uid46655	NC_013941, NC_013942
Escherichia_coli_O55_H7_RM12579_uid162153	NC_017653, NC_017655, NC_017657, NC_017654, NC_017656, NC_017658
Escherichia_coli_O7_K1_CE10_uid162115	NC_017646, NC_017647, NC_017648, NC_017649, NC_017650
Escherichia_coli_O83_H1_NRG_857C_uid161987	NC_017634, NC_017659
Escherichia_coli_P12b_uid162061	NC_017663
Escherichia_coli_PMV_1_uid219679	NC_022370, NC_022371

Escherichia_coli_S88_uid62979	NC_011742, NC_011747
Escherichia_coli_SE11_uid59425	NC_011407, NC_011411, NC_011415, NC_011419, NC_011408, NC_011413, NC_011416
Escherichia_coli_SE15_uid161939	NC_013654, NC_013655
Escherichia_coli_SMS_3_5_uid58919	NC_010485, NC_010486, NC_010487, NC_010488, NC_010498
Escherichia_coli_UM146_uid162043	NC_017630, NC_017632
Escherichia_coli_UMN026_uid62981	NC_011739, NC_011749, NC_011751
Escherichia_coli_UMNK88_uid161991	NC_017639, NC_017641, NC_017643, NC_017640, NC_017642, NC_017645
Escherichia_coli_UTI89_uid58541	NC_007941, NC_007946
Escherichia_coli_W_uid162011	NC_017635, NC_017636, NC_017637
Escherichia_coli_W_uid162101	NC_017662, NC_017664, NC_017665
Escherichia_coli_Xuzhou21_uid163995	NC_017903, NC_017906, NC_017907

V.II.- Material biológico y medios de cultivo

Las cepas bacterianas utilizadas en este trabajo son: la cepa de *Escherichia coli* derivada de la cepa K-12, BW25113 (*lac*^F *rrnB*_{T14} Δ *lacZ*_{WJ16} *hsdR*₅₁₄ Δ *araBAD*_{AH33} Δ *rhaBAD*_{LD78}) (Datsenko and Wanner, 2000), CFT073 (aislado silvestre, UPEC) (Welch et al., 2002) y HS (aislado comensal silvestre) (Rasko et al., 2008). Las cepas fueron generalmente crecidas a 30° o a 37°C, sobre medio sólido o líquido Luria Bertani (LB) con agitación a 225 rpm, cuando se requería se suplementaron con los siguientes antibióticos: apramicina (Apra) a una concentración de 60 µg/ml, cloranfenicol (Cm) a una concentración de 25 µg/ml, tetraciclina (Tc) a una concentración de 15 µg/ml.

V.III.- Técnicas de manipulación de DNA

El DNA cromosomal se extrajo con el paquete de reactivos de *DNeasy Blood & Tissue* (QIAGEN). Las enzimas de restricción y ligación de DNA se obtuvieron de New England Biolabs. Las reacciones en cadena de la polimerasa (PCR) se realizaron con la enzima DNA polimerasa de alta fidelidad *Phusion Hot Start II* (Thermo SCIENTIFIC). Los productos de PCR y los fragmentos de DNA de geles de agarosa se purificaron con el *PCR purification kit* (Jena Bioscience). Los plásmidos se trabajaron en las cepas BW25113 y CFT073 y se extrajeron utilizando el paquete de *Fast-n-Easy Plasmid Mini-Prep*. La síntesis de oligonucleótidos y secuenciación de DNA se llevó a cabo en la unidad de síntesis y secuenciación de DNA (USSDNA) del Instituto de Biotecnología, UNAM.

Tabla 3 | Oligonucleótidos utilizados en este trabajo

Nombre	Secuencia 5´- 3´
bcsADEFw	GAGTATCCTGACCCGGTGGTTGCTTATCCCGCCGGTCAAGGTTAGCTCCTTCGGTCCTC
bcsADERv	CATTGTTGAGCCAAAGCCTGATCCGATGGTTGTGCCGTTTGCTTCCTCGCTCACTGACTC
bcsBDEFw	GAAAAGAAAAC TATTCTGGATTGTGCAGTGGCTATGGGTATTCCGGGGATCCGTCGAC
bcsBDERv	CTCGTTATCCGGGTTAAGACGACGACGACTAATAATTCGCTGTAGGCTGGAGCTGCTTC
bcsBDEFw	GGCACAACCATCGGATCAGGCTTTGGCTCAACAATGATGATAACGCGATGTATTCCGGGGATCCGTCGAC
bcsBDEKRv	ATTCCACTACGCAACACATTCATCTTCAATTACTCGTTATCCGGGTTAAGTGTAGGCTGGAGCTGCTTC
bcsHSFw	AAAGAACTGGAGATGGGCCG
bcsHSRv	CGCAATGAGGAAGATTGCCG
bcsAcIFw	GGAATCCCTGGCGGCTGAAGAAACTG

bcsAclRv	GCTCTAGAGAGCCACTGCACAAATCCAG
bcsBclFw	GGAATTCCGTATCGTTTATTCCGCGCCG
bcsBclRv	GCTCTAGACATCGTCACGAGTCCACTAC
bcsBchkFw	CTATTCTGGATTTGTGCAGTG
bcsBchkRv	GACGACGACGACTAATAATTC

Las letras en negrita en los primeros seis oligonucleótidos (a-f) representan las secuencias del cebador utilizadas para amplificar por PCR a los casetes de resistencia a antibióticos en cada caso (a-b, tetraciclina y c-f, apramicina), el resto de la secuencia es la región de recombinación del cebador. Los oligonucleótidos g) y h) fueron utilizados para amplificar la región de la delección del operón *bcs* en la cepa HS y los oligonucleótidos i) a l) fueron utilizados para la clonación de los genes *bcsA* y *bcsB*, m) y n) se utilizaron para amplificar el gen *bcsB* clonado en el plásmido pUCP20.

V.IV.- Construcción de plásmidos

Los fragmentos de DNA de 2749 pares de bases y 5091 pares de bases correspondientes a los genes *bcsA* y *bcsAB* sin sus correspondientes promotores, respectivamente, y flanqueados por sitios de restricción EcoRI y XbaI, fueron amplificados, digeridos y ligados en el plásmido pUCP20 (West et al., 1996) previamente digerido con las enzimas correspondientes, resultando en la construcción de los vectores pUCP20-*bcsA* o pUCP-*bcsAB*, con los genes *bcsA* o *bcsAB* bajo el control del promotor *lac*. Para la expresión de los genes bajo el control del promotor *lac* se utilizó el inductor isopropil- β -D-1-tiogalactopiranosido (IPTG) a una concentración de 0.5 mM. Los plásmidos utilizados en este trabajo se pueden consultar en el anexo.

V.V.- Selección de cepas mutantes en *bcsB* y *bcsA* en *E. coli* BW25113 y CFT073

Los oligonucleótidos utilizados para construir las cepas mutantes por delección, tienen una región con extensión de 39 o 40 nucleótidos en el extremo 5´ idéntica a la región 3´ o 5´ al gen blanco y una región de 20 nucleótidos en la extensión 3´ que se utilizó para amplificar mediante PCR, el cassette de resistencia a los antibióticos seleccionados. La cepa con la delección cromosomal $\Delta bcsB$ fue generada con oligonucleótidos con una extensión total de 70 nucleótidos donde la región 5´ tiene una extensión de 50 nucleótidos homólogos al gen *bcsB*. Como producto de PCR se tenía el cassette de resistencia al antibiótico flanqueado por secuencias homólogas que promueven la recombinación específica con el gen blanco de la delección, para así poder reemplazar el gen con el casete de resistencia al antibiótico (Datsenko and Wanner, 2000).

Para llevar a cabo la delección de los genes *bcsA* y *bcsB*, las cepas BW25113 y CFT073 que ya llevaban el plásmido pIJ790, un derivado del plásmido pKD20 (Gust et al., 2000) con resistencia a Cm y que expresa el sistema de recombinasa lambda Red, se crecieron en medio líquido en presencia de este antibiótico y una concentración de 1mM de L-arabinosa a 30°C y hasta alcanzar una densidad óptica (D.O.₆₀₀) de 0.5 y luego concentradas por centrifugación y lavadas tres veces con agua fría con una concentración de glicerol al 10%. Los productos de PCR fueron purificados en geles de agarosa al 1% y suspendidos en buffer de elución.

Las células electrocompetentes fueron transformadas con 1 µg de producto purificado de PCR. Las células electroporadas fueron resuspendidas en 1 ml de medio LB con una concentración de L-arabinosa 1 mM, después incubadas 2 horas a 30°C y luego concentradas por centrifugación y dispersadas en cajas de

Petri con medio sólido LB para seleccionar las transformantes con resistencia a Tc o Apra, según correspondía, en substitución del gen seleccionado. Las deleciones se confirmaron por medio de PCR, utilizando DNA total y esperando productos de PCR del tamaño del casete de resistencia al antibiótico utilizado en cada caso, en lugar de productos que representen el tamaño completo del gen a substituir. La doble deleción de la cepa CFT073 se realizó aplicando en dos veces consecutivas la misma estrategia, de tal manera que primero se realizó la deleción del gen *bcsA* y sobre ésta la deleción del gen *bcsB*. El número de mutantes reportadas en este trabajo es equivalente a la media \pm desviación estándar del número de colonias obtenidas en tres experimentos independientes.

V.VI.- Evaluación de la producción de celulosa

Las células de *E. coli* se crecieron toda la noche en medio líquido LB bajo aireación a 225 rpm y a 37°C, posteriormente 5 μ l de este cultivo fueron depositados sobre medio sólido LB sin sal (NaCl) y adicionado con los colorantes Rojo Congo (40 μ g/ml) y azul de Coomassie (20 μ g/ml) en cajas de Petri y una vez que las gotas estaban secas, posteriormente fueron incubados a 28°C por 7 días, los experimentos individuales se realizaron por triplicado (Serra et al., 2013).

V.VII.- Análisis de los ortólogos de los genes esenciales de P. aeruginosa en las cepas PAO1, PA14 y en A. vinelandii

Los datos que se utilizaron para realizar los análisis de los genes esenciales de *P. aeruginosa* fueron obtenidos del NCBI, los proteomas fueron los siguientes: *Pseudomonas aeruginosa* PAO1 NC_002516, *P. aeruginosa* PA14 NC_008463, y *Azotobacter vinelandii* DJ CP001157. Los conjuntos de secuencias consultadas en los proteomas anteriores fueron a) los 210 genes esenciales predichos en *P. aeruginosa* PA14, que se reportó que estaban ausentes en el genoma de la cepa PAO1 (Skurnik et al., 2013); b) el conjunto de genes esenciales reportados en la cepa PAO1 y comparados con los de la PA14 para crecer en esputo de pacientes con fibrosis quística (Turner et al., 2015); c) el conjunto de 352 genes esenciales generales de *P. aeruginosa* PAO1 (Lee et al., 2015), estos 352 también fueron buscados en el genoma de *A. vinelandii* DJ.

Los conjuntos de genes esenciales en *P. aeruginosa* PAO1 y PA14 propuestos por los tres trabajos fueron agrupados, traducidos a proteínas y utilizados para realizar las búsquedas con BLASTP versión 2.2.26+, utilizando los proteomas completos de las cepas PAO1, PA14 y/o de *A. vinelandii* como bases de datos en cada caso, con un valor de corte (e-value) de 1e-5. En éste caso se consideró que una secuencia era homóloga cuando si tenía al menos un 90% de la longitud de la secuencia consultada y al menos un 80% de identidad en la secuencia de aminoácidos (Altschul et al., 1997). La búsqueda de cada proteína fue realizada utilizando la secuencia proteica codificada por cada gen esencial propuesto en el proteoma de la otra cepa. Es decir, que los genes esenciales propuestos en la cepa PAO1 se buscaron en la cepa PA14 y viceversa.

En cada una de las búsquedas, la mejor cobertura y el primer resultado se definió como candidato homólogo, se analizó la cobertura completa de cada secuencia para estar seguros de la homología. El análisis de los genes esenciales entre las dos cepas de *P. aeruginosa* y la búsqueda de los 352 genes esenciales generales de la cepa PAO1 en *A. vinelandii* también se complementó con un análisis sinténico para cada uno de los genes individualmente, utilizando el programa MAUVE versión snapshot_2015-02-13 (<http://darlinglab.org/mauve>) (Darling et al., 2004). La comparación de los genes esenciales reportados en diferentes estudios

en las cepas PAO1 y PA14 se realizó utilizando el paquete upsetR (Figura 11) (Lex et al., 2014). La clasificación de los genes esenciales compartidos por los diferentes estudios se realizó de acuerdo a la clasificación de COGs (Tatusov et al., 1997).

VI. RESULTADOS

Con la finalidad de llegar a los objetivos planteados, se utilizaron varios enfoques, tanto experimentales como bioinformáticos.

VI.I.- Estudio de los genes esenciales de E. coli

VI.I.I.- Búsqueda de los ortólogos de los 303 genes esenciales en los 63 genomas completos

Con el objetivo de encontrar genes esenciales en *E. coli* K-12 ausentes en genomas de otras cepas de la misma especie, como un posible primer indicio de ejemplos del fenómeno de DGNO se buscaron ortólogos de los 303 genes esenciales definidos anteriormente (Baba et al., 2006), en los 63 genomas completos de *E. coli* de mejor calidad de secuencia y anotación (mismos que se tomaron de la base de datos del NCBI), utilizando TBLASTN. Como se esperaba, la mayoría de los genes esenciales se encuentran presentes en casi todos los genomas, pero sólo 4 de los genomas (cepas: DH1, W3110, MG1655 y BW2952) tienen los 303 genes esenciales, resultado sorprendente, pues aunque se esperaba obtener genes esenciales ausentes en algunos genomas, la cantidad de genomas que albergan a los 303 genes esenciales es muy bajo. De esta manera, varios de los genes esenciales parecían estar ausentes en uno o varios genomas,

como puede observarse en la Figura 4. Con ayuda de la escala de colores, puede observarse una franja de color gris oscuro que representa a los genes ausentes en los genomas mostrados en la gráfica y algunos puntos aislados, también de ese color. Se encontraron 11 genes ausentes en más de uno de los genomas analizados y además otros 11 genes ausentes en sólo uno de los genomas, Figura 4.

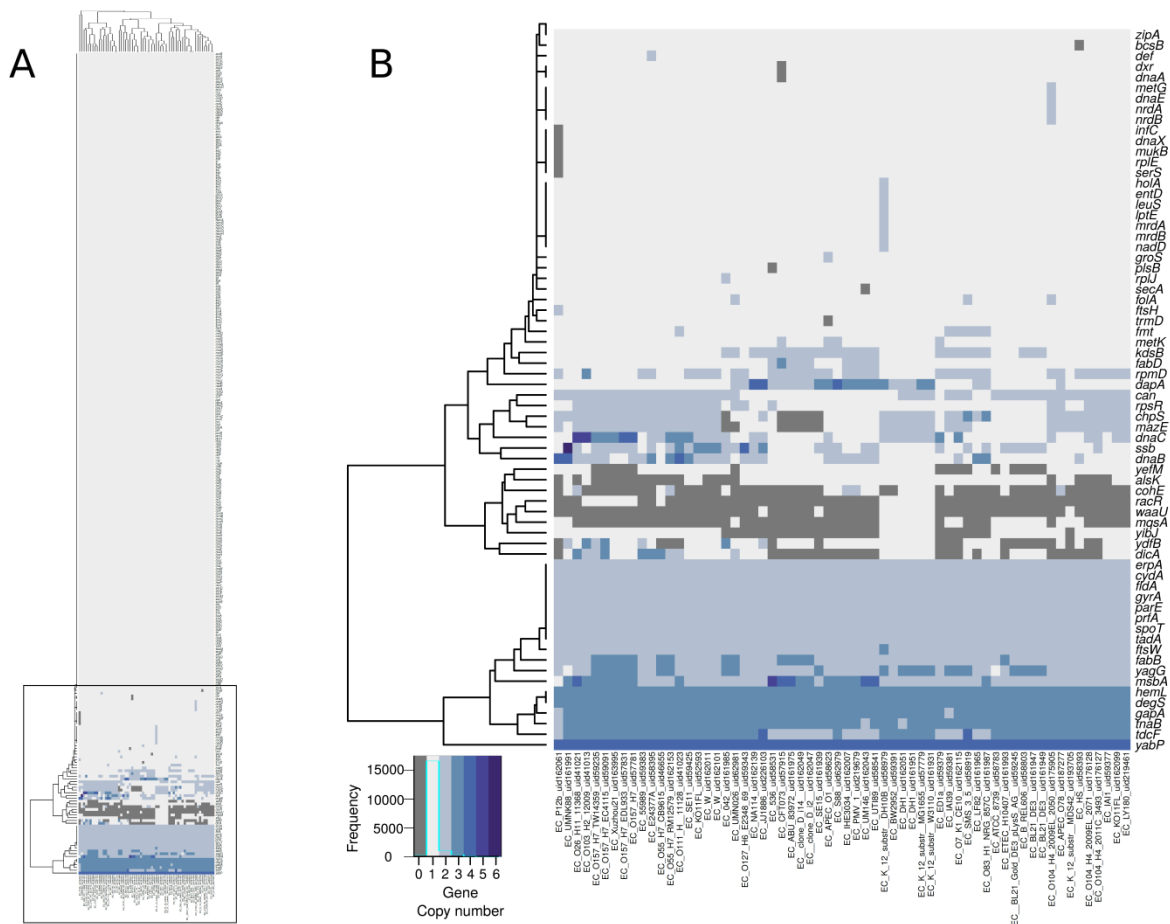


Figura 4 | Representación esquemática (mapa de colores) de la presencia de los 303 genes esenciales propuestos por Baba et al., en 63 genomas de diferentes cepas de *E. coli*. La escala de colores indica el número de copias de cada gen en cada genoma. En el panel A se muestra el gráfico completo y en B se observa el acercamiento y el código de colores.

Se observó también que muchos de los genes presentan más de una copia en los cromosomas y que la mayoría de los genes esenciales están en la mayor parte de los genomas, como puede apreciarse en la Figura 4 y en la Tabla 4, donde se enlistan los genes esenciales que están ausentes en uno o más genomas. Puede apreciarse en la Tabla 4, que los genes que se encuentran ausentes en más de un genoma, tienden a ser genes que codifican para proteínas que participan en sistemas toxina – antitoxina o como represores de fagos, mientras que los genes ausentes en sólo un genoma codifican proteínas con funciones metabólicas más diversificadas, entre las que destacan algunos como *bcsB*, que participa en la biosíntesis de celulosa, *dxr*, que participa en la síntesis de isoprenoides, *dnaA*, que codifica para una proteína iniciadora de la replicación cromosomal, *infC*, que participa en la traducción de proteínas o *dnaX*, que codifica a una proteína que forma parte de la DNA polimerasa III. El análisis posterior, con mayor detalle, continuó con aquellos genes esenciales que se reportaron ausentes en sólo un genoma, para de esta manera reducir la posibilidad de que debido a errores en las secuencias de los genomas es que hayan resultado ausentes, además de que los genes ausentes en un solo genoma son reales candidatos para la búsqueda de ejemplos de DGNO.

Tabla 4 | Genes esenciales definidos por Baba et al., que están ausentes en el genoma de alguna o varias cepas analizadas.

Gen	Función asignada	Número de cepas en las que está ausente
<i>bcsB</i>	Biosíntesis de celulosa	1 (HS)
<i>dxR</i>	Participa en la síntesis de isoprenoides, catalizando la conversión de 1-desoxi-D-xilulosa-5-fosfato (DXR) a 2-metil-D-eritrol-4-fosfato (MEP).	1 (CFT073)
<i>dnaA</i>	Proteína iniciadora de la replicación cromosomal DnaA	1 (CFT073)

<i>infC</i>	Factor 3 para el inicio de la traducción de proteínas	1 (P12B)
<i>dnaX</i>	DNA polimerasa III / Factor de elongación de DNA III / Subunidades tau y gamma de la DNA polimerasa III	1 (P12B)
<i>mukB</i>	Condensina del cromosoma MukBEF, ATPasa y subunidad de unión a DNA	1 (P12B)
<i>rplE</i>	Proteína ribosomal E de la subunidad del ribosoma grande	1 (P12B)
<i>serS</i>	Serina-tRNA sintetasa	1 (P12B)
<i>plsB</i>	Glicerol-3-fosfato aciltransferasa, cataliza el primer paso en la síntesis de fosfolípidos	1 (536)
<i>secA</i>	Acopla la hidrólisis de ATP con la secreción de proteínas a través de la membrana celular	1 (UM146)
<i>trmD</i>	tRNA metiltransferasa m(1)G37 dependiente de SAM	1 (APEC 01)
<i>chpS</i>	ChpS antitoxina; toxina ChpB	6
<i>maze</i>	MazE antitoxina; toxina MazF	7
<i>yelM</i>	YelM antitoxina; toxina YoeB	15
<i>alsK</i>	D-alosa cinasa	21
<i>cohE</i>	Represor del fago e14 tipo Ci	45
<i>racR</i>	Represor del profago Rac	50
<i>waaU</i>	Biosíntesis central de lipopolisacáridos	55
<i>mqsA</i>	MqsA antitoxina; toxina MqsR	50
<i>yibJ</i>	Putativa proteína perteneciente a la familia RHS	24
<i>ydiB</i>	Proteína del profago Qin	18
<i>dicA</i>	Represor transcripcional de dicB, profago Qin	24

La búsqueda detallada de los genes esenciales ausentes en un solo genoma en la primer parte del análisis, analizando las secuencias de los genes en los genomas aparentemente ausentes y mediante el análisis con Artemis (Carver et al., 2012) y MAUVE (Darling et al., 2004), dio como resultado que la gran mayoría de los genes esenciales que restaban (Tabla 4) realmente se encontraban presentes en los genomas y sorprendentemente conservan la longitud completa del gen, así

como la sintenia. Como ejemplo, es el caso de los genes *secA*, *infC* y *dxr* (Figuras 5 y 6), que a pesar de parecer ausentes cuando se buscan mediante TBLASTN, en los genomas de las cepas UM146, P12b y CFT073, respectivamente, realmente se encuentran los genes completos, sin embargo presentan errores de secuenciación, que causan una aparente interrupción de marco de lectura de modo que las secuencias de estos genes parecen truncadas.

En el caso del gen *dxr* de la cepa CFT073, la presencia del gen y la secuencia fueron corroborados amplificando la región por PCR y por secuenciación. La secuencia del gen *dxr* presenta errores al introducir bases nucleotídicas que cambian la fase de lectura y esto originó la incapacidad para encontrarlo con la primer estrategia de búsqueda (Figura 6). Cuando se corrigen los errores de secuenciación en la secuencia reportada de *dxr* en el genoma de la cepa CFT073 (Figura 6), puede observarse que se recupera la fase de lectura y la secuencia de la proteína deja de parecer truncada.

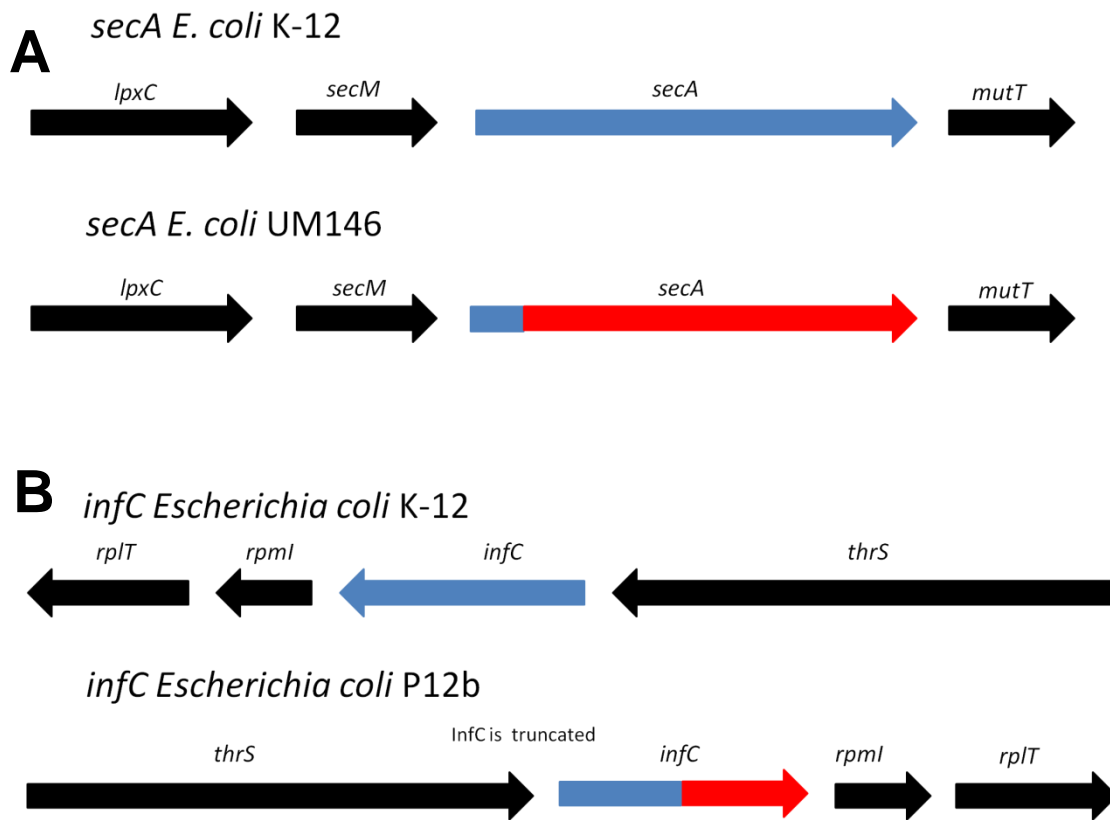


Figura 5 | Representación sintética de los genes *secA* (A) e *infC* (B) en las cepas UM146 y P12b, respectivamente, comparado con la cepa referencia K-12 en cada caso. Los genes *secA* e *infC* se encuentran presentes, parece que errores en la secuencia de los genes originan cambios en las fases de lectura.

Un ejemplo de los genes ausentes en más de un genoma, ausente en un total de 21 genomas, es el gen *alsK*, que participa en la fosforilación de la alosa hexosa; presentó tres patrones distintos en los 63 genomas analizados: presente en 42 genomas y ausente en 21 (Figura 7). En algunos de los genomas estaba ausente porque dichos genomas tienen una delección que abarca a todo el operón *als*, como ejemplo se tiene a la cepa 042. Sin embargo, en otros genomas el loci al que corresponde estar al operón *als* simplemente se encuentra ocupado por otros

genes. El operón *als* está ausente también en estos genomas, el genoma de la cepa O55:H7 CB9615 representa esta situación (Figura 7). Esta variación en la presencia del gene *alsK* sugiere que en realidad no se trata de un gen esencial como se había reportado.



Figura 6 | Representación sinténica del gen *dxr* en la cepa CFT073 en comparación con la cepa de referencia K-12 (A) y alineamiento de las secuencias del gen *dxr* de las cepas CFT073 y cepa K-12 (B), se encierran en rojo los huecos producto de errores de secuenciación.

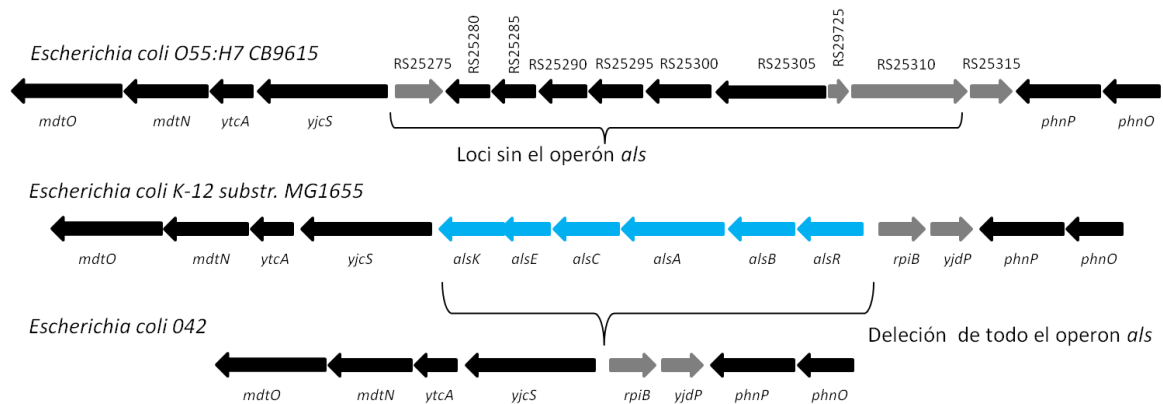


Figura 7 | Representación sintética de la región del gen *alsK* en tres cepas representativas de las tres composiciones observadas en los 63 genomas. La cepa de referencia *E. coli* K-12 tiene todo el operón (azul), mientras que en la cepa 042 se observa la delección completa del operón *als* y en la cepa O55:H7 CB9615 se observa que el loci donde debería estar el operón *als* está ocupado por otros genes, los genes *als* no se encuentran en este genoma.

De los 11 genes ausentes en sólo un genoma, resultado de la primer parte del análisis, 10 realmente parecen estar presentes en sus respectivos genomas, pero con errores de secuenciación. La única excepción es el gen *bcsB*, que con los análisis detallados con ARTEMIS y MAUVE se corroboró que realmente está ausente en el genoma de la cepa HS. La revisión detallada de la región genómica del loci *bcs* con los programas ARTEMIS y MAUVE, indicaba una aparente delección de aproximadamente 12.5 Kb en la cepa HS, lo que ocasiona la ausencia del gen *bcsB* y casi todo el operón (Figura 8). Se puede observar que en la cepa HS, sólo se encuentra una pequeña parte del gen *bcsG* y de *bcsC*, estando 7 genes totalmente ausentes en esa cepa, incluidos *bcsB* y *bcsA*.

Para corroborar la delección que involucra al gen *bcsB* en la cepa HS, propuesto como esencial en el estudio realizado por Baba y colaboradores (2006), se amplificó por PCR la región del operón *bcs* con oligonucleótidos específicos y comunes (*bcsHSFw* y *bcsHSRv*, ver secuencias en el anexo) que se encuentran tanto en el cromosoma de las cepas CFT073 (cepa de referencia, que sí tiene los

genes del operón *bcs*) y el de la HS, que presenta la delección. Como se muestra en la Figura 9, en la cepa HS hay una amplificación de sólo 2.1 Kb, significativamente menor que el producto de PCR en la cepa CFT073 que es de 13.9 Kb y que coincide con lo esperado, por lo tanto se corrobora la delección del operón *bcs* en la cepa de *E. coli* HS.

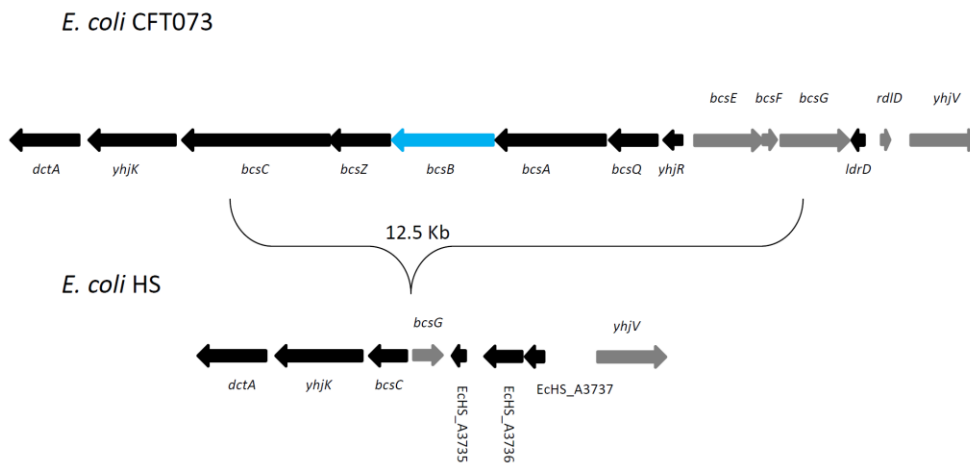


Figura 8 | Representación sintética de la delección del operón *bcs* (involucrado en la síntesis de celulosa) en la cepa HS que abarca aproximadamente 12.5 Kb, en comparación con la cepa de referencia CFT073, que parece tener el operón completo y funcional.

VI.I.II.- Evaluación de la esencialidad del gen *bcsB*

Con los enfoques bioinformáticos y experimentales que se utilizaron, observamos que el gen *bcsB* es el único de los 303 genes esenciales que se encuentra ausente en solo uno de los 63 genomas completos utilizados para en este trabajo. El gen *bcsB* en *E. coli* forma parte del operón *bcs* que se encarga de la biosíntesis y traslocación de celulosa (McNamara et al., 2015) como material de construcción de biopelículas. Partiendo de reportes de que *bcsB* participa en la traslocación de

celulosa, sería posible que la esencialidad de este gen sea atribuida a la necesidad de traslocar la celulosa producida en el interior de la célula (por el complejo formado por BcsA y otras proteínas) hacia el exterior de la misma y así evitar la acumulación del polímero y la consecuente toxicidad. Nos planteamos que era muy posible que en la cepa comensal HS el gen *bcsB* no fuese esencial ya que no cuenta con este gen ni con los genes del operón *bcs* que participan en la síntesis de celulosa debido a la delección de más de 12 Kb y por lo tanto no resulta vital traslocar celulosa que no se está biosintetizando.

En estudios previos fue reportado que las cepas de *E. coli* derivadas de la cepa K-12 no producen celulosa debido a un polimorfismo en una base (SNP) en el gen *bcsQ*, al inicio del operón *bcs* (Serra et al., 2013), debido a este fenómeno es que no se puede explicar la esencialidad de *bcsB* en la cepa BW25113, que fue en la que se reportaron los genes esenciales de *E. coli* (Baba et al., 2006). Para evaluar la esencialidad de *bcsB*, se planteó evaluar la capacidad de aislar mutantes en *bcsB* en una cepa productora de celulosa, la cepa uropatogénica CFT073, con respecto a la cepa BW25113 que no la produce y analizar la producción de celulosa en la cepa HS. Con base en lo anterior, realizamos las delecciones en los genes *bcsA* y *bcsB* en la cepa uropatogénica CFT073 que no contiene el SNP en *bcsQ* (Tabla 5), realizando primero la delección en *bcsA* y sobre este contexto genómico, la delección de *bcsB*, también se intentó hacer la delección sencilla de *bcsB* en la cepa CFT073, pero como se observa en la Tabla 5, no se obtuvieron este tipo de mutantes lo que sugiere que la delección es letal para la célula. Las delecciones en *bcsA* o *bcsB* en la cepa BW25113 son posibles y se obtienen fácilmente debido a que no expresa estos genes (Tabla 5).

Este resultado que concuerda con lo reportado por Serra et al., (2013) pues al no expresar el operón *bcs*, las cepas derivadas de la cepa K-12 no producen celulosa y por lo tanto *bcsB* no es esencial en ese contexto genómico, pues no es necesario traslocar celulosa.

Se observó que la cepa CFT073 doble mutante $\Delta bcsA \Delta bcsB$ puede ser transformada y complementada en *trans* por un plásmido que contiene a ambos genes, pero no se tienen transformantes cuando se complementa sólo con el plásmido que expresa *bcsB* (Tabla 5). Se obtuvieron 75.3 ± 14.6 transformantes / μg de plásmido pUCP-*bcsA* usando la cepa silvestre como receptora, mientras que sólo 0.66 ± 1.1 transformantes utilizando a la doble mutante con la misma cantidad de plásmido (pUCP-*bcsA*), ninguna de estas fue capaz de producir celulosa, por lo que pensamos que son mutantes espontáneas Amp^r . Se obtuvieron 92 ± 17 transformantes / μg cuando se utilizó a la doble mutante CFT073 $\Delta bcsA \Delta bcsB$ como receptora del plásmido pUCP-*bcsAB*, estas transformantes fueron capaces de producir celulosa.

El fenotipo de las cepas puede observarse claramente en la Figura 10, ninguna de las dos mutantes derivadas de la cepa BW25113 producen celulosa, pero sí curli, que es lo que da la coloración café a las colonias. Así pues, la cepa BW25113 con la delección del gen *bcsA* se utilizó para corroborar el fenotipo celulosa⁻ al no expresar el gen; la cepa HS no produce ni celulosa ni curli, y los fenotipos de las cepas CFT073 WT, CFT073 $\Delta bcsA \Delta bcsB$ y CFT073 $\Delta bcsA \Delta bcsB$ + pUCP-*bcsAB* son: curli-, celulosa +; curli -, celulosa – y curli-, celulosa +, respectivamente. Con estos experimentos queda bastante claro que la esencialidad del gen *bcsB* depende de la producción de celulosa en la célula (Martínez-Carranza et al., 2018b).

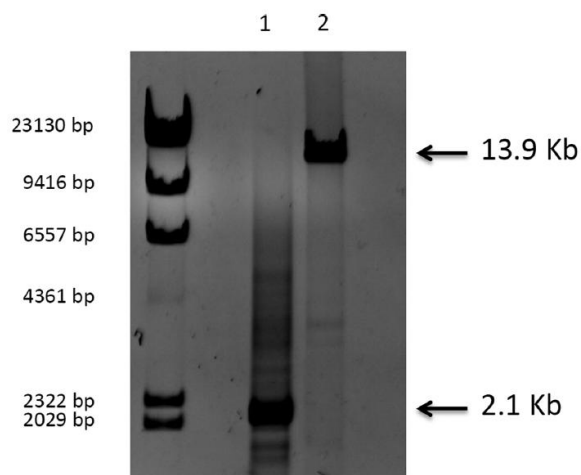


Figura 9 | Identificación de la delección del operón *bcs* en la cepa HS que abarca aproximadamente 12.5 Kb, en comparación con la cepa de referencia CFT073. La región del operón *bcs* se amplificó por PCR utilizando los oligonucleótidos *bcs*HSFw y *bcs*HSRv. Como indicaba el análisis bioinformático, existe una delección en la cepa HS (2.1 Kb), mientras que en la cepa CFT073 el producto de PCR (13.9 Kb) indica que el operón *bcs* está completo. El marcador de peso molecular corresponde al fago lambda digerido con la enzima de restricción HindIII.

Tabla 5 | Relación del número de candidatas al hacer la delección de los genes *bcsA* y *bcsB* en las cepas CFT073 y BW25113 (A) y número de transconjugantes al transformar con los plásmidos pUCP-*bcsA* y pUCP-*bcsAB* (B) para evaluar la esencialidad de *bcsB* en las cepas CFT073 y BW25113. La expresión de los genes se realiza en presencia de IPTG 0.5 mM.

A

Número de candidatas:

Cepa	Delección	Media	Fenotipo
CFT073 WT	<i>bcsB</i>	0 ± 0	-
CFT073 WT	<i>bcsA</i>	98 ± 7	CEL-
BW25113 WT	<i>bcsB</i>	199 ± 13	CEL-
BW25113 $\Delta bcsA$	<i>bcsB</i>	127 ± 12	CEL-

B

Número de transconjugantes:

Cepa	Plásmido	Número	Fenotipo
CFT073 WT	pUCP- <i>bcsA</i>	75 ± 14	CEL+
CFT073 $\Delta bcsA-B$	pUCP- <i>bcsA</i>	0,6 ± 1	-
CFT073 $\Delta bcsA-B$	pUCP- <i>bcsA-B</i>	92 ± 16	CEL +

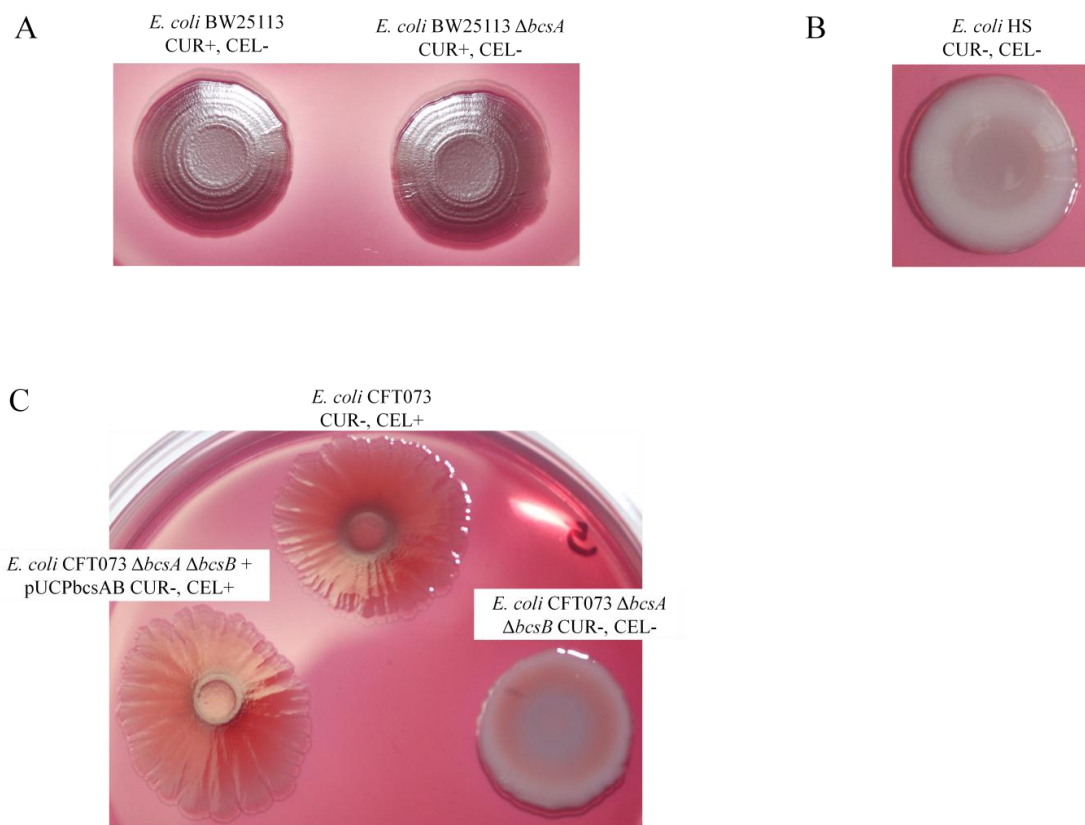


Figura 10 | Evaluación de la producción de celulosa por las cepas BW25113, HS y CFT073. En la Figura se observa a las cepas BW25113 y la mutante $\Delta bcsA$ (A), la cepa HS (B) y la cepa silvestre CFT073 y sus derivadas, la doble mutante CFT073 $\Delta bcsA$ - $\Delta bcsB$, y la doble mutante complementada para la producción de celulosa con el plásmido pUCP-bcsAB (C). Las colonias se crecieron por 7 días, sobre medio LB modificado sin NaCl y con colorantes rojo Congo y azul de Coomassie. Los morfotipos de las colonias concuerdan con la clasificación de morfotipos en enterobacterias (Römling, 2005; Steenackers et al., 2012). Los fenotipos están indicados en cada colonia. Tomada de (Martínez-Carranza et al., 2018b).

VI.II.- Estudio de la variabilidad de los genes esenciales en las cepas PAO1 y PA14 de *P. aeruginosa*

Para comprender mejor la variabilidad de los genes esenciales entre las cepas de *P. aeruginosa* PAO1 y PA14, se analizaron los genes esenciales que son compartidos por los estudios de mayor relevancia que se mencionaron anteriormente (Figura 11), se observó que sólo 48 genes son compartidos por los 4 estudios, es decir, que sólo este pequeño grupo de genes son propuestos como esenciales en los 4 estudios evaluados (Lee et al., 2015; Liberati et al., 2006; Skurnik et al., 2013; Turner et al., 2015). Se buscó además la categoría funcional a la que pertenecen estos 48 genes, como puede observarse en la Figura 11, la mayoría de estos genes participan en la traducción, estructura ribosomal y biogénesis, seguido por los genes que participan en estructura y biogénesis de la pared celular.

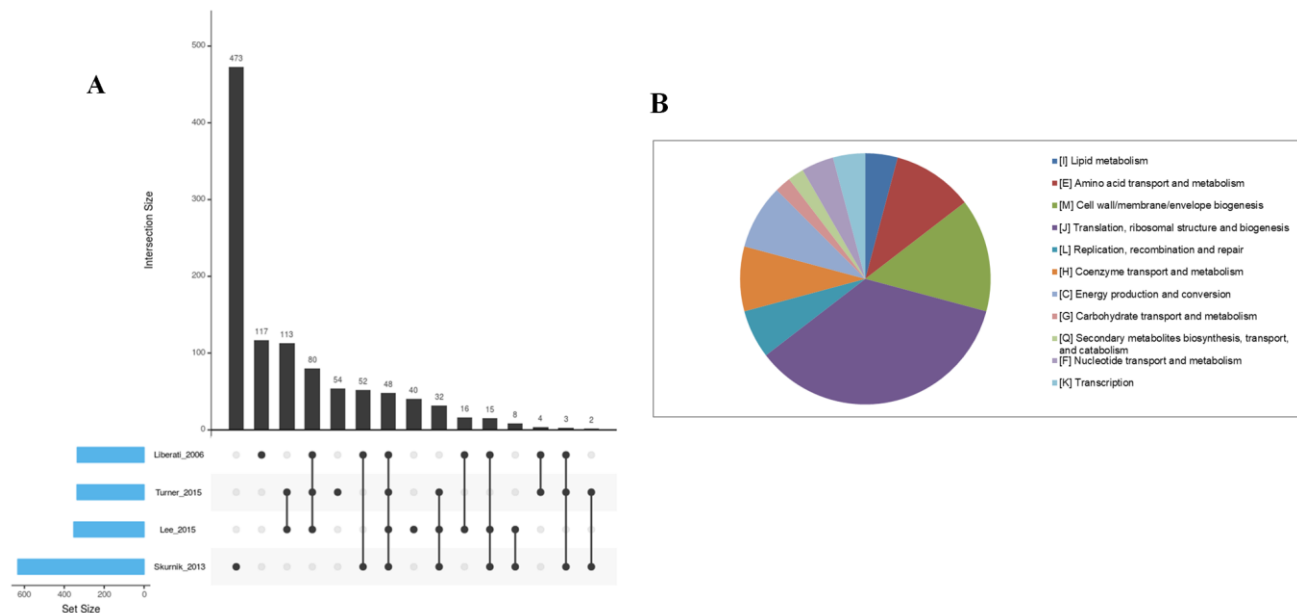


Figura 11 | Representación de los genes esenciales que se comparten en los 4 estudios analizados. Se observa que sólo 48 genes esenciales son compartidos por los 4 reportes, también

se observa el número de genes únicos y los genes que se comparten por dos o más estudios (A). Categorías funcionales (COGs) a las que pertenecen los 48 genes esenciales compartidos por los 4 estudios (B).

VI.II.I.- Análisis de los 210 genes esenciales de *P. aeruginosa* PA14 reportados como ausentes en el genoma de PAO1

Con la finalidad de analizar la alta variabilidad de los genes esenciales de *P. aeruginosa* PA14 reportada, en el sentido de que 210 genes esenciales descritos en esta cepa se encuentran ausentes en el genoma de la cepa PAO1 (Skurnik et al., 2013), se realizó una búsqueda detallada de todo el conjunto de estos 210 genes por métodos bioinformáticos, como se describe en la sección de materiales y métodos. Las secuencias de los genes esenciales fueron traducidas a secuencias protéicas y buscadas en el proteoma de la cepa PAO1. Se encontró que de los 210 genes esenciales, sólo 8 genes están realmente ausentes en el genoma de la cepa PAO1 (Tabla 6), lo que representa sólo un 3.8% de los genes, además, estos 8 genes (Tabla 6) podrían en realidad no ser esenciales, ya que son genes de tamaño muy pequeño (entre 147 y 390 nucleótidos) y es menos probable obtener inserciones de transposones en estos genes. Además, las funciones para las que codifican estos 8 genes no dejan claro que podrían codificar para funciones esenciales, de los 8 genes, 4 son proteínas hipotéticas y la esencialidad de las funciones de los otros 4 es cuestionable (Tabla 6).

Tabla 6 | Conjunto de genes esenciales reportados en *P. aeruginosa* PA14 (Skurnik et al., 2013), ausentes en el genoma de *P. aeruginosa* PAO1.

Anotación del gen	Función
PA14_59110	Proteína hipotética
PA14_58980	Proteína hipotética
PA14_58940	Proteína hipotética
PA14_31250	Proteína hipotética
PA14_31080	Peptidasa de la familia S26
PA14_15530	<i>uhpA</i> – Regulador de respuesta de sistemas de dos componentes
PA14_15470	<i>merP</i> – Proteína del sistema de resistencia a mercurio
PA14_14310	Probable regulador transcripcional de la familia ArsR

VI.II.II.- Análisis de la presencia en la cepa PA14 de los genes esenciales de *P. aeruginosa* PAO1 para el crecimiento en esputo de pacientes con fibrosis quística

En este trabajo se buscaron los 336 genes esenciales reportados en la cepa PAO1 para su crecimiento en esputo de pacientes con fibrosis quística (Turner et al., 2015) en el genoma de la cepa PA14, de los 336 genes, se logró identificar a 12 genes esenciales en la cepa PAO1 y ausentes en la cepa PA14. Estos 12 genes son parte de un grupo que se encuentran en la misma región genómica de

la cepa PAO1 y que participan en la biosíntesis de lipopolisacáridos del antígeno O (Tabla 7). El lipopolisacárido de ambas cepas es diferente por lo que es de esperarse que los genes que intervienen en su biosíntesis no sean ortólogos. De nueva cuenta, este análisis se observa baja variabilidad de los genes esenciales propuestos entre las dos cepas.

Tabla 7 | Genes esenciales en *P. aeruginosa* PAO1 (Turner et al., 2015), ausentes en el genoma de *P. aeruginosa* PA14.

Gen	Función
<i>wbpL</i>	Glicosiltransferasa / participa en la biosíntesis de lipopolisacáridos
<i>wbpJ</i>	Probable glicosiltransferasa
<i>wbpI</i>	UDP-N-acetilglucosamina 2-epimerasa / participa en la biosíntesis de lipopolisacáridos
<i>wbpH</i>	Probable glicosiltransferasa / Participa en la biosíntesis de antígeno O
<i>wbpG</i>	Proteína de biosíntesis de lipopolisacáridos
<i>hisF2</i>	Participa en la biosíntesis de antígeno O / glicerol-fosfato sintasa
<i>wzx</i>	Participa en la biosíntesis de antígeno O / translocasa de antígeno O
<i>wzy</i>	Participa en la biosíntesis de lipopolisacáridos / participa en la biosíntesis de antígeno O
<i>wbpE</i>	Transaminasa de ácido UDP-2-acetamido-2-dideoxi-d-ribo-hex-3-ulurónico / Participa en la biosíntesis de antígeno O
<i>wbpD</i>	Participa en la biosíntesis de lipopolisacáridos / Transaminasa de ácido UDP-2-acetamido-2-dideoxi-d-ribo-hex-3-ulurónico
<i>wbpB</i>	Transaminasa de ácido UDP-2-acetamido-2-dideoxi-d-ribo-hex-3-ulurónico / Participa en la biosíntesis de antígeno O

wbpA	UDP-N-acetil-d-glucosamina 6-Deshidrogenasa / Participa en la biosíntesis de lipopolisacáridos
-------------	--

VI.II.III.- Análisis de los 352 genes esenciales generales de *P. aeruginosa* PAO1 en el genoma de la cepa PA14

Un estudio reciente reportó los genes esenciales generales de *P. aeruginosa*, para lo que empleó varias condiciones de cultivo (Lee et al., 2015) y propuso 352 genes esenciales comunes a las condiciones de cultivo empleadas utilizando la técnica de mutagénesis por transposones saturada. Con el objetivo de evaluar si este conjunto de 352 genes esenciales generales en la cepa PAO1 representa a los genes esenciales de *P. aeruginosa*, evaluamos la conservación de estos genes en la cepa PA14. Se encontró que sólo dos de los 352 genes están ausentes en la cepa PA14 (*wbpB* e *imm2*), debido a la función anotada a estos dos genes, se podría cuestionar la esencialidad de los mismos en la cepa PA14, ya que *wbpB* participa en la biosíntesis de antígeno O e *imm2* es una proteína de inmunidad a la piocina S2 (Tabla 8).

Tabla 8 | Genes esenciales generales en el genoma de *P. aeruginosa* PAO1 propuestos por Lee (2015), ausentes en el genoma de *P. aeruginosa* PA14.

Gene name	Function
<i>imm2</i>	Proteína de inmunidad a piocina S2 / inmunidad a bacteriocinas
<i>wbpB</i>	Transaminasa de ácido UDP-2-acetamido-2-dideoxi-d-ribo-hex-3-ulurónico / Participa en la biosíntesis de antígeno O

VI.II.IV.- Los genes esenciales generales en *P. aeruginosa* PAO1 están conservados en *A. vinelandii*

En este trabajo se evaluó la conservación de los 352 genes esenciales generales de *P. aeruginosa* PAO1 en el genoma de *A. vinelandii* DJ, para lo que se realizó la búsqueda de los genes por secuencia y sintenia. En la Figura 12 se muestra el alineamiento de los genomas de *P. aeruginosa* PAO1 y *A. vinelandii* DJ, donde se aprecia la alta cantidad de rearrreglos cromosómicos que existen cuando se comparan ambos genomas, sin embargo, se conservan muchas regiones genómicas en ambos, que se pueden observar en las líneas transversales que señalan las regiones homólogas en ambos genomas.

La presencia de cada gen, encontrada por secuencia, se confirmó con el análisis sinténico para cada uno de los genes (Figura 13). En este caso se muestra la conservación del gen y de su contexto genómico para el gen *dnaE*. De los 352 genes esenciales generales, encontramos que únicamente 10 genes no están presentes en el genoma de *A. vinelandii* (Tabla 9). Sin embargo, en dos de ellos la esencialidad de los genes es cuestionable ya que codifican para funciones que para *A. vinelandii* pueden no ser necesarias, es el caso de los genes *wbpB* y *imm2* que participan en la síntesis de lipopolisacáridos del antígeno O y en la resistencia a piocinas, respectivamente, y que no pueden ser considerados genes esenciales generales para *P. aeruginosa*, ya que, como se muestra en la sección anterior, no se conservan en *P. aeruginosa* PA14. Adicionalmente, en otros dos genes, que son *nadE* y *PA0092* se han reportado mutantes por transposones (Jacobs et al., 2003), por lo que no pueden ser considerados esenciales. Adicionalmente, los genes *PA4674*, *PA0906* y *prtN* parecen ser genes esenciales condicionales, ya que codifican proteínas que participan en sistemas toxina–antitoxina, represores de fagos y represores de genes de piocinas, respectivamente (Tabla 9). Estos

genes requieren de la presencia de las toxinas o los fagos para ser esenciales, mismos que no se encuentran en *A. vinelandii*, y por lo tanto podrían ser eliminados de la lista de los genes esenciales ausentes en *A. vinelandii*, ya que no podrían ser esenciales en esta especie, quedando sólo 3 genes (PA0442, PA4405 y PA4685, están anotados como proteínas hipotéticas) en la lista de los genes esenciales generales de *P. aeruginosa* PAO1 que están ausentes en el genoma de *A. vinelandii*.

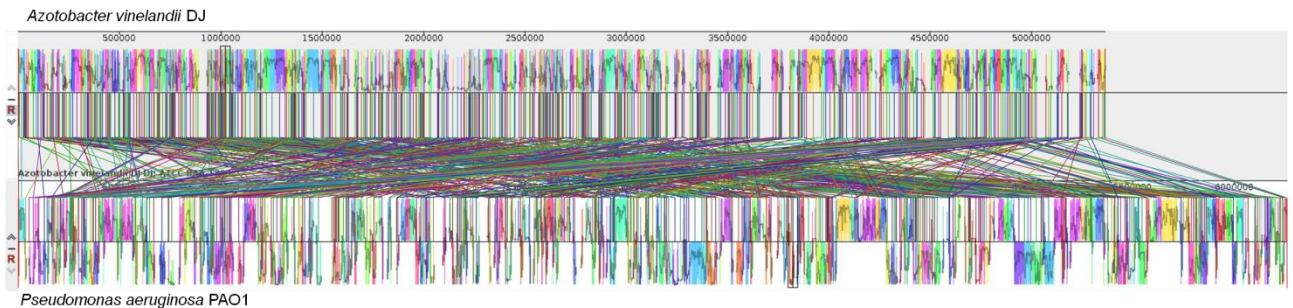


Figura 12 | Alineamiento de los genomas de *P. aeruginosa* PAO1 y *A. vinelandii* DJ con el programa MAUVE. Se observa una gran cantidad de rearrreglos cromosómicos entre ambos genomas, sin embargo se comparten muchas regiones del cromosoma tienen homología.

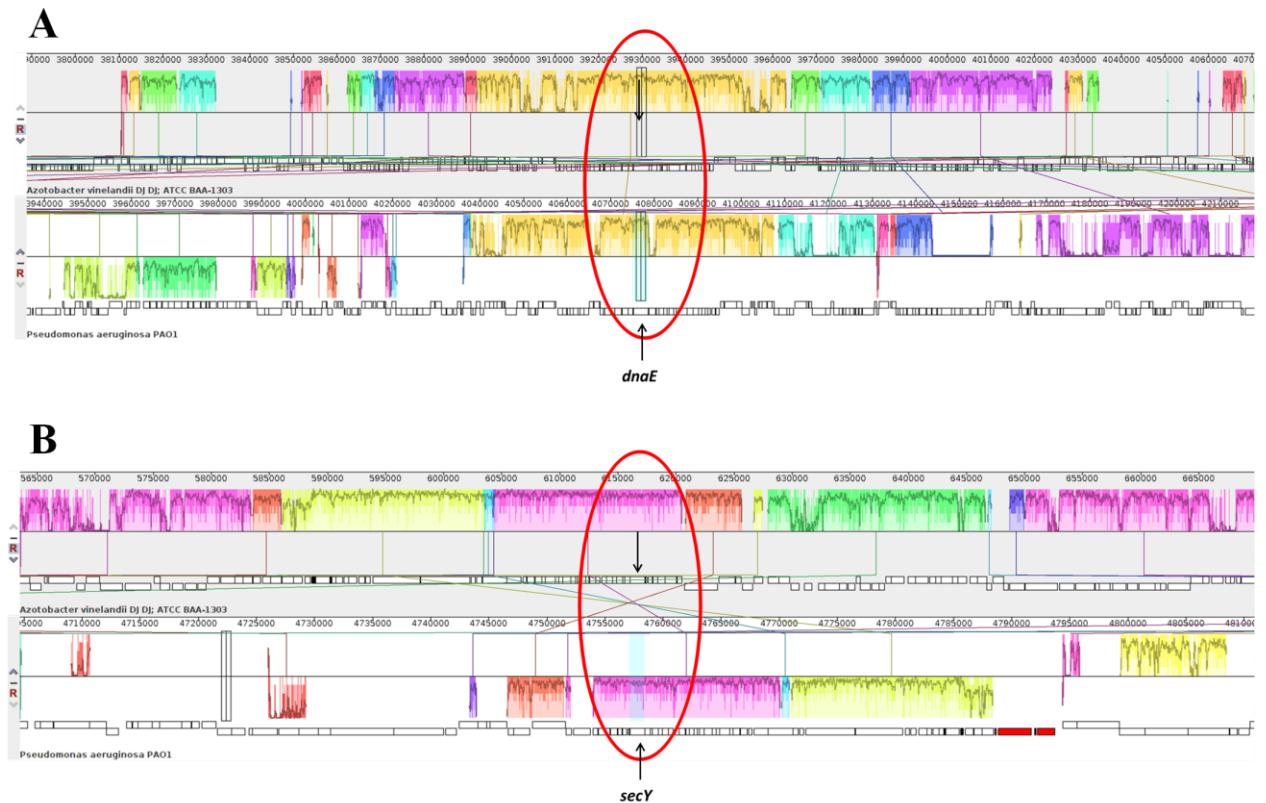


Figura 13 | Comparación sinténica de los loci de los genes esenciales generales propuestos por Lee (2015) de *P. aeruginosa* en *A. vinelandii*. La región del gen *dnaE* (A) de *P. aeruginosa* tiene homología en el genoma de *A. vinelandii*, la homología puede visualizarse por los colores en el gráfico, de la misma manera el gen *secY* (B).

Tabla 9 | Genes esenciales generales propuestos por Lee (2015), ausentes en el genoma de *A. vinelandii*.

Nombre	Función
PA0092*	Proteína de secreción del sistema de secreción tipo IV / Proteína de unión toxina-antitoxina
<i>nadE</i> *	Sintetasa de NAD dependiente de NH ₃ / Participa en la síntesis de NAD

<i>prtR</i>	Regulador transcripcional PrtR / Participa en la regulación negativa de procesos biosintéticos de metabolitos secundarios.
PA0906	Represor de fenotipo de lisis / respuesta celular al daño en el DNA / regulación negativa de la transcripción.
<i>imm2</i>	Proteína de inmunidad a piocina S2 / inmunidad a bacteriocinas
PA4674	Antitoxina HigA / toxina-antitoxina par tipo II
<i>wbpB</i>	Participa en procesos biosintéticos de antígeno O
PA4405	Proteína hipotética
PA4685	Proteína hipotética
PA0442	Proteína hipotética

VII.- DISCUSIÓN

La alta variabilidad de los genes esenciales entre bacterias cercanamente relacionadas representa un reto a la teoría de la evolución bacteriana actual (Ish-Am et al., 2015; Luo et al., 2015; Martínez-Carranza et al., 2018a, 2018b), en la que se ha postulado que los genes esenciales se encuentran en el genoma central y son los genes básicos sobre los que se basa la vida celular, por lo tanto, si los genes esenciales son variables entre especies o entre cepas de una misma especie, son genes que no forman parte del genoma central, sino del genoma accesorio. Sin embargo esta problemática no se reconoce en los trabajos que estudian los genes esenciales, pero fue la motivación primordial para llevar a cabo esta tesis.

Varios estudios han reportado diferentes conjuntos de genes esenciales en *E. coli* y en *P. aeruginosa*, que se diferencian cualitativamente y cuantitativamente uno del otro dependiendo del enfoque utilizado para la determinación de los genes esenciales.

En *E. coli* se han utilizado varias estrategias para definir el conjunto de genes esenciales para el desarrollo de esta especie bacteriana en medio rico. Entre los trabajos realizados a la fecha destacan algunos por la precisión con la que han definido a los genes esenciales, uno de los trabajos más importantes es el de Baba et al., que hace deleciones específicas en cada uno de los marcos de lectura y evita efectos de polaridad que pueda arrojar resultados falsos positivos.

Como puede observarse en la Figura número 2, de los estudios más importantes realizados hasta el momento, donde se identifican diferentes conjuntos de genes como esenciales en *E. coli*, sólo 164 genes son compartidos por todos los estudios. Estos 164 genes representan el 3.8% del genoma de *E. coli*, que consta de 4218 genes, muy posiblemente estos 164 genes representen gran parte de los genes esenciales generales en todas las cepas de *E. coli*, sin embargo aún se puede observar mucha variación de los genes esenciales reportados en los diferentes estudios que puede deberse en gran medida a los enfoques experimentales y a las condiciones de cultivo utilizadas.

Se ha reportado que los genes esenciales son variables entre especies bacterianas (Sato and Miyazaki, 2017; Zhaxybayeva et al., 2009) y también entre cepas de una misma especie (Skurnik et al., 2013; Turner et al., 2015). La variación de los genes esenciales puede explicarse bajo el fenómeno de DGNO (Koonin et al., 1997) o por los errores en las secuencias genómicas y bases de datos (Charlebois and Doolittle, 2004; Cook and Ussery, 2013), que originan variación implícita en los genes esenciales. Además, se ha reportado una alta tasa de THG entre bacterias de la misma o de distinta especie, entre bacterias cercanas y lejanas filogenéticamente, donde se ha observado la transferencia no

solo de genes accesorios, sino también de genes esenciales. También se han hecho más evidentes los casos de incongruencias genómicas en varias especies bacterianas (González-Casanova et al., 2014; Sato and Miyazaki, 2017; Zhaxybayeva et al., 2009).

E. coli es considerada un organismo modelo, en cuyo funcionamiento celular y molecular se basa mucho de nuestro conocimiento (Gopal and Kumar, 2013; Juhas et al., 2011) y ha sido fundamental para el avance de las ciencias biológicas, es por ello que se decidió analizar detalladamente los genes esenciales de esta especie en los genomas de diferentes cepas. En el trabajo de Baba et al., definieron un conjunto de 303 genes esenciales en *E. coli* K-12, específicamente la cepa BW25113.

En este trabajo se buscaron los 303 genes esenciales propuestos por Baba et al., en los 63 genomas completos de *E. coli* mejor anotados que están presentes en la base de datos NCBI con el objetivo de encontrar casos de DGNO y analizar la variabilidad de los genes esenciales entre las 63 cepas seleccionadas. El hallazgo de casos de DGNO en los genomas de *E. coli*, también apoyaría a el modelo del banco de semillas (Blath et al., 2016; González-Casanova et al., 2014), en el que se explica que la alta variabilidad genética encontrada en los genomas bacterianos puede deberse a la generación de una poza genética originada por organismos que llevan a cabo procesos de enquistamiento y/o esporulación, además de los fagos.

De los 303 genes esenciales estudiados en 63 genomas, 11 genes están ausentes en varios genomas, otros genes 11 genes están ausentes en sólo 1 de los genomas, sin embargo, como se observa en la Figura 4 y en la Tabla 4, la mayoría de los genes ausentes en más de un genoma codifican para sistemas de toxina-antitoxina o represores de fagos. Por lo tanto parece que estos genes son esenciales en la cepa BW25113 debido a que esta cepa posee la correspondiente toxina y/o los fagos que son contrarrestados con los productos de estos genes, en

los genomas en que se encuentran ausentes estos genes parecen no estar las toxinas ni los fagos. Como se mencionaba en los resultados, el gen *alsK*, que participa en la fosforilación de la hexosa alosa está ausente en 21 de los 63 genomas analizados y es difícil de entender porqué este gen es esencial en la cepa BW25113. En otros trabajos se ha propuesto a este gen como no esencial (Goodall et al., 2018) y nuestros resultados concuerdan con esta propuesta. Como se puede observar en la Tabla 4, la mayor parte de los genes ausentes en varios de los genomas analizados, parecen ser esenciales en contextos genómicos específicos (como sería la presencia de un gen que codifica para una toxina o de un fago lisogénico), además por definición, estos genes forman parte del genoma accesorio, pues no se encuentran en todos los genomas de *E. coli*.

Los genes que en primera instancia parecían ausentes en sólo un genoma, se seleccionaron y se analizaron detalladamente para corroborar su ausencia y/o presencia en el genoma específico, sin embargo en la mayoría de los casos se trató de errores de secuencia de los genomas en los que parecían estar ausentes.

Únicamente el gen *bcsB* se demostró estar ausente en el genoma de la cepa HS. En *E. coli*, el gen *bcsB* forma parte del operón *bcs*, que contiene a los genes que codifican a proteínas encargadas de la biosíntesis y traslocación de celulosa. En la cepa HS el operón *bcs* está casi totalmente ausente debido a una delección de 12.5 Kb. Debido a que *bcsB* participa en la biosíntesis y traslocación de la celulosa (McNamara et al., 2015; Morgan et al., 2013, 2014, 2016; Serra et al., 2013), nos propusimos la hipótesis de que *bcsB* es esencial cuando hay producción de celulosa, proceso que se lleva a cabo por todo el operón *bcs* y es por esta razón que en la cepa HS el gen *bcsB* no es esencial, porque no produce celulosa debido a la delección del operón casi completo. Trabajos al respecto indican que las cepas derivadas de *E. coli* K-12 no producen celulosa naturalmente debido a un SNP en el gen *bcsQ* que impide la expresión correcta del operón *bcs* al ocasionar un cambio de fase de lectura de este gen y a un efecto de polaridad (Serra et al., 2013).

La celulosa es el polímero producido naturalmente más abundante en la naturaleza y forma el principal componente de la pared celular de plantas vasculares y algas, aunque también es producido por bacterias (Morgan et al., 2016; Zimmer, 2019). Sirve como componente de las biopelículas de muchas bacterias Gram-negativas (McNamara et al., 2015; Steenackers et al., 2012). En la literatura hay varios trabajos publicados que tratan de comprender la síntesis de celulosa en bacterias, y aunque el mecanismo productor de celulosa aún no está del todo caracterizado, se tiene cada vez más información acerca de la síntesis de celulosa en varias especies bacterianas como *Rhizobium leguminosarum*, *Burkholderia spp.*, *Pseudomonas putida*, *Dickeya dadantii*, *Erwinia chrysanthemi*, *Agrobacterium tumefaciens*, *Escherichia coli* y *Salmonella entérica* (Römling and Galperin, 2015). Sin embargo, la más eficiente bacteria productora de celulosa es *Gluconacetobacter xylinus* (Buldum et al., 2018). La celulosa bacteriana es altamente apreciada por su grado de pureza (Buldum et al., 2018), que es superior al de las plantas, por lo que recientemente se ha tratado de producir celulosa en bacterias recombinantes (Buldum et al., 2018).

Por esta razón evaluamos la esencialidad del gen *bcsB* en una cepa que sí sintetiza celulosa, en este caso la cepa CFT073. En este estudio se evaluó la esencialidad de *bcsB* mediante el aislamiento de mutantes sencillas de $\Delta bcsB$, y de dobles mutantes $\Delta bcsA-\Delta bcsB$ y la viabilidad de su complementación con un plásmido que expresa sólo *bcsA*. Si nuestra hipótesis es correcta no se obtendrían deleciones sencillas en $\Delta bcsB$ pero sí dobles mutantes $\Delta bcsA-\Delta bcsB$. Como se esperaba, en la cepa CFT073, la deleción del gen *bcsB* es letal (Tabla 5), no así cuando se hacen deleciones consecutivas, primero se hace la deleción en el gen *bcsA* (que participa en la síntesis de celulosa) y luego en el gen *bcsB* ($\Delta bcsA-\Delta bcsB$). En la cepa BW25113, en la que no se expresan los genes del operón *bcs*, las mutaciones en los genes *bcsA* o *bcsB* son viables, ya que no produce celulosa. Como se observa en la sección de resultados, en la Figura 10, la cepa CFT073 produce celulosa, sin embargo la cepa con la doble deleción no la

produce pero, cuando es transformada con el plásmido pUCP-bcsAB que expresa los genes *bcsA* y *bcsB*, recupera el fenotipo celulosa⁺.

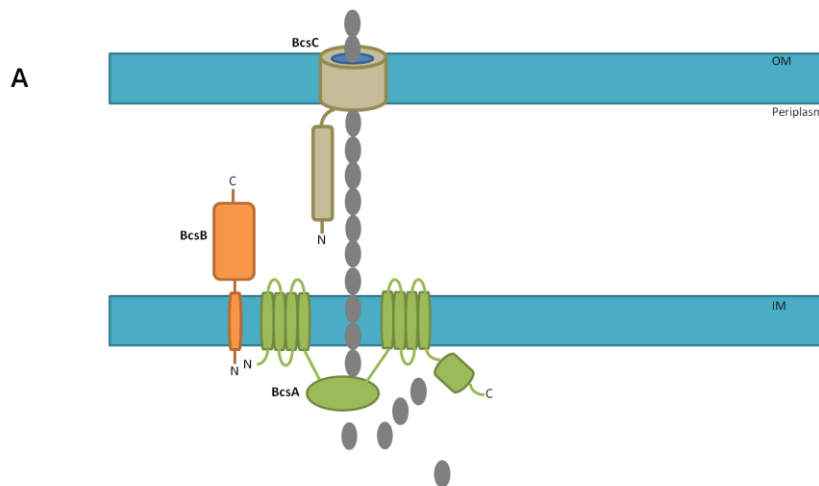
La cepa BW25113 no produjo celulosa, como se esperaba, al no expresarse los genes del operón *bcs*; la cepa HS tampoco produce celulosa, ya que los genes del operón *bcs* están ausentes (Figura 10). Sugerimos que la incapacidad de aislar mutantes en el gen *bcsB* de la cepa BW25113 (Baba et al., 2006) probablemente ocurrió porque la cepa que ellos utilizaron sí era productora de celulosa por tratarse de una revertante espontánea en el gene *bcsQ*, aunque pertenece al linaje productor de celulosa (cepa K-12).

Con este análisis de la esencialidad de *bcsB* en *E. coli*, se aporta evidencia a la visión de que la definición de los genes esenciales solo con el criterio de la incapacidad de aislar mutantes viables es cuestionable y depende del contexto genético y de las condiciones de cultivo, ya que la inactivación de algunos genes se puede llevar a cabo sólo en determinados fondos genéticos. Hay muchos ejemplos de genes esenciales condicionales en bacterias, como los genes que participan a sistemas toxina-antitoxina o los que proveen inmunidad a agentes tóxicos o a fagos (Martínez-Carranza et al., 2018a; Melnicáková et al., 2015).

En la Figura 14 se muestra una representación esquemática de nuestra hipótesis de la esencialidad de *bcsB*, que es un gen esencial condicional de la producción de celulosa, ya que *bcsB* participa en la translocación de del polímero y por lo tanto la delección del gen *bcsB* en una cepa capaz de producir celulosa es letal, pues no es posible translocar la celulosa y ésta se acumula en el interior de la célula produciendo un efecto tóxico. El gen *bcsB* no es considerado esencial en el último trabajo sobre los genes esenciales en *E. coli* K-12 (Goodall et al., 2018), debido a que el operón *bcs* no es expresado en estas cepas, esto es lo esperado.

Recientemente se han desarrollado nuevas estrategias para la identificación de genes esenciales condicionales en bacterias, estas técnicas aprovechan la tecnología de mutagénesis por inserción de transposones (Mobegi et al., 2017;

Zhao et al., 2017). Con estas metodologías, se han estudiado los genes esenciales de muchas especies bacterianas. En nuestro trabajo proponemos que *bcsB* es un gen esencial condicional, que depende de la producción de celulosa en la célula, aunque había sido propuesto como un gen esencial (Baba et al., 2006), con nuestro trabajo se ha concluido que la esencialidad depende del contexto intrínseco de la célula bacteriana. Sin embargo, no es el único caso, recientemente se han publicado otros trabajos de genes que habían sido considerados esenciales en otras especies, es el caso del regulador de la captación de hierro (Fur) en *P. aeruginosa*, en la que varios intentos sin éxito para obtener mutantes aumentaban la creencia de la esencialidad de ese gen en *P. aeruginosa*, hasta que recientemente se probó que *fur* es esencial para el crecimiento de la bacteria en medio sólido y es necesario para el crecimiento en colonias (Pasqua et al., 2017). Los genes que participan en sistemas toxina – antitoxina también son genes esenciales condicionales de la producción de la toxina, ya que de otro modo dejan de ser esenciales.



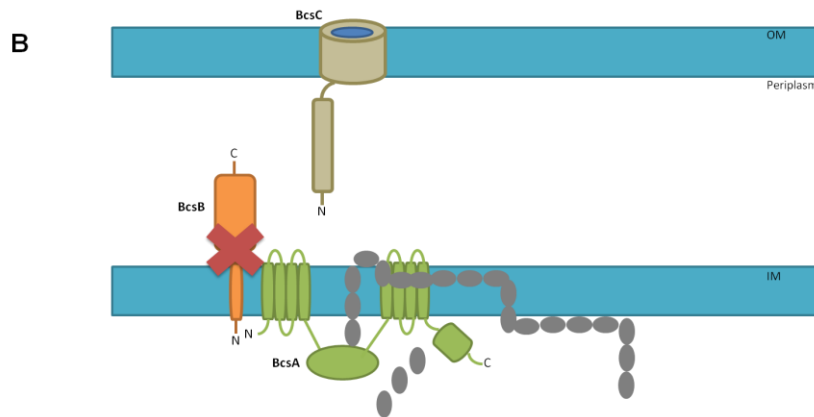


Figura 14 | Esquema de los componentes del sistema de síntesis funcional de celulosa en *E. coli*, cuando se produce celulosa y la célula es capaz de traslocarla (A), modificada de McNamara et al., (2015). Sin embargo, cuando BcsB, que participa en la traslocación de celulosa no es funcional, la célula no puede traslocar la celulosa y esto produce toxicidad a la célula, lo que es letal.

En *P. aeruginosa* son también variados los trabajos que han postulado conjuntos de genes esenciales en varias condiciones de cultivo y con diferentes estrategias experimentales (Gallagher et al., 2011; Jacobs et al., 2003; Juhas, 2015; Lee et al., 2015; Rusmini et al., 2014; Skurnik et al., 2013; Turner et al., 2015). En algunos de estos trabajos se ha reportado que los genes esenciales son muy variables de la cepa PA14 a la cepa PAO1 (Skurnik et al., 2013; Turner et al., 2015), lo que es muy impresionante debido a la alta conservación de los genomas entre las cepas. Como se mencionaba anteriormente estos postulados contrastan con la teoría de evolución de los genomas bacterianos, ya que los genes esenciales formarían parte del genoma accesorio en el caso de la alta variabilidad en los genes esenciales. Es muy importante entonces encontrar ejemplos fidedignos de la variabilidad de los genes esenciales en bacterias, así como se han reportado casos de THG de genes esenciales y especies bacterianas con incoherencias genómicas muy importantes (González-Casanova et al., 2014; Sato and Miyazaki, 2017; Zhaxybayeva et al., 2009).

En este trabajo se analizó la variabilidad de los genes esenciales reportada en los trabajos de *P. aeruginosa* entre las cepas PAO1 y PA14. Se ha reportado que la cepa de *P. aeruginosa* PA14 contiene 210 genes esenciales que están ausentes en el genoma de la cepa PAO1 (Skurnik et al., 2013), sin embargo, pudimos observar que de estos 210 genes sólo 8 genes están realmente ausentes en el genoma de la cepa PAO1. Esto significa que la variabilidad de los genes esenciales reportada entre estas dos cepas en realidad es mucho menor de lo que se ha postulado, con nuestro análisis, la variabilidad de los 210 genes esenciales reportada puede ser descartada.

Es muy probable que se hayan reportado esos 210 genes esenciales como ausentes en la cepa PAO1 debido al tipo de análisis bioinformático que realizaron en ese trabajo, buscando únicamente por anotación. Debe resaltarse la falta de anotación gran cantidad de estos genes en el genoma de la cepa PAO1, que es uno de los genomas de mejor calidad de secuenciación y anotación, sobreestimando la variabilidad.

Otro estudio que analizó los genes esenciales en *P. aeruginosa* PAO1 para su crecimiento en esputo de pacientes con fibrosis quística (Turner et al., 2015), reportó un total de 336 genes esenciales en estas condiciones, pero además analizó también los genes esenciales en estas condiciones en la cepa PA14, y determinó que 434 de 5906 genes son esenciales, y que el crecimiento en medio MOPS-esputo reveló otros 122 genes esenciales adicionales, reportó que 20 genes esenciales de los 434 no tienen ortólogo en la cepa PAO1.

En nuestro trabajo analizamos detalladamente los 336 genes esenciales propuestos en el trabajo de Turner et al., (2015) para el crecimiento de la cepa PAO1 en esputo de pacientes con fibrosis quística en el genoma de la cepa PA14, en total se identificaron 12 genes ausentes en esta cepa, aunque como se comenta en los resultados, estos 12 genes forman parte de un grupo, se encuentran en la misma región genómica y participan en la síntesis de

lipopolisacáridos del antígeno O, lo que podría indicar la nula esencialidad de esos 12 genes en el genoma de la cepa PA14, puesto que esta cepa puede producir los lipopolisacáridos por medio de otras vías (Whitfield and Trent, 2014). Este resultado de nueva cuenta muestra la baja variabilidad de los genes esenciales en *P. aeruginosa*.

En un estudio que analizó los genes esenciales de *P. aeruginosa* PAO1 por inserción saturada de transposones y en un total de 6 condiciones de cultivo (Lee et al., 2015), se propusieron 352 genes esenciales generales. Este conjunto de genes esenciales generales podría ser considerado como los genes esenciales de *P. aeruginosa* y de esta manera estar conservados en la cepa PA14, pero para analizar su conservación, hicimos una búsqueda de los 352 en el genoma de la cepa PA14. Como resultado de nuestro análisis se encontró que sólo dos genes de los 352 propuestos están ausentes en el genoma de la cepa PA14: el gen *wbpB* y el gen *imm2* que participan en la síntesis de lipopolisacáridos del antígeno O y en la resistencia a piocinas S2. La esencialidad de ambos genes podría considerarse que es nula en la cepa PA14, ya que como se discutía anteriormente, la síntesis de lipopolisacáridos puede realizarse por diversas vías (Whitfield and Trent, 2014) y si esta cepa no produce la piocina, la resistencia no es necesaria. Considerando que ambos genes pueden eliminarse del conjunto de genes esenciales generales debido a que no se conservan en la cepa PA14, se puede proponer a estos 350 genes como los genes esenciales de *P. aeruginosa*.

Varios estudios han postulado que *A. vinelandii* debería ser considerada como un miembro del género *Pseudomonas* basados en análisis filogenéticos de RNAr 16S y también con proteínas conservadas (Özen and Ussery, 2012; Rediers et al., 2004). En otros estudios (Setubal et al., 2009) se ha propuesto a *A. vinelandii* como un grupo externo de *Pseudomonas*, sin embargo, es muy interesante que dos especies bacterianas tan distintas fisiológica y fenotípicamente puedan llegar a confundirse en estudios filogenéticos. El género *Pseudomonas* se distingue por ser una bacteria ubicua y metabólicamente versátil e incluso con algunas especies

como *P. aeruginosa* que son patógenos de plantas y animales mientras que *A. vinelandii* es una bacteria fijadora de nitrógeno y formadora de quistes. Recientemente se ha reportado que casi la mitad de los genes de *A. vinelandii* codifican proteínas con un grado de homología superior al 67% con *Pseudomonas* que podría considerarse el esqueleto genómico que comparte con *Pseudomonas* y la otra mitad de los genes tiene origen polifilético.

En este trabajo tratamos de apoyar la hipótesis de que *P. aeruginosa* y *A. vinelandii* comparten un esqueleto genómico obtenido por herencia vertical de un ancestro en común, así que realizamos la búsqueda de los 352 genes esenciales generales de *P. aeruginosa* propuestos por Lee et al., (2015) en *A. vinelandii*. De los 352 genes esenciales generales, encontramos que sólo 10 genes están ausentes en *A. vinelandii*, sin embargo, dos genes (*nadE* y PA0092) han podido ser mutados por transposones en estudios previos (Jacobs et al., 2003), por lo que no pueden considerarse esenciales, los productos de otros 5 genes (*wbpB*, *imm2*, PA4674, PA0906 y *prtN*) participan en actividades biológicas poco relevantes en *A. vinelandii*, como son la síntesis de lipopolisacáridos, inmunidad a piocinas, antitoxina HigA o represores de genes de fagos o piocinas (Tabla 9). Los tres genes esenciales generales restantes están anotados como proteínas hipotéticas y por lo tanto se desconoce su función.

Por lo tanto, con este análisis se apoya la hipótesis de que *P. aeruginosa* y *A. vinelandii* comparten un esqueleto genómico heredado verticalmente y que se correlaciona con el alto grado de conservación de genes esenciales y alto grado de homología en el casi 50% de los genes de *A. vinelandii* con *Pseudomonas* (González-Casanova et al., 2014). Aunque la estructura de los cromosomas de ambas especies es interdispersa cuando se alinean (Figura 12), comparte muchas regiones homólogas y conserva la sintenia en regiones pequeñas, en las que se encuentran gran parte de los 342 genes esenciales de *P. aeruginosa* que comparten. Así pues, aunque no conserva la estructura sinténicamente a nivel general en todo el cromosoma, las regiones homólogas que conserva si tienen

conservación de microsintenia en casi todos los ortólogos de los genes esenciales de *Pseudomonas* (Figura 13), ya que únicamente los genes *lptE*, *lptD*, *dxs*, *yidC* y PA0728 de los 342 genes compartidos por ambas especies no comparten la microsintenia.

VIII.- CONCLUSIONES

- No se encontraron casos de “desplazamiento de genes no ortólogos” en los genomas de *Escherichia coli* ni en *Pseudomonas aeruginosa*.
- *bcsB* es un gen esencial en el contexto de la síntesis funcional de celulosa en *Escherichia coli*.
- Los genes esenciales de *P. aeruginosa* no son variables como se ha reportado, están muy conservados entre las cepas PAO1 y PA14.
- Los genes esenciales de *P. aeruginosa* están muy conservados en el genoma de *A. vinelandii*.

IX.- ANEXOS

Tabla 10 | Conjunto de 210 genes esenciales reportados por Skurnik et al., 2013, como ausentes en el genoma de *P. aeruginosa* PAO1.

Genes esenciales de <i>P. aeruginosa</i> PA14 reportados por Skurnik et al., 2013				
PA14_73420	PA14_55840	PA14_38910	PA14_19090	PA14_06090
PA14_73280	PA14_55660	PA14_38395	PA14_18890	PA14_05120

PA14_73240	PA14_55637	PA14_38310	PA14_18320	PA14_04970
PA14_73230	PA14_55635	PA14_37840	PA14_17710	PA14_04460
PA14_71400	PA14_55634	PA14_37470	PA14_17340	PA14_02610
PA14_70900	PA14_55631	PA14_36940	PA14_17180	PA14_02570
PA14_70450	PA14_55530	PA14_36920	PA14_17070	PA14_02490
PA14_70440	PA14_55460	PA14_36860	PA14_16700	PA14_00800
PA14_69820	PA14_55340	PA14_36790	PA14_16510	PA14_00290
PA14_68955	PA14_55290	PA14_36660	PA14_16200	PA14_00050
PA14_68700	PA14_54790	PA14_36620	PA14_15960	
PA14_68210	PA14_54730	PA14_36460	PA14_15880	
PA14_68170	PA14_54660	PA14_36270	PA14_15530	
PA14_68010	PA14_54420	PA14_35970	PA14_15470	
PA14_67960	PA14_54350	PA14_34780	PA14_15340	
PA14_66910	PA14_53870	PA14_34640	PA14_14975	
PA14_66750	PA14_52850	PA14_34300	PA14_14890	
PA14_66250	PA14_52640	PA14_34070	PA14_14460	
PA14_66240	PA14_52550	PA14_34020	PA14_14310	
PA14_66220	PA14_51230	PA14_34000	PA14_13730	
PA14_66210	PA14_49510	PA14_33550	PA14_13320	
PA14_65380	PA14_49230	PA14_33160	PA14_12470	
PA14_65210	PA14_49220	PA14_33060	PA14_12230	
PA14_65180	PA14_48980	PA14_32600	PA14_12050	
PA14_64530	PA14_48900	PA14_32570	PA14_11570	
PA14_64000	PA14_48320	PA14_32220	PA14_11430	
PA14_63750	PA14_46950	PA14_31530	PA14_11380	
PA14_63530	PA14_46160	PA14_31250	PA14_10780	

PA14_63420	PA14_45300	PA14_31080	PA14_10490
PA14_63360	PA14_45020	PA14_30720	PA14_10270
PA14_63050	PA14_44670	PA14_28880	PA14_09450
PA14_62960	PA14_43970	PA14_28720	PA14_09050
PA14_62860	PA14_43690	PA14_28660	PA14_09020
PA14_62800	PA14_43370	PA14_28190	PA14_08990
PA14_62120	PA14_43040	PA14_27610	PA14_08880
PA14_62080	PA14_42610	PA14_27210	PA14_08870
PA14_61780	PA14_42600	PA14_27190	PA14_08860
PA14_61760	PA14_42530	PA14_26485	PA14_08850
PA14_60160	PA14_42510	PA14_26480	PA14_08820
PA14_59110	PA14_42490	PA14_26340	PA14_08790
PA14_58980	PA14_42410	PA14_25500	PA14_08690
PA14_58940	PA14_41340	PA14_25490	PA14_08670
PA14_58130	PA14_41290	PA14_24410	PA14_08660
PA14_57960	PA14_41210	PA14_24220	PA14_08580
PA14_57930	PA14_41200	PA14_24120	PA14_08110
PA14_57810	PA14_40750	PA14_24050	PA14_07960
PA14_57460	PA14_40610	PA14_21560	PA14_07550
PA14_57450	PA14_39925	PA14_20940	PA14_07190
PA14_57370	PA14_39700	PA14_19580	PA14_07130
PA14_57300	PA14_39260	PA14_19510	PA14_07090

Tabla 11 | Conjunto de 352 genes esenciales generales reportados en *P. aeruginosa* PAO1 (Lee et al., 2015).

<i>Genes esenciales generales en P. aeruginosa PAO1 propuestos por Lee et al., 2015</i>					
<i>dnaA</i>	<i>lpdG</i>	<i>frr</i>	<i>rplB</i>	PA4674	<i>mpA</i>
<i>dnaN</i>	<i>acnB</i>	<i>pyrH</i>	<i>rplW</i>	PA4679	<i>rpmH</i>
<i>gyrB</i>	PA1792	<i>tsf</i>	<i>rplD</i>	PA4685	
PA0006	<i>glnS</i>	<i>rpsB</i>	<i>rplC</i>	<i>pssA</i>	
<i>glyS</i>	<i>cysS</i>	<i>dapD</i>	<i>rpsJ</i>	<i>folK</i>	
<i>glyQ</i>	<i>folD</i>	PA3685	<i>fusA1</i>	<i>pnp</i>	
<i>fmt</i>	<i>rnhA</i>	<i>adk</i>	<i>rpsG</i>	<i>infB</i>	
<i>def</i>	<i>dnaQ</i>	<i>lysS</i>	<i>rpsL</i>	<i>nusA</i>	
PA0022	<i>pgsA</i>	<i>prfB</i>	<i>rpoC</i>	<i>tpiA</i>	
<i>hemF</i>	PA2605	<i>rplS</i>	<i>rpoB</i>	<i>glmM</i>	
PA0092	PA2606	<i>rpsP</i>	<i>rplL</i>	<i>folP</i>	
<i>rpiA</i>	PA2607	<i>ffh</i>	<i>rplJ</i>	<i>dapB</i>	
<i>lgt</i>	PA2608	<i>guaA</i>	<i>rplA</i>	<i>dnaK</i>	
<i>thyA</i>	<i>serS</i>	<i>guaB</i>	<i>rplK</i>	<i>grpE</i>	
<i>folA</i>	<i>lolA</i>	PA3799	<i>secE</i>	<i>fur</i>	
<i>coaD</i>	<i>ftsK</i>	<i>hisS</i>	PA4279	<i>omlA</i>	
<i>ftsY</i>	<i>infA</i>	<i>gcpE</i>	<i>birA</i>	<i>accB</i>	
<i>rpoH</i>	<i>trmU</i>	<i>fdx2</i>	<i>sodB</i>	<i>accC</i>	
PA0404	<i>purB</i>	<i>hscA</i>	<i>groEL</i>	<i>nadE</i>	
PA0442	<i>pheT</i>	<i>hscB</i>	<i>groES</i>	<i>dnaB</i>	
<i>bioF</i>	<i>pheS</i>	<i>iscA</i>	<i>secA</i>	<i>rpsR</i>	
<i>metK</i>	<i>rplT</i>	<i>iscU</i>	PA4405	<i>rpsF</i>	
<i>tktA</i>	<i>rpml</i>	<i>iscS</i>	<i>lpxC</i>	<i>purA</i>	

<i>fda</i>	<i>infC</i>	<i>secF</i>	<i>ftsZ</i>	<i>PA4948</i>
<i>rpoD</i>	<i>thrS</i>	<i>secD</i>	<i>ftsA</i>	<i>psd</i>
<i>dnaG</i>	<i>etfA</i>	<i>PA3827</i>	<i>ftsQ</i>	<i>parC</i>
<i>rpsU</i>	<i>etfB</i>	<i>PA3828</i>	<i>murC</i>	<i>parE</i>
<i>gcp</i>	<i>PA2953</i>	<i>valS</i>	<i>murG</i>	<i>waaA</i>
<i>pdxA</i>	<i>holB</i>	<i>hemL</i>	<i>ftsW</i>	<i>PA4991</i>
<i>surA</i>	<i>tmk</i>	<i>PA3982</i>	<i>murD</i>	<i>PA4992</i>
<i>lptD</i>	<i>acpP</i>	<i>Int</i>	<i>mraY</i>	<i>rfaE</i>
<i>rpe</i>	<i>fabG</i>	<i>leuS</i>	<i>murF</i>	<i>msbA</i>
<i>prtR</i>	<i>fabD</i>	<i>lptE</i>	<i>murE</i>	<i>PA4998</i>
<i>PA0655</i>	<i>rne</i>	<i>holA</i>	<i>ftsI</i>	<i>PA5006</i>
<i>PA0661</i>	<i>murB</i>	<i>lis</i>	<i>ftsL</i>	<i>PA5008</i>
<i>PA0665</i>	<i>kdsB</i>	<i>lipB</i>	<i>gmhA</i>	<i>waaP</i>
<i>PA0728.1</i>	<i>lpxK</i>	<i>rodA</i>	<i>rpsI</i>	<i>waaG</i>
<i>PA0759</i>	<i>PA2986</i>	<i>pbpA</i>	<i>rplM</i>	<i>waaC</i>
<i>mucA</i>	<i>PA2987</i>	<i>ppa</i>	<i>trpS</i>	<i>waaF</i>
<i>lepB</i>	<i>PA2988</i>	<i>dxs</i>	<i>murA</i>	<i>hemE</i>
<i>era</i>	<i>topA</i>	<i>thiL</i>	<i>kdsD</i>	<i>argS</i>
<i>pdxJ</i>	<i>PA3088</i>	<i>nusB</i>	<i>PA4459</i>	<i>ubiE</i>
<i>alas</i>	<i>folC</i>	<i>ribE</i>	<i>PA4460</i>	<i>ubiB</i>
<i>lysC</i>	<i>accD</i>	<i>ribB</i>	<i>PA4461</i>	<i>fbp</i>
<i>PA0906</i>	<i>asd</i>	<i>ribC</i>	<i>mreD</i>	<i>glnA</i>
<i>PA0947</i>	<i>gltX</i>	<i>ribD</i>	<i>mreC</i>	<i>PA5221</i>
<i>proS</i>	<i>wbpB</i>	<i>PA4068</i>	<i>mreB</i>	<i>ubiH</i>
<i>aspS</i>	<i>rpsA</i>	<i>ssb</i>	<i>gatC</i>	<i>PA5237</i>
<i>tolQ</i>	<i>serC</i>	<i>rplQ</i>	<i>gatA</i>	<i>rho</i>

<i>tolR</i>	<i>gyrA</i>	<i>rpoA</i>	<i>gatB</i>	<i>hemB</i>
<i>tolA</i>	<i>ubiG</i>	<i>rpsD</i>	<i>comL</i>	<i>hemD</i>
<i>tolB</i>	<i>fpr</i>	<i>rpsK</i>	<i>lytB</i>	<i>hemC</i>
<i>oprL</i>	<i>PA3481</i>	<i>rpsM</i>	<i>lspA</i>	<i>dapF</i>
<i>dapA</i>	<i>metG</i>	<i>secY</i>	<i>ileS</i>	<i>rpmB</i>
<i>pdxH</i>	<i>ygbB</i>	<i>rplO</i>	<i>ribF</i>	<i>coaC</i>
<i>imm2</i>	<i>ygbP</i>	<i>rpmD</i>	<i>PA4562</i>	<i>gmk</i>
<i>nrdB</i>	<i>PA3634</i>	<i>rpsE</i>	<i>rpsT</i>	<i>ubiA</i>
<i>nrdA</i>	<i>kdsA</i>	<i>rplR</i>	<i>obg</i>	<i>PA5492</i>
<i>dapE</i>	<i>pyrG</i>	<i>rplF</i>	<i>rpmA</i>	<i>glmS</i>
<i>pcpS</i>	<i>PA3638</i>	<i>rpsH</i>	<i>rplU</i>	<i>glmU</i>
<i>zipA</i>	<i>accA</i>	<i>rpsN</i>	<i>ispB</i>	<i>atpC</i>
<i>lig</i>	<i>dnaE</i>	<i>rplE</i>	<i>PA4636</i>	<i>atpD</i>
<i>dnaX</i>	<i>lpxB</i>	<i>rplX</i>	<i>hemH</i>	<i>atpG</i>
<i>PA1564</i>	<i>lpxA</i>	<i>rplN</i>	<i>murI</i>	<i>atpA</i>
<i>sdhC</i>	<i>fabZ</i>	<i>rpsQ</i>	<i>prfA</i>	<i>atpH</i>
<i>sdhD</i>	<i>lpxD</i>	<i>rpmC</i>	<i>hemA</i>	<i>atpF</i>
<i>sdhA</i>	<i>opr86</i>	<i>rplP</i>	<i>PA4668</i>	<i>atpE</i>
<i>sdhB</i>	<i>dxr</i>	<i>rpsC</i>	<i>ipk</i>	<i>atpB</i>
<i>sucA</i>	<i>cdsA</i>	<i>rplV</i>	<i>prs</i>	<i>atpI</i>
<i>sucB</i>	<i>uppS</i>	<i>rpsS</i>	<i>PA4672</i>	<i>yidC</i>

Tabla 12 | Conjunto de genes esenciales en *P. aeruginosa* PAO1 para su desarrollo en medio con esputo de fibrosis quística (Turner et al., 2015).

<i>Genes esenciales en P. aeruginosa PAO1 propuestos por Turner et al., 2015</i>				
<i>dnaA</i>	<i>sigX</i>	<i>lpxD</i>	<i>nusG</i>	<i>nadE</i>
<i>dnaN</i>	<i>acnB</i>	<i>opr86</i>	<i>secE</i>	<i>dnaB</i>
<i>gyrB</i>	PA1792	<i>dxr</i>	PA4279	<i>rpsF</i>
PA0006	<i>glnS</i>	<i>cdsA</i>	PA4333	<i>purA</i>
<i>glyS</i>	<i>cysS</i>	<i>uppS</i>	<i>sodB</i>	<i>psd</i>
<i>glyQ</i>	<i>folD</i>	<i>frr</i>	<i>groEL</i>	<i>parC</i>
<i>fmt</i>	<i>rnhA</i>	<i>pyrH</i>	<i>secA</i>	<i>parE</i>
<i>def</i>	<i>dnaQ</i>	<i>tsf</i>	PA4405	<i>cpdA</i>
PA0022	PA1969	<i>dapD</i>	<i>lpxC</i>	<i>waaA</i>
<i>hemF</i>	PA2024	PA3685	<i>ftsZ</i>	PA4991
PA0123	PA2156	<i>lysS</i>	<i>ftsA</i>	PA4992
<i>rpiA</i>	<i>pgsA</i>	<i>prfB</i>	<i>murC</i>	<i>rfaE</i>
<i>lgt</i>	PA2605	<i>wspF</i>	<i>murG</i>	<i>msbA</i>
<i>thyA</i>	<i>serS</i>	<i>nalC</i>	<i>ftsW</i>	PA4998
<i>folA</i>	<i>lolA</i>	<i>rimM</i>	<i>murD</i>	<i>waaL</i>
<i>coaD</i>	<i>ftsK</i>	PA3767	<i>mraY</i>	PA5006
<i>rpoH</i>	<i>trmU</i>	<i>guaA</i>	<i>murF</i>	PA5008
PA0404	<i>purB</i>	<i>guaB</i>	<i>murE</i>	<i>waaP</i>
<i>codB</i>	PA2726	PA3799	<i>ftsI</i>	<i>waaG</i>
<i>tkiA</i>	<i>pheT</i>	<i>hisS</i>	PA4420	<i>waaC</i>
<i>epd</i>	<i>pheS</i>	<i>gcpE</i>	PA4425	<i>waaF</i>
<i>fda</i>	<i>rplT</i>	<i>hscA</i>	PA4429	<i>hemE</i>
<i>rpoD</i>	<i>infC</i>	<i>hscB</i>	<i>rpsI</i>	<i>argS</i>

<i>dnaG</i>	<i>thrS</i>	<i>iscU</i>	<i>rplM</i>	<i>ubiE</i>
<i>gcp</i>	PA2798	<i>iscS</i>	<i>trpS</i>	<i>ubiB</i>
<i>folB</i>	PA2950	<i>secD</i>	<i>murA</i>	<i>fbp</i>
<i>pdxA</i>	<i>etfA</i>	PA3827	PA4457	<i>glnA</i>
<i>surA</i>	<i>etfB</i>	PA3828	PA4459	<i>rmlB</i>
<i>rpe</i>	PA2953	<i>valS</i>	PA4460	<i>rmlD</i>
<i>priR</i>	<i>tmk</i>	<i>hemL</i>	PA4461	<i>rmlA</i>
PA0655	<i>fabG</i>	PA3982	<i>mreD</i>	<i>rmlC</i>
PA0665	<i>murB</i>	<i>Int</i>	<i>mreC</i>	PA5174
PA0678	<i>kdsB</i>	<i>leuS</i>	<i>mreB</i>	PA5221
PA0759	<i>lpxK</i>	<i>holA</i>	<i>gatA</i>	<i>ubiH</i>
<i>lepB</i>	PA2986	<i>lipB</i>	<i>gatB</i>	PA5237
<i>era</i>	PA2987	<i>rodA</i>	<i>coaE</i>	<i>rho</i>
<i>pdxJ</i>	PA2988	<i>pbpA</i>	<i>lytB</i>	<i>hemB</i>
<i>alaS</i>	<i>topA</i>	PA4019	<i>lspA</i>	<i>hemD</i>
<i>lysC</i>	PA3088	<i>ppa</i>	<i>ileS</i>	<i>hemC</i>
PA0906	<i>accD</i>	<i>ispA</i>	<i>ribF</i>	<i>dapF</i>
PA0947	<i>asd</i>	<i>dxs</i>	PA4562	<i>rep</i>
<i>proS</i>	<i>gltX</i>	<i>thiL</i>	<i>ispB</i>	<i>coaC</i>
<i>aspS</i>	<i>wbpM</i>	<i>nusB</i>	<i>glyA3</i>	<i>dut</i>
<i>tolQ</i>	<i>wbpL</i>	PA4068	PA4636	<i>gmk</i>
<i>tolR</i>	<i>wbpJ</i>	PA4093	<i>hemH</i>	<i>ubiA</i>
<i>tolA</i>	<i>wbpI</i>	<i>rplQ</i>	<i>hemK</i>	PA5492
<i>tolB</i>	<i>wbpH</i>	<i>rpoA</i>	<i>prfA</i>	<i>glmS</i>
<i>dapA</i>	<i>wbpG</i>	<i>rpsD</i>	PA4668	<i>glmU</i>
PA1056	<i>hisF2</i>	<i>rpsK</i>	<i>ipk</i>	<i>atpD</i>

<i>PA1059</i>	<i>wzx</i>	<i>rpsM</i>	<i>prs</i>	<i>atpG</i>
<i>nrdB</i>	<i>wzy</i>	<i>secY</i>	<i>PA4672</i>	<i>atpA</i>
<i>nrdA</i>	<i>wbpE</i>	<i>rplO</i>	<i>PA4679</i>	<i>atpH</i>
<i>dapE</i>	<i>wbpD</i>	<i>rpsE</i>	<i>PA4685</i>	<i>atpF</i>
<i>pdxB</i>	<i>wbpB</i>	<i>rplF</i>	<i>folK</i>	<i>atpB</i>
<i>ccmA</i>	<i>wbpA</i>	<i>rpsH</i>	<i>pnp</i>	<i>PA5568</i>
<i>zipA</i>	<i>rpsA</i>	<i>rplE</i>	<i>infB</i>	<i>rnpA</i>
<i>lig</i>	<i>serC</i>	<i>rplP</i>	<i>nusA</i>	
<i>dnaX</i>	<i>gyrA</i>	<i>rpsC</i>	<i>PA4746</i>	
<i>sdhC</i>	<i>ubiG</i>	<i>rplB</i>	<i>tpiA</i>	
<i>sdhD</i>	<i>fpr</i>	<i>rplD</i>	<i>glmM</i>	
<i>sdhA</i>	<i>PA3481</i>	<i>rplC</i>	<i>folP</i>	
<i>sdhB</i>	<i>ygbB</i>	<i>fusA1</i>	<i>ftsH</i>	
<i>sucA</i>	<i>kdsA</i>	<i>rpsG</i>	<i>ftsJ</i>	
<i>sucB</i>	<i>pyrG</i>	<i>rpsL</i>	<i>dapB</i>	
<i>lpdG</i>	<i>PA3638</i>	<i>rpoC</i>	<i>dnaK</i>	
<i>sucC</i>	<i>accA</i>	<i>rpoB</i>	<i>fur</i>	
<i>sucD</i>	<i>dnaE</i>	<i>rplL</i>	<i>omlA</i>	
<i>fabB</i>	<i>lpxB</i>	<i>rplJ</i>	<i>PA4767</i>	
<i>fabA</i>	<i>lpxA</i>	<i>rplA</i>	<i>accB</i>	
<i>aroC</i>	<i>fabZ</i>	<i>rplK</i>	<i>accC</i>	

Tabla 13 | Plásmidos utilizados en este trabajo.

Plásmido	Característica	Referencia
pUCP20	Vector de clonación	(West et al., 1996)
pIJ773	Origen del casete de resistencia a apramicina	(Gust et al., 2003)
pIJ790	Recombinasa lambda Red; Cm ^R	(Gust et al., 2000)
pUCP-<i>bcsAB</i>	<i>bcsA</i> y <i>bcsB</i> clonados en el vector pUCP20	Este trabajo
pUCP-<i>bcsA</i>	<i>bcsA</i> clonado en el vector pUCP20	Este trabajo

X.- PUBLICACIONES

Producto de la presente tesis, se lograron publicar dos artículos, uno en la revista *Frontiers in Microbiology* y otro en la revista *FEMS Microbiology Letters*, un tercer artículo se encuentra sometido a la revista *Microbiology*. Los artículos se anexan a continuación.

XI.- REFERENCIAS

- Acevedo-Rocha, C. G., Fang, G., Schmidt, M., Ussery, D. W., and Danchin, A. (2013). From essential to persistent genes: A functional approach to constructing synthetic life. *Trends Genet.* 29, 273–279. doi:10.1016/j.tig.2012.11.001.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi:10.1093/nar/25.17.3389.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* 2. doi:10.1038/msb4100050.
- Bergmiller, T., Ackermann, M., and Silander, O. K. (2012). Patterns of evolutionary conservation of essential genes correlate with their compensability. *PLoS Genet.* 8. doi:10.1371/journal.pgen.1002803.
- Blath, J., Casanova, A. G., Kurt, N., and Wilke-Berenguer, M. (2016). A new coalescent for seed-bank models. *Ann. Appl. Probab.* 26, 857–891. doi:10.1214/15-AAP1106.
- Bobay, L. M., and Ochman, H. (2017). The evolution of bacterial genome architecture. *Front. Genet.* 8, 1–6. doi:10.3389/fgene.2017.00072.
- Buldum, G., Bismarck, A., and Mantalaris, A. (2018). Recombinant biosynthesis of bacterial cellulose in genetically modified *Escherichia coli*. *Bioprocess Biosyst. Eng.* 41, 265–279. doi:10.1007/s00449-017-1864-1.
- Carver, T., Harris, S. R., Berriman, M., Parkhill, J., and McQuillan, J. A. (2012). Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28, 464–469. doi:10.1093/bioinformatics/btr703.
- Chao, M. C., Abel, S., Davis, B. M., and Waldor, M. K. (2016). The design and analysis of transposon insertion sequencing experiments. *Nat. Rev. Microbiol.* 14, 119–128. doi:10.1038/nrmicro.2015.7.
- Charlebois, R. L., and Doolittle, W. F. (2004). Computing prokaryotic gene ubiquity: Rescuing the core from extinction. *Genome Res.* 14, 2469–2477. doi:10.1101/gr.3024704.

- Ciccarelli, F. D., Doerks, T., Von Mering, C., Creevey, C. J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science (80-)*. 311, 1283–1287. doi:10.1126/science.1123061.
- Cook, H., and Ussery, D. W. (2013). Sigma factors in a thousand *E. coli* genomes. *Environ. Microbiol.* 15, 3121–3129. doi:10.1111/1462-2920.12236.
- Cordero, O. X., and Hogeweg, P. (2009). The impact of long-distance horizontal gene transfer on prokaryotic genome size. *Proc. Natl. Acad. Sci.* 106, 21748–21753. doi:10.1073/pnas.0907584106.
- Dagan, T., Artzy-Randrup, Y., and Martin, W. (2008). Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc. Natl. Acad. Sci.* 105, 10039–10044. doi:10.1016/j.ejphar.2005.08.004.
- Dame, R. T., Espéli, O., Grainger, D. C., and Wiggins, P. A. (2012). Multidisciplinary perspectives on bacterial genome organization and dynamics. *Mol. Microbiol.* 86, 1023–1030. doi:10.1111/mmi.12055.
- Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements. 1394–1403. doi:10.1101/gr.2289704.
- Darmon, E., and Leach, D. R. F. (2014). Bacterial Genome Instability. *Microbiol. Mol. Biol. Rev.* 78, 1–39. doi:10.1128/MMBR.00035-13.
- Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci.* 97, 6640–6645. doi:10.1073/pnas.120163297.
- de Berardinis, V., Besnard, M., Durot, M., Schächter, V., Cruaud, C., Döring, V., et al. (2008). A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* 4. doi:10.1038/msb.2008.10.
- Deng, J., Tan, L., Lin, X., Lu, Y., and Lu, L. J. (2012). Exploring the optimal strategy to predict essential genes in microbes. *Biomolecules* 2, 1–22. doi:10.3390/biom2010001.
- Dorman, C. J. (2013). Genome architecture and global gene regulation in bacteria: Making progress towards a unified model? *Nat. Rev. Microbiol.* 11, 349–355. doi:10.1038/nrmicro3007.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science (80-)*. 269, 496–512. doi:10.1126/science.7542800.

- Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G., and Hanage, W. P. (2009). The Bacterial Species Challenge: Ecological Diversity. *Science* 323, 741–746. doi:10.1126/science.1159388.
- French, C. T., Lao, P., Matthews, B. T., Yu, H., Loraine, A. E., and Dybvig, K. (2008). Large-scale transposon mutagenesis of *Mycoplasma pulmonis*. *Mol. Microbiol.* 69, 67–76. doi:10.1111/j.1365-2958.2008.06262.x.
- Gallagher, L. A., Shendure, J., and Manoil, C. (2011). Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using Tn-seq. *MBio* 2, 1–8. doi:10.1128/mBio.00315-10.
- Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balázsi, G., Ravasz, E., Daugherty, M. D., et al. (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.* 185, 5673–5684. doi:10.1128/JB.185.19.5673-5684.2003.
- Gil, R., Silva, F. J., Peretó, J., and Pereto, J. (2004). Determination of the Core of a Minimal Bacterial Gene Set. *Microbiol. Mol. Biol. Rev.* 68, 518–537. doi:10.1128/MMBR.68.3.518.
- Glass, J. I., Assad-garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., et al. (2006). Essential genes of a minimal bacterium. 103, 425–430.
- Goh, S., Hohmeier, A., Stone, T. C., Offord, V., Sarabia, F., Garcia-Ruiz, C., et al. (2015). Silencing of essential genes within a highly coordinated operon in *Escherichia coli*. *Appl. Environ. Microbiol.* 81, 5650–5659. doi:10.1128/AEM.01444-15.
- González-Casanova, A., Aguirre-von-Wobeser, E., Espín, G., Servín-González, L., Kurt, N., Spanò, D., et al. (2014). Strong seed-bank effects in bacterial evolution. *J. Theor. Biol.* 356, 62–70. doi:10.1016/j.jtbi.2014.04.009.
- Goodall, E. C. A., Robinson, A., Johnston, I. G., Jabbari, S., Turner, K. A., Cunningham, A. F., et al. (2018). The essential genome of *Escherichia coli* K-12. *MBio* 9, 1–18. doi:10.1128/mBio.02096-17.
- Gopal, G. J., and Kumar, A. (2013). Strategies for the production of recombinant protein in *Escherichia coli*. *Protein J.* 32, 419–425. doi:10.1007/s10930-013-9502-5.
- Gust, B., Challis, G. L., Fowler, K., Kieser, T., and Chater, K. F. (2003). PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci.* 100, 1541–1546. doi:10.1073/pnas.0337542100.
- Gust, B., Kieser, T., and Chater, K. (2000). PCR targeting system in *Streptomyces*

coelicolor A3(2). *Norwich Res. Park* 3, 1–39.

- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. *Nat. Microbiol.* 1, 1–6. doi:10.1038/nmicrobiol.2016.48.
- Hutchison, C. A., Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science (80-)*. 351, 1414–1420. doi:10.1126/science.aad6253.
- Hutchison, C. A., Iii, C. A. H., Peterson, S. N., Gill, S. R., Smith, H. O., and Venter, J. C. (1999). Global Transposon Mutagenesis and a Minimal *Mycoplasma* Genome. *Science (80-)*. 286, 2165–2169. doi:10.1126/science.286.5447.2165.
- Hutchison, E. A., Miller, D. A., and Angert, E. R. (2014). Sporulation in Bacteria: Beyond the Standard Model. *Microbiol. Spectr.* 2, 1–15. doi:10.1128/microbiolspec.tbs-0013-2012.
- Ibberson, C. B., Stacy, A., Fleming, D., Dees, J. L., Rumbaugh, K., Gilmore, M. S., et al. (2017). Co-infecting microorganisms dramatically alter pathogen gene essentiality during polymicrobial infection. *Nat. Microbiol.* 2, 1–6. doi:10.1038/nmicrobiol.2017.79.
- Ish-Am, O., Kristensen, D. M., and Ruppin, E. (2015). Evolutionary conservation of bacterial essential metabolic genes across all bacterial culture media. *PLoS One* 10, 1–15. doi:10.1371/journal.pone.0123785.
- Jacobs, M. A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., et al. (2003). Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci.* 100, 14339–14344. doi:10.1073/pnas.2036282100.
- Jang, J., Yan, T., Ishii, S., Hur, H.-G., Sadowsky, M. J., and Byappanahalli, M. N. (2017). Environmental *Escherichia coli*: ecology and public health implications-a review. *J. Appl. Microbiol.* 123, 570–581. doi:10.1111/jam.13468.
- Ji, Y., Zhang, B., Van Horn, S. F., Warren, P., Woodnutt, G., Burnham, M. K. R., et al. (2001). Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science (80-)*. 293, 2266–2269. doi:10.1126/science.1063566.
- Jordan, I. K., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V. (2002). Essential Genes Are More Evolutionarily Conserved Than Are Nonessential Genes in Bacteria. *Genome Res.* 12, 962–968. doi:10.1101/gr.87702.

- Judson, N., and Mekalanos, J. J. (2000). Transposon-based approaches to identify essential bacterial genes. *Trends Microbiol.* 8, 521–526. doi:10.1016/S0966-842X(00)01865-5.
- Juhas, M. (2015). *Pseudomonas aeruginosa* essentials: An update on investigation of essential genes. *Microbiol. (United Kingdom)* 161, 2053–2060. doi:10.1099/mic.0.000161.
- Juhas, M., Eberl, L., and Church, G. M. (2012). Essential genes as antimicrobial targets and cornerstones of synthetic biology. *Trends Biotechnol.* 30, 601–607. doi:10.1016/j.tibtech.2012.08.002.
- Juhas, M., Eberl, L., and Glass, J. I. (2011). Essence of life: Essential genes of minimal genomes. *Trends Cell Biol.* 21, 562–568. doi:10.1016/j.tcb.2011.07.005.
- Juhas, M., Reuß, D. R., Zhu, B., and Commichau, F. M. (2014). *Bacillus subtilis* and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. *Microbiol. (United Kingdom)* 160, 2341–2351. doi:10.1099/mic.0.079376-0.
- Kato, J. I., and Hashimoto, M. (2007). Construction of consecutive deletions of the *Escherichia coli* chromosome. *Mol. Syst. Biol.* 3, 1–7. doi:10.1038/msb4100174.
- Klockgether, J., Cramer, N., Wiehlmann, L., Davenport, C. F., and Tümmler, B. (2011). *Pseudomonas aeruginosa* genomic structure and diversity. *Front. Microbiol.* 2, 1–18. doi:10.3389/fmicb.2011.00150.
- Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Arnaud, M., et al. (2003). Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci.* 100, 4678–4683. doi:10.1073/pnas.0730515100.
- Koonin, E. V. (2000). How many genes can make a cell: the minimal-gene-set concept. *Annu. Rev. Genomics Hum. Genet.*, 99–116.
- Koonin, E. V., Mushegian, A. R., Galperin, M. Y., and Walker, D. R. (1997). Comparison of archaeal and bacterial genomes: Computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Mol. Microbiol.* 25, 619–637. doi:10.1046/j.1365-2958.1997.4821861.x.
- Koskella, B., and Brockhurst, M. A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* 38, 916–931. doi:10.1111/1574-6976.12072.
- Lagesen, K., Ussery, D. W., and Wassenaar, T. M. (2010). Genome Update: the 1000th genome - a cautionary tale. *Microbiol. (United Kingdom)* 156, 603–608.

doi:10.1099/mic.0.038257-0.

- Lan, R., and Reeves, P. R. (1996). Gene transfer is a major factor in bacterial evolution. *Mol. Biol. Evol.* 13, 47–55. doi:10.1093/oxfordjournals.molbev.a025569.
- Land, M., Hauser, L., Jun, S. R., Nookaew, I., Leuze, M. R., Ahn, T. H., et al. (2015). Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* 15, 141–161. doi:10.1007/s10142-015-0433-4.
- Lang, A. S., Zhaxybayeva, O., and Beatty, J. T. (2012). Gene transfer agents: Phage-like elements of genetic exchange. *Nat. Rev. Microbiol.* 10, 472–482. doi:10.1038/nrmicro2802.
- Langridge, G. C., Maskell, D. J., Phan, M.-D., Charles, I., Turner, D. J., Parts, L., et al. (2009). Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Res.* 19, 2308–2316. doi:10.1101/gr.097097.109.
- Lawrence, J. G. (1999). Gene transfer, speciation, and the evolution of bacterial genomes. *Curr. Opin. Microbiol.* 2, 519–523. doi:10.1016/S1369-5274(99)00010-7.
- Lee, S. A., Gallagher, L. A., Thongdee, M., Staudinger, B. J., Lippman, S., Singh, P. K., et al. (2015). General and condition-specific essential functions of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci.* 112, 5189–5194. doi:10.1073/pnas.1422186112.
- Lex, A., Gehlenborg, N., Strobel, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* 20, 1983–1992. doi:10.1109/TVCG.2014.2346248.
- Liberati, N. T., Urbach, J. M., Miyata, S., Lee, D. G., Drenkard, E., Wu, G., et al. (2006). An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc. Natl. Acad. Sci.* 103, 2833–2838. doi:10.1073/pnas.0511100103.
- Lister, P. D., Wolter, D. J., and Hanson, N. D. (2009). Antibacterial-resistant *Pseudomonas aeruginosa*: Clinical impact and complex regulation of chromosomally encoded resistance mechanisms. *Clin. Microbiol. Rev.* 22, 582–610. doi:10.1128/CMR.00040-09.
- Loman, N. J., Constantinidou, C., Chan, J. Z. M., Halachev, M., Sergeant, M., Penn, C. W., et al. (2012). High-throughput bacterial genome sequencing: An embarrassment of choice, a world of opportunity. *Nat. Rev. Microbiol.* 10, 599–606. doi:10.1038/nrmicro2850.
- Loman, N. J., and Pallen, M. J. (2015). Twenty years of bacterial genome

- sequencing. *Nat. Rev. Microbiol.* 13, 787–794. doi:10.1038/nrmicro3565.
- Luo, H., Gao, F., and Lin, Y. (2015). Evolutionary conservation analysis between the essential and nonessential genes in bacterial genomes. *Sci. Rep.* 5, 1–8. doi:10.1038/srep13210.
- Luo, H., Lin, Y., Gao, F., Zhang, C. T., and Zhang, R. (2014). DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res.* 42, 574–580. doi:10.1093/nar/gkt1131.
- Martínez-Carranza, E., Barajas, H., Alcaraz, L. D., Servín-González, L., Ponce-Soto, G. Y., and Soberón-Chávez, G. (2018a). Variability of bacterial essential genes among closely related bacteria: The case of *Escherichia coli*. *Front. Microbiol.* 9, 1–7. doi:10.3389/fmicb.2018.01059.
- Martínez-Carranza, E., Servín-González, L., Cocotl-Yañez, M., and Soberón-Chávez, G. (2018b). The *Escherichia coli* bcsB gene is a conditional essential gene in the context of functional cellulose synthesis. *FEMS Microbiol. Lett.* 365, 1–6. doi:10.1093/femsle/fny232.
- McCutcheon, J. P., and Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26. doi:10.1038/nrmicro2670.
- McNamara, J. T., Morgan, J. L. W., and Zimmer, J. (2015). A Molecular Description of Cellulose Biosynthesis. *Annu. Rev. Biochem.* 84, 895–921. doi:10.1146/annurev-biochem-060614-033930.
- Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594. doi:10.1016/j.gde.2005.09.006.
- Mekalanos, J. J., Akerley, B. J., Amaya, K., Judson, N., Novick, V. L., and Rubin, E. J. (2002). A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc. Natl. Acad. Sci.* 99, 966–971. doi:10.1073/pnas.012602299.
- Melnicáková, J., Becárová, Z., Makroczyová, J., and Barák, I. (2015). Analysis of the *Bacillus cereus* SpoIIIS antitoxin-toxin system reveals its three-component nature. *Front. Microbiol.* 6, 1–11. doi:10.3389/fmicb.2015.00808.
- Meng, J., Kanzaki, G., Meas, D., Lam, C. K., Crummer, H., Tain, J., et al. (2012). A genome-wide inducible phenotypic screen identifies antisense RNA constructs silencing *Escherichia coli* essential genes. *FEMS Microbiol. Lett.* 329, 45–53. doi:10.1111/j.1574-6968.2012.02503.x.
- Mira, A., Ochman, H., and Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet.* 17, 589–596. doi:10.1016/S0168-

9525(01)02447-7.

- Mira, A., Pushker, R., Legault, B. A., Moreira, D., and Rodríguez-Valera, F. (2004). Evolutionary relationships of *Fusobacterium nucleatum* based on phylogenetic analysis and comparative genomics. *BMC Evol. Biol.* 4, 1–17. doi:10.1186/1471-2148-4-50.
- Mobegi, F. M., Zomer, A., de Jonge, M. I., and van Hijum, S. A. F. T. (2017). Advances and perspectives in computational prediction of microbial gene essentiality. *Brief. Funct. Genomics* 16, 70–79. doi:10.1093/bfpg/elv063.
- Moradali, M. F., Ghods, S., and Rehm, B. H. A. (2017). *Pseudomonas aeruginosa* Lifestyle: A Paradigm for Adaptation, Survival, and Persistence. *Front. Cell. Infect. Microbiol.* 7. doi:10.3389/fcimb.2017.00039.
- Morgan, J. L. W., McNamara, J. T., Fischer, M., Rich, J., Chen, H. M., Withers, S. G., et al. (2016). Observing cellulose biosynthesis and membrane translocation in crystallo. *Nature* 531, 329–334. doi:10.1038/nature16966.
- Morgan, J. L. W., McNamara, J. T., and Zimmer, J. (2014). Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nat. Struct. Mol. Biol.* 21, 489–496. doi:10.1038/nsmb.2803.
- Morgan, J. L. W., Strumillo, J., and Zimmer, J. (2013). Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* 493, 181–186. doi:10.1038/nature11744.
- Moriel, D., Tan, L., Goh, K., Phan, M., Ipe, D., Lo, A., et al. (2016). A Novel Protective Vaccine Antigen from the Core *Escherichia coli* Genome. *mSphere* 1, 1–13. doi:10.1128/mSphere.00326-16.Editor.
- Mosquera-Rendón, J., Rada-Bravo, A. M., Cárdenas-Brito, S., Corredor, M., Restrepo-Pineda, E., and Benítez-Páez, A. (2016). Pangenome-wide and molecular evolution analyses of the *Pseudomonas aeruginosa* species. *BMC Genomics* 17, 1–14. doi:10.1186/s12864-016-2364-4.
- Özen, A. I., and Ussery, D. W. (2012). Defining the *Pseudomonas* Genus: Where Do We Draw the Line with *Azotobacter*? *Microb. Ecol.* 63, 239–248. doi:10.1007/s00248-011-9914-8.
- Pasqua, M., Visaggio, D., Lo Sciuto, A., Genah, S., Banin, E., Visca, P., et al. (2017). Ferric Uptake Regulator Fur Is Conditionally Essential in *Pseudomonas aeruginosa*. *J. Bacteriol.* 199, 1–17.
- Penadés, J. R., Chen, J., Quiles-Puchalt, N., Carpena, N., and Novick, R. P. (2015). Bacteriophage-mediated spread of bacterial virulence genes. *Curr. Opin. Microbiol.* 23, 171–178. doi:10.1016/j.mib.2014.11.019.

- Peng, C., and Gao, F. (2014). Protein localization analysis of essential genes in prokaryotes. *Sci. Rep.* 4, 1–7. doi:10.1038/srep06001.
- Peng, C., Lin, Y., Luo, H., and Gao, F. (2017). A comprehensive overview of online resources to identify and predict bacterial essential genes. *Front. Microbiol.* 8, 1–13. doi:10.3389/fmicb.2017.02331.
- Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P., et al. (2008). The pangenome structure of *Escherichia coli*: Comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* 190, 6881–6893. doi:10.1128/JB.00619-08.
- Rediers, H., Vanderleyden, J., and De Mot, R. (2004). *Azotobacter vinelandii*: a *Pseudomonas* in disguise? *Microbiology* 150, 1117–9. doi:10.1099/mic.0.27096-0.
- Römling, U. (2005). Characterization of the rdar morphotype, a multicellular behaviour in Enterobacteriaceae. *Cell. Mol. Life Sci.* 62, 1234–1246. doi:10.1007/s00018-005-4557-x.
- Römling, U., and Galperin, M. Y. (2015). Bacterial cellulose biosynthesis: Diversity of operons, subunits, products, and functions. *Trends Microbiol.* 23, 545–557. doi:10.1016/j.tim.2015.05.005.
- Rosselló-Mora, R., and Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol. Rev.* 25, 39–67. doi:10.1016/S0168-6445(00)00040-1.
- Rouli, L., Merhej, V., Fournier, P. E., and Raoult, D. (2015). The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* 7, 72–85. doi:10.1016/j.nmni.2015.06.005.
- Rusmini, R., Vecchiotti, D., Macchi, R., Vidal-Aroca, F., and Bertoni, G. (2014). A shotgun antisense approach to the identification of novel essential genes in *Pseudomonas aeruginosa*. *BMC Microbiol.* 14, 1–9. doi:10.1186/1471-2180-14-24.
- Salama, N. R., Shepherd, B., and Falkow, S. (2004). Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*. *J. Bacteriol.* 186, 7926–7935. doi:10.1128/JB.186.23.7926-7935.2004.
- Sato, M., and Miyazaki, K. (2017). Phylogenetic network analysis revealed the occurrence of horizontal gene transfer of 16S rRNA in the genus Enterobacter. *Front. Microbiol.* 8, 1–10. doi:10.3389/fmicb.2017.02225.
- Schneiker, S., Perlova, O., Kaiser, O., Gerth, K., Alici, A., Altmeyer, M. O., et al. (2007). Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* 25, 1281–1289. doi:10.1038/nbt1354.

- Serra, D. O., Richter, A. M., and Hengge, R. (2013). Cellulose as an architectural element in spatially structured *Escherichia coli* biofilms. *J. Bacteriol.* 195, 5540–5554. doi:10.1128/JB.00946-13.
- Setubal, J. C., Dos Santos, P., Goldman, B. S., Ertesvåg, H., Espin, G., Rubio, L. M., et al. (2009). Genome sequence of *Azotobacter vinelandii*, an obligate aerobe specialized to support diverse anaerobic metabolic processes. *J. Bacteriol.* 191, 4534–4545. doi:10.1128/JB.00504-09.
- Shoemaker, W. R., and Lennon, J. T. (2018). *Evolution with a seed bank: The population genetic consequences of microbial dormancy.* doi:10.1111/eva.12557.
- Skurnik, D., Roux, D., Aschard, H., Cattoir, V., Yoder-Himes, D., Lory, S., et al. (2013). A Comprehensive Analysis of In Vitro and In Vivo Genetic Fitness of *Pseudomonas aeruginosa* Using High-Throughput Sequencing of Transposon Libraries. *PLoS Pathog.* 9. doi:10.1371/journal.ppat.1003582.
- Soberón-Chávez, G., Alcaraz, L. D., Morales, E., Ponce-Soto, G. Y., and Servín-González, L. (2017). The transcriptional regulators of the CRP family regulate different essential bacterial functions and can be inherited vertically and horizontally. *Front. Microbiol.* 8, 1–8. doi:10.3389/fmicb.2017.00959.
- Soucy, S. M., Huang, J., and Gogarten, J. P. (2015). Horizontal gene transfer: Building the web of life. *Nat. Rev. Genet.* 16, 472–482. doi:10.1038/nrg3962.
- Steenackers, H., Hermans, K., Vanderleyden, J., and De Keersmaecker, S. C. J. (2012). *Salmonella* biofilms: An overview on occurrence, structure, regulation and eradication. *Food Res. Int.* 45, 502–531. doi:10.1016/j.foodres.2011.01.038.
- Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., et al. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406, 959–964. doi:10.1038/35023079.
- Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997). A genomic perspective on protein families. *Science* (80-.). 278, 631–637. doi:10.1126/science.278.5338.631.
- Tatusova, T., Ciufo, S., Federhen, S., Fedorov, B., McVeigh, R., O'Neill, K., et al. (2015). Update on RefSeq microbial genomes resources. *Nucleic Acids Res.* 43, D599–D605. doi:10.1093/nar/gku1062.
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome.” *Proc. Natl. Acad. Sci.*

102, 13950–13955. doi:10.1073/pnas.0506758102.

- Tettelin, H., Riley, D., Cattuto, C., and Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 11, 472–477. doi:10.1016/j.mib.2008.09.006.
- Touchon, M., and Rocha, E. P. C. (2007). Causes of insertion sequences abundance in prokaryotic genomes. *Mol. Biol. Evol.* 24, 969–981. doi:10.1093/molbev/msm014.
- Touchon, M., and Rocha, E. P. C. (2016). Coevolution of the organization and structure of prokaryotic genomes. *Cold Spring Harb. Perspect. Biol.* 8, 1–18. doi:10.1101/cshperspect.a018168.
- Turner, K. H., Wessel, A. K., Palmer, G. C., Murray, J. L., and Whiteley, M. (2015). Essential genome of *Pseudomonas aeruginosa* in cystic fibrosis sputum. *Proc. Natl. Acad. Sci.* 112, 4110–4115. doi:10.1073/pnas.1419677112.
- Valot, B., Guyeux, C., Rolland, J. Y., Mazouzi, K., Bertrand, X., and Hocquet, D. (2015). What it takes to be a *Pseudomonas aeruginosa*? The core genome of the opportunistic pathogen updated. *PLoS One* 10, 126468. doi:10.1371/journal.pone.0126468.
- Welch, R. A., Burland, V., Iii, G. P., Redford, P., Roesch, P., Rasko, D., et al. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci.* 99, 17020–17024. doi:10.1073/PNAS.252529799.
- West, S. E. H., Schweizer, H. P., Dall, C., Sample, A. K., and Runyen-Janecky, L. J. (1996). Construction of improved *Escherichia-Pseudomonas* shuttle vectors derived from pUC 18/19 and sequence of the region required for their replication in *Pseudomonas aeruginosa*. *Behav. Interv.* 11, 141–146. doi:10.1002/(SICI)1099-078X(199607)11:3<141::AID-BRT161>3.0.CO;2-K.
- Whitfield, C., and Trent, M. S. (2014). Biosynthesis and Export of Bacterial Lipopolysaccharides. *Annu. Rev. Biochem.* 83, 99–128. doi:10.1146/annurev-biochem-060713-035600.
- Wiens, J. R., Vasil, A. I., Schurr, M. J., and Vasil, M. L. (2014). Iron-regulated expression of alginate production, mucoid phenotype, and biofilm formation by *Pseudomonas aeruginosa*. *MBio* 5, 1–12. doi:10.1128/mBio.01010-13.
- Zeigler, D. R. (2003). Gene sequences useful for predicting relatedness of whole genomes in bacteria. *Int. J. Syst. Evol. Microbiol.* 53, 1893–1900. doi:10.1099/ijs.0.02713-0.
- Zhang, X., Peng, C., Zhang, G., and Gao, F. (2015). Comparative analysis of essential genes in prokaryotic genomic islands. *Sci. Rep.* 5, 1–9.

doi:10.1038/srep12561.

- Zhao, L., Anderson, M. T., Wu, W., Harry, H. L., and Bachman, M. A. (2017). TnseqDiff: Identification of conditionally essential genes in transposon sequencing studies. *BMC Bioinformatics* 18, 1–11. doi:10.1186/s12859-017-1745-2.
- Zhaxybayeva, O., Swithers, K. S., Lapierre, P., Fournier, G. P., Bickhart, D. M., Deboy, R. T., et al. (2009). On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *PNAS* 106, 5865–5870.
- Zimmer, J. (2019). Structural features underlying recognition and translocation of extracellular polysaccharides. doi:10.1098/rsfs.2018.0060.