



UNIVERSIDAD DE CHILE – FACULTAD DE CIENCIAS – ESCUELA DE PREGRADO

Análisis de los genes que codifican RNAs de transferencia (tDNAs) como sitios de integración de islas genómicas en *Klebsiella pneumoniae*.

Seminario de Título entregado a la Universidad de Chile en cumplimiento parcial de los requisitos para optar al Título de Ingeniero en Biotecnología Molecular.

Camilo Ignacio Berríos Pastén

Directores del Seminario de Título:

Dra. Rosalba Lagos Mónaco.

Dr. Andrés Marcoleta Caldera.

Octubre 2018

Santiago – Chile



INFORME DE APROBACIÓN SEMINARIO DE TÍTULO

Se informa a la Escuela de Pregrado de la Facultad de Ciencias, de la Universidad de Chile, que el Seminario de Título presentado por el Sr. Camilo Ignacio Berríos Pastén.

Análisis de los genes que codifican RNAs de transferencia (tDNAs) como sitios de integración de islas genómicas en *Klebsiella pneumoniae*.

Ha sido aprobado por la Comisión de Evaluación, en cumplimiento parcial de los requisitos para optar al Título de Ingeniero en Biotecnología Molecular.

Directores Seminario de Título:

Dra. Rosalba Lagos M.

Dr. Andrés Marcoleta C.

Comisión Revisora y Evaluadora:

Presidente Comisión

Dra. Julieta Orlando

Evaluador

Dr. Nicolás Guiliani G.

Santiago de Chile, 26 de octubre de 2018

BIOGRAFÍA

Nací un día 15 de enero de 1993. Mis padres, Elizabeth Pastén y Marcos Berríos, decidieron nombrarme Camilo Ignacio. Mis primeros recuerdos son en Vicuña, en el corazón del Valle del Elqui. Junto a mis padres vivimos siempre un poco fuera de la ciudad, recuerdo con cariño esas tardes tranquilas en las tibias tierras que me vieron crecer.

Cursé mis estudios de enseñanza básica en la Escuela Lucila Godoy Alcayaga de Vicuña. Desde mis primeros años de estudios siempre tuve preferencia por las ciencias naturales, probablemente porque mi madre ha trabajado en el área desde antes de que yo naciera. Durante este periodo llegó a alegrar nuestras vidas, Felipe, mi hermano menor.

En busca de mejores oportunidades, antes de acabar la enseñanza básica, migré a la ciudad de La Serena. Mis padrinos, con mucho cariño, me acogieron en su hogar y me permitieron luego completar mis estudios en el Colegio Gerónimo Rendic, lugar donde hice muchos amigos y terminé de consolidar mi gusto por las ciencias naturales.

Al finalizar la enseñanza media tuve que dar forma a mi futuro. La desesperación llegó, ya que en todo el tiempo de la enseñanza media no había decidido que estudiar. Finalmente, el día de la postulación, dejé que el azar decidiera por mí. Mi primera opción fue la carrera que ahora estoy terminando. Si el azar hubiera dicho otra cosa, quizás estuviera terminando la carrera de Bioquímica o Derecho (mis opciones eran muy amplias).

Dejar a mi familia en busca de conseguir una carrera fue una decisión difícil, me trasladé a una ciudad que me resultaba del todo desconocida. Sin embargo, fui muy afortunado y recibí siempre los consejos precisos y las advertencias necesarias de amigos y familiares, sin ellos quizás esta historia no hubiese llegado a buen puerto.

Con este acto, estoy concluyendo mis estudios de pregrado, el camino recorrido fue duro, pero satisfactorio. Veo el fin de esta etapa más bien como una nueva aventura que está recién comenzando.

*A mis padres y a mi hermano.
Con mucho cariño les dedico este trabajo y
espero algún día recuperemos el tiempo perdido.*

*Dedicado a la memoria de Doña Gladys Aracena y Don Sergio Cepeda
Antes de dejar este mundo, su compañía fue clave
en mis primeros años de universidad.*

*Apresúrate a dejar plantado el
semblante de tu alma en la faena.
No quedaran más retratos tuyos
verdaderos que ese que haces
sin saberlo (...)*

— *Gabriela Mistral*

AGRADECIMIENTOS

En esta instancia, no me queda más que agradecer la labor de mis directores de Seminario de Título, Dra. Rosalba Lagos y Dr. Andrés Marcoleta. Lo que he aprendido junto a ustedes permanecerá conmigo por siempre. No sólo debo agradecer la dirección de este trabajo, también de todas las oportunidades que he tenido de desarrollar distintas habilidades en el contexto de nuestra actividad.

A los miembros de la comisión evaluadora, en especial a la Dra. Julieta Orlando, por su revisión exhaustiva de este trabajo y consejos pertinentes respecto de la evaluación de algunos análisis.

A mis compañeros de universidad, por ser muchas veces el soporte que tanto necesitaba. Franco, Constanza, Bianca, Sebastián, Anita y Macarena, tengo con ustedes los mejores recuerdos de esta etapa de mi vida.

A mis compañeros del 4°B de la generación 2010 del Colegio Gerónimo Rendic. En especial a Constanza, Felipe, Karla y, en menor medida, a Matías. El recuerdo de su amistad me acompañará por siempre.

A todo el equipo del Laboratorio de Biología Molecular y Estructural (BEM), han hecho que nuestro lugar de trabajo se sienta más como un hogar. Agradezco al Dr. Octavio Monasterio todos los consejos y enseñanzas que ha compartido conmigo, y la posibilidad de ser ayudante en distintas instancias que ha coordinado. También, quisiera agradecer a Macarena Collao, Nicole Molina y Don Pedro Núñez, que siempre me han ayudado cordialmente y su labor hace que nuestro laboratorio este en marcha día a día.

A todos quienes han formado parte de mi querida Siberia. Con cariño agradezco a Ramiro y Leonardo su grata compañía y esas largas conversaciones que hacían llevadero nuestro trabajo. También a Álvaro, gracias por tu ayuda y también por tu compañía. A Rodolfo, quién llegó a ayudarme a terminar algunos de los trabajos de este Seminario, pero que al poco andar se volvió indispensable para el desarrollo de las proyecciones de este trabajo. También a Johanna y José Ignacio, grandes compañeros de trabajo y amigos.

También en el grupo de Siberianos, no puedo dejar de agradecer a Josefina. Por esas cosas extrañas de la vida, de alguna manera siempre nos encontramos durante la carrera, pero desde que entramos a trabajar en el laboratorio, he tenido la oportunidad de conocer a una maravillosa persona. No solo trabajar codo a codo y apoyarnos cuando las cosas parecían que no resultaban fue muy importante. Me has enseñado tanto de la vida, de la justicia y del mundo en que nos toca vivir, que de alguna manera me has hecho una mejor persona. Me has hecho valiente, sensible y optimista. Estemos siempre orgullosos de lo que hemos conseguido. Muchas gracias, Josefina.

A mi familia, en especial a mis padres, por ser el soporte de toda esta aventura. A mis abuelitas, tías, tíos, primas y primos, porque su recuerdo me conforta, a pesar de la distancia. También a mis padrinos, el sueño de esta carrera nació dentro del hogar que me brindaron por tantos años.

A todos y todas que han formado parte de mi vida, a quienes me han enseñado de lo bueno y de lo malo. Ya no hay rencor, solo la esperanza de un futuro mejor.

ÍNDICE DE CONTENIDOS

1. INTRODUCCIÓN	1
1.1 <i>Klebsiella pneumoniae</i>: una amenaza urgente para la salud humana	1
1.1.1 <i>Klebsiella pneumoniae</i>	1
1.1.2 Patogénesis y epidemiología de infecciones de <i>K. pneumoniae</i>	2
1.1.3 Factores de virulencia en <i>K. pneumoniae</i>	2
1.1.4 Cepas hipervirulentas y multirresistentes de <i>K. pneumoniae</i>	5
1.2 La plasticidad del genoma de <i>K. pneumoniae</i>	9
1.2.1 El genoma accesorio y los elementos que lo determinan.....	9
1.3 Islas Genómicas: Generalidades y los determinantes de su integración en los cromosomas bacterianos	12
1.3.1 Definición de Isla Genómica.....	12
1.3.2 Algoritmos de identificación de GIs	14
1.3.3 tDNAs y tmDNAs como sitios de integración de GIs	15
1.3.4 El contexto genómico y su rol en la integración de GIs.....	16
2. OBJETIVOS	19
2.1 Objetivo general	19
2.2 Objetivos específicos	19
3. MATERIALES Y MÉTODOS	20
3.1 Secuencias analizadas.....	20
3.2 Identificación manual de Islas Genómicas	20
3.3 Alineamientos de tDNAs y su contexto.	20
3.4 Identificación y análisis de Islas Genómicas	21
3.5 Análisis de curvas de acumulación de tipos de tDNAs y tipos de GIs.....	21
3.6 Búsqueda de motivos conservados	22
3.7 Mapeo de los sitios de unión de proteínas del nucleóide en <i>K. pneumoniae</i>	22
3.8 Análisis Pangenómicos	23
3.9 Análisis Filogenético	23
4.0 Análisis Estadísticos	24
4. RESULTADOS	25
4.1 Identificación de tDNAs y tmDNAs en las cepas de <i>K. pneumoniae</i>	25
4.2 Comparación entre la frecuencia de codones y de anticodones.	25
4.3 Una nueva nomenclatura para los tDNAs de <i>K. pneumoniae</i>	30
4.4 Prevalencia y conservación de los tDNAs de <i>K. pneumoniae</i>	36
4.5 El contexto genómico de los tDNAs: definición del <i>locus</i> virgen	38
4.6 Identificación de GIs.....	40
4.7 Frecuencia de inserción de MGEs en los tDNAs de <i>K. pneumoniae</i>	42
4.8 El contexto genómico de los tDNAs y la disparidad en las frecuencias de uso como sitios de integración.....	47
4.9 Las proteínas asociadas al nucleóide bacteriano y el uso de tDNAs como sitios de integración de MGEs.	50
4.10 Clasificación y generalidades de las GIs identificadas.....	51
4.11 Las funciones codificadas en los MGEs identificadas.....	56
4.12 Un nuevo concepto operacional para comparar genomas de <i>K. pneumoniae</i> y potencialmente otras especies relacionadas: el isotipo	58
4.13 Las integrasas codificadas en las islas y su rol sobre la preferencia por algunos sitios.	60
4.14 Análisis pangenómico de las cepas y el aporte de las GIs al genoma accesorio.....	62
5. DISCUSIÓN Y PROYECCIONES	64
6. CONCLUSIONES	71
7. REFERENCIAS	73
8. ANEXOS	78

ÍNDICE DE FIGURAS

Figura 1: Esquema de los principales factores de virulencia de <i>K. pneumoniae</i>	3
Figura 2: Cronología del uso de distintas familias de antibióticos contra <i>K. pneumoniae</i> y la aparición de cepas resistentes.....	8
Figura 3: El concepto del pangenoma y los mecanismos de evolución del genoma procarionte	11
Figura 4: Características frecuentes de las Islas Genómicas (GIs).....	13
Figura 5: Cantidad de tDNAs identificados por cepa y el tamaño promedio de estos tDNAs	26
Figura 6: Comparación entre la frecuencia de anticodones codificados en los tDNAs de <i>K. pneumoniae</i> , y la frecuencia de uso de los codones complementarios	28
Figura 7: Esquema de la curación manual de los <i>loci</i> analizados y del nuevo sistema de nomenclatura propuesto.....	32
Figura 8: Esquemas de los contextos de 39 <i>loci</i> de tDNAs seleccionados.....	34
Figura 9: Prevalencia y conservación de los distintos tipos de tDNAs identificados en las cepas analizadas	37
Figura 10: Conservación promedio de los elementos que constituyen los <i>loci</i> vírgenes de los tDNAs analizados	39
Figura 11: Esquema del método de identificación de GIs basado en la continuidad del <i>loci</i> vírgenes	41
Figura 12: Distribución de las islas identificadas entre las cepas analizadas y la frecuencia con que los tDNAs son utilizados como sitios de integración en <i>K. pneumoniae</i>	43
Figura 13: Las frecuencias de uso de los tDNAs representado en el cromosoma y comparación de las frecuencias de inserción de <i>loci</i> idénticos dentro del cromosoma	46
Figura 14: Análisis de características de los tDNAs y su contexto en función de la frecuencia de inserción de GIs y otros MGEs	49
Figura 15: Estrategia para la identificación de posibles regiones de unión de proteínas del nucleóide bacteriano a partir de experimentos realizados en la cepa de <i>E. coli</i> K12 MG655	52
Figura 16: Relación entre la ocupación de proteínas asociadas al nucleóide bacteriano en <i>K. pneumoniae</i> , mapeado a partir de experimentos de ChIP-Seq en <i>E. coli</i> , y la frecuencia de uso de los tDNAs como sitios de integración	53
Figura 17: Curvas de proyección de la identificación de nuevos tipos de tDNAs y de nuevos tipos de GIs en las cepas analizadas de <i>K. pneumoniae</i>	55
Figura 18: Los MGEs que se insertan en los <i>loci</i> de tDNAs y las funciones codificadas en las islas genómicas en <i>K. pneumoniae</i>	57
Figura 19: Comparación de la clasificación de las cepas a partir de los genes del <i>core genome</i> y a partir de los isotipos	59
Figura 20: Relación filogenética entre las integrasas codificadas en las distintas clases de islas asociadas a tDNAs de <i>K. pneumoniae</i>	61
Figura 21: Análisis del pangenoma de las cepas de <i>K. pneumoniae</i> analizados	63
Figura 22: Los contextos genómicos de los tDNAs <i>asn1</i> poseen elementos comunes entre ocho especies de enterobacterias	65

ÍNDICE DE TABLAS

Anexo I: Cepas analizadas en este trabajo.	78
Anexo II: Resumen de las frecuencias de los anticodones (f_{ant}) de los tDNAs y de sus codones complementarios (f_{cod}).	80
Anexo III: Las 97 clases de MGEs identificadas en tDNAs y tmDNAs.....	81
Anexo IV: Islotipos de las cepas analizadas.....	85

ABREVIATURAS

DNA	Ácido desoxirribonucleico
RNA	Ácido ribonucleico
tRNA	RNA de transferencia
mRNA	RNA mensajero
tmRNA	RNA mensajero y de transferencia
tDNA	Gen que codifica un tRNA
tmDNA	Gen que codifica un tmRNA
MGE	Elemento genético móvil
CAI	Índice de adaptación de codones
ICE	Elemento conjugativo integrativo
GI	Isla genómica
ORF	Marco de lectura abierto
COG	Grupos de genes ortólogos

CÓDIGO DE AMINOÁCIDOS

Ala	Alanina	Lys	Lisina
Arg	Arginina	Met	Metionina
Asn	Asparagina	Phe	Fenilalanina
Asp	Ácido Aspártico	Pro	Prolina
Cys	Cisteína	Sec	Selenocisteína
Gln	Glutamina	Ser	Serina
Glu	Ácido Glutámico	Thr	Treonina
Gly	Glicina	Trp	Triptófano
His	Histidina	Tyr	Tirosina
Ile	Isoleucina	Val	Valina
Leu	Leucina		

RESUMEN

En las últimas dos décadas, la especie bacteriana *Klebsiella pneumoniae* se ha convertido en uno de los principales focos de atención de las autoridades mundiales de salud, debido principalmente a la rápida aparición de cepas hipervirulentas y multiresistentes. La aparición de estas cepas estaría relacionada con el gran dinamismo del genoma de *K. pneumoniae*, promovido por una alta frecuencia de adquisición y pérdida de elementos genéticos móviles, entre los cuales las Islas Genómicas (GIs) jugarían un rol fundamental. Las GIs son segmentos de DNA que presentan un contenido GC distinto al del resto del cromosoma, y que pueden escindirarse o integrarse en distintas posiciones de éste, normalmente en genes que codifican RNAs de transferencia (tDNAs). No obstante la importancia propuesta de las GIs en la evolución de *K. pneumoniae*, hasta ahora el número de GIs conocidas en esta especie es bajo, y poco se conoce respecto al uso de los tDNAs como sitios de integración de estos elementos en ésta y otras especies bacterianas. En este trabajo se caracterizó el set completo de tDNAs presentes en el cromosoma de 50 cepas de *K. pneumoniae*, identificando las GIs insertadas en dichos *loci*. Así, se identificaron y anotaron 97 clases de GIs, las que codifican sobre un 50% de proteínas con función desconocida. Además, se observó que solo un grupo reducido de tDNAs son ocupados como sitios de integración de GIs en esta especie, proponiendo y sometiendo a prueba posibles mecanismos o patrones que podrían explicar en parte este sesgo. Los resultados muestran que la dinámica de integración de GIs en *K. pneumoniae* es un área de investigación poco explorada, y que la evaluación del rol de las GIs en la hipervirulencia y resistencia a antibióticos de *K. pneumoniae* requerirá caracterizar una gran cantidad de proteínas codificadas en las mismas, cuya función se desconoce.

ABSTRACT

In the last two decades, the bacterial species *Klebsiella pneumoniae* has attracted the attention of global health authorities, mainly due to the dissemination of hypervirulent and multi-resistant strains, which are responsible for more severe infections and are more difficult to treat with the available antibiotics. The rapid development of these strains would be explained by the high dynamism of the *K. pneumoniae* genome, promoted by a high frequency of gain and loss of mobile genetic elements. Among them, genomic islands (GIs) were proposed to play a major role. GIs are DNA segments harboring a GC content that differs from the average for the host chromosome, and that can excise from and integrate into different locations, mainly into genes encoding transfer RNAs (tDNAs). In spite of the proposed role of GIs in *K. pneumoniae* evolution, a small number of these elements have been described in this species. Moreover, little is known regarding the usage of tDNAs as integration sites for these elements in bacteria. In this work we characterized the whole set of tDNAs typically present in the *K. pneumoniae* chromosome, searching for GIs integrated at each of them among 50 *K. pneumoniae* strains. This way, we identified a total of 97 classes of GIs, where more than 50% of the encoded genes have no predicted function. Additionally, we observed that only a small subset of tDNAs are used as integration sites for GIs in this species, and tested possible hypotheses explaining, at least partially, this bias in the selection of the integration site. Our results indicate that the dynamics of GIs integration in bacteria is an underexplored field, and that evaluating the impact of GIs over *K. pneumoniae* virulence and drug resistance will require efforts directed to address the functions of a high number of proteins encoded in their GIs.

1. INTRODUCCIÓN

1.1 *Klebsiella pneumoniae*: una amenaza urgente para la salud humana

1.1.1 *Klebsiella pneumoniae*

Klebsiella pneumoniae, descrita por vez primera en 1882 (Trevisan, 1885; Von Frisch, 1882), es una bacteria Gram-negativo perteneciente al grupo de las Enterobacterias. Microscópicamente, tiene morfología de bacilo encapsulado y no motil, y macroscópicamente forma colonias blanquecinas que, dependiendo de la cepa, presentan distintos grados de mucosidad (Montgomerie, 1979). Aunque originalmente fue aislada a partir de muestras clínicas de hospederos animales y humanos, se han encontrado miembros del género *Klebsiella* en distintos ambientes naturales e intervenidos, incluyendo muestras de agua, suelo, desechos industriales y plantas (Fouts, et al., 2008).

En humanos, *Klebsiella* spp. es parte de la microbiota nasofaríngea y gastrointestinal. Se estima que cerca del 30% de la población posee *Klebsiella* spp. en sus heces, mientras que solo entre el 3% y el 4% de los individuos posee especies de dicho género en muestras provenientes de la garganta. La incidencia de *Klebsiella* spp. entre los individuos hospitalizados es ligeramente superior en ambos lugares del cuerpo (Davis & Matsen, 1974). Sin embargo, tratamientos con antibióticos y hospitalizaciones prolongadas en el tiempo se correlacionan con un incremento en la colonización de *Klebsiella* spp. (Podschun & Ullmann, 1998; Pollack, et al., 1972). Además, se ha comprobado que los serotipos capsulares identificados en infecciones del tracto respiratorio y del tracto urinario son idénticos a los serotipos identificados en el tracto gastrointestinal y faríngeo (Selden, et al., 1971). En este mismo estudio fue comprobado que un porcentaje menor de los trabajadores en contacto con los pacientes es portador

de *K. pneumoniae* y que rara vez es posible aislar esta especie desde el ambiente inanimado del hospital (pisos, superficies e instrumentos) lo que sugiere que las infecciones intrahospitalarias asociadas a *K. pneumoniae* se originan de reservorios presentes previamente en los pacientes.

1.1.2 Patogénesis y epidemiología de infecciones de *K. pneumoniae*

En el caso de *K. pneumoniae*, el resultado de sus interacciones con hospederos humanos es variable, desde portadores asintomáticos, hasta patologías asociadas a infecciones intrahospitalarias, o adquiridas en la comunidad (Brisse, et al., 2006).

La neumonía asociada a *K. pneumoniae* tiene un cuadro clínico característico muy violento, el que incluye la rápida aparición de fiebre y esputo gelatinoso de color que varía entre rojo y rosado. También es característico que las radiografías de tórax de los pacientes presenten anomalías como bultos en la región superior de los pulmones (Gil, et al., 2005), y en algunos casos las lesiones pulmonares pueden ser del tipo necrotizante (Erlj, et al., 2013). Sin embargo, las infecciones con *K. pneumoniae* no están limitadas al órgano pulmonar. Dentro de las infecciones que se pueden conseguir en el contexto intrahospitalario, las que afectan al tracto urinario son las más comunes. Otro tipo de infección que destaca en el grupo de las contraídas en este contexto son infecciones de *K. pneumoniae* que afectan el torrente sanguíneo, cuyo origen está asociado a la infección de heridas quirúrgicas y la capacidad de la bacteria de sobrevivir en los catéteres, por medio de la formación de biopelículas.

1.1.3 Factores de virulencia en *K. pneumoniae*

K. pneumoniae posee factores que son claves para su virulencia, permitiendo la colonización e infección de distintos hospederos (Fig. 1). En primer lugar, la mayoría

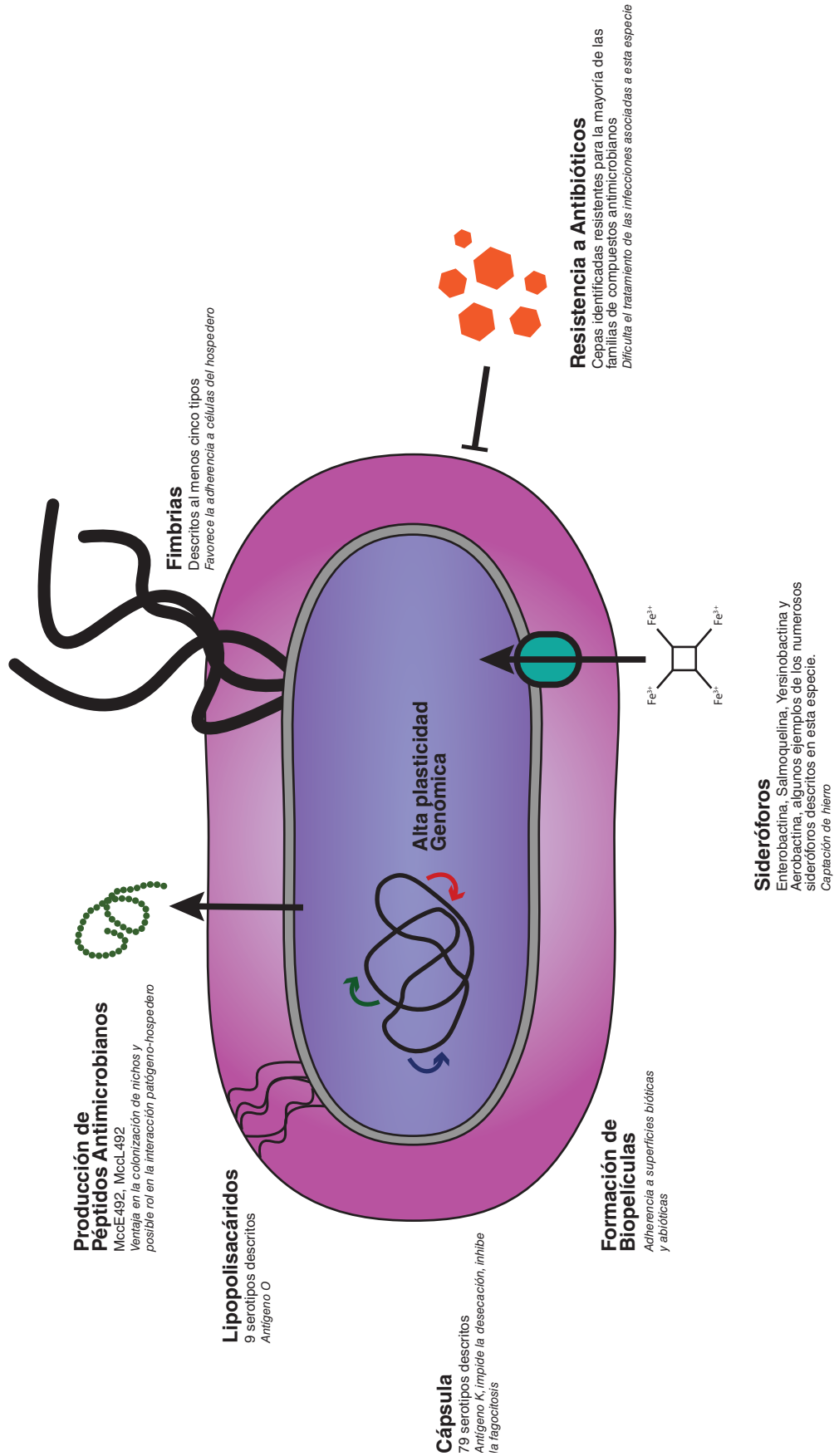


Figura 1: Esquema de los principales factores de virulencia de *K. pneumoniae*.
Adaptado de Podschun, R. & Ullmann, U. (1998).

de los aislados clínicos de esta especie posee una cápsula viscosa que envuelve a la célula, la cual es responsable de la apariencia mucosa de las colonias en placas con agar y medio de cultivo. La cápsula está conformada por polisacáridos ácidos complejos, que consisten en la repetición de subunidades de cuatro a seis azúcares e incluyen una subunidad de ácido urónico (Corsaro, et al., 2005). La complejidad de estas subunidades y su combinatoria de unión hacen que la cápsula posea distintas variantes en cuanto a su composición bioquímica. Entre las cepas aisladas de *K. pneumoniae*, han sido descritos 78 serotipos capsulares (Johnson, 2018; Ørskov & Ørskov, 1984). Por ejemplo, los serotipos K21a, K36, y K50 poseen residuos di-manosa/ramnosa, mientras que los serotipos K2, K8 y K55 no los poseen (Sahly, et al., 2009). El serotipo más frecuente entre las cepas de *K. pneumoniae* asociada a cuadros infecciosos graves es el K1, el que posee fucosa en su composición, a diferencia de todos los demás serotipos identificados (Pan, et al., 2011). En cuanto al rol de la cápsula en la patogénesis, se ha comprobado que ésta inhibe la fagocitosis por parte de las células del sistema inmune del hospedero (Williams, et al., 1986).

Otro factor clave para la virulencia y éxito adaptativo de *K. pneumoniae* es la captación de hierro desde el ambiente, elemento que es requerido como cofactor para muchas reacciones del metabolismo bacteriano. Se ha descrito un amplio abanico de sistemas de captación de hierro en esta especie. Estudios previos con cepas de los serotipos K1 y K2 indican que la producción del sideróforo aerobactina se correlaciona con una mayor virulencia (Nassif & Sansonetti, 1986), mientras que en otros serotipos es común la producción del sideróforo enteroquelina (Tarkkanen, et al., 1992). Otros sistemas de captación de hierro codificados en el genoma de distintas cepas incluyen el sistema citrato férrico, el sistema ferricromo, y la producción de salmoquelina y yersiniabactina, entre otros sideróforos (Hsieh, et al., 2008)

En cuanto a la colonización, la adherencia de las células bacterianas a las células y tejidos del hospedero está asociada a las fimbrias. Han sido descritos once grupos de genes posiblemente relacionados con la producción de fimbrias, sin embargo, los estudios se han concentrado en un grupo reducido de ellos (Clegg & Murphy, 2016). En particular, dos tipos de fimbrias han sido ampliamente descritas: la fimbria de Tipo 1 (FimA y FimH) y la fimbria de Tipo 3 (MrkH). Estos tipos de fimbrias han sido identificadas en cepas que ocasionan infecciones urinarias, y posiblemente determinan su resistencia al *clearance* de las vías urinarias mediado por la orina. La adherencia de *K. pneumoniae* a superficies bióticas y abióticas está también determinada por la capacidad de algunas cepas de formar biopelículas. Las fimbrias y otros factores determinan las características de las biopelículas en *K. pneumoniae*, lo que ha sido descrito no sólo para la infección de las vías urinarias, sino que también en la contaminación de catéteres, instrumentos colonoscópicos y otros implementos clínicos (Jagnow & Clegg, 2003; Langstraat, et al., 2001; Ong, et al., 2008; Schroll, et al., 2010).

1.1.4 Cepas hipervirulentas y multirresistentes de *K. pneumoniae*

En las últimas dos décadas, se han descrito numerosas cepas hipervirulentas de *K. pneumoniae*. La hipervirulencia está determinada por 1) la capacidad de estas cepas de generar infecciones graves en individuos sanos y sin antecedentes de susceptibilidad a estos cuadros; 2) la capacidad de infectar tejidos u órganos inusuales, es decir, que no son blancos habituales de infección por parte de este patógeno; y 3) la habilidad de diseminar las infecciones hacia distintos órganos. Además, estas cepas poseen una aumentada producción de cápsula y elementos que favorecen su adhesión, lo que se conoce como fenotipo hiper mucoviscoso. Sin embargo, la virulencia de estas cepas no se limita a la adhesión, por lo que el término hipervirulento es el más adecuado para estas cepas (Shon, et al., 2013).

A principios del siglo XX, en el continente asiático, fueron reportados una serie de casos de abscesos hepáticos piógenos, es decir, zonas llenas de pus en el interior del hígado causado por infecciones de cepas de *K. pneumoniae*. Clásicamente, este tipo de manifestación patológica es resultado de infecciones simultáneas con múltiples bacterias. No obstante, se ha observado que ciertas cepas de *K. pneumoniae* pueden causar este efecto por sí solas. Más aún, en este tipo de infecciones concurren una serie de condiciones que han hecho de los abscesos hepáticos vinculados a *K. pneumoniae* una preocupación sanitaria a nivel mundial: 1) una alta mortalidad, 2) un difícil diagnóstico e identificación del patógeno, y 3) una alta complejidad en el tratamiento, considerando que las cepas responsables de estos cuadros son generalmente resistentes a antibióticos. Inicialmente, los abscesos hepáticos piógenos estaban restringidos a algunas regiones de Asia. Sin embargo, en los últimos años se ha reportado un aumento de casos de este tipo de infección en Europa y América del Norte (Lederman & Crum, 2005; Rossi, et al., 2018). Estos reportes también dan cuenta de que el contagio de las cepas causantes de los abscesos no está limitado al ambiente intrahospitalario, sino que también puede darse por el contagio en la comunidad. Algo que es común entre los pacientes con abscesos hepáticos piógenos ligados a *K. pneumoniae* es la propagación de la infección a otros órganos, lo que ha sido denominado metástasis séptica. La infección puede propagarse y causar abscesos o embolias pulmonares, abscesos cerebrales, endoftalmítis, abscesos prostáticos o artritis séptica, entre otras infecciones (Chang & Chou, 1995; Cheng, et al., 1991; Yeh, et al., 1998).

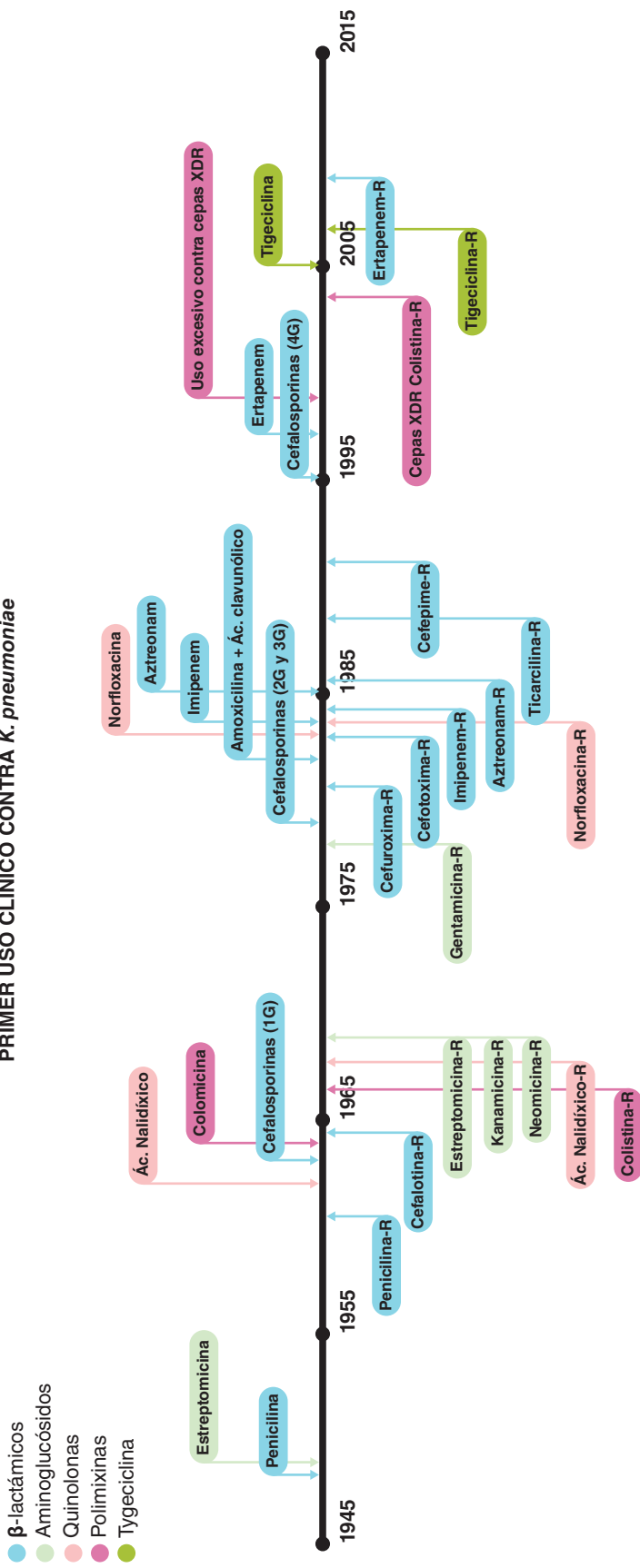
En cuanto a la resistencia a antibióticos, una alta proporción de cepas de *Klebsiella spp.* poseen en su cromosoma genes que codifican para distintas β -lactamasas, lo que les confiere distintos grados de resistencia a aminopenicilinas y

carboxipenicilinas tales como la ampicilina y la carbenicilina, respectivamente (Chaves, et al., 2001). Además, se ha identificado la transferencia horizontal de genes que median y potencian esta resistencia (Bradford, 2001; Evans & Amyes, 2014), los cuales están codificados en distinto tipo de elementos genéticos móviles incluyendo islas genómicas, integrones y plásmidos (Jacoby, 2009).

K. pneumoniae ha conseguido desarrollar resistencia para gran parte de los antibióticos que se utilizan en el contexto clínico (Fig. 2), lo cual es otra arista de la problemática asociada a esta especie. Al incremento en el número de cepas hipervirulentas descritas se suma la cada vez más frecuente identificación de cepas multirresistentes. Lo anterior no se limita a los antibióticos frecuentemente utilizados, sino también a la última línea de antibióticos disponibles actualmente. De lo anterior destaca la identificación de cepas resistentes a los carbapenémicos y cefalosporinas de tercera generación, antibióticos pertenecientes al grupo de los β -lactámicos y que son la última línea de ataque en contra de infecciones bacterianas. Estas cepas, además de ser resistentes a carbapenémicos, son por lo general extremadamente resistentes a varios antibióticos (Clegg & Murphy, 2016), por lo que las opciones de tratamiento son muy limitadas, explicando la alta mortalidad de los pacientes infectados con estas cepas (Patel, et al., 2008). La situación es bastante preocupante considerando que se han descrito cepas de *K. pneumoniae* resistentes a antibióticos como las polimixinas (entre ellas la colistina), normalmente usadas como último recurso en contra de las infecciones bacterianas y de uso reservado para evitar la aparición de cepas resistentes (World Health Organization, 2017).

La rápida aparición de cepas de *K. pneumoniae* resistentes es motivo de que esta especie fuera incorporada al grupo de patógenos ESKAPE, que incluye a las especies bacterianas *Enterococcus faecium*, *Staphylococcus aureus*, *K. pneumoniae*

PRIMER USO CLÍNICO CONTRA *K. pneumoniae*



AISLAMIENTO CLÍNICO DE CEPAS DE *K. pneumoniae* RESISTENTES

Figura 2: Cronología del uso de distintas familias de antibióticos contra *K. pneumoniae* y la aparición de cepas resistentes. Línea del tiempo indicando el primer uso de antibióticos de distintas familias (superior) y el primer aislamiento de cepas de *K. pneumoniae* resistente a estos antibióticos en contextos clínicos (inferior). XDR: Extremadamente resistentes a antibióticos. Adaptado de Navon-Venezia, S., et al (2016)

Acinetobacter baumannii, *Pseudomonas aeruginosa*, *Enterobacter spp.* Los esfuerzos se han concentrado en el desarrollo de estrategias para la identificación de estas especies y su diagnóstico, además del desarrollo de estrategias terapéuticas, principalmente la búsqueda de agentes antibacterianos específicos contra estos patógenos (Pendleton, et al., 2013; Santajit & Indrawattana, 2016).

1.2 La plasticidad del genoma de *K. pneumoniae*

1.2.1 El genoma accesorio y los elementos que lo determinan

La aparición de cepas hipervirulentas y multirresistentes está estrechamente ligada con la alta plasticidad del genoma de *K. pneumoniae* (Holt, et al., 2015). Podemos entender este fenómeno desde el punto de vista del gran repertorio de genes accesorios que posee esta especie bacteriana. La colección completa de genes de una especie (o de un grupo de cepas) corresponde al pangenoma, y dentro de éste podemos distinguir tres grupos de genes (Fig. 3A): el genoma núcleo (*core genome*) que es compartido por todas las cepas, el genoma accesorio (*accessory genome*) que es compartido por algunas cepas y los genes cepa-específicos (*strain-specific genes*), presentes solo en una cepa (Soucy, et al., 2015).

La variación y arquitectura del genoma están determinadas por mecanismos dependientes de eventos de recombinación homóloga y/o de mutaciones que permiten la ganancia o pérdida de funciones, las cuales pueden estar asociadas a la virulencia o la adaptación de las cepas a distintos nichos (Dobrindt & Hacker, 2001). Entre los mecanismos que permiten la incorporación de material genético se encuentran la transformación, la transducción y la conjugación (Fig. 3B).

Los elementos genéticos móviles (MGEs) son fragmentos de DNA susceptibles de cambiar de posición dentro de un mismo cromosoma o bien ser movilizados hacia un

nuevo hospedero mediante distintos mecanismos de transferencia. Existen distintos tipos de MGEs, entre ellos, los elementos conjugativos-integrativos, transposones, bacteriófagos temperados, e islas genómicas (Fig. 3B). Comúnmente estos elementos codifican factores de virulencia y de resistencia a antibióticos, y su alta dinámica causa una gran variabilidad de los genes portados (Frost, et al., 2005). Los transposones y secuencias de inserción son MGEs que codifican transposasas, enzimas que catalizan la recombinación entre las secuencias repetidas e invertidas que limitan el elemento, resultando en la escisión y posterior integración de éste en un sitio diferente del genoma (transposición del DNA). En *K. pneumoniae* se han descrito transposones que codifican resistencias a antibióticos (Tolmasky, et al., 1988; Tolmasky & Crosa, 1987). Los transposones que sólo codifican las proteínas requeridas para la transposición son denominados secuencias de inserción (Mahillon & Chandler, 1998). En el caso de los bacteriófagos, el análisis genómico de distintas cepas de *K. pneumoniae* ha comprobado que es frecuente la identificación de profagos, es decir, DNA de bacteriófagos en fase lisogénica. Por ejemplo, fueron identificados al menos cinco en la cepa hipervirulenta NTUH-K2044 (Wu, et al., 2009). Finalmente, los elementos conjugativos-integrativos (ECIs) poseen una naturaleza intermedia entre los plásmidos conjugativos y los bacteriófagos, en el sentido en que pueden movilizarse a través de conjugación y pueden insertarse en el genoma como el material genético de los bacteriófagos (Burrus & Waldor, 2004; Germon, et al., 2007). Previamente, se han reportado ECIs en *K. pneumoniae*, por ejemplo aquellas que codifican los determinantes genéticos para la producción de colibactina (Putze, et al., 2009).

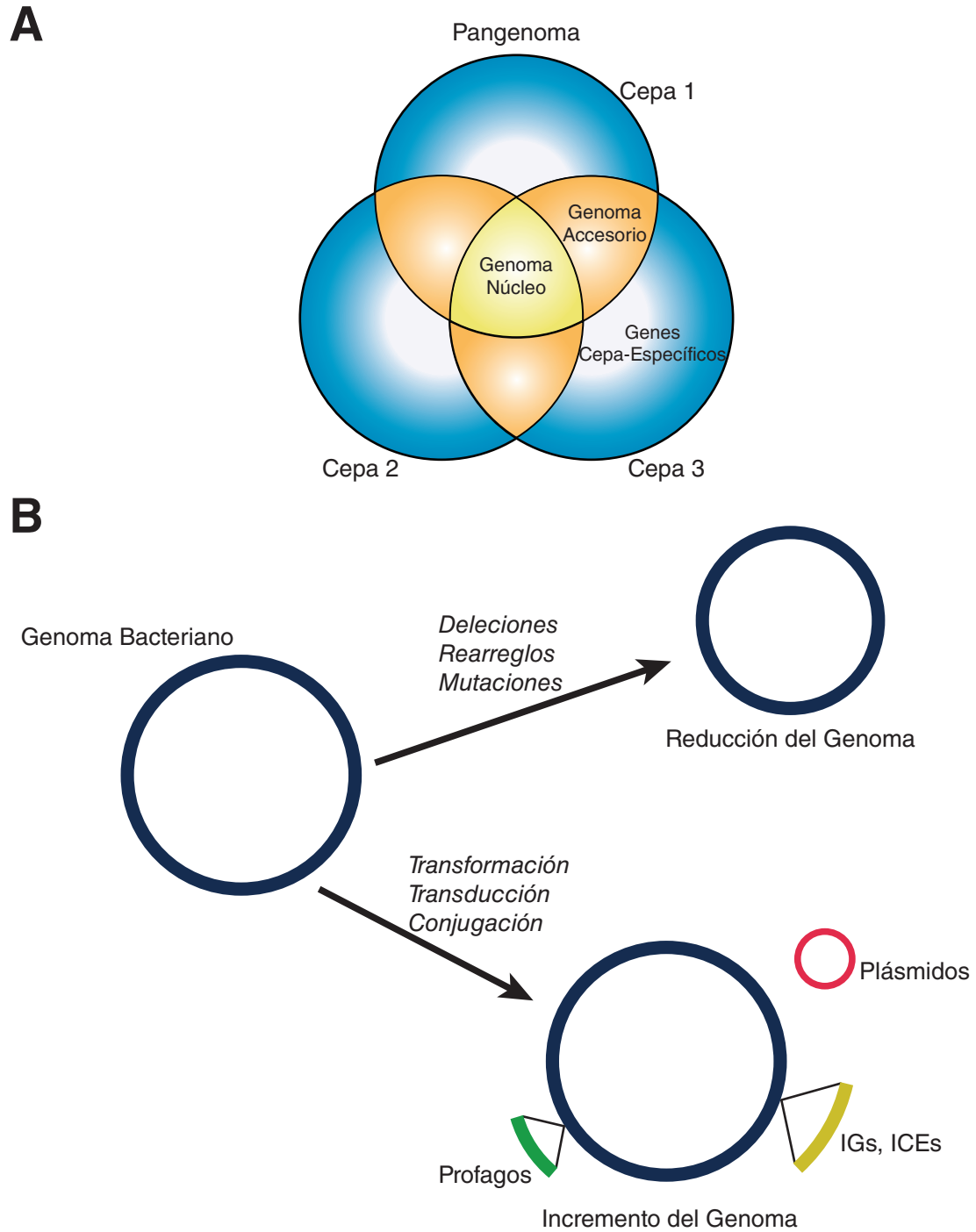


Figura 3: El concepto del pangenoma y los mecanismos de evolución del genoma procarionte.

(A) Representación del pangenoma y sus componentes. (B) La evolución del genoma bacteriano está determinado por la adquisición y pérdida de genes. Los EGMs median la ganancia de genes a través de distintos fenómenos. Adaptado de Dobrindt, U. & Hacker, J. (2001).

1.3 Islas Genómicas: Generalidades y los determinantes de su integración en los cromosomas bacterianos.

1.3.1 Definición de Isla Genómica

El foco de este trabajo está en las Islas Genómicas. Las islas genómicas (GIs) son segmentos cromosómicos de entre 10 kpb y 200 kpb, que difieren del resto del cromosoma por su contenido GC (o contenido G+C, %GC), y la tendencia del uso de codones, generalmente analizado por medio del índice de adaptación de codones (CAI), que mide la desviación del uso de codones con respecto a un conjunto de proteínas de referencia (Fig. 4). Generalmente, las GIs codifican genes de integrasas que median su integración y escisión del cromosoma, y están flanqueadas por repetidos directos de extensión variable (15 pb a 100 pb), que se originan como producto de la recombinación sitio-específica que da lugar a la integración del elemento (Juhas, et al., 2009; Marcoleta, et al., 2016).

Además de los elementos genéticos que median su movilidad, las GIs pueden comprender una gran cantidad de genes que le otorgan a su hospedero nuevas capacidades metabólicas o adaptativas. Entre ellas, se han identificado islas que codifican factores de virulencia y de resistencia a antibióticos (islas de patogenicidad), así como también islas cuyos productos génicos participan en la fijación de nitrógeno, distintos tipos de simbiosis, y degradación de fuentes de carbono alternativas, entre otras funciones (Juhas et al., 2009). En el caso de *K. pneumoniae*, se ha propuesto que las GIs constituyen uno de los principales reservorios de factores de virulencia en esta especie (Holt, et al., 2015; Marcoleta, et al., 2016; Struve, et al., 2015).

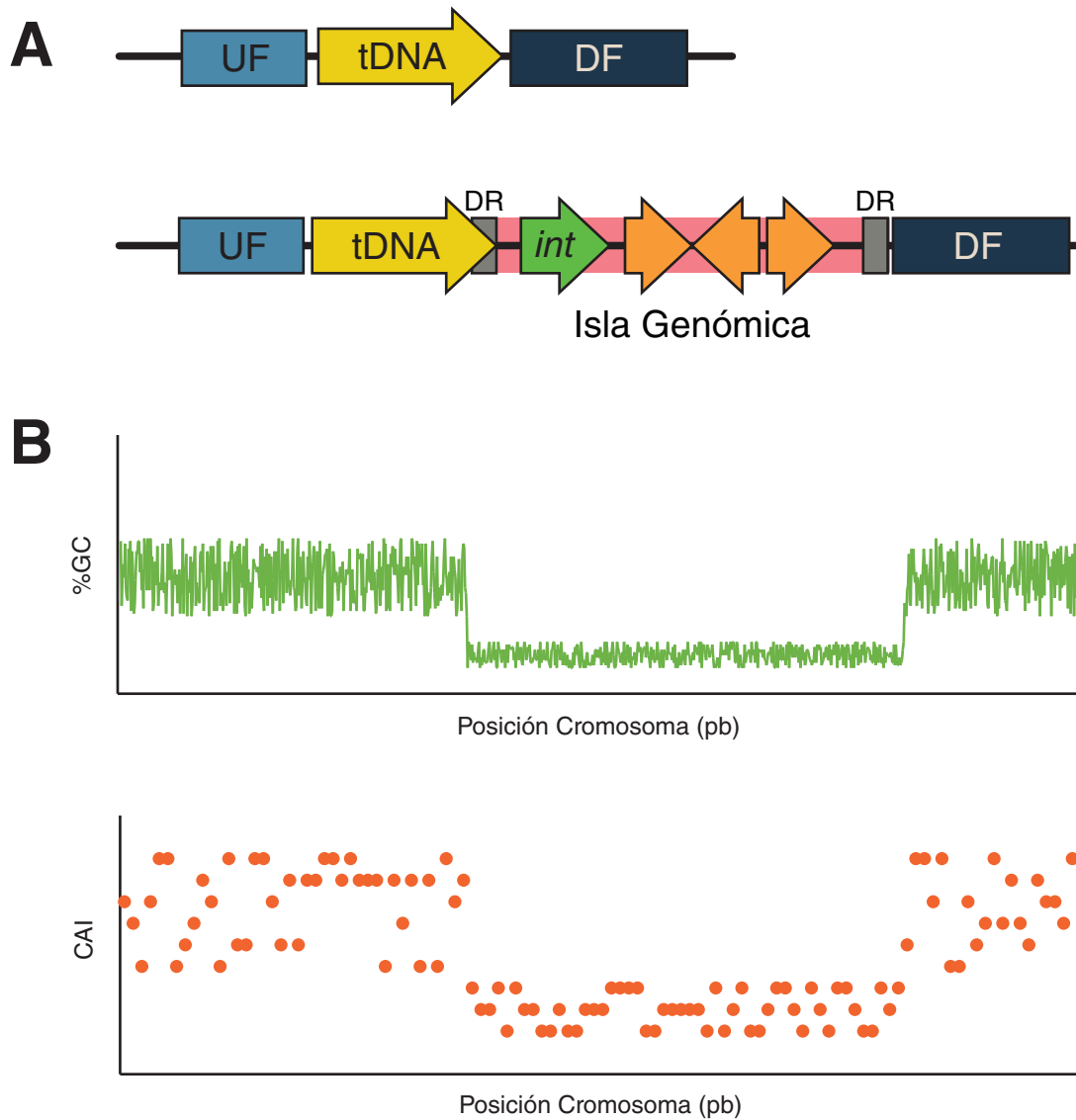


Figura 4: Características frecuentes de las Islas Genómicas (GIs)

(A) Esquema de la inserción de una IG en un cromosoma hospedero. Mientras que la región río arriba (UF) del tDNA queda intacta, la región río abajo (DF) es desplazada producto de la inserción. Las IGs frecuentemente codifican una integrasa (*int*) ubicada en el extremo más próximo al tDNA. (B) Las islas fueron originalmente identificadas producto de la diferencia en el contenido G-C (%GC) y el índice de adaptación de codones (CAI) que presentan los genes codificados en estos elementos, en comparación al resto del cromosoma.

1.3.2 Algoritmos de identificación de GIs

Dada la importancia de las GIs en la propagación de factores benéficos para la adaptación de las bacterias, así como también la creciente cantidad de genomas secuenciados disponibles de distintas especies bacterianas, se han desarrollado numerosas estrategias para la identificación de GIs mediante herramientas bioinformáticas.

Las herramientas desarrolladas para este fin explotan la identificación de características comunes a las GIs. Una de las primeras estrategias consistió en el alineamiento del genoma de cepas filogenéticamente relacionadas, identificando fragmentos presentes solo en algunas de ellas. Entre las herramientas disponibles basadas en esta estrategia se encuentran Mauve (Darling, et al., 2004), que permite el alineamiento de hasta una decena de cromosomas bacterianos completos, y la herramienta derivada de Mauve, tRNAcc, que compara *loci* de tDNAs que son definidos por los usuarios contra una secuencia de referencia, que también debe ser definida (Ou, et al., 2006).

Otro grupo de herramientas permite identificar GIs analizando un único cromosoma. En estos casos, la estrategia consiste en detectar características de las islas inherentes a su secuencia, tanto a nivel de composición de nucleótidos como de los elementos codificados en la secuencia. Por ejemplo, el programa SIGI-HMM (Waack, et al., 2006) mide la tendencia del uso de codones de todos los genes de un genoma, de modo que cuando genes contiguos poseen un uso de codones dispar al valor determinado para la especie, el fragmento es anotado como una IG. El principal desafío de este tipo de herramientas es que, con el progreso de la evolución, las GIs tienden a tener características similares al cromosoma hospedero, por lo que el umbral de diferencia requerido para identificar una IG no siempre es alcanzado. Por ello, muchas

veces es necesario reducir el mínimo de variación requerida para diferenciar una IG, lo que afecta la precisión de estos programas y causa la detección de numerosos falsos positivos. Otras herramientas utilizadas en la identificación de GIs son la búsqueda de genes que codifican integrasas utilizando BLAST (Altschul, et al., 1990) y la búsqueda de los repetidos directos que limitan una IG, como el algoritmo REputer (Kurtz, et al., 2001).

Gracias a la utilización de las distintas herramientas de identificación mencionadas anteriormente, se han construido bases de datos de GIs. Una de las más populares es Islander (Hudson, et al., 2015), que ofrece un algoritmo que combina distintos métodos para la identificación de GIs asociadas a tDNAs. Los fragmentos del cromosoma de las especies bacterianas analizadas son filtrados según la presencia de genes de integrasas y repetidos directos. Además, esta base de datos es curada y actualizada regularmente.

1.3.3 tDNAs y tmDNAs como sitios de integración de GIs

Las GIs frecuentemente se integran en genes que codifican tRNAs (tDNAs) y genes que codifican tmRNAs (tmDNAs). Los tRNAs son un componente esencial de la maquinaria de traducción de proteínas en todos los dominios de la vida (Shepherd & Ibba, 2015). Su rol es transferir los aminoácidos que son cargados en su estructura durante la traducción de RNAs mensajeros (mRNAs) en el ribosoma, cuando se produce el correcto apareamiento entre el anticodón presente en la secuencia del tRNA y la secuencia complementaria del codón presente en el mRNA.

Los RNAs con función dual de transferencia y mensajeros (tmRNAs) participan en el control de calidad del proceso de traducción. En primer lugar, actúan como un tRNA, uniéndose a ribosomas en los que la traducción se ha estancado. Luego, el tmRNA funciona como un mRNA, añadiendo una señal peptídica a la región C-terminal.

Esta señal funciona como indicador de una traducción incorrecta de la proteína, la cual es destinada a la degradación proteolítica (Janssen & Hayes, 2012).

En las enterobacterias, la evidencia demuestra que los tDNAs y tmDNAs son *hotspots* de integración de GIs y otros MGEs (Juhas, et al., 2009). Si bien se han propuesto varias hipótesis al respecto, no es claro el por qué estos genes son sitios preferidos para la integración de dichos elementos. Por un lado, los tDNAs habitualmente se encuentran en múltiples copias por genoma, por lo que la posible interrupción de una de las copias producto de la integración no resultaría en la pérdida de la función del tRNA. Sin embargo, se ha observado que habitualmente la integración de GIs no causa una alteración en la secuencia del tDNA, por lo que dicha explicación no satisface a una gran cantidad de casos. Por otro lado, algunos autores han propuesto que en *E. coli* existe cierto sesgo por la búsqueda de MGEs en *loci* de tDNAs, pero que eventualmente pueden integrarse en otros sitios (Touchon, et al., 2009). Tal como se describirá en las siguientes secciones, como parte de este trabajo se ponen a prueba y discuten las principales hipótesis que podrían determinar la preferencia de los *loci* de tDNA como sitios de integración de GIs.

1.3.4 El contexto genómico y su rol en la integración de GIs

Se ha propuesto que el uso de un determinado tDNA como sitio de integración dependería principalmente de su secuencia, parte de la cual es reconocida por la integrasa. Dicha secuencia es sustrato de recombinación sitio-específica con una secuencia homóloga presente en la forma extracromosómica de la isla, la cual se transforma en los repetidos directos que flanquean la isla una vez integrada. Sin embargo, esto no explica fenómenos como el observado para el caso de los tDNAs de asparagina (asn-tDNA) en *K. pneumoniae*. El análisis del genoma de distintas cepas de *K. pneumoniae* realizado por nuestro grupo de investigación reveló que existen cuatro

copias del asn-tDNA, las que, si bien son 100% idénticas, son usadas como sitio de integración de islas genómicas con frecuencias muy disímiles (Marcoleta, et al., 2016). Por lo tanto, el uso de los tDNAs como sitio de integración no solo dependería de su secuencia nucleotídica, sino además de factores hasta ahora no esclarecidos. En este sentido, uno de los elementos que diferencia a las copias idénticas de los tDNAs es su contexto genómico, es decir, las secuencias de nucleótidos situadas río arriba y río abajo de los mismos. En base a esto, se ha propuesto que el contexto genómico de algunos tDNAs podría propiciar la inserción de GIs, aunque a la fecha no se ha reportado ningún estudio global de los tDNAs como sitios de integración de GIs en una especie bacteriana. Más aún, a la fecha no existe un método sistemático de nomenclatura de los tDNAs de *K. pneumoniae*, lo que dificulta los análisis genómicos centrados en dichos genes.

Para intentar resolver el rol que tiene el contexto genómico en cuanto a la frecuencia de inserción de islas genómicas, en el presente trabajo se comparó el contexto de cada uno de los tDNAs presentes en un grupo de cepas de *K. pneumoniae*, de modo de establecer patrones comunes. Adicionalmente, a partir de la caracterización del set de tDNAs propio de esta especie se desarrolló una nomenclatura sistemática que permite individualizar no sólo distintos tDNAs sino además copias de un mismo tDNA situadas en distintos contextos genómicos. Por otra parte, el análisis de los asn-tDNA de *K. pneumoniae* realizado previamente indicó que el contexto río arriba de cada tDNA permanece inalterado luego de la integración de GIs, a la vez que el contexto río abajo del tDNA es desplazado quedando situado inmediatamente contiguo al extremo 3' de la isla. Esta propiedad fue aprovechada para desarrollar un método preciso de identificación de GIs que se basa en la definición del contexto genómico virgen de cada tDNA (aquel donde no ha ocurrido integración), deducido al comparar regiones equivalentes en los cromosomas de más de 50 cepas de esta especie (Marcoleta et al.,

2016). De esta manera se identificaron una gran cantidad de islas genómicas asociadas a cada uno de los *asn*-tDNA, estableciendo la frecuencia de uso de cada uno de estos *loci*.

Replicando lo realizado con los *asn*-tDNA, en este trabajo se describe la curación de los contextos genómicos de todos los tDNAs presentes en las cepas analizadas de *K. pneumoniae*, además de la identificación y caracterización de todas las GIs asociadas a estos genes. Adicionalmente, se buscó una explicación a la asimetría que existe entre la frecuencia de inserción de GIs entre tDNAs que poseen la misma secuencia de nucleótidos, desafiando las principales hipótesis propuestas anteriormente sobre los fundamentos biológicos del uso de tDNAs y tmDNAs como sitio de integración de MGEs.

2. OBJETIVOS

2.1 Objetivo General

Identificar y caracterizar los tDNAs presentes en una colección de cepas de *K. pneumoniae* y las islas genómicas integradas en dichos genes.

2.2 Objetivos Específicos

- 2.2.1 Identificar y caracterizar el conjunto de tDNAs típicamente encontrado en miembros de la especie *K. pneumoniae*.
- 2.2.2 Identificar y caracterizar las Islas Genómicas asociadas a los tDNAs analizados.
- 2.2.3 Buscar patrones en los tDNAs, o en sus contextos genómicos, que expliquen la disparidad en su uso como sitios de integración de Islas Genómicas.

3. MATERIALES Y MÉTODOS

3.1 Secuencias Analizadas

Las secuencias de los genomas de las cepas de *K. pneumoniae* analizadas en este trabajo se obtuvieron desde la base de datos de secuencias de nucleótidos del Centro Nacional para la Información Biotecnológica (NCBI), que es parte de los Institutos Nacionales de la Salud (NIH) de Estados Unidos (NCBI Resource Coordinators, 2017). Se consideraron cincuenta genomas que cumplieron la condición de estar ensamblados y anotados. El número de acceso en la base de datos y una selección de los metadatos de las cepas seleccionadas se muestran en Anexo I.

3.2 Identificación Manual de Islas Genómicas

La identificación primaria de tDNAs para todos los genomas se realizó usando la herramienta ARAGORN (Laslett & Canback, 2004). De esta forma todos los tDNAs analizados en este trabajo se predijeron bajo un mismo criterio. Luego, se analizaron las secuencias río arriba y río abajo de cada tDNA, de modo de identificar su contexto genómico característico y la presencia de islas genómicas. Dicho contexto genómico se definió arbitrariamente como los cuatro genes que se sitúan río arriba y río abajo del tDNA, considerando su identidad (secuencia nucleotídica de la región codificante) y organización (ordenamiento). En caso de ser necesario, la identidad de marcos de lectura abierta (ORFs) no anotados se resolvieron mediante el uso de BLAST (Altschul, et al., 1990), comparando con la base de datos de nucleótidos de la NCBI o la base de datos de proteínas UniProtKB (Pundir, et al., 2017). La inspección visual de los genomas se llevó a cabo con el software SnapGene Viewer® (GSL Biotec).

3.3 Alineamientos de tDNAs y su contexto.

La evaluación de la conservación a nivel de secuencia de los tDNAs y las regiones que los flanquean se llevó a cabo mediante alineamientos múltiples. Para ello,

las secuencias se extrajeron de los genomas utilizando un script desarrollado utilizando Biopython (Cock, et al., 2009) y los alineamientos se realizaron con ClustalW (Thompson, et al., 1994). Los porcentajes de identidad presentados se calcularon como el promedio del porcentaje de identidad de todos los alineamientos de a pares realizados por ClustalW. En el caso de los contextos genómicos, se extrajeron 6.000 o 1.000 pb desde las regiones adyacentes a los tDNAs y utilizadas para realizar alineamientos múltiples como se describió anteriormente. Por otra parte, los tDNAs se anotaron siguiendo los lineamientos de la nueva nomenclatura sistemática propuesta para estos genes como parte de este trabajo (ver sección de Resultados).

3.4 Identificación y Análisis de Islas Genómicas

Las islas genómicas se identificaron durante la curación manual de los tDNAs identificados en el cromosoma de cada cepa de *K. pneumoniae*. En este sentido, las islas genómicas se identificaron evaluando el corrimiento del contexto conservado río abajo de cada tDNA (provocado por la inserción de la isla). Las secuencias identificadas como GIs se extrajeron para identificar los genes codificados en ellas y clasificarlas. Para ello, se utilizó la herramienta Blast2Go (Götz, et al., 2008).

3.5 Análisis de Curvas de Acumulación de tipos de tDNAs y tipos de GIs

La acumulación de nuevos tipos de contextos de los tDNAs (considerando su contexto genómico) y de nuevos tipos de islas genómicas se calculó utilizando el paquete *vegan* de R, así como se ha descrito previamente (Holt, et al., 2015). Para evaluar si la cantidad de tipos tiende a aumentar o a converger en un valor, las curvas de acumulación obtenidas se ajustaron según la ecuación de la Ley de Heaps (Tettelin, et al., 2008), con la ecuación $N(x) = kN^r$ donde $N(x)$ es la cantidad de tipos identificados en función del

número de cepas analizadas (M) y el exponente de crecimiento se utilizó para determinar si el conjunto es abierto ($\gamma \geq 0$) o si tiende a estabilizarse en un valor ($\gamma = 1$).

3.6 Búsqueda de Motivos Conservados

Para la búsqueda de motivos conservados entre las secuencias analizadas se utilizaron dos estrategias. La primera, utilizando la plataforma de análisis MEME (Bailey, et al., 2002) que identifica y calcula lo conservado que está un determinado motivo en una colección de secuencias. El programa se utilizó con los parámetros definidos por defecto. La segunda estrategia fue utilizar la herramienta de análisis DREME (Bailey, 2011), que permite no sólo identificar algún motivo conservado en un grupo de secuencias de DNA, sino también compararlos contra una o más secuencias donde no se espera encontrar algún motivo. De esta forma, el programa analiza también un control negativo de la presencia de una secuencia conservada. Ambas herramientas pueden encontrarse en la plataforma MEME de búsqueda de motivos en secuencias de DNA o proteínas (Bailey, et al., 2015).

3.7 Mapeo de los sitios de unión de proteínas del nucleoide en el genoma de *K. pneumoniae*

No existe en la literatura evidencia respecto de los sitios de unión de proteínas asociadas al nucleoide bacteriano en *K. pneumoniae*. Para aproximar la dinámica de las proteínas Fis y HNS en *K. pneumoniae*, la unión a distintas regiones del cromosoma de dichas proteínas se mapearon a partir de datos provenientes de experimentos de inmunoprecipitación de cromatina seguida de secuenciación masiva del DNA (ChIP-Seq) en la cepa *E. coli* K12 MG1655 previamente publicados (Kahramanoglou, et al., 2011). Para este análisis, la cepa *K. pneumoniae* 1084 se utilizó como referencia. Ambos cromosomas se alinearon utilizando la herramienta Mauve (Darling, et al., 2004)

que permite identificar bloques conservados entre ambos genomas. Luego, la información de estos bloques conservados se consideró para ubicar las señales de unión de las proteínas del nucleoide desde el cromosoma de *E. coli* y el de *K. pneumoniae* utilizando scripts de Python escritos para resolver esta tarea. Finalmente, se promediaron las señales de uso de las proteínas en la extensión de los tDNAs para comparar la señal de la unión de proteínas asociadas al nucleoide y la frecuencia de uso de los tDNAs como sitios de integración de MGEs.

3.8 Análisis Pangenómicos

Para los análisis pangenómicos se utilizó el programa Roary (Page, A. et al., 2015), que compara los genes codificados en los cromosomas de las cepas analizadas, calculando las frecuencias con las que cada gen está presente en el grupo de genomas analizados. Además, luego de la identificación del genoma núcleo y del genoma accesorio de la especie, las proteínas fueron utilizadas para determinar a qué grupo de genes ortólogos (COGs) pertenecen, utilizando la herramienta en línea de eggNOG (Jensen, et al., 2008).

3.9 Análisis de Filogenia

Los análisis filogenéticos se realizaron a partir de alineamientos múltiples o colecciones de genes concatenados y también alineados. Para resolver las relaciones filogenéticas entre las secuencias nucleotídicas o proteicas, fueron construidos árboles con algoritmo de máxima verosimilitud (*maximum likelihood*) utilizando el programa RAxML (Stamatakis, A., 2014). Para evaluar las agrupaciones obtenidas se calculó el puntaje de Bootstrap a partir de 1.000 árboles construidos con la herramienta.

3.10 Análisis Estadísticos

Los análisis estadísticos fueron realizados utilizando el software GraphPad Prism 7. Los detalles de cada análisis realizado se encuentran especificados junto a los resultados. Para las búsquedas de correlaciones se testearon índices paramétricos o no-paramétricos, en búsqueda de cualquier relación entre los datos.

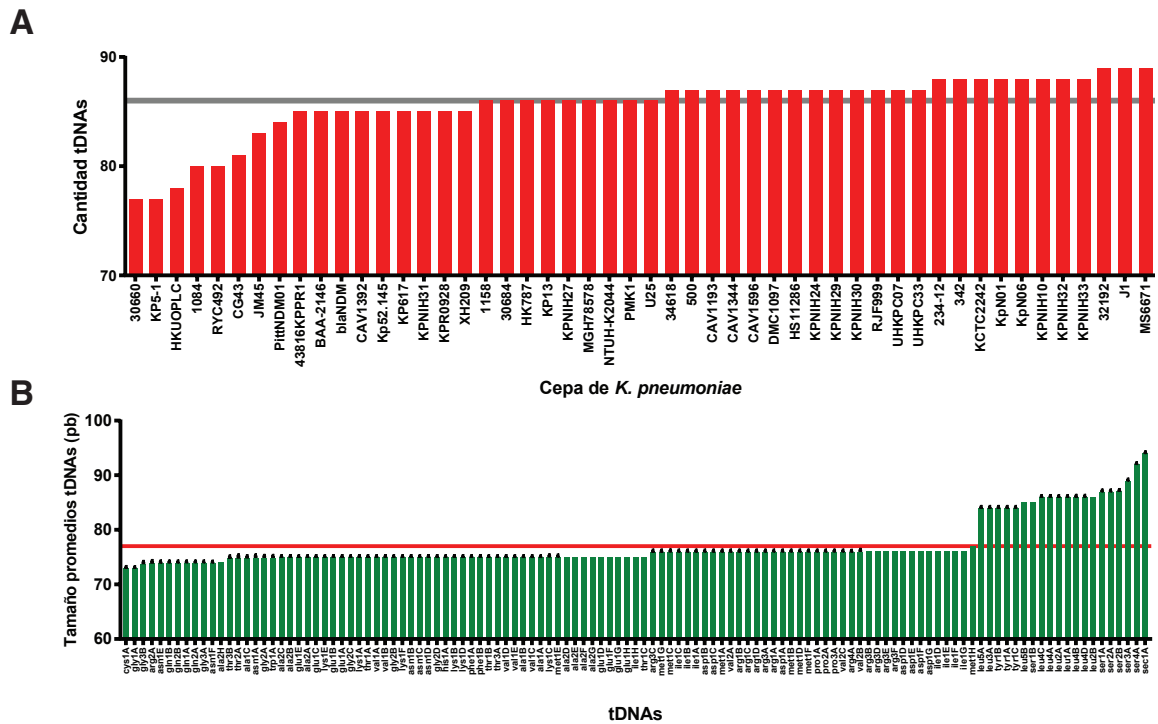
4. RESULTADOS

4.1 Identificación de tDNAs y tmDNAs en las cepas de *K. pneumoniae*

Con el objetivo de investigar el número, tipo y organización típica de los genes que codifican para tRNAs y tmRNAs en *K. pneumoniae*, dichos *loci* se identificaron en el cromosoma de 50 cepas de esta especie, cuyo genoma había sido previamente secuenciado y ensamblado. Para ello, se utilizó la herramienta ARAGORN y la posterior curación manual de los datos. De esta manera, se identificaron un total de 4.281 *loci* de tDNAs. En promedio, los cromosomas de *K. pneumoniae* incluyen 86 tDNAs, con un mínimo de 76 y un máximo de 89 (Fig.5A). Dicho grupo comprende tDNAs que se encuentran en copia única, y otros que se encuentran en múltiples copias (ubicadas en distintas posiciones del genoma). Por otra parte, el tamaño promedio de los tDNAs analizados resultó ser de 76 ± 4 pb (rango entre 73 y 94 pb) (Fig. 5B). En el caso de los tmDNAs, todos los genomas poseían sólo una copia, la cual tiene una extensión promedio de 363 ± 2 pb.

4.2 Comparación entre la frecuencia de codones y de anticodones.

Al analizar en detalle el set de tDNAs presentes en la mayoría de los cromosomas de *K. pneumoniae* analizados, se observó que éste codifica solo una fracción de los anticodones requeridos para decodificar el total de codones posibles. De los 62 anticodones del código genético bacteriano, 22 no están presentes en los genomas analizados (Anexo II). Si bien se sabe que la carencia de genes que codifican tRNAs con ciertos anticodones puede ser suplida por mecanismos de edición post-transcripcional del anticodón (Hou, et al., 2015; Manickam, et al., 2016; Vold, 1978), no se tiene claridad respecto al por qué durante la evolución de esta especie se ha seleccionado este grupo particular de tDNAs por sobre el resto posible. Una posibilidad es que el set de tRNAs



codificados en el cromosoma correspondan a aquellos con anticodón complementario a los codones más frecuentemente usados. En vista de lo anterior, quisimos determinar si el set de tDNAs presentes en *K. pneumoniae* se correlaciona con el uso preferencial de codones típico de esta especie. Para ello, se determinaron los anticodones portados por cada uno de los tDNAs identificados usando ARAGORN. Por otra parte, también se calcularon los índices de uso de codones de los genes presentes en los genomas de las cepas analizadas con la herramienta GCUA (McInerney, 1998), usando como referencia la cepa *K. pneumoniae* 1084. La frecuencia de anticodón se calculó como el número de tDNAs que codifican un mismo anticodón sobre el total de tDNAs para un mismo aminoácido (Anexo II). Por ejemplo, el aminoácido glutamina puede ser transferido por tRNAs que portan un anticodón UUG o CUG. Al analizar el set de tDNAs que conforma el pangenoma de *K. pneumoniae*, se observó que el 50% de los tDNAs de glutamina codifican el anticodón UUG y el 50% restante CUG (Fig. 6). Por otro lado, de acuerdo con el uso preferencial de codones calculado para el genoma de *K. pneumoniae*, el codón complementario a UUG corresponde al usado en el 15% de las veces en que se especifica el aminoácido glutamina, mientras que el complementario al anticodón CUG se usa en el 85% de los casos. Lo anterior evidencia que no necesariamente los codones más usados son decodificados por tRNAs cuyos genes se encuentren en un mayor número de copias. Dicho de otro modo, a pesar de que el tRNA que porta el anticodón UUG es usado en el 84% de los casos en que se adiciona un residuo glutamina, el número de tDNAs por cromosoma que codifican dicho tRNA no es superior al de los tDNAs que codifican el otro codón posible, y que es usado con mucho menor frecuencia.

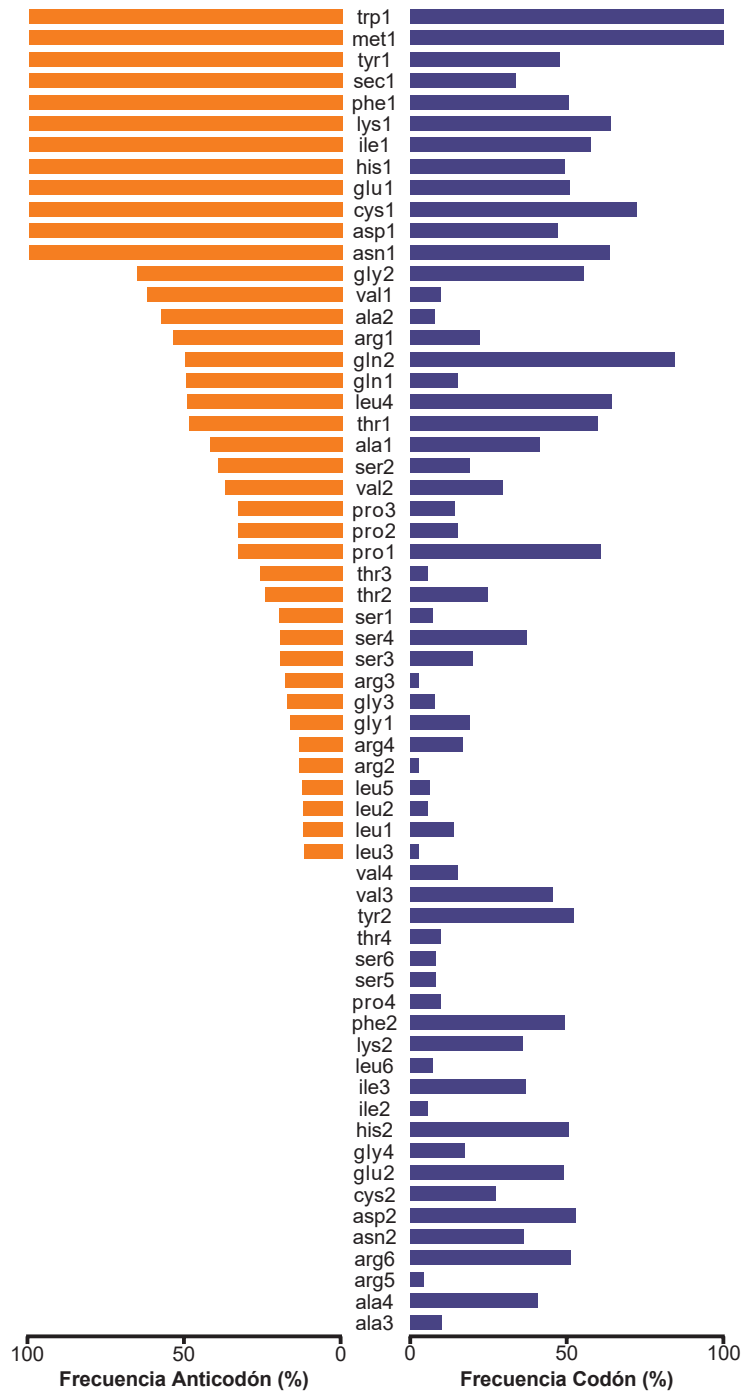


Figura 6: Comparación entre la frecuencia de anticodones codificados en los tDNAs de *K. pneumoniae*, y la frecuencia de uso de los codones complementarios.

Gráfico piramidal donde se comparan las frecuencias de anticodones codificados en los tDNAs de *K. pneumoniae* y el uso preferencial de codones típico de esta especie. Notar la asimetría entre ambas, donde existen 22 anticodones que no están codificados por ningún tDNA del genoma de *K. pneumoniae*, algunos de ellos requeridos para decodificar codones usados con una frecuencia superior al 50%.

Por otra parte, existe un grupo sorprendentemente alto (35%) de codones que son utilizados con frecuencias entre el 4% y el 50%, pero cuyos anticodones complementarios no están codificados en ninguno de los tDNAs de *K. pneumoniae* (frecuencia del anticodón equivalente a cero). Un caso destacable es el de arginina, la cual puede ser decodificada por 6 posibles codones, siendo el más usado CGC ($f_{\text{cod}}=51\%$). Sin embargo, en el genoma de *K. pneumoniae* no existen tDNAs que codifiquen el tRNA que porta el anticodón GCG complementario a dicho codón ($f_{\text{ant}}=0$). En contraste, más de la mitad de los arg-tDNAs presentes en el cromosoma codifican el anticodón ACG ($f_{\text{ant}}=54\%$), el cual permite decodificar un codón usado con una frecuencia de sólo un 22%. Lo anterior motiva una pregunta altamente relevante: ¿por qué se ha seleccionado evolutivamente que el tRNA con el anticodón más frecuentemente requerido para decodificar un determinado aminoácido no esté codificado en el genoma, mientras que otros que se necesitan menos frecuentemente sí lo estén? Más aún, ¿cuáles son las fuerzas evolutivas que han llevado a la selección de este set particular de tDNAs por sobre otro? La evidencia presentada anteriormente indica que la selección evolutiva del set de tDNAs presentes en el cromosoma de *K. pneumoniae* no estaría explicada de manera evidente por el uso preferencial de codones característico de esta especie.

A nivel de la traducción, la carencia de algunos tDNAs implica que no habrían tRNAs con determinados anticodones, y que por lo tanto habría codones que no se podrían decodificar, haciendo imposible la elongación de algunas proteínas. Sin embargo, el sistema traductor es flexible a distintas escalas. En primer lugar, existe evidencia que indica que el anticodón de algunos tRNAs puede ser editado post-transcripcionalmente y la hipótesis del balanceo o *woobling* (Crick, 1966), que incluye que la posición 5' del anticodón no requiera aparear en la interacción codón-anticodón y

además las adeninas en el sitio del anticodón pueden ser convertidas en inosina (Alseth, et al., 2014), este cambio permite que la inosina en aquella posición pueda aparearse con citosina, uracilo o adenina, lo que permite suplir la falta de ciertos tRNAs. Por otra parte, se ha demostrado que la complementariedad requerida entre el codón y anticodón para la transferencia de aminoácidos en el ribosoma no excluye la posibilidad de que algunos de los nucleótidos no aparezcan (Hou, et al., 2015; Zhang, et al., 2016). Finalmente, se ha propuesto que esta disparidad entre la presencia de los anticodones y codones podría estar asociada a mecanismos de regulación de la traducción, por ejemplo, afectando la velocidad de traducción (Gorgoni, et al., 2014).

4.3 Una nueva nomenclatura para los tDNAs de *K. pneumoniae*

Para el análisis preciso de los tDNAs identificados en cuanto a su secuencia nucleotídica y contexto genómico, fue necesario organizarlos y clasificarlos, de modo de poder sistematizar las comparaciones. Actualmente no existe un consenso respecto a cómo se deben anotar o nombrar los tDNAs en procariontes, causando que la nomenclatura varíe entre distintas cepas. La práctica más habitual es considerar como referencia la nomenclatura que fue utilizada para la anotación de los tDNAs en el genoma de la cepa *E. coli* K12 (Blattner, et al., 1997), la que no es más que el uso del código de tres letras para los aminoácidos y una letra que indica la posición del tDNA a partir de un origen determinado. Por ejemplo, el gen *thrW* de esta cepa de *E. coli* fue identificado como sitio de integración de un profago durante la secuenciación de su genoma. Luego, este sistema fue utilizado para algunas otras especies bacterianas, principalmente debido a que muchos proyectos de secuenciación y ensamblaje usan como referencia el genoma de *E. coli*. No obstante, esta nomenclatura no es sistematizada, no es universalmente utilizada, y su aplicabilidad a otras especies bacterianas no ha sido validada. Finalmente, dicha nomenclatura no permite individualizar tDNAs 100%

idénticos ubicados en distintos contextos genómicos, y por lo tanto impide su comparación certera entre múltiples cepas. Considerando que la dinámica de integración y escisión de GIs podría estar influenciada por el contexto genómico del sitio de recombinación y ante la falta de un sistema de nomenclatura sistemática para dar cuenta de los tDNAs y su contexto, como parte de este trabajo se desarrolló una nueva nomenclatura para tDNAs en *K. pneumoniae*, la que eventualmente podría ser adaptada y utilizada en cualquier genoma procarionte.

El sistema identifica los tDNAs siguiendo las siguientes directrices (Fig. 7):

- (i)** Utilizando una herramienta de predicción de tDNAs, como ARAGORN o tRNAScan, deben determinarse las coordenadas de los tDNAs en la secuencia blanco del análisis.
- (ii)** La primera parte de la nomenclatura se compone de dos elementos, el primero es la identificación del aminoácido transferido por el tRNA, usando el código de tres letras para la identificación de aminoácidos. Luego, el anticodón que posee el tDNA es identificado con un número arábigo, que es seleccionado a partir de una tabla construida con este fin. El orden de los números y de los anticodones se seleccionó de forma arbitraria.
- (iii)** Finalmente, el contexto genómico del tDNA es determinado, considerando tres genes río arriba y río abajo del tDNA. Esta instrucción es relativamente flexible, y en realidad son considerados tantos genes como sean necesarios para poder diferenciar los contextos entre ellos. En general, se consideró hasta un máximo de 10 kpb, por ejemplo, algunos tDNAs están rodeados por genes que codifican rRNAs, los que a su vez se encuentran en múltiples copias en el genoma. El contexto identificado es luego representado con una letra, que es derivada de una lista de contextos previamente caracterizados.

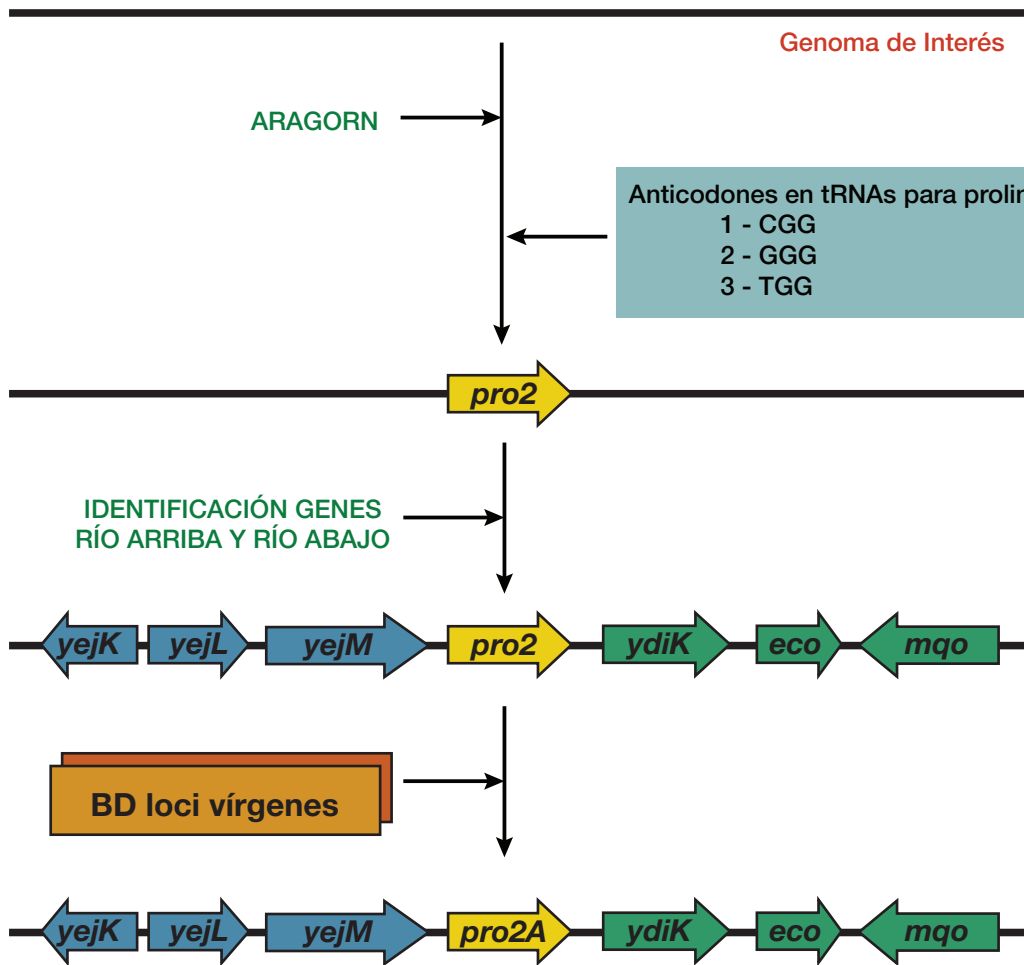


Figura 7: Esquema de la curación manual de los loci analizados y del nuevo sistema de nomenclatura propuesto.

En este caso se presenta la asignación de la nomenclatura para el tDNA *pro2A*. En el primer paso, el gen del tDNA fue identificado mediante la herramienta ARAGORN, con ello también se identificó el anticodón portado en el gen. En este caso, el gen codifica para un tDNA de prolina (pro) y que porta el anticodón GGG (2). Luego, se identifica en los contextos río arriba y río abajo del gen, los que finalmente son comparados con los contextos en la base de datos de *loci* vírgenes. En el ejemplo, la identidad y configuración de los genes coincide con un *locus* virgen previamente con la letra A. Sumado lo anterior, el nombre del tDNA es *pro2A*.

(iv) Respecto del paso anterior conviene aclarar dos puntos. Primero, la letra a utilizar se corresponderá con algún contexto previamente caracterizado en el caso de que exista. Si no se ha identificado algún contexto con las características del analizado, deberá utilizarse la letra que le sigue en el orden del alfabeto inglés luego del último contexto descrito. Otro punto que considerar es que la integración de GIs u otros MGEs genera un corrimiento de los elementos río debajo de los tDNAs, por lo que el determinante de la identificación del contexto son los genes río arriba y debiesen ser suficientes para identificar y definir el contexto.

Lo anterior permite entonces que en la clave de anotación de los tDNAs se pongan de manifiesto sus características primordiales: el aminoácido transferido por el tRNA codificado, el anticodón portado, y el contexto genómico de dicho tDNA. Como se menciona en las instrucciones, los contextos son comparados con una base de datos de contextos conocidos (previamente definidos e identificados), denominada base de datos de *loci vírgenes*, concepto al que se hará alusión más adelante.

Considerando este nuevo sistema de nomenclatura, fue posible clasificar los 4.281 tDNAs analizados dentro de 119 tipos diferentes, 39 de estos son presentados como referencia (Fig. 8). Entonces, estos tipos de loci de tDNAs difieren en al menos una de las siguientes características: 1) el aminoácido que porta el tRNA codificado, 2) el anticodon de dicho tRNA, o 3) su contexto genómico. Esta clasificación permitió estudiar algunas características de los tDNAs que se mencionan a continuación.

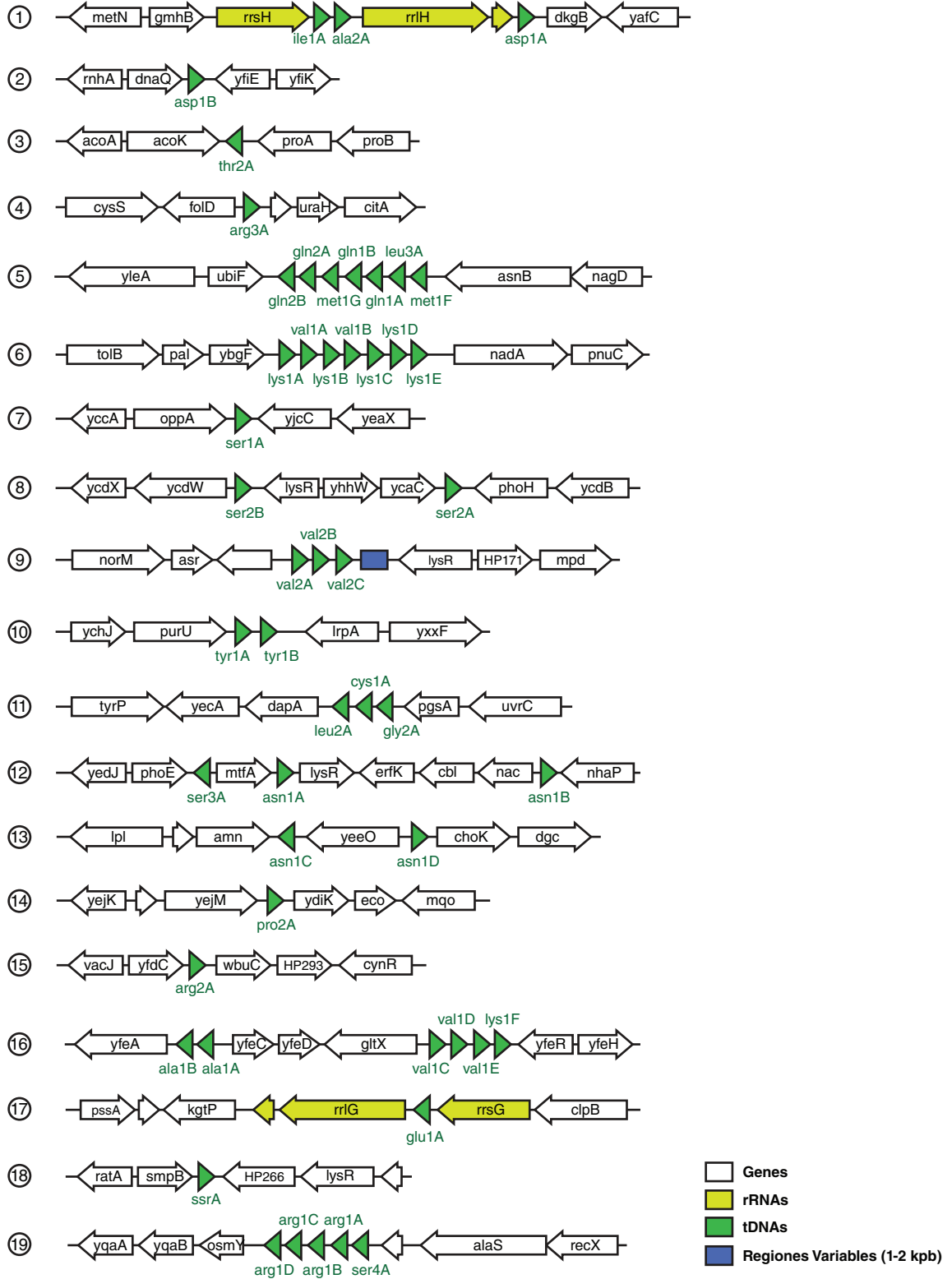


Figura 8: Esquemas de los contextos de 39 loci de tDNAs seleccionados
 Continúa en la página anterior.

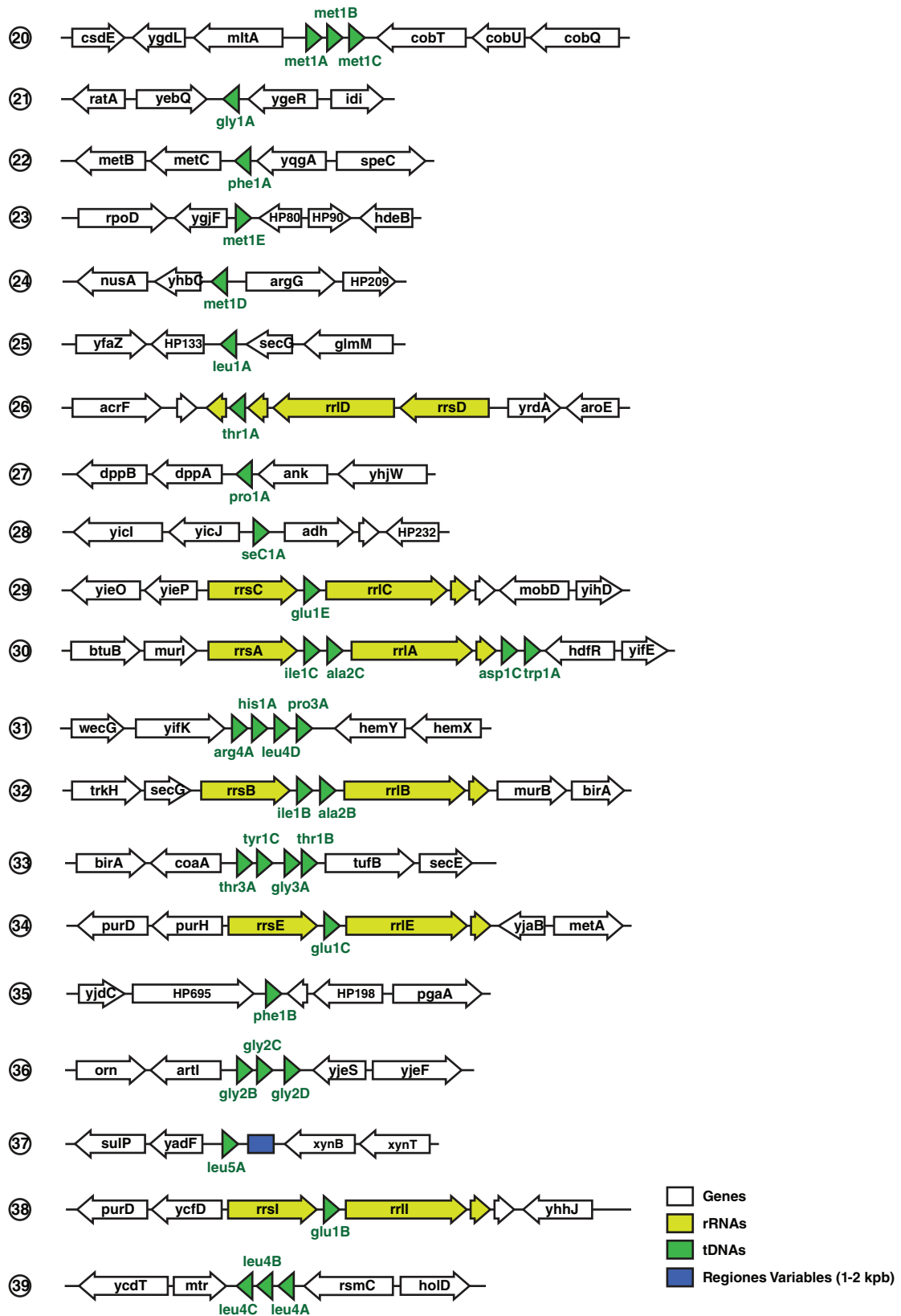


Figura 8: Esquemas de los contextos de 39 *loci* de tDNAs seleccionados
 Sigue desde la página anterior.

4.4 Prevalencia y conservación de los tDNAs de *K. pneumoniae*

Considerando la nueva clasificación de los tDNAs, se determinó la prevalencia de cada uno de ellos entre las cepas de *K. pneumoniae* analizadas. Ésta fue calculada como el porcentaje de cepas que posee un determinado tDNA (Fig. 9A). De esta manera, se identificó un set de tDNAs presentes en más del 80% de las cepas, el cual se denominó “colección núcleo de tDNAs” (tDNA core set). De los 119 tipos de tDNAs identificados, 86 forman parte de este grupo, lo que coincide con el número promedio de tDNAs determinado para la muestra de cepas analizadas. Otro grupo observado corresponde a los tDNAs raros, cuya prevalencia es menor al 30%. Entre estos tDNAs se encuentran tDNAs presentes en profagos y en GIs, y también contextos que se encuentran en regiones del cromosoma que presentan anomalías en la organización génica y por lo tanto podrían ser producto de errores en el proceso de ensamblaje del cromosoma. Por ello, los tDNAs raros fueron excluidos de los análisis presentados en las secciones sucesivas de este estudio.

Para evaluar el grado de conservación de la secuencia de los distintos tDNAs entre las distintas cepas de *K. pneumoniae*, estos se compararon a través de alineamientos múltiples, calculando los porcentajes de identidad promedio (Fig. 9B). Destaca de este análisis la alta conservación de secuencia de los tDNAs, 98 de los cuales mostraron una identidad promedio superior al 98%. Los tDNAs menos conservados se encontraron asociados a profagos y/o a contextos raros, por lo que la evidencia sugiere que la colección núcleo de tDNAs es altamente conservada. Conviene señalar que para este análisis se consideraron sólo los tDNAs identificados en al menos dos cepas, dado que es imposible calcular el promedio de identidad de una única secuencia.

4.5 El contexto genómico de los tDNAs: definición del *locus virgen*

Como se mencionó previamente, la integración de GIs en los *loci* de tDNAs ocasiona el corrimiento del contexto ubicado río abajo de dichos genes. Teniendo en cuenta esta propiedad, al comparar el contexto genómico de un mismo tDNA en múltiples cromosomas es posible deducir la configuración de esa región del genoma en ausencia de la integración de GIs. Dicha configuración encontrada en los *loci* que no han sido blanco de la integración de GIs u otros MGEs, se definió como el *locus virgen*. En términos formales, el *locus virgen* corresponde a la colección de genes y a su organización en torno a un tDNA donde no se ha insertado una IG. La colección de estos contextos vírgenes como parte de una base de datos es la que sirve de referencia para la identificación de contextos y nombramiento de tDNAs según la nomenclatura propuesta.

Con el objetivo de evaluar la conservación del contexto genómico de cada uno de los tipos de tDNAs identificados, se realizó un alineamiento múltiple de las regiones río arriba y río abajo de cada tDNA (Fig. 10). Al considerar un segmento de hasta 6 kpb, se observó que ambas regiones poseen identidades promedio superiores al 90% en la mayoría de los casos. No obstante, la región río arriba resultó ser altamente conservada para prácticamente la totalidad de los tDNAs, mientras que se observaron varios casos de tDNAs cuya región río abajo resultó ser más variable. Considerando lo anterior, para la identificación e individualización de cada tDNA en el cromosoma de *K. pneumoniae* es suficiente la secuencia de la región situada río arriba del mismo, la cual es altamente conservada entre distintas cepas de esta especie. Luego, la comparación del contexto río abajo de un mismo tDNA en distintos genomas permite deducir de manera directa la configuración virgen de dicho *locus*.

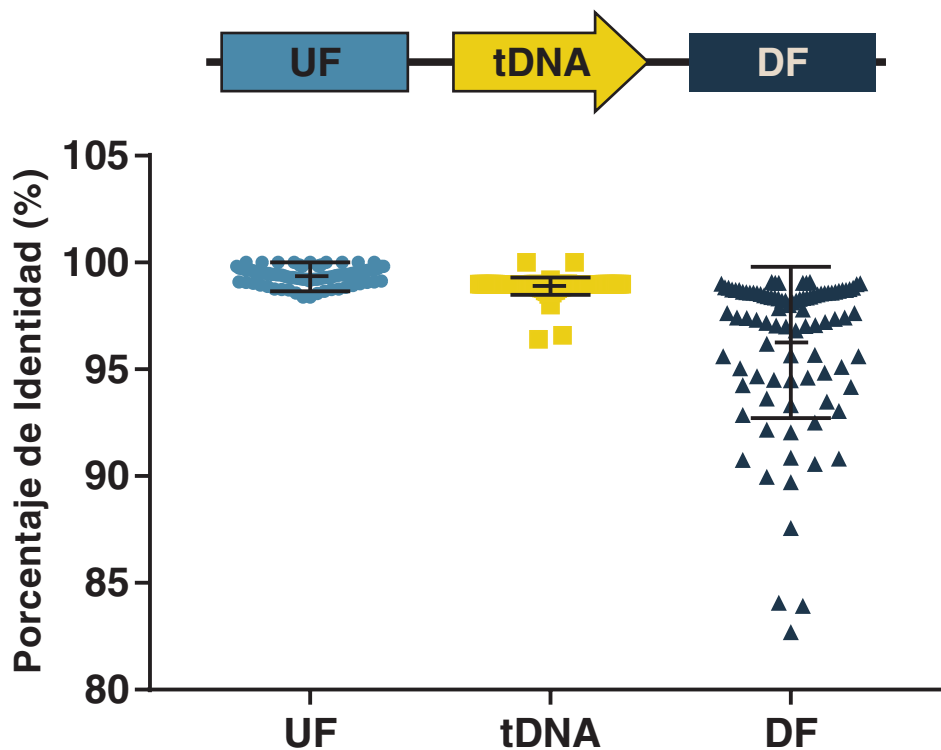


Figura 10: Conservación promedio de los elementos que constituyen los loci virgen de los tDNAs analizados.

El gráfico presenta el promedio de porcentaje de identidad de los distintos elementos del *locus* virgen (representado en la zona superior): la región río arriba (UF), el tDNA y la región río abajo (DF). Cada punto representa el promedio del porcentaje de identidad de un tipo de tDNA distinto y las barras representan el promedio de la muestra (línea media) y la desviación estándar del promedio de la muestra (extremos).

Al analizar el contexto en el que se encuentran los distintos tDNAs en el cromosoma de *K. pneumoniae*, se observaron básicamente tres situaciones: 1) tDNAs adyacentes a genes que codifican proteínas; 2) tDNAs adyacentes a genes que codifican rRNAs; 3) tDNAs adyacentes a otros tDNAs, formando estructuras similares a operones. Respecto de este último grupo, existe evidencia que sugiere que algunos de estos clústeres se transcriben de forma policistrónica (Mohanty & Kushner, 2007). Sin embargo, no existe evidencia directa de que en *K. pneumoniae* la transcripción de los tDNAs dispuestos en estructuras similares a operones sea de forma policistrónica. Análisis futuros serán requeridos para determinar la naturaleza de estas unidades transcripcionales.

4.6 Identificación de GIs

Como parte del presente trabajo se desarrolló una nueva estrategia de identificación de islas genómicas en *K. pneumoniae*, la cual se basa en la colección de *loci* vírgenes definida en base al análisis multicromosómico descrito anteriormente. Considerando la conservación de los contextos río arriba y río abajo de cada tDNA, así como el corrimiento del contexto río abajo producto de la integración de una IG, las islas genómicas (y otros MGEs) pueden detectarse como la interrupción de la continuidad del *locus* virgen. Por ejemplo, el *locus* virgen del tDNA *pro2A* está definido por seis genes, en el contexto río arriba los genes *yejK*, *yejL*, *yejM* y en la región río abajo los genes *ydiK*, *eco* y *mgo* (Fig. 11). En la cepa *K. pneumoniae* PMK1 dicha configuración de *locus* virgen se encontró interrumpida por un segmento de DNA con características típicas de una GI: un repetido directo de 21 pb, un gen que codifica una integrasa y siete genes que podrían codificar una función biológica.

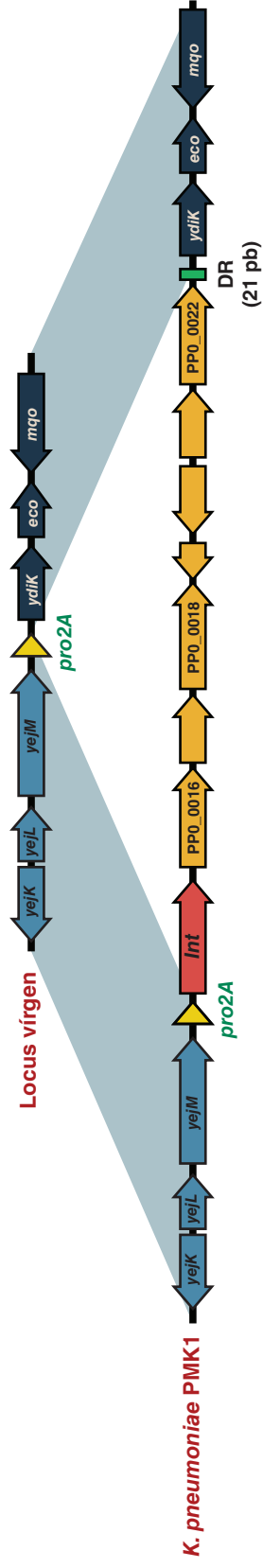


Figura 11: Esquema del método de identificación de GIs basado en la continuidad del *loci* virgen.

En el ejemplo, el *locus pro2A* en la cepa *K. pneumoniae* PMK1 es comparada con el *locus* virgen descrito para este tipo de tDNA. Notar que la integración, en este caso de una isla genómica, ocasiona el corrimiento de los genes en el contexto río abajo (azul), dejando inalterada la continuidad entre el contexto río arriba (celeste) y el tDNA (amarillo). En el caso de esta isla, se identificó un gen codificando una integrasa (*Int*), un repetido directo (DR) y ORFs que codifican proteínas de función desconocida.

Utilizando este sistema de análisis, se analizaron los 4.281 *loci* de tDNAs presentes en las cepas analizadas y se identificaron 307 MGEs, entre los que hay GIs, profagos, transposones y secuencias de inserción.

Cabe destacar que el método de análisis utilizado permite la búsqueda de GIs en el cromosoma sin la necesidad de análisis filogenéticos y no depende de la composición de la secuencia de nucleótidos ni de la presencia de genes que codifican integrasas u otros. Además, es posible determinar con precisión los límites de la IG, dado que comienza en el extremo 3' del tDNA y termina en el punto en que es posible identificar la secuencia río abajo desplazada por la IG o por la presencia de los repetidos directos.

4.7 Frecuencia de inserción de MGEs en los tDNAs de *K. pneumoniae*

Luego de identificar todos los MGEs asociados a tDNAs, se analizó la distribución de estos elementos en las distintas cepas. Se observó que la totalidad de las cepas analizadas posee al menos 3 MGEs integrados en alguno de sus tDNAs. Las cepas HKUOPLC, J1 y KPNIH27 fueron las que tuvieron la menor cantidad de MGEs, mientras que la cepa 342 presentó el mayor número de GIs, nueve en total (Fig. 12A). El promedio calculado para el set de cepas de *K. pneumoniae* analizado resultó ser de 6 GIs por cromosoma aproximadamente.

Al evaluar y cuantificar la presencia de GIs en los tDNAs de las distintas cepas de *K. pneumoniae* se observó que, de los 119 tipos de tDNAs caracterizados, sólo 18 se encontraron asociados a GIs. Es decir, sólo una fracción minoritaria (15%) de los tDNAs presentes en el cromosoma de esta especie son utilizados como sitios de integración de islas genómicas y otros MGEs. Para indagar más al respecto, se calculó la frecuencia de ocupación por GIs para cada tipo de tDNA, definida como el número de cepas que poseen una IG integrada en dicho tDNA dividido por el número total de cepas (Fig. 12B).

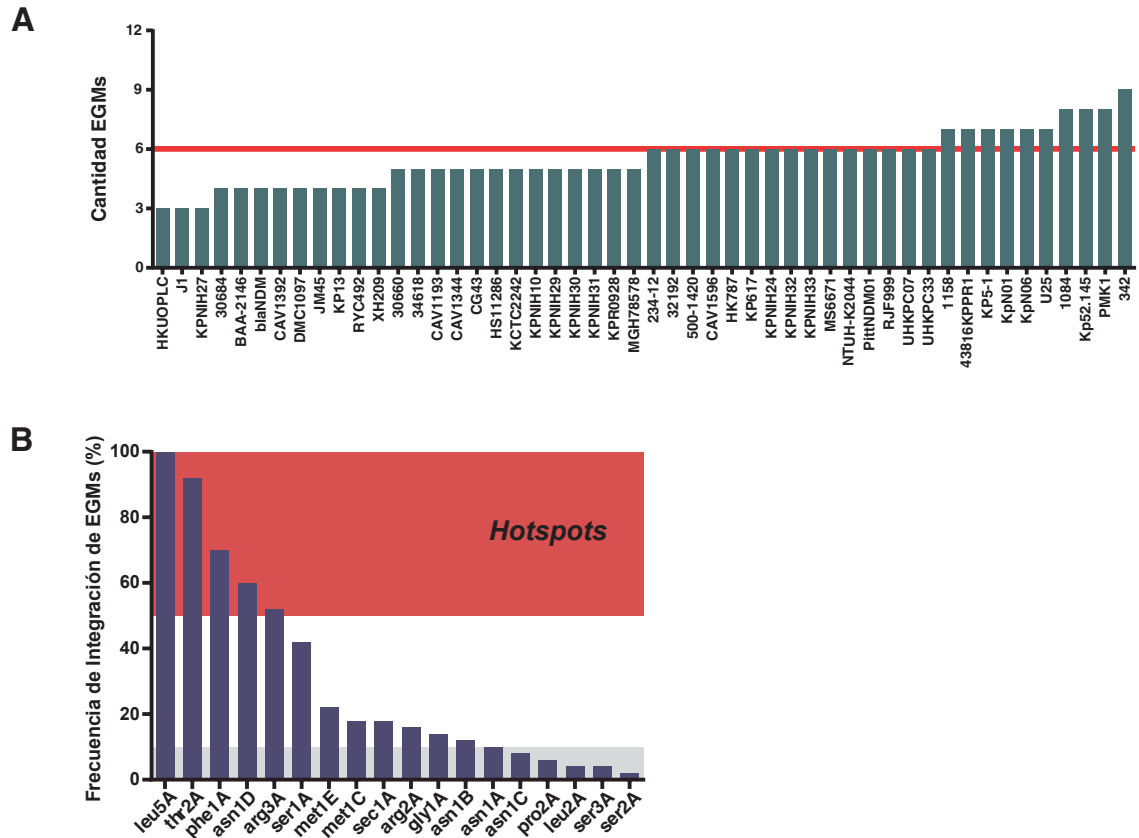


Figura 12: Distribución de las islas identificadas entre las cepas analizadas y la frecuencia con que los tDNAs son utilizados como sitios de integración en *K. pneumoniae*.

(A) Número de islas genómicas identificadas en cada cepa analizada. La línea representa la cantidad promedio de islas entre las cepas. (B) Frecuencia de uso de distintos tDNAs como sitio de integración de IGs. Las frecuencias fueron calculadas como el porcentaje de veces que los *loci* estaban ocupados por un EGM sobre el total de cepas analizadas. La región roja del gráfico muestra los *hotspots* de integración de IGs y otros EGMs, definidos como frecuencias por sobre el 50%. En el otro extremo se encuentran los tDNAs menos usados, seis de los sitios de integración identificados poseen frecuencia de uso menor a un 10% (región gris).

En base a las frecuencias calculadas se identificaron los tDNAs que corresponden a *hotspots* de integración de MGEs, definidos como aquellos que se encontraban ocupados por GIs en sobre el 50% de las cepas. Cinco tDNAs están incluidos dentro de este grupo: *arg3A*, *asn1D*, *phe1A*, *thr2A* y *leu5A*. De estos, algunos previamente habían sido descritos como *hotspots*, este es el caso de los *loci phe1A* (van Aartsen, 2008), *thr2A* (Liu & Zhu, 2010) y *asn1D* (Marcoleta, et al., 2016). Destaca el caso de *leu5A*, que corresponde al tDNA más utilizado como sitio de integración, encontrándose islas integradas en este *locus* en el 100% de las cepas estudiadas. En el caso de la única copia del gen que codifica el tmRNA encontrada en *K. pneumoniae* (anotado como *ssrA*), su frecuencia de uso resultó ser de un 48%. Otros tDNAs son utilizados con frecuencias inferiores al 50% de los casos, en el extremo inferior está *ser2A*, donde sólo una de las cepas presentó una IG integrada en este *locus*.

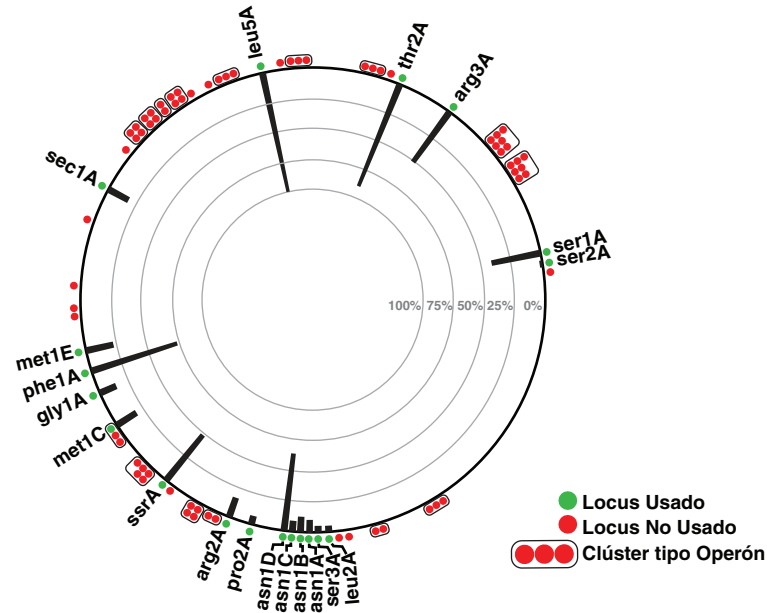
Tal como se mencionó anteriormente, es comúnmente aceptado que el uso de un tDNA en particular como sitio de integración estaría determinado principalmente por la secuencia de dicho tDNA, parte de la cual corresponde al sitio reconocido por la integrasa que cataliza la integración. No obstante, en un estudio reciente publicado por nuestro grupo donde se identificaron GIs asociadas a tDNAs de asparagina, se observó que si bien existen 4 copias idénticas de dicho tDNA en el cromosoma de *K. pneumoniae*, éstas son usadas como sitio de integración con una frecuencia muy disímil (Marcoleta, AE., et al. 2016). Con el objetivo de indagar si este fenómeno ocurre con otros tipos de tDNAs, se compararon las frecuencias de uso como sitios de integración de los tDNAs que se encuentran en múltiples copias idénticas dentro del cromosoma (Fig. 13). Respecto de la ubicación de los tDNAs utilizados como sitios de integración en los cromosomas analizados, pueden resolverse algunas generalidades respecto de su organización (Fig. 13A). En primer lugar, los tDNAs utilizados como sitios

de integración son en su mayoría tDNAs solitarios, es decir, que no están rodeados por otros tDNAs o rRNAs. El único tDNA que no cumple con lo anterior es *met1C*, que es el último elemento de una estructura de tipo operón. Además, los tDNAs utilizados como sitios de integración están repartidos en el cromosoma, salvo por una región donde se concentran ocho tDNAs usados (*leu2A*, *ser3A*, *asn1A*, *asn1B*, *asn1C*, *asn1D*, *pro2A* y *arg2A*).

La disparidad de las frecuencias de integración reportadas previamente para los *asn*-tDNAs se replica para los otros *hotspots* de integración. Por ejemplo, en el caso de los tDNAs que transfieren fenilalanina, de las dos copias idénticas presentes en los cromosomas, sólo una es utilizada como sitio de integración. Similar es el caso de los tDNAs que portan el anticodón para metionina, de las siete copias idénticas que se encuentran regularmente en el cromosoma de *K. pneumoniae*, sólo dos sirven como sitio de integración (*met1C* y *met1E*). Puede notarse en los resultados también lo dispar que es el uso para los cuatro tDNAs que transfieren asparagina, donde *asn1D* es utilizado como sitio de integración preferentemente, por sobre los otros *asn*-tDNAs codificados en el cromosoma.

Ahora bien, es necesario tener en cuenta que las frecuencias aquí presentadas fueron calculadas para una muestra de 50 cepas. No es posible descartar que al analizar nuevas cepas de *K. pneumoniae*, otros tipos de tDNAs sean utilizados o que las frecuencias varíen al considerar una muestra más amplia de cepas para analizar. Sin embargo, cabe destacar que no encontramos en la literatura otro estudio que utilice múltiples cepas para evaluar el set completo de tDNAs del cromosoma de una especie bacteriana en cuanto a su uso como sitios de integración, por lo que los resultados que aquí se presentan son una primera aproximación para el análisis de las tendencias de integración de MGEs en *K. pneumoniae*.

A



B

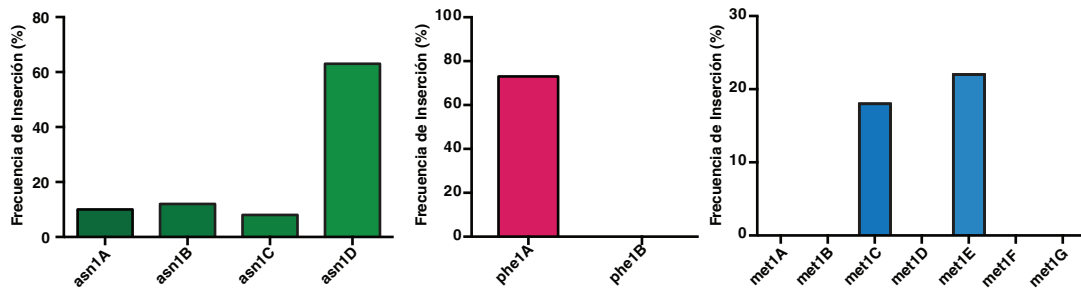


Figura 13: Las frecuencias de uso de los tDNAs representado sobre un cromosoma de referencia y comparación de las frecuencias de inserción en *loci* idénticos dentro del cromosoma.

(A) Las frecuencias de integración de islas genómicas en grafico de barras radial, dispuesto sobre la configuración de los tDNAs en el cromosoma de la cepa *K. pneumoniae* MGH78578. Luego, comparación de las frecuencias de integración de *loci* idénticos dentro del cromosoma: (B) *loci asn1*, (C) *loci phe1* y (D) *loci met1*.

4.8 El contexto genómico de los tDNAs y la disparidad en las frecuencias de uso como sitios de integración

Con el propósito de encontrar algún patrón que permita explicar la asimetría en la frecuencia de uso de distintos tDNAs como sitios de integración, se analizaron diferentes características de los tDNAs y su contexto, intentando establecer potenciales correlaciones. En primer lugar, se observó que aquellos tDNAs utilizados como sitios de integración parte del grupo más prevalente (Fig. 14A). Asimismo, los tDNAs utilizados como sitios de integración corresponden a tDNAs altamente conservados entre las cepas, aunque no se observó una correlación entre la frecuencia de inserción de GIs y la conservación de los tDNAs (Fig. 14B). Tampoco se observó una correlación entre la conservación del contexto genómico río arriba o río abajo del tDNA y la frecuencia de inserción de MGEs en dicho *locus* (Fig. 14C-D). Por lo tanto, solo puede evidenciarse que los tDNAs utilizados como sitios de integración son tDNAs prevalentes y altamente conservados entre las cepas, aunque un gran número de tDNAs no asociados a GIs también cumplen con ambas características.

Una posibilidad es que exista algún tipo de elemento o patrón de secuencia característico de los tDNAs frecuentemente usados (o de sus contextos) y que esté ausente en aquellos no usados. Para evaluar esta posibilidad se buscaron motivos de secuencia conservados que pudiesen dar cuenta de la preferencia de algunos sitios por sobre otros. Este análisis se realizó comparando las regiones río arriba, río abajo y los tDNAs propiamente tales. Sin embargo, la estrategia de búsqueda empleada, basada en las herramientas MEME y DREME, no entregó ningún resultado significativo (resultados no mostrados).

La integración de GIs en una determinada región del cromosoma podría verse afectada por la actividad de una serie de factores que interactúan con el DNA, entre ellas

la maquinaria transcripcional. En este sentido, una mayor actividad transcripcional en las cercanías de un determinado tDNA podría favorecer o evitar la integración de MGEs. Considerando lo anterior, ideamos una aproximación para evaluar si existe alguna correlación entre la frecuencia de inserción de GIs en un determinado tDNA y la actividad transcripcional de la región del cromosoma donde éste se ubica. Para ello, se utilizaron los datos de expresión génica de un estudio anterior, donde se determinó por medio de RNA-seq el transcriptoma de 17 cepas de *K. pneumoniae* (Bruchmann, et al., 2015). Para estimar y comparar la actividad transcripcional de las vecindades de cada tDNA, se calcularon los promedios de los niveles de expresión de los tres genes inmediatamente río arriba y río debajo de los tDNAs (Fig. 14E-F). El análisis no evidenció ningún grado de correlación entre la actividad transcripcional de la región río arriba o río abajo de los tDNAs y la frecuencia de inserción de GIs. Sin embargo, es preciso señalar que este análisis está limitado a las condiciones en las que fue realizado el experimento de transcriptómica, dado que son muestras de cultivos bacterianos en medio rico en nutrientes que alcanzaron la fase exponencial de crecimiento bacteriano. Por lo tanto, los perfiles de expresión de los genes en los contextos de los tDNAs podrían variar al considerar condiciones menos favorables para el desarrollo de las células, las que podrían fomentar la integración/escisión de MGEs.

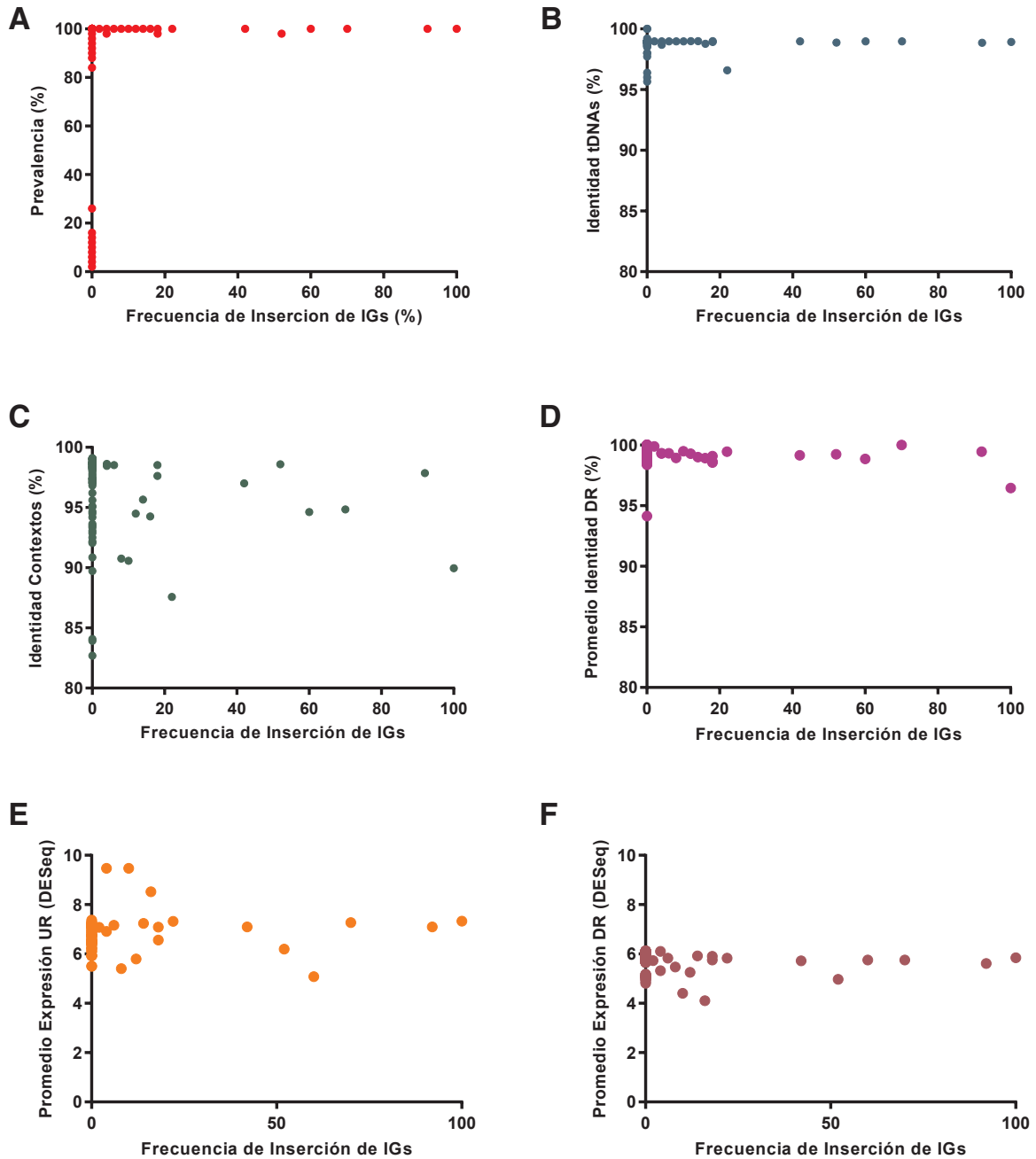


Figura 14: Análisis de características de los tDNAs y su contexto en función de la frecuencia de inserción de GIs y otros MGEs.

La correlación entre la frecuencia de inserción de GIs y (A) la prevalencia de los tDNAs, (B) la identidad de promedio de los tDNAs entre las cepas, (C) la identidad promedio de los contextos río arriba y (D) río abajo de los tDNAs. Además, se analizaron los perfiles de transcripción de los genes del contexto río arriba (E) y río abajo (F) de los tDNAs. Se calcularon los índices de correlación de Pearson, en ninguno de los casos la correlación entre las variables es estadísticamente significativa (dato no mostrado).

4.9 Las proteínas asociadas al nucleoide bacteriano y el uso de tDNAs como sitios de integración de MGEs.

Siguiendo con la idea de que la actividad de proteínas que se unen al DNA y modifican su topología podría afectar la integración de MGEs dirigiéndola hacia ciertos tDNAs por sobre otros, ideamos una aproximación para evaluar si los tDNAs preferidos se ubican en una región del cromosoma con una frecuencia anómala de sitios de unión de proteínas asociadas a la remodelación del nucleoide bacteriano. Para ello, las señales de unión de las proteínas asociadas al nucleoide HNS y Fis, determinadas a través de experimentos de ChIP-Seq en distintas etapas del crecimiento bacteriano, fueron mapeadas en el cromosoma de la cepa *K. pneumoniae* 1084. Conviene destacar que en el proceso de identificación de los bloques conservados entre la cepa de *E. coli* considerada y la referencia de *K. pneumoniae*, la suma en extensión de estos bloques representa el 58.52% del cromosoma de la cepa de *K. pneumoniae*, lo que da cuenta de la relación filogenética entre ambas cepas. Además, la identidad promedio de los bloques es de $60,04 \pm 15,84\%$, lo que apoya el punto anterior. Por ejemplo, en el caso del tDNA *met1C* (Fig. 15) el bloque homólogo entre ambas cepas abarca el tDNA mismo y su región río abajo, pero no la región río arriba.

Luego, se comparó la frecuencia de integración de MGEs y la densidad de unión de estas proteínas en las vecindades de los tDNAs usados como sitio de integración (Fig. 16). Los datos de transcriptómica considerados proviene de cultivos en distintas etapas del crecimiento bacteriano, de esta forma se consigue una correlación apenas significativa de la frecuencia de integración de MGEs con la señal de unión de la proteína HNS en fase estacionaria del crecimiento bacteriano (Fig. 16A). En las otras etapas del crecimiento consideradas para la proteína HNS, aunque existe alguna tendencia, la correlación entre la señal de unión promedio de HNS en los contextos inmediatos de los

tDNAs y la frecuencia de inserción de MGEs no resultó ser estadísticamente significativa. Tampoco hubo una correlación entre las señales de unión de la proteína Fis y la frecuencia de uso (Fig. 16B). Además, fue evaluado el contenido AT de los contextos y comparado con la frecuencia de uso de GIs, comparación que tampoco resultó en algún tipo de correlación significativa.

En resumen, de los puntos anteriores, este trabajo no consiguió identificar elementos en los tDNAs o en sus contextos que permitan dar cuenta de la preferencia de algún tDNA por sobre sus equivalentes dentro del cromosoma. Por lo tanto, se desconocen los mecanismos genéticos y/o moleculares que dirigen la integración de GIs y otros MGEs hacia ciertos tDNAs, así como también el significado biológico de dicho fenómeno.

4.10 Clasificación y generalidades de las GIs identificadas

Para ahondar en las características de las GIs identificadas en las cepas de *K. pneumoniae* analizadas, sus secuencias se extrajeron desde los genomas, se clasificaron, caracterizaron y utilizaron para construir una base de datos.

En primer lugar, las GIs fueron comparadas mediante alineamientos múltiples, permitiendo su clasificación por identidad de secuencia. De esta manera, los 302 MGEs identificados fueron clasificados en 97 clases (Anexo II).

El sitio de integración con mayor variedad de islas fue *thr2A* con 16 clases de GIs, seguido del tmRNA y *leu5*, ambos con 14 clases de GIs identificadas. El dato anterior es interesante, dado que el *locus leu5A* es aquel ocupado por GIs en el 100% de las cepas, y la gran diversidad de islas en este *locus* evidencia su alta dinámica, y que distintas GIs han utilizado este *locus* como sitio de integración en eventos de integración independientes. En el otro extremo están los *loci leu2A* y *ser2A*, en ambos

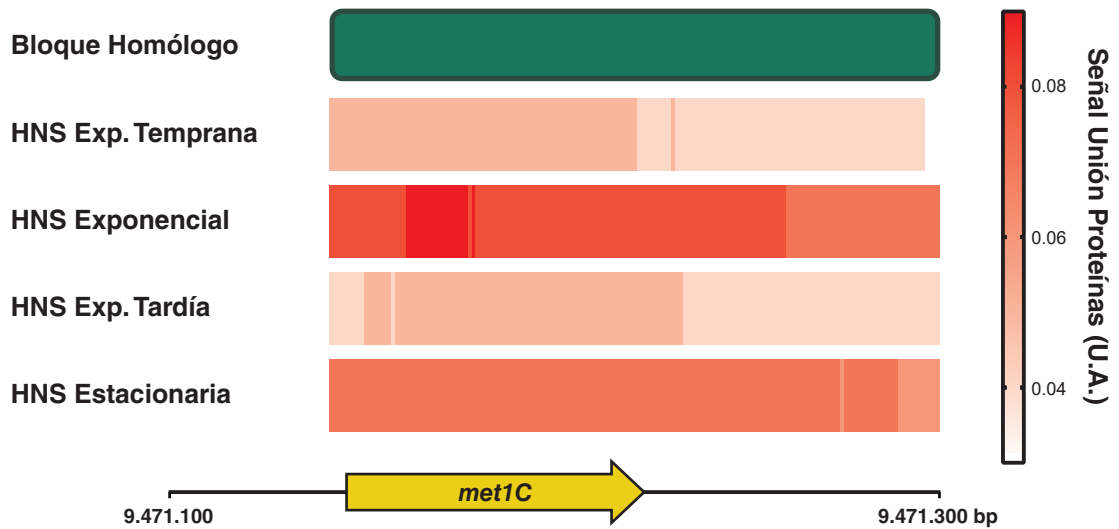


Figura 15: Estrategia para la identificación de posibles regiones de unión de proteínas del nucleóide bacteriano a partir de experimentos realizados en la cepa de *E. coli* K12 MG655. Ejemplo del mapeo realizado para el tDNA *met1C*. A partir del bloque homólogo compartido entre el cromosoma de *E. coli* y el cromosoma de *K. pneumoniae* 1084. La intensidad de la señal de unión de las proteínas asociadas al nucleóide (HNS en el ejemplo) fue extrapolada sobre el cromosoma de *K. pneumoniae* de referencia. Luego, la señal promedio sobre el tDNA se calculó y considero para los análisis siguientes. En este caso puede notarse que la señal de unión de proteínas del nucleóide sobre el locus de *met1C* varía según la fase del crecimiento bacteriano considerado.

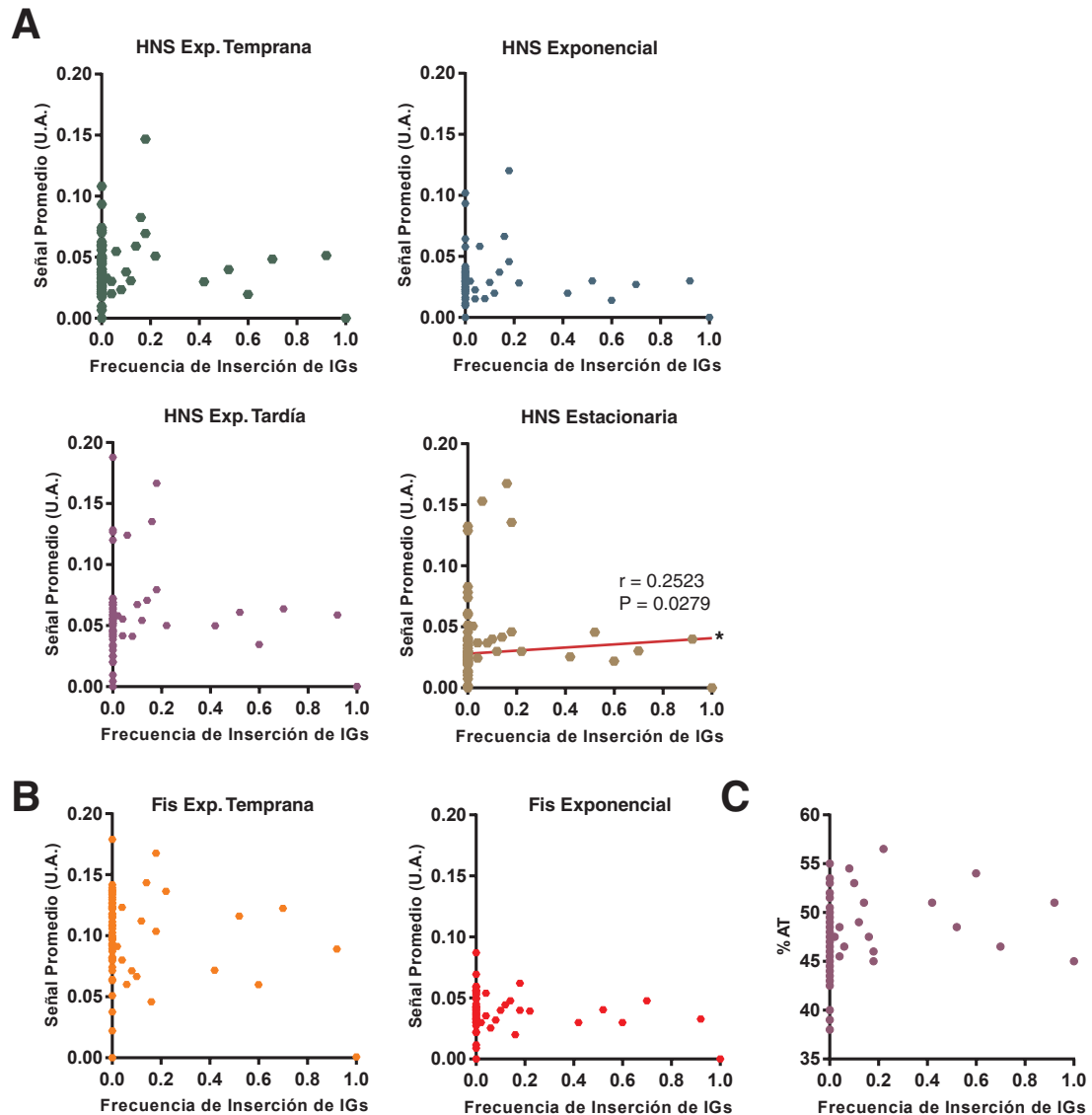


Figura 16: Relación entre la ocupación de proteínas asociadas al nucleóide bacteriano en *K. pneumoniae* mapeado a partir de experimentos de CHIP-Seq en *E. coli*, y la frecuencia de uso de los tDNAs como sitios de integración.

Las frecuencias de inserción de MGEs en los distintos tDNAs analizados se compararon con con las señales de unión promedio de proteínas asociadas al nucleóide. (A) La unión de la proteína HNS se mapeó según datos de cuatro etapas del crecimiento bacteriano y (B) la proteína Fis se evaluó solo en dos etapas del crecimiento. Además, se comparó la composición AT (%AT) de los tDNAs y su entorno inmediato con las frecuencias de uso. Se llevaron a cabo análisis de correlación no paramétrica de Spearman, identificándose sólo una correlación estadísticamente significativa (*: $P < 0.05$).

solo se identificó una clase de GI; no obstante, en el caso de *ser2A* esto es de esperar porque sólo fue identificada una única IG.

A través del análisis manual de los cromosomas, se identificaron los repetidos directos presentes en el conjunto de GIs asociadas a tDNAs. Es importante destacar que en algunas islas no se encontraron repetidos directos, lo que podría deberse a delección de esta región de forma posterior a su integración. Al respecto, nuestro grupo de investigación ya había descrito en *K. pneumoniae* la existencia de GIs asociadas a asn-tDNAs que carecen de repetidos directos (Marcoleta et al., 2016). En muchos de esos casos se encontraron secuencias de inserción en las cercanías del límite 3' de la isla, sugiriendo que se había producido la delección de parte del elemento, probablemente por recombinación homóloga entre dos secuencias de inserción idénticas. Respecto de los repetidos directos identificados, la mayoría de éstos corresponden a la región 3' de los tDNAs y tmDNAs analizados y su tamaño promedio bordea las 20 pb, aunque en algunos casos menos frecuentes pueden superar las 100 pb.

De manera previa a este estudio, en la literatura se habían descrito un total de 24 GIs de *K. pneumoniae* en la base de datos Islander, la cual corresponde a una de las más completas bases de datos de islas genómicas bacterianas (Hudson, et al., 2015). Con nuestro método de identificación fue posible aumentar a 97 el número de GIs conocidas de esta especie. Con el propósito de estimar la fracción que representa este conjunto de islas conocidas respecto del total de islas que conformarían el pangenoma de *K. pneumoniae*, se construyeron curvas de acumulación o rarefacción (Fig. 17A). Lo mismo se realizó para la acumulación de los distintos tipos de contextos de tDNAs identificados (Fig. 17B). Del ajuste de las curvas y del valor del factor de crecimiento de las curvas, es posible predecir que, con el análisis de más cepas, tanto la cantidad

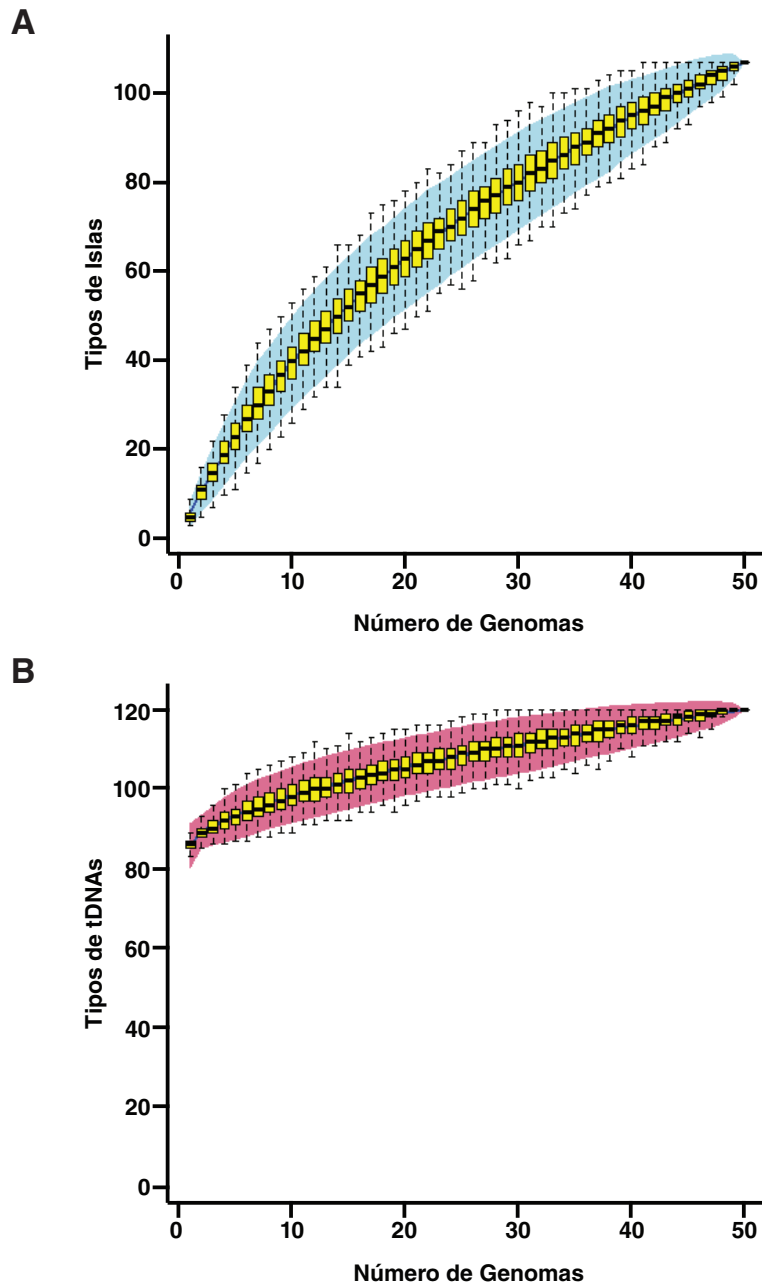


Figura 4.12: Curvas de proyección de la identificación de nuevos tipos de tDNAs y de nuevos tipos de GIs en las cepas analizadas de *K. pneumoniae*. Utilizando el paquete de análisis *vegan* de R, fue iterada 1.000 veces la acumulación de (A) nuevos tipos de islas genómicas, esto incluye otros EGMs, y (B) de nuevos tipos de tDNAs. La estadística de cada punto está representada por el gráfico de caja (*box plot*) y la zona sombreada en azul o rojo muestra las líneas obtenidas en cada iteración.

de tipos de islas como la cantidad de tipos de contextos de tDNAs aumentará, no pudiendo predecirse un valor a converger con estos datos. Sin embargo, comparando los parámetros del ajuste, es posible predecir que el número de islas aumentará más rápidamente que la cantidad de contextos de tDNAs, dado que el factor de crecimiento para los tDNAs ($\gamma = 0,625 \pm 0,004$) es más cercano a 1 que en el caso de los tipos de islas ($\gamma = 0,094 \pm 0,002$). Por lo tanto, ambos son conjuntos abiertos, pero el conjunto de las islas tiende a crecer más rápidamente que el de los contextos de tDNAs.

4.11 Las funciones codificadas en los MGEs identificados

Para dar cuenta de las funciones codificadas en los MGEs identificados, estos se analizaron para resolver su identidad. Lo anterior, respecto a si son GIs propiamente tal u otros MGEs, como profagos o transposones. En este sentido, sólo el *locus met1E* es utilizado como sitio de integración de GIs exclusivamente, en los demás *loci* los MGEs son variados (Fig. 18A).

Las secuencias aminoacídicas de las proteínas codificadas en los MGEs identificados se extrajeron y clasificaron según su eventual función, utilizando la base de datos eggNOG.

Como resultado, de las 2.598 proteínas presentes en todos los MGEs, solo 1.112 poseen un ortólogo en la base de datos (aproximadamente un 43%). Es decir, más de la mitad de las proteínas codificadas en MGEs asociados a tDNAs de *K. pneumoniae* poseen una función desconocida. Las GIs identificadas se clasificaron por las funciones más relevantes codificadas en ellas. Así fue posible identificar islas que codifican funciones asociadas al transporte y captación de hierro, formación de fimbrias y de pilus, transporte de azúcares, y resistencia a drogas, entre otras (Fig. 18B). Dada la alta proporción de proteínas con función desconocida, hubo un grupo importante de MGEs a los que no fue posible asignar una función putativa.

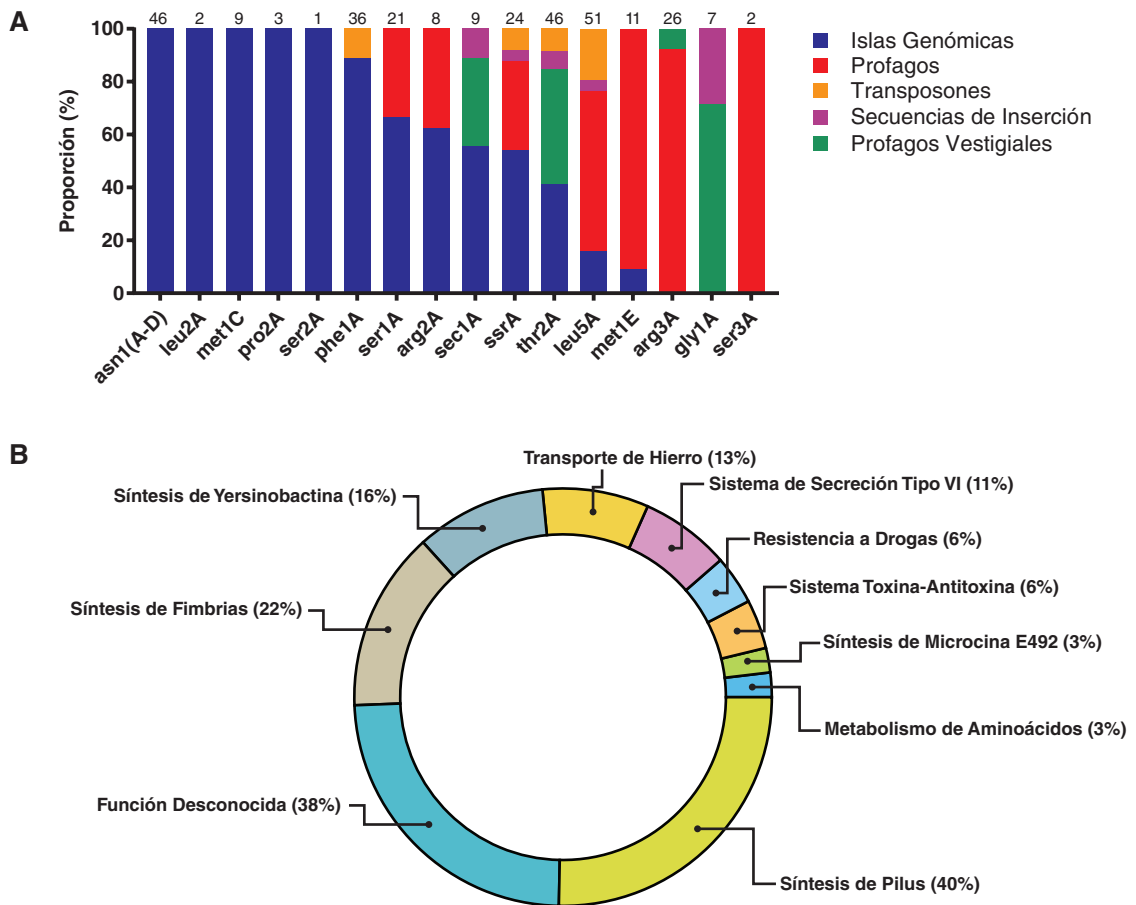


Figura 4.13: Los MGEs que se insertan en los *loci* de tDNAs y las funciones codificadas en las islas genómicas en *K. pneumoniae*.

(A) Para cada *locus* de tDNA utilizado como sitio de integración de GIs se determinó la cantidad total de MGEs que se encontraron en estos *loci* (número sobre la barra), los que fueron clasificados según la anotación funcional de los genes codificados en las islas. (B) Las islas genómicas identificadas (158 en estricto rigor) se analizaron para determinar las funciones biológicas predichas según la anotación funcional. El gráfico muestra la proporción de cada función sobre el total de islas. En el caso de las islas que poseen más de una función codificada, se consideró la función con mayor relevancia para la patogénesis de las cepas.

4.12 Un nuevo concepto operacional para comparar genomas de *K. pneumoniae* y potencialmente otras especies relacionadas: el isotipo

En base a la clasificación de las GIs descrita anteriormente, hemos desarrollado un nuevo concepto operacional útil en el análisis genómico de *K. pneumoniae*: el *islotipo* de las cepas analizadas. El isotipo corresponde a la combinación de GIs presentes en una cepa, considerando el sitio de integración y la clase de isla que se integró en dicho *locus*. Conviene señalar que, aunque se identificaron distintos tipos de MGEs, en el caso de los isotipos todos los MGEs se consideraron como GIs. El isotipo de todas las cepas analizadas puede ser encontrado en Anexo III.

Los isotipos de las cepas se utilizaron para construir relaciones filogenéticas que representan cómo se ha modificado la presencia de distintas islas en la evolución de estas cepas. De alguna forma, lo realizado fue clasificar las cepas sólo desde el punto de vista de las GIs (Fig. 19). Esta reconstrucción fue comparada con la filogenia de las cepas calculadas a partir de los genes que forman parte del genoma núcleo. Al comparar las agrupaciones obtenidas, puede comprobarse que algunas cepas permanecen relacionadas, lo que sugiere que en esos casos las islas compartidas fueron adquiridas por expansión clonal (herencia vertical). En contraste, varias de las agrupaciones obtenidas a través del *core genoma* se reorganizan al comparar los isotipos. La existencia de cepas filogenéticamente cercanas según el análisis del genoma núcleo, pero que poseen un isotipo diferente, constituye una evidencia de que en dichos casos la adquisición de las islas no compartidas fue por transferencia genética horizontal.

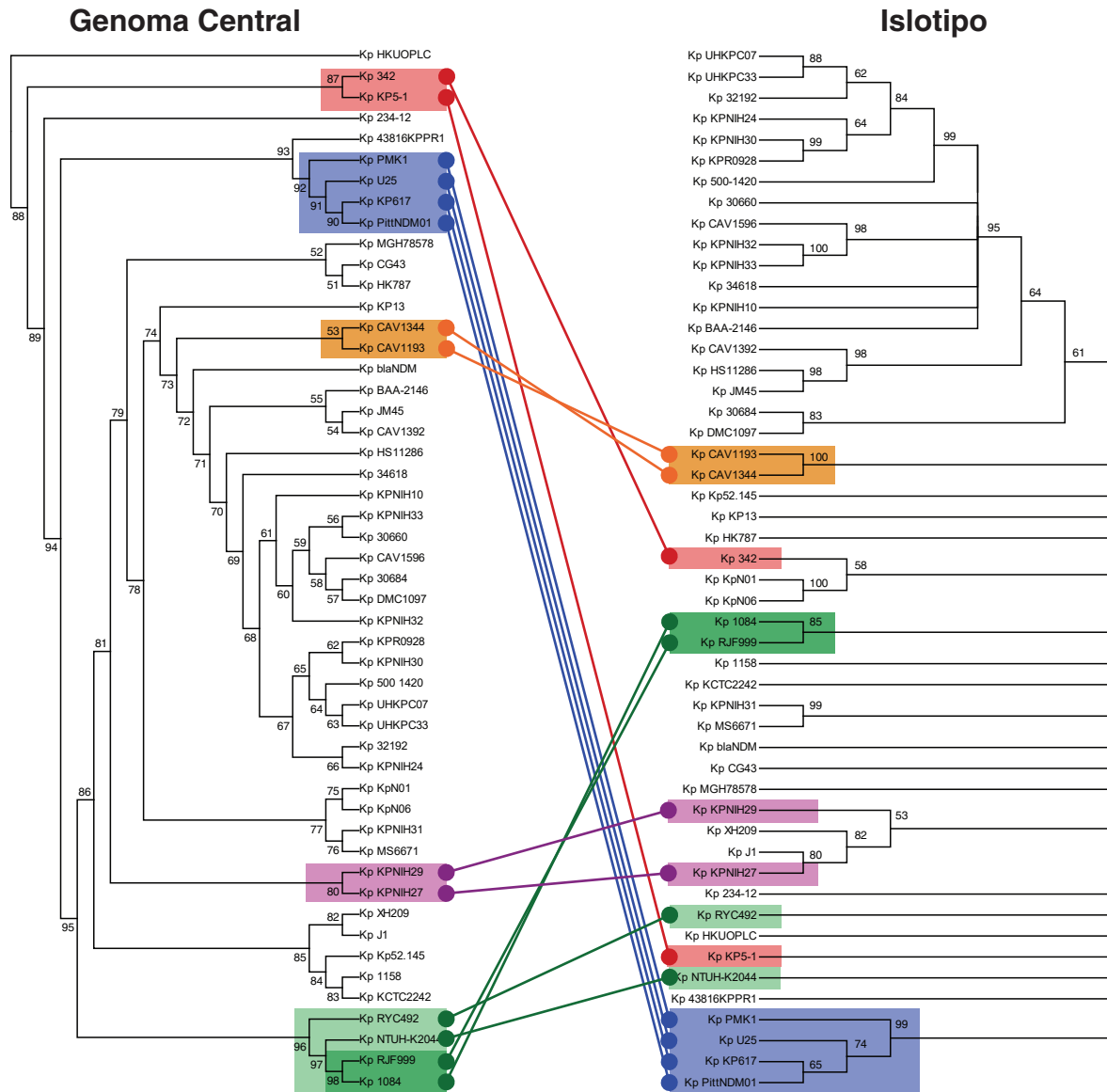


Figura 19: Comparación de la clasificación de las cepas a partir de los genes del core genome y a partir de los isotipos.

Comparación de los dendrogramas obtenidos de la clasificación de las cepas según los genes presentes en el genoma núcleo de las cepas y según la carga de islas (islotipo). En la figura debe notarse que, aunque la agrupación de algunos clados se conserva en algunos casos, existen nuevos agrupamientos cuando se considera el islotipo. El número junto a los nodos indica el porcentaje obtenido a partir de bootstrap realizado sobre la construcción de 1.000 árboles, solo son indicados valores superiores al 50%.

4.13 Las integrasas codificadas en las islas y su rol sobre la preferencia por algunos sitios.

Las secuencias de las proteínas de tipo integrasas codificadas en las distintas GIs se extrajeron desde los genomas y se compararon mediante alineamientos múltiples. En términos globales, los genes de las integrasas en las islas del mismo tipo poseen un alto porcentaje de identidad (superior al 80%). Sin embargo, al comparar las integrasas de islas de distinto tipo, el grado de identidad disminuye. Además, al analizar el dendrograma generado a partir de los alineamientos, se observó que en general las integrasas se agrupan de acuerdo con el sitio de integración, independiente de si las islas son o no del mismo tipo (Fig. 20). Esto es altamente relevante debido a que este agrupamiento podría dar cuenta de la selección de un grupo de tDNAs como sitios de integración de GIs. Las secuencias de aminoácidos en los distintos dominios de las integrasas determinan las interacciones proteína-DNA que median el mecanismo bioquímico que consigue la integración de intermediarios circulares en los cromosomas, y como distintos tDNAs ofrecen distintos sitios de integración, las integrasas pudieron haber evolucionado para unirse preferentemente en estos sitios, lo que se traduce en que la secuencia aminoacídica de las integrasas que usan el mismo sitio de integración se altamente similar. Sin embargo, conviene hacer notar que, aunque la naturaleza de la interacción entre las integrasas y el cromosoma podría explicar la selección de ciertos tDNAs como sitios de integración, lo anterior no resuelve la asimetría que existe entre el uso de tDNAs idénticos dentro de un mismo cromosoma.



Figura 20: Relación filogenética entre las integrasas codificadas en las distintas clases de islas asociadas a tDNAs en *K. pneumoniae*.

En total, 80 integrasas provenientes de las distintas clases de GIs identificadas en los cromosomas de *K. pneumoniae* analizados fueron comparadas a través de alineamiento múltiple utilizando ClustalW. A partir de las distancias calculadas se construyó el dendrograma presentado. Las regiones coloreadas indican coincidencia entre el agrupamiento y el tDNA asociado a las islas, con una única excepción en el grupo del tmRNA (*ssrA*). EL número indicado en los nodos indica el porcentaje resultado del bootstrap de un total de 1.000 árboles.

4.14 Análisis pangénomico de las cepas y el aporte de las GIs al genoma accesorio

Con el objetivo de delimitar el genoma central, genoma accesorio y pangenoma de la especie *K. pneumoniae* en base las cepas analizadas, sus genomas se analizaron utilizando el programa Roary (Page, A., et al., 2015). Para ello, los genomas deducidos de cada cepa se compararon generando grupos de proteínas que comparten una identidad de secuencia tal que son consideradas como variantes de una sola proteína. Luego, se cuantificó el conjunto de genes no redundante (eliminando aquellos que corresponden a variantes), y se determinó su frecuencia de aparición entre las distintas cepas. Aquellas presentes en el 99% de las cepas se consideraron como parte del genoma núcleo y el resto como parte del genoma accesorio (Fig. 21). De este análisis, se identificaron 12.973 genes distintos entre los genomas de las cepas analizadas en este trabajo, esta es entonces la extensión del pangenoma de estas cepas. Dentro del pangenoma, 3.638 proteínas pertenecen al genoma núcleo. La colección de genes pertenecientes al genoma núcleo en este análisis es superior al reportado anteriormente, de aproximadamente 2.000 genes (Martin & Bachman, 2018). La discrepancia en estos valores está relacionada a dos factores, el primero es el método utilizado y también la cantidad de cepas analizadas. Habitualmente, al considerar un mayor número de cepas, la cantidad de genes compartidos por todas ellas tiende a disminuir. Por otra parte, 9.335 proteínas fueron clasificadas dentro del genoma accesorio. Los MGEs asociados a tDNAs identificados en este estudio codifican un total de 2.598 proteínas distintas. En consecuencia, nuestros resultados indican que aproximadamente el 28% de los genes en el genoma accesorio de *K. pneumoniae* corresponde a genes codificados en este tipo de elementos.

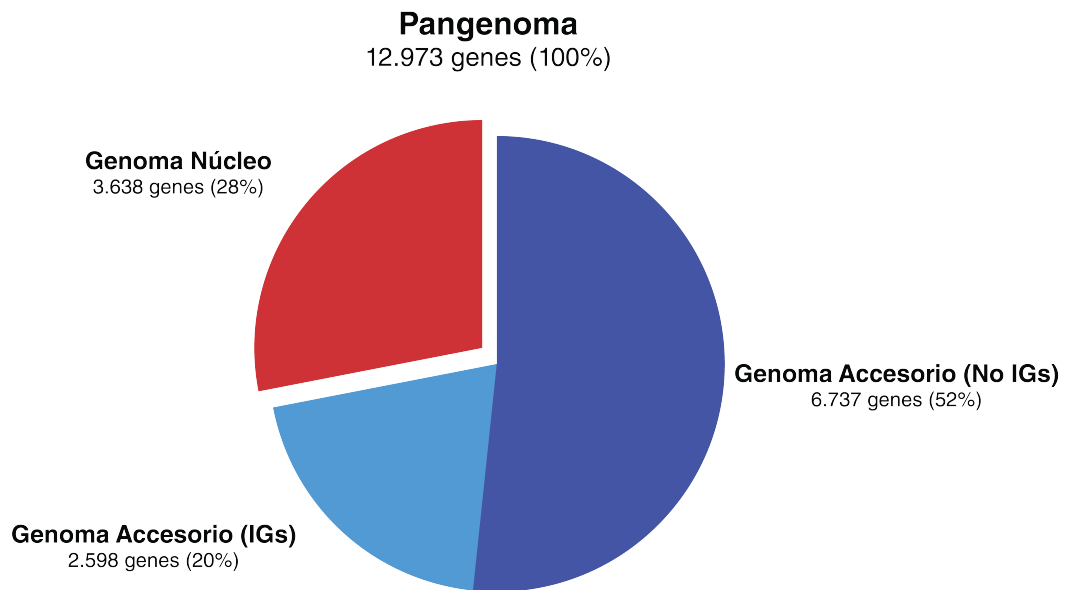
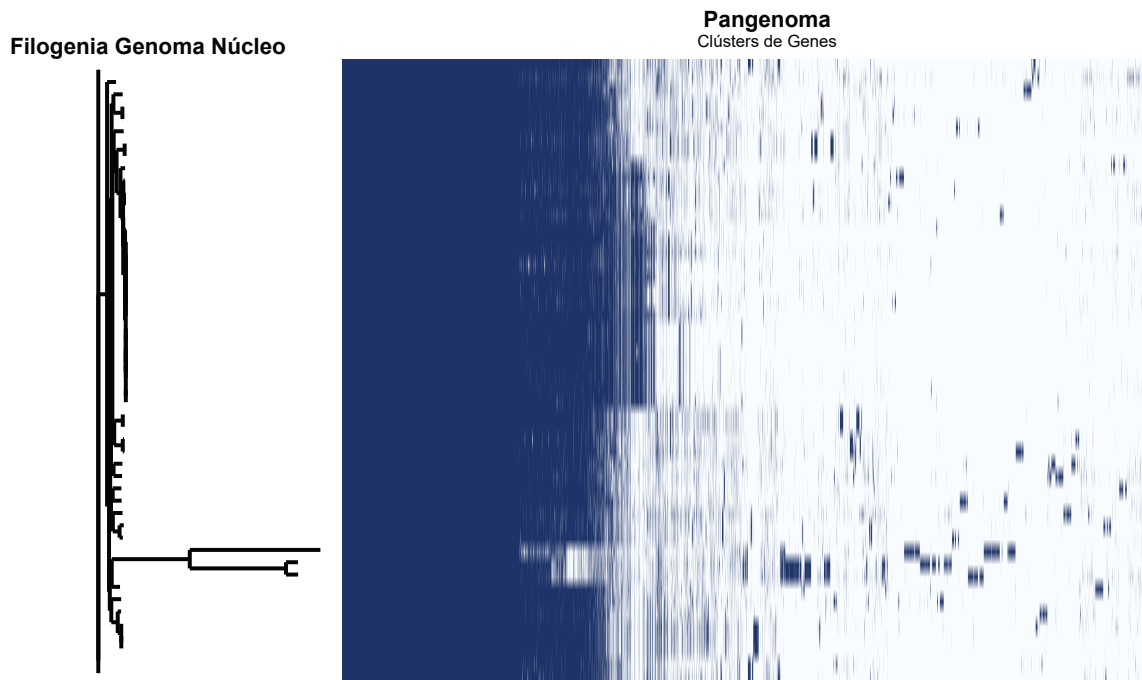


Figura 21: Análisis del pangenoma de las cepas de *K. pneumoniae* analizadas. Los genomas se analizaron con Roary, obteniéndose los resultados de la parte superior. En la matriz, las bandas horizontales representan cada cepa analizada, y el eje vertical cada cluster de proteínas identificados. En la parte inferior, un resumen de lo obtenido a través del análisis.

5. DISCUSIÓN Y PROYECCIONES

Los tDNAs y tmDNAs, en extensión, representan cerca del 2% del cromosoma de *K. pneumoniae*. A pesar de su representación minoritaria, existe consenso en que la integración de muchos tipos de MGEs descritos en distintas enterobacterias se concentra en estos *loci*. En base a los resultados obtenidos en el presente estudio y a la evidencia recopilada en la literatura, es posible distinguir algunas características de estos sitios que permiten explicar, aunque aun parcialmente, la preferencia en la selección de estos sitios. En primer lugar, el análisis global del set núcleo de tDNAs y tmDNAs de *K. pneumoniae* indica que estos genes son extremadamente conservados entre cepas de esta especie. Más aún, es muy probable que este grado de conservación se mantenga dentro del género *Klebsiella* e incluso otros géneros filogenéticamente relacionados. Por ejemplo, los tDNAs *asn1* identificados en *K. pneumoniae* poseen sobre un 90% de identidad de secuencia con sus contrapartes de especies de los géneros *Escherichia*, *Shigella*, *Salmonella*, *Serratia* y *Enterobacter*. Además, existe cierto grado de conservación de algunos elementos que flanquean estos tDNAs (Fig. 22). La alta conservación de tDNAs y tmDNAs entre linajes que comparten un mismo nicho (como en el caso de las enterobacterias), favorecería la herencia horizontal de los MGEs, en el sentido en que dichos elementos encontrarían sitios de integración equivalentes en cromosomas de distintos hospederos, posibilitando su incorporación al genoma y con ello su posterior diseminación vertical y horizontal. La alta conservación de estos genes estaría determinada por su rol clave como parte de la maquinaria traduccional, el cual es primordial para la sobrevivencia de las bacterias. Sin perjuicio de lo anterior, nuestros análisis indican que a pesar de que la gran mayoría de los tDNAs/tmDNAs son altamente conservados, sólo una pequeña fracción de ellos serían

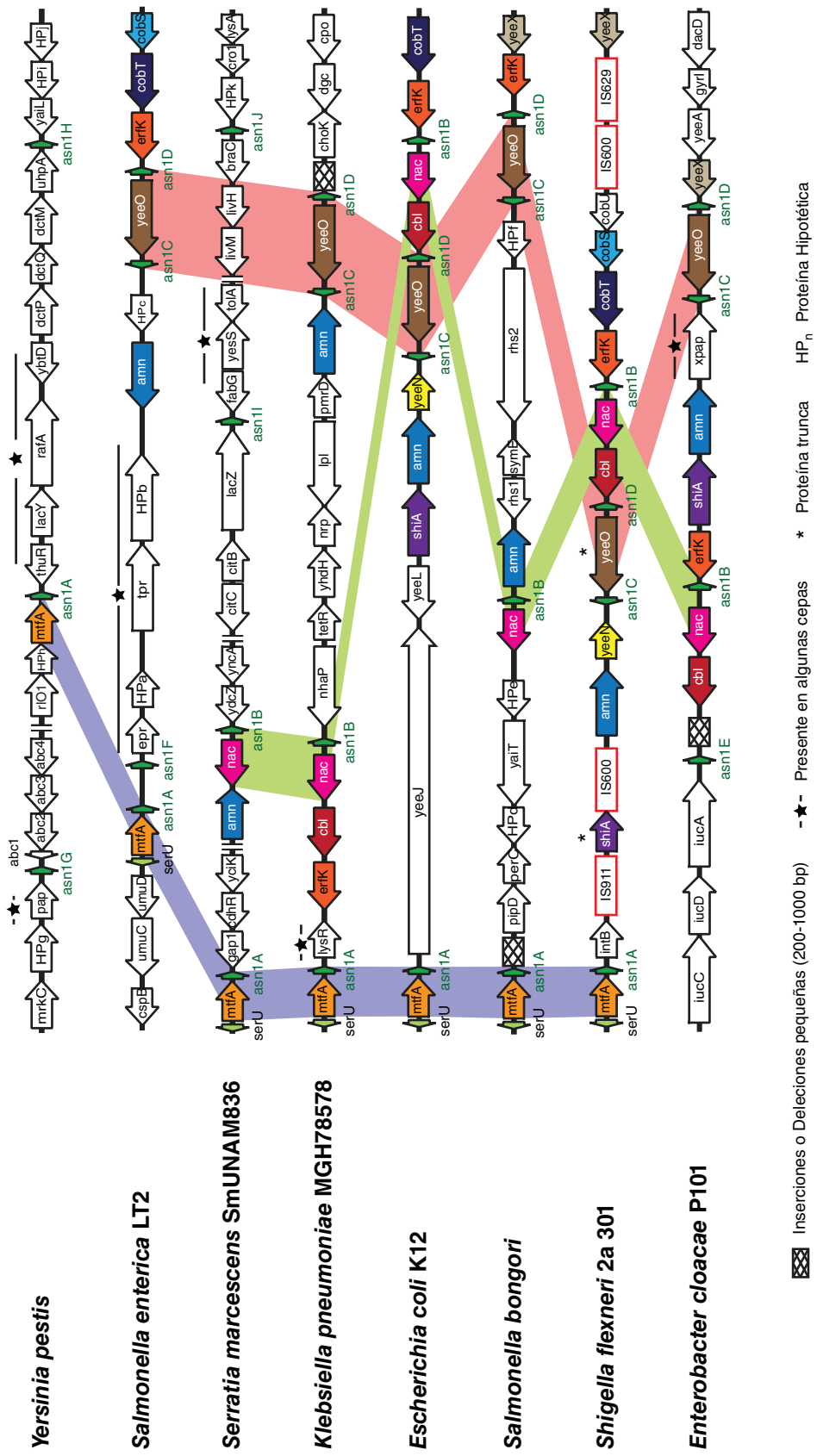


Figura 22: Los contextos genómicos de los tDNAs *asn1* poseen elementos comunes entre ocho especies de enterobacterias. Representación de fragmentos de los cromosomas de ocho especies de enterobacterias en los que se indican la posición de los tDNAs para asparagina (*asn1A-J*) en distintos contextos. Notar que, en los distintos cromosomas, un grupo de tDNAs se encuentra flanqueado río arriba por uno o más genes que se conservan (coloreados), mientras que otros elementos no están conservados (sin color).

usados como sitios de integración de MGEs. A nuestro saber, este es el primer estudio en el que se analiza de manera global el set completo de este tipo de genes y su rol como sitios de integración de MGEs en bacterias. Más aún, es el primer reporte de la marcada disparidad que existe en la frecuencia de integración de MGEs en distintos tDNAs, incluso entre copias idénticas ubicadas en distintas regiones del cromosoma.

Los tDNAs y tmDNAs utilizados como sitios de integración de MGEs son altamente prevalentes dentro de las cepas analizadas. Esto sugiere que la maquinaria de integración y escisión de los MGEs evolucionó de forma de reconocer y utilizar los sitios que frecuentemente se encuentran entre las distintas cepas o especies. Aunque la alta prevalencia y conservación de los tDNAs y tmDNAs favorecería su uso como sitios de integración, dichas propiedades son compartidas también por los genes que codifican RNAs ribosomales y diversas proteínas del metabolismo central. De hecho, cualquier característica inherente al tDNA o tmDNA propiamente tal no permitiría explicar por qué algunos tDNAs son usados como sitios de integración y otros no. Por lo tanto, dichos tDNAs y tmDNAs deben poseer alguna característica adicional. Al respecto, conviene señalar que los tRNAs son sujetos a la modificación post-transcripcional de sus extremos 5' y 3' (Shepherd & Ibba, 2015). Además, los tRNAs que se encuentran en clústeres pueden ser transcritos como operones, y estos transcritos son luego procesados mediante clivaje para producir los tRNAs individuales. Lo anterior es relevante porque, en primer lugar, los tDNAs utilizados como sitios de integración son siempre tDNAs solitarios, o, dicho de otra manera, no se encuentran flanqueados por otros tDNAs. En segundo lugar, al analizar los repetidos directos de las GIs, así como de otros MGEs, estos repetidos están localizados generalmente en la región 3' de los tDNAs, por lo que la integración de los MGEs altera de alguna manera la región 3' del tDNA, pero no la región 5'. Por lo tanto, la evidencia sugiere que los tDNAs seleccionados como sitios de

integración son tDNAs en los que la integración, aunque afectara la región 3' de la unidad transcripcional, puede subsanarse con los mecanismos propios de la bacteria y que sean tDNAs solitarios evita que se afecten irremediablemente las unidades transcripcionales de tDNAs que podrían ser transcritos de forma policistónica, afectando el balance de tRNAs en el citoplasma.

Según los resultados expuestos, las integrasas codificadas en las GIs de *K. pneumoniae* poseen secuencias aminoacídicas que se agrupan de acuerdo con el sitio de integración de dichos elementos. Dicho de otro modo, GIs que poseen integrasas altamente similares se encuentran integradas en los mismos tipos de tDNAs, a pesar de que el resto de los elementos genéticos que conforman la isla sean completamente diferentes. La presencia de tDNAs en profagos podría ser clave para entender la evolución de las integrasas. Algunos autores proponen que los tDNAs presentes en el material genético de los bacteriófagos favorecen la traducción de los genes requeridos para la propagación de esta clase de virus. Sin embargo, la evidencia al respecto es más bien dispar al apoyar o rechazar esta hipótesis.

Si bien distintos indicios apuntan a que el contexto genómico en el que se encuentra cada tDNA sería determinante en sus propiedades como sitio de integración, las estrategias empleadas en este trabajo no permitieron dar cuenta de algún factor que explique la desigual integración de GIs en tDNAs idénticos. No obstante, no es posible descartar el efecto de la actividad transcripcional y de la ocupación de los contextos genómicos por parte de proteínas asociadas al nucleóide sobre las frecuencias de inserción de MGEs. Al respecto, se ha descrito previamente el rol de las proteínas asociadas al nucleóide IHF y H-NS, en el mecanismo de integración del DNA de algunos bacteriófagos (Landy, 2015). Estas proteínas provocan el doblamiento de la molécula de DNA del cromosoma y del bacteriófago en regiones que flanquean los sitios de

integración, lo que permite que tetrámeros formados por las integrasas acerquen el cromosoma y el intermediario circular del profago, permitiendo luego la integración del MGE. Es de esperar que este fenómeno sea transversal a MGEs cuya integración éste mediada por estos tipos de integrasas. Sin embargo, es complejo intentar comprobar que los sitios de integración de MGEs están siempre flanqueados por sitios de unión de proteínas asociadas al nucleóide. Lo anterior dado a que la unión de estas proteínas es dinámica y no depende sólo de la secuencia del tDNA, sino también del estado metabólico en el que se encuentre la bacteria, lo cual influye sobre la concentración de estas proteínas en el citoplasma celular y con ello, su frecuencia de unión al DNA (Dillon & Dorman, 2010). Sin embargo, otros trabajos indican que la frecuencia de integración estaría determinada esencialmente por la secuencia del sitio de recombinación (Wilde, et al., 2008). Por esto, el tema de resolver los determinantes de la preferencia de algunos sitios para la integración de MGEs es un campo de investigación abierto.

Gracias a la identificación masiva de islas genómicas de *K. pneumoniae* y a su clasificación y caracterización, como parte del presente estudio hemos definido el concepto de isotipo. Éste corresponde a la colección de MGEs asociados a tDNAs/tmDNAs que presenta el cromosoma de una determinada cepa. El estudio de la identidad de todos los elementos que conforman dicha colección permite hacer inferencias respecto al potencial fenotípico de dicha cepa, incluyendo rasgos clínicos relevantes como son la producción de factores de virulencia o de resistencia a antibióticos. Además, el análisis de cómo varía el isotipo entre un grupo de cepas y su contraste con las relaciones filogenéticas determinadas en base al genoma núcleo permite hacer inferencias respecto a fenómenos asociados a la transmisión de GIs. Específicamente, permite inferir si las GIs fueron adquiridas verticalmente desde sus antecesores (es decir, mediante expansión clonal), o si éstas fueron obtenidas por algún

mecanismo de transferencia genética horizontal. Nuestros resultados indican que, en el universo de cepas muestreado, es posible evidenciar la ocurrencia de ambas clases de adquisición de genes. Por ello, si bien estos dos tipos de transferencia a veces se entienden como antagónicas, probablemente operan de manera paralela contribuyendo a la variación de los cromosomas en cuanto a su composición y estructura

Finalmente, el mecanismo manual de identificación de GIs presentado en este trabajo está siendo traducido en un programa que permita la automatización de este proceso. El algoritmo detrás de este programa sigue la misma secuencia lógica utilizada para el análisis de forma manual. En primer lugar, realiza una identificación primaria de todos los tDNAs y tmDNAs presentes en el cromosoma problema, mediante el programa ARAGORN. Luego, la secuencia del contexto río arriba de cada tDNA/tmDNA es analizada en búsqueda de claves que permitan su identificación, comparándola con una base de datos que alberga las secuencias de los *loci* vírgenes definidos en este estudio para cada tipo de tDNA de *K. pneumoniae*. En caso de que un contexto no se asimile a alguno presente en la base de datos, el programa identificará dicho contexto como nuevo y permitirá al usuario solicitar la incorporación del nuevo contexto a la base de datos de *loci* vírgenes. Una vez que el contexto río arriba es identificado, el tDNA es nombrado de acuerdo con el sistema de nomenclatura propuesto en el presente estudio. Luego, el contexto río abajo “esperado” de acuerdo con la estructura del *locus* virgen es buscado a través de alineamientos, recorriendo el cromosoma. En el caso de la presencia de una IG en el *locus*, la distancia entre el tDNA y el alineamiento con mejor puntaje tendrá que ser superior a 10 kpb para que la GI sea identificada. Posteriormente, las islas identificadas podrán ser comparadas con una base de datos de GIs previamente identificadas. Se espera que en el futuro próximo esta herramienta de análisis de genomas, así como también la base de datos de GIs e información relevante asociada

a dichas islas, forme parte de la *Klebsiella pneumoniae* tDNA-associated Genomic Islands Database (KleptGI-DB), una plataforma online proyectada a partir de los resultados de este trabajo, que se espera poner a disposición de la comunidad científica a través de internet.

6. CONCLUSIONES

1.- El presente estudio corresponde al primer análisis global de los tDNAs y tmDNAs presentes en el cromosoma de una especie bacteriana y de su uso como sitios de integración de MGEs. No existe en la literatura un análisis de MGEs asociados a tDNAs en *K. pneumoniae* de esta envergadura.

2.- El análisis detallado del genoma de 50 cepas de *K. pneumoniae* permitió desarrollar una nueva nomenclatura para sus tDNAs, la que fue fundamental para el estudio de sus propiedades. Esta nueva nomenclatura permitió clasificar los 4281 *loci* de tDNAs identificados en 119 tipos distintos, y con ello, individualizar y comparar de manera sistemática copias idénticas de un mismo tDNA.

3.- El análisis sistemático de los tDNAs de *K. pneumoniae* individualizados gracias al nuevo sistema de nomenclatura propuesto permitió la identificación y clasificación del contexto virgen para cada tDNA, el cual es el elemento fundamental de un nuevo método de identificación de MGEs asociados a tDNAs y tmDNAs. La utilidad de este método quedó de manifiesto tras la identificación de los 302 MGEs asociados a dichos *loci*, presentes en las 50 cepas analizadas.

4.- El análisis y clasificación de los MGEs asociados a tDNAs y tmDNAs permitió la identificación de 97 clases distintas de elementos, incluyendo profagos, profagos vestigiales, e Islas genómicas propiamente tal. La alta variabilidad observada en las clases de MGEs asociados a tDNAs es reflejo de la alta flexibilidad del genoma accesorio de *K. pneumoniae*.

5.- La preferencia en el uso de tDNAs o tmDNAs como sitios de integración de MGEs por sobre otros tipos de genes podría estar explicado por dos factores: 1) la capacidad de la maquinaria transcripcional de corregir la extensión de los tRNAs, dado que la integración de los MGEs podría afectar la secuencia de terminación de las unidades transcripcionales de los tRNAs y 2) la evolución de las integrasas para reconocer estos sitios. Aunque es materia de investigación, estos factores podrían explicar la tendencia de preferir estos sitios por otros genes, ya sean codificantes de proteínas o no.

6.- Copias idénticas de un mismo tipo de tDNA ubicadas en distintos contextos genómicos son usadas como sitios de integración de MGEs con frecuencias muy dispares. Lo anterior indica que la selección del sitio de integración no está determinada exclusivamente por la secuencia de dicho sitio, y que de alguna manera el contexto podría estar favoreciendo o desfavoreciendo la integración de MGEs.

7.- Como parte de este estudio, no fue posible identificar ningún factor específico que pudiese explicar el uso preferencial de ciertos tDNAs por sobre otros, como sitios de integración. Ni las características de las secuencias de los tDNAs y tmDNAs, ni la expresión de los genes en los contextos o la posible ubicación de proteínas asociadas a la formación del nucleoide pudieron ser correlacionadas con el uso dispar de sitios equivalentes.

8.- Los MGEs asociados a tDNAs representan una parte importante del genoma accesorio de *K. pneumoniae*. El análisis de una mayor cantidad de cepas permitirá la identificación de nuevos MGEs que forman parte del pangenoma de esta especie.

7. REFERENCIAS

- Alseth, I., Dalhus, B., & Bjørås, M. (2014). Inosine in DNA and RNA. *Current Opinion in Genetics & Development*, 26, 116–123.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410.
- Bailey, T. L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics*, 27(12), 1653–1659.
- Bailey, T. L., Bailey, & L., T. (2002). Discovering Novel Sequence Motifs with MEME. In *Current Protocols in Bioinformatics* (p. 2.4.1-2.4.35). Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, 43(W1), W39–W49.
- Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., ... Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science (New York, N.Y.)*, 277(5331), 1453–62.
- Bradford, P. A. (2001). Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. *Clinical Microbiology Reviews*, 14(4), 933–51, table of contents.
- Brisse, S., Grimont, F., & Grimont, P. A. D. (2006). The Genus *Klebsiella*. In *The Prokaryotes. Volume 6: Proteobacteria: Gamma Subclass*.
- Bruchmann, S., Muthukumarasamy, U., Pohl, S., Preusse, M., Bielecka, A., Nicolai, T., ... Häussler, S. (2015). Deep transcriptome profiling of clinical *Klebsiella pneumoniae* isolates reveals strain and sequence type-specific adaptation. *Environmental Microbiology*, 17(11), 4690–4710.
- Burrus, V., & Waldor, M. K. (2004). Shaping bacterial genomes with integrative and conjugative elements. *Research in Microbiology*, 155(5), 376–386.
- Chang, F., & Chou, M. (1995). Comparison of pyogenic liver abscesses caused by *Klebsiella pneumoniae* and non-K. pneumoniae pathogens. *Journal of the Formosan Medical Association = Taiwan Yi Zhi*, 94(5), 232–7.
- Chaves, J., Ladona, M. G., Segura, C., Coira, A., Reig, R., & Ampurdanés, C. (2001). SHV-1 beta-lactamase is mainly a chromosomally encoded species-specific enzyme in *Klebsiella pneumoniae*. *Antimicrobial Agents and Chemotherapy*, 45(10), 2856–61.
- Cheng, DL., Liu, YC., Yen, MY., Liu, CY., & Wang, RS. (1991). Septic Metastatic Lesions of Pyogenic Liver Abscess. *Archives of Internal Medicine*, 151(8), 1557.
- Clegg, S., & Murphy, C. N. (2016). Epidemiology and Virulence of *Klebsiella pneumoniae*. *Microbiology Spectrum*, 4(1).
- Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., ... de Hoon, M. J. L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422–1423.
- Corsaro, M. M., De Castro, C., Naldi, T., Parrilli, M., Tomás, J. M., & Regué, M. (2005). 1H and 13C NMR characterization and secondary structure of the K2 polysaccharide of *Klebsiella pneumoniae* strain 52145. *Carbohydrate Research*, 340(13), 2212–2217.
- Cosentino, S., Larsen, M. V., Aarestrup, F. M., & Lund, O. (2013). PathogenFinder: distinguishing friend from foe using bacterial whole genome sequence data. *PloS one*, 8(10), e77302.
- Crick, F. H. C. (1966). Codon-anticodon pairing: The wobble hypothesis. *Journal of Molecular Biology*, 19(2), 548–555.
- Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, 14(7), 1394–403.
- Dillon, S. C., & Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology*, 8(3), 185–195.
- Dobrindt, U., & Hacker, J. (2001). Whole genome plasticity in pathogenic bacteria. *Current Opinion in Microbiology*, 4(5), 550–557.

- Erlj, Daniel, Michalland H., S., Neira Q., O., Wolff C., V., Jara B., V., & Meneses C., M. (2013). Diagnóstico diferencial de nódulos pulmonares cavitados: a propósito de un caso clínico. *Revista Chilena de Enfermedades Respiratorias*, 29(1), 39–42.
- Evans, B. A., & Amyes, S. G. B. (2014). OXA β -lactamases. *Clinical Microbiology Reviews*, 27(2), 241–63.
- Fouts, D. E., Tyler, H. L., DeBoy, R. T., Daugherty, S., Ren, Q., Badger, J. H., ... Methé, B. A. (2008). Complete genome sequence of the N₂-Fixing broad host range endophyte *Klebsiella pneumoniae* 342 and virulence predictions verified in mice. *PLoS Genetics*, 4(7), e1000141.
- Frost, L. S., Leplae, R., Summers, A. O., & Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology*, 3(9), 722–732.
- Germon, P., Roche, D., Melo, S., Mignon-Grasteau, S., Dobrindt, U., Hacker, J., ... Moulin-Schouleur, M. (2007). tDNA *locus* polymorphism and ecto-chromosomal DNA insertion hot-spots are related to the phylogenetic group of *Escherichia coli* strains. *Microbiology*, 153(3), 826–837.
- Gil, R., Fernández, P., & Sabbagh, E. (2005). Diagnóstico clínico-radiológico de la neumonía del adulto adquirida en la comunidad. *Revista Chilena de Infectología*, 22, s26-231.
- Gorgoni, B., Marshall, E., McFarland, M. R., Romano, M. C., & Stansfield, I. (2014). Controlling translation elongation efficiency: tRNA regulation of ribosome flux on the mRNA. *Biochemical Society Transactions*, 42(1), 160–165.
- Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., ... & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research*, 36(10), 3420-3435.
- Holt, K. E., Wertheim, H., Zadoks, R. N., Baker, S., Whitehouse, C. A., Dance, D., ... Thomson, N. R. (2015). Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proceedings of the National Academy of Sciences of the United States of America*, 112(27), E3574-81.
- Hou, Y.-M., Gamper, H., & Yang, W. (2015). Post-transcriptional modifications to tRNA: a response to the genetic code degeneracy. *RNA (New York, N.Y.)*, 21(4), 642–4.
- Hsieh, P., Lin, T., Lee, C., Tsai, S., & Wang, J. (2008). Serum-induced iron-acquisition systems and TonB contribute to virulence in *Klebsiella pneumoniae* causing primary pyogenic liver abscess. *The Journal of Infectious Diseases*, 197(12), 1717–1727.
- Hudson, C. M., Lau, B. Y., & Williams, K. P. (2015). Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Research*, 43(Database issue), D48-53.
- Jacoby, G. A. (2009). AmpC beta-lactamases. *Clinical Microbiology Reviews*, 22(1), 161–82, Table of Contents.
- Jagnow, J., & Clegg, S. (2003). *Klebsiella pneumoniae* MrkD-mediated biofilm formation on extracellular matrix- and collagen-coated surfaces. *Microbiology*, 149(9), 2397–2405.
- Janssen, B. D., & Hayes, C. S. (2012). The tmRNA ribosome-rescue system. *Advances in Protein Chemistry and Structural Biology*, 86, 151–91.
- Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., & Bork, P. (2008). eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Research*, 36(Database issue), D250-4.
- Johnson, D. I. (2018). *Klebsiella* spp. In *Bacterial Pathogens and Their Virulence Factors* (pp. 269–277). Cham: Springer International Publishing.
- Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., & Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews*, 33(2), 376–393.
- Kahramanoglou, C., Seshasayee, A. S. N., Prieto, A. I., Ibberson, D., Schmidt, S., Zimmermann, J., ... Luscombe, N. M. (2011). Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Research*, 39(6), 2073–2091.
- Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., & Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research*, 29(22), 4633–42.

- Landy, A. (2015). The λ Integrase Site-specific Recombination Pathway. *Microbiology Spectrum*, 3(2), MDNA3-0051-2014.
- Langstraat, J., Bohse, M., & Clegg, S. (2001). Type 3 fimbrial shaft (MrkA) of *Klebsiella pneumoniae*, but not the fimbrial adhesin (MrkD), facilitates biofilm formation. *Infection and Immunity*, 69(9), 5805–12.
- Laslett, D., & Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, 32(1), 11–16.
- Lederman, E. R., & Crum, N. F. (2005). Pyogenic liver abscess with a focus on *Klebsiella pneumoniae* as a primary pathogen: An emerging disease with unique clinical characteristics. *The American Journal of Gastroenterology*, 100(2), 322–331.
- Liu, H.-L., & Zhu, J. (2010). Analysis of the 3' ends of tRNA as the cause of insertion sites of foreign DNA in *Prochlorococcus*. *Journal of Zhejiang University. Science. B*, 11(9), 708–18.
- Mahillon, J., & Chandler, M. (1998). Insertion sequences. *Microbiology and Molecular Biology Reviews: MMBR*, 62(3), 725–74.
- Manickam, N., Joshi, K., Bhatt, M. J., & Farabaugh, P. J. (2016). Effects of tRNA modification on translational accuracy depend on intrinsic codon-anticodon strength. *Nucleic Acids Research*, 44(4), 1871–81.
- Marcoleta, A. E., Berríos-Pastén, C., Nuñez, G., Monasterio, O., & Lagos, R. (2016). *Klebsiella pneumoniae* asparagine tDNAs are integration hotspots for different genomic islands encoding Microcin E492 production determinants and other putative virulence factors present in hypervirulent strains. *Frontiers in Microbiology*, 7, 849.
- Martin, R. M., & Bachman, M. A. (2018). Colonization, infection, and the accessory genome of *Klebsiella pneumoniae*. *Frontiers in Cellular and Infection Microbiology*, 8, 4.
- McInerney, J. O. (1998). GCUA: general codon usage analysis. *Bioinformatics (Oxford, England)*, 14(4), 372–3.
- Mohanty, B. K., & Kushner, S. R. (2007). Ribonuclease P processes polycistronic tRNA transcripts in *Escherichia coli* independent of ribonuclease E. *Nucleic Acids Research*, 35(22), 7614–25.
- Montgomerie, J. Z. (1979). Epidemiology of *Klebsiella* and Hospital-Associated Infections. *Clinical Infectious Diseases*, 1(5), 736–753.
- Nassif, X., & Sansonetti, P. J. (1986). Correlation of the virulence of *Klebsiella pneumoniae* K1 and K2 with the presence of a plasmid encoding aerobactin. *Infection and Immunity*, 54(3), 603–8.
- NCBI Resource Coordinators. (2017). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 45(D1), D12–D17.
- Ong, C.-L. Y., Ulett, G. C., Mabbett, A. N., Beatson, S. A., Webb, R. I., Monaghan, W., ... Schembri, M. A. (2008). Identification of type 3 fimbriae in uropathogenic *Escherichia coli* reveals a role in biofilm formation. *Journal of Bacteriology*, 190(3), 1054–63.
- Ørskov, I., & Ørskov, F. (1984). Serotyping of *Klebsiella*. *Methods in Microbiology*, 14, 143–164.
- Ou, H.-Y., Chen, L.-L., Lonnen, J., Chaudhuri, R. R., Thani, A. Bin, Smith, R., ... Rajakumar, K. (2006). A novel strategy for the identification of genomic islands by comparative analysis of the contents and contexts of tRNA sites in closely related bacteria. *Nucleic Acids Research*, 34(1), e3.
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., ... & Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31(22), 3691–3693.
- Pan, P.-C., Chen, H.-W., Wu, P.-K., Wu, Y.-Y., Lin, C.-H., & Wu, J. H. (2011). Mutation in fucose synthesis gene of *Klebsiella pneumoniae* affects capsule composition and virulence in mice. *Experimental Biology and Medicine*, 236(2), 219–226.
- Patel, G., Huprikar, S., Factor, S. H., Jenkins, S. G., & Calfee, D. P. (2008). Outcomes of carbapenem-resistant *Klebsiella pneumoniae* infection and the impact of antimicrobial and adjunctive therapies. *Infection Control & Hospital Epidemiology*, 29(12), 1099–1106.

- Pendleton, J. N., Gorman, S. P., & Gilmore, B. F. (2013). Clinical relevance of the ESKAPE pathogens. *Expert Review of Anti-Infective Therapy*, 11(3), 297–308.
- Podschun, R., & Ullmann, U. (1998). *Klebsiella* spp. as nosocomial pathogens: Epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clinical Microbiology Reviews*, 11(4), 589–603.
- Pollack, M., Nieman, R., Reinhardt, J., Charache, P., Jett, M., & Hardy, P. (1972). Factors influencing colonisation and antibiotic-resistance patterns of gram-negative bacteria in hospital patients. *The Lancet*, 300(7779), 668–671.
- Pundir, S., Martin, M. J., & O'Donovan, C. (2017). UniProt Protein Knowledgebase (pp. 41–55). Humana Press, New York, NY.
- Putze, J., Hennequin, C., Nougayrède, J.-P., Zhang, W., Homburg, S., Karch, H., ... Dobrindt, U. (2009). Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. *Infection and Immunity*, 77(11), 4696–703.
- Rossi, B., Gasperini, M. L., Leflon-Guibout, V., Gioanni, A., de Lastours, V., Rossi, G., ... Lefort, A. (2018). Hypervirulent *Klebsiella pneumoniae* in cryptogenic liver abscesses, Paris, France. *Emerging Infectious Diseases*, 24(2), 221–229.
- Sahly, H., Keisari, Y., & Ofek, I. (2009). Manno(rhamno)biose-containing capsular polysaccharides of *Klebsiella pneumoniae* enhance opsono-stimulation of human polymorphonuclear leukocytes. *Journal of Innate Immunity*, 1(2), 136–44.
- Santajit, S., & Indrawattana, N. (2016, May 5). Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. *BioMed Research International*. Hindawi.
- Schroll, C., Barken, K. B., Krogfelt, K. A., & Struve, C. (2010). Role of type 1 and type 3 fimbriae in *Klebsiella pneumoniae* biofilm formation. *BMC Microbiology*, 10(1), 179.
- Selden, R., Lee, S., Wang, W., Bennet, J., & Eickhoff, T. (1971). Nosocomial *Klebsiella* infections: intestinal colonization as a reservoir. *Annals of Internal Medicine*, 74(5), 657–664.
- Shepherd, J., & Ibbá, M. (2015). Bacterial transfer RNAs. *FEMS Microbiology Reviews*, 39(3), 280–300.
- Shon, A. S., Bajwa, R. P. S., & Russo, T. A. (2013). Hypervirulent (hypermucoviscous) *Klebsiella pneumoniae*. *Virulence*, 4(2), 107–118.
- Soucy, S. M., Huang, J., & Gogarten, J. P. (2015). Horizontal gene transfer: building the web of life. *Nature Reviews Genetics*, 16(8), 472–482.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313.
- Struve, C., Roe, C. C., Stegger, M., Stahlhut, S. G., Hansen, D. S., Engelthaler, D. M., ... Krogfelt, K. A. (2015). Mapping the evolution of hypervirulent *Klebsiella pneumoniae*. *MBio*, 6(4), e00630–15.
- Tarkkanen, A. M., Allen, B. L., Williams, P. H., Kauppi, M., Haahtela, K., Siitonen, A., ... Korhonen, T. K. (1992). Fimbriation, capsulation, and iron-scavenging systems of *Klebsiella* strains associated with human urinary tract infection. *Infection and Immunity*, 60(3), 1187–92.
- Tettelin, H., Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, 11(5), 472–477.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673–80.
- Tolmasky, M. E., Chamorro, R. M., Crosa, J. H., & Marini, P. M. (1988). Transposon-mediated amikacin resistance in *Klebsiella pneumoniae*. *Antimicrobial Agents and Chemotherapy*, 32(9), 1416–20.
- Tolmasky, M. E., & Crosa, J. H. (1987). Tn1331, a novel multiresistance transposon encoding resistance to amikacin and ampicillin in *Klebsiella pneumoniae*. *Antimicrobial Agents and Chemotherapy*, 31(12), 1955–60.
- Touchon, M., Hoede, C., Tenailon, O., Barbe, V., Baeriswyl, S., Bidet, P., ... Denamur, E. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genetics*, 5(1), e1000344.

- Trevisan, V. (1885). Caratteri di alcuni nuovi generi di Batteriacee. *Atti Della Accademia Fisio-Medico-Statistica in Milano. Series, 4(3)*, 92–107.
- van Aartsen, J. J. (2008). The *Klebsiella* pheV tRNA locus: a hotspot for integration of alien genomic islands. *Bioscience Horizons, 1(1)*, 51–60.
- Vold, B. S. (1978). Post-transcriptional modifications of the anticodon loop region: alterations in isoaccepting species of tRNA's during development in *Bacillus subtilis*. *Journal of Bacteriology, 135(1)*, 124–32.
- Von Frisch, A. (1882). Zur Aetiologie des Rhinoskleroms. *Wiener Medizinische Wochenschrift, 32*, 969–972.
- Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W., ... Merkl, R. (2006). Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics, 7(1)*, 142.
- Wilde, C., Mazel, D., Hochhut, B., Middendorf, B., Le Roux, F., Carniel, E., ... Hacker, J. (2008). Delineation of the recombination sites necessary for integration of pathogenicity islands II and III into the *Escherichia coli* 536 chromosome. *Molecular Microbiology, 68(1)*, 139–151.
- Williams, P., Lambert, P. A., Haigh, C. G., & Brown, M. R. W. (1986). The influence of the O and K antigens of *Klebsiella aerogenes* on surface hydrophobicity and susceptibility to phagocytosis and antimicrobial agents. *Journal of Medical Microbiology, 21(2)*, 125–132.
- World Health Organization. (2017). *20th WHO Model List of Essential Medicines*. http://www.who.int/medicines/publications/essentialmedicines/EML2015_8-May-15.pdf.
- Wu, KM., Li, LH., Yan, JJ., Tsao, N., Liao, TL., Tsai, HC., ... Tsai, SF. (2009). Genome sequencing and comparative analysis of *Klebsiella pneumoniae* NTUH-K2044, a strain causing liver abscess and meningitis. *Journal of Bacteriology, 191(14)*, 4492–501.
- Yeh, TS., Jan, YY., Jeng, LB., Hwang, TL., Chao, TC., Chien, RN., & Chen, MF. (1998). Pyogenic liver abscesses in patients with malignant disease. *Archives of Surgery, 133(3)*, 242–245.
- Zhang, D., Chen, D., Cao, L., Li, G., & Cheng, H. (2016). The effect of codon mismatch on the protein translation system. *PloS One, 11(2)*, e0148302.

8. ANEXOS

Anexo I: Cepas de *K. pneumoniae* analizadas en este trabajo.

Cepa	NCBI Acceso	Año	País	Fuente Aislado	Mucosidad	Pathogen Finder ¹
1084	NC_018522.1	2002	Taiwan	Cepa Clínica, Hospedero Humano, Absceso Hepático	Hipermucosa	0,874
1158	NZ_CP006722.1	2002	Taiwan	Cepa Clínica, Hospedero Humano	NR ²	0,892
234-12	NZ_CP011313.1	2011	Germany	Cepa Clínica, Hospedero Humano, Sangre	NR	0,913
30660	NZ_CP006923.1	2010	USA	Cepa Clínica, Hospedero Humano, Tracto Urinario	Mucoso	0,806
30684	NZ_CP006918.1	2010	USA	Cepa Clínica, Hospedero Humano, Tracto Urinario	No Mucoso	0,906
32192	CP010361.1	2010	USA	Cepa Clínica, Hospedero Humano, Espruto	NR	0,904
342	CP000964.1	2008	USA	Cepa Endofítica Maíz	NR	0,711
34618	CP010392.1	2011	USA	Aislado Clínico, Hospedero Humano, Fluido Corporal	NR	0,903
43816KPPR1	NZ_CP009208.1	2000	USA	Cepa de Laboratorio, Resistente a Rifampicina	NR	0,901
500_1420	CP011980.1	2014	USA	Aislado Clínico, Hospedero Humano	NR	0,903
BAA-2146	NZ_CP006659.2	2010	USA	Cepa Clínica, Hospedero Humano, Tracto Urinario	NR	0,898
BlaNDM-1	CP009114.1	2012	USA	Cepa Clínica, Hospedero Humano, Tracto Urinario	NR	0,886
CAV1193	CP013322.1	2010	USA	Cepa Clínica, Hospedero Humano, Tracto Respiratorio	NR	0,894
CAV1344	CP011624.1	2010	USA	Cepa Clínica, Hospedero Humano, Tracto Urinario	NR	0,903
CAV1392	CP011578.1	2011	USA	Cepa Clínica, Hospedero Humano, Espruto	NR	0,895
CAV1596	CP011647.1	2012	USA	Cepa Clínica, Hospedero Humano, Región Perirectal	NR	0,906
CG43	CP006648.1	2012	Taiwán	Cepa Clínica, Hospedero Humano, Absceso Hepático	NR	0,902
DMC1097	CP011976.1	2010	USA	Cepa Clínica, Hospedero Humano, Sangre	NR	0,902
HK787	NZ_CP006738.1	2013	Taiwán	Cepa Clínica, serotipo K2 (sugiere Absceso Hepático)	NR	0,901
HKUOPLC	CP012300.1	2013	Hong Kong	Cepa Veterinaria, Hospedero panda, Heces	NR	-
HS11286	CP003200.1	2011	China	Cepa Clínica, Hospedero Humano, Espruto	NR	0,772
J1	CP013711.1	2014	China	Aislado ambiental, Aguas Servidas	NR	0,890
JM45	CP006656.1	2010	China	Cepa Clínica, Hospedero Humano, Sangre	NR	0,897
KCTC2242	CP002910.1	2010	República de Corea	Aislado ambiental, Productora de 2,3-butanodiol	NR	-

KP13	NZ_CP003999.1	2009	Brasil	Cepa Clínica, Hospedero Humano, Sangre	NR	0,901
KP5-1	CP008700.1	2005	USA	Aislado ambiental, Hospedero <i>Nezara viridula</i>	NR	0,422
Kp52.145	NZ_FO834906.1	1935	Indonesia	Cepa Clínica, Hospedero Humano, Tracto Respiratorio	Mucoso	0,897
KP617	CP012753.1	2013	Corea del Sur	Cepa Clínica, Hospedero Humano	NR	0,893
KpN01	CP012987.1	2013	Canadá	Cepa Clínica, Hospedero Humano, Tracto Urinario	NR	0,893
KpN06	CP012992.1	2013	Canadá	Cepa Clínica, Hospedero Humano, Sangre	NR	0,893
KPNIH10	NZ_CP007727.1	2011	USA	Cepa Clínica, Hospedero Humano, Región Inguinal	NR	0,905
KPNIH24	NZ_CP008797.1	2012	USA	Cepa Clínica, Hospedero Humano, Nasofaríngeo	NR	0,902
KPNIH27	NZ_CP007731.1	2012	USA	Cepa Clínica, Hospedero Humano, Región Inguinal	NR	0,910
KPNIH29	NZ_CP009863.1	2013	USA	Cepa Clínica, Hospedero Humano, Región Perirectal	NR	0,900
KPNIH30	NZ_CP009872.1	2013	USA	Cepa Clínica, Hospedero Humano, Región Perirectal	NR	0,903
KPNIH31	NZ_CP009876.1	2013	USA	Cepa Clínica, Hospedero Humano, Tracto Urinario	NR	0,792
KPNIH32	NZ_CP009775.1	2013	USA	Cepa Clínica, Hospedero Humano, Región Perirectal	NR	0,894
KPNIH33	NZ_CP009771.1	2013	USA	Cepa Clínica, Hospedero Humano, Región Perirectal	NR	0,894
KPR0928	NZ_CP008831.1	2012	USA	Cepa Clínica, Hospedero Humano, Espruto	NR	0,902
MGH78578	CP000647.1	1994	USA	Cepa Clínica, Hospedero Humano, Espruto	NR	0,874
MS6671	NZ_LN824133.1	2014	Emiratos Árabes	Cepa Clínica, Hospedero Humano, Tracto Urinario	NR	0,904
NTUH-K2044	AP006725.1	2008	Taiwán	Cepa Clínica, Hospedero Humano, Meningitis	NR	0,872
PittNDM01	CP006798.1	2013	USA	Cepa Clínica, Hospedero Humano, Tracto Urinario	NR	0,892
PMK1	NZ_CP008929.1	2011	Nepal	Cepa Clínica, Hospedero Humano, Tracto Urinario	NR	0,887
RJF999	CP014010.1	2015	China	Cepa Clínica, Hospedero Humano, Tracto Urinario	NR	0,821
RYC492	APGM01000001.1	1984	España	Cepa Clínica, Hospedero Humano, Heces	NR	0,887
U25	CP012043.1	2010	India	Cepa Clínica, Hospedero Humano, Tracto Urinario	NR	0,893
UHKPC07	CP011985.1	2015	USA	Cepa Clínica, Hospedero Humano	NR	0,902
UHKPC33	CP011989.1	2014	USA	Cepa Clínica, Hospedero Humano	NR	0,902
XH209	CP009461.1	2014	China	Cepa Clínica, Hospedero Humano, Sangre	NR	0,904

¹Probabilidad de que la cepa sea un patógeno humano. Calculado como la probabilidad que, según las proteínas codificadas en el genoma, la cepa pertenezca a un grupo de cepas caracterizadas como patógenos (Cosentino, S., et al., 2013).

²NR: Mucosidad no reportada en la literatura.

Anexo II: Resumen de las frecuencias de los anticodones (f_{ant}) de los tDNAs y de sus codones complementarios (f_{cod}).

Aminoácido	Anticodón	N ^{o2}	f_{ant} (%) ¹	f_{cod} (%) ¹
ala	GGC	1	42.19	41.23
	TGC	2	57.81	8.02
	AGC	3	0³	9.98
	CGC	4	0	40.76
arg	ACG	1	54.12	22.14
	CCT	2	13.74	2.61
	TCT	3	18.41	2.62
	CCG	4	13.74	16.69
	TCG	5	0	4.5
	GCG	6	0	51.45
asn	GTT	1	100	63.57
	ATT	2	0	36.43
cys	GCA	1	100	72.5
	ACA	2	0	27.5
gln	TTG	1	49.75	15.38
	CTG	2	50.25	84.62
glu	TTC	1	100	51.07
	CTC	2	0	48.93
gly	CCC	1	16.67	18.93
	GCC	2	65.67	55.55
	TCC	3	17.87	7.92
	ACC	4	0	17.6
his	GTG	1	100	49.44
	ATG	2	0	50.56
ile	GAT	1	100	57.63
	TAT	2	0	5.61
	AAT	3	0	36.76
pro	CGG	1	33.33	60.84
	GGG	2	33.33	15.32
	TGG	3	33.33	14.17
	AGG	4	0	9.67

Aminoácido	Anticodón	N ^{o2}	f_{ant} (%) ¹	f_{cod} (%) ¹
asp	GTC	1	100	47.14
	ATC	2	0	52.86
leu	GAG	1	12.59	13.92
	TAA	2	12.59	5.58
	TAG	3	12.34	2.69
	CAG	4	49.62	64.35
	CAA	5	12.85	6.14
	AAG	6	0	7.31
lys	TTT	1	100	64.01
	CTT	2	0	35.99
met	CAT	1	100	100
	GAA	1	100	50.65
phe	AAA	2	0	49.35
	TGA	1	20.32	7.08
ser	GGA	2	39.84	19.11
	CGA	3	19.92	19.95
	GCT	4	19.92	37.27
	ACT	5	0	8.32
	AGA	6	0	8.26
	thr	GGT	1	49.01
CGT		2	24.75	24.87
TGT		3	26.24	5.63
AGT		4	0	9.65
trp	CCA	1	100	100
	GTA	1	100	47.66
tyr	ATA	2	0	52.34
	TAC	1	62.47	9.81
val	GAC	2	37.53	29.73
	CAC	3	0	45.52
	AAC	4	0	15.05
	TCA	1	100	33.67

¹Las frecuencias se calcularon como la cantidad de codones o anticodones por sobre el total de codones o anticodones para un aminoácido dado.

²Numeración empleada para la resolución de una nueva nomenclatura para tDNAs (ver Sección 4.3).

³Se destacan los anticodones con frecuencias de anticodón igual a 0% (rojo) e igual a un 100% (azul).

Anexo III: Las 97 clases de MGEs identificadas en tDNAs y tmDNAs.

<i>Locus</i>	<i>Clase</i>	<i>Tamaño MGE (kb)</i>	<i>Tamaño DR¹ (pb)</i>	<i>Presencia Integrada</i>	<i>Tipo MGE</i>	<i>Función Predominante²</i>
arg2A	I	47.11	-	Sí	Profago	Infección y Replicación de Bacteriófago
	II	45.44	-	Sí	Isla Genómica	Transporte de azúcares y Proteínas de Fimbrias
	III	40.93	-	No	Isla Genómica	Secreción Tipo VI
	IV	42.77	23	Sí	Isla Genómica	Desconocida
	V	43.89	25	Sí	Profago	Infección y Replicación de Bacteriófago
	VI	37.27	-	Sí	Isla Genómica	Secreción Tipo VI
arg3A	I	49.34	210	Sí	Profago	Infección y Replicación de Bacteriófago
	II	47.20	46	Sí	Profago	Infección y Replicación de Bacteriófago
	III	21.10	21	Sí	Profago (vestigio)	Desconocida
	IV	48.55	47	Sí	Profago	Infección y Replicación de Bacteriófago
	V	50.05	45	Sí	Profago	Infección y Replicación de Bacteriófago
	VI	47.52	213	Sí	Profago	Infección y Replicación de Bacteriófago
	VII	43.42	210	Sí	Profago	Infección y Replicación de Bacteriófago
	VIII	41.30	206	Sí	Profago	Infección y Replicación de Bacteriófago
asn1 ³	I	19.36	21	Sí	Isla Genómica	Resistencia a drogas
	II	29.84	24	Sí	Isla Genómica	Ensamblaje de Pilus
	III	138.85	40	Sí	Isla Genómica	Síntesis de Yersinobactina
	IV	76.20	21	Sí	Isla Genómica	Síntesis de Yersinobactina
	V	66.22	20	-	Isla Genómica	Síntesis de Yersinobactina
	VI	62.18	21	Sí	Isla Genómica	Síntesis de Yersinobactina
	VII	22.29	20	Sí	Isla Genómica	Síntesis de Microcina E492
	VIII	6.10	106	No	Isla Genómica	Sistema Toxina-Antitoxina
	IX	4.64	36	No	Isla Genómica	Sistema Toxina-Antitoxina
	X	14.85	19	Sí	Isla Genómica	Resistencia a drogas
	XI	63.42	20	Sí	Isla Genómica	Ensamblaje de Pilus
	XII	6.74	36	No	Isla Genómica	Metabolismo de Aminoácidos
gly1A	I	6.61	23	Sí	Isla Genómica	Desconocido
	II	11.78	50	Sí	Isla Genómica	Desconocido
	III	14.37	23	No	Isla Genómica	Desconocido
leu2A	I	6.70	125	Sí	Profago (vestigio)	Desconocido
leu5A	I	11.21	24	Sí	Profago (vestigio)	Desconocido
	II	79.19	59	Sí	Profago	Infección y Replicación de Bacteriófago
	III	44.55	-	Sí	Profago	Infección y Replicación de Bacteriófago
	IV	24.46	-	No	Isla Genómica	Desconocida
	V	29.24	74	Sí	Isla Genómica	Sistema Toxina-Antitoxina
	VI	20.65	130	Sí	Isla Genómica	Desconocida
	VII	43.69	76	Sí	Profago (vestigio)	Desconocida
	VIII	45.20	60	Sí	Isla Genómica	Resistencia a drogas
	IX	13.62	65	Sí	Secuencia de Inserción	Ninguna
	X	56.96	75	Sí	Transposón	Desconocida

	XI	12.59	37	Sí	Isla Genómica	Regulación de Fimbrias
	XII	26.13	-	No	ECl	Desconocida
leu5A	XIII	33.66	133	Sí	Isla Genómica	Sistema Toxina-Antitoxina
	XIV	5.91	-	Sí	Isla Genómica	Desconocida
met1C	I	21.00	88	No	Isla Genómica	Sistema Secreción Tipo VI
	II	23.89	79	No	Isla Genómica	Sistema Secreción Tipo VI
	I	32.88	100	Sí	Profago	Infección y Replicación de Bacteriófago
met1E	II	35.26	101	Sí	Profago	Infección y Replicación de Bacteriófago
	III	33.58	50	Sí	Profago	Infección y Replicación de Bacteriófago
	IV	33.48	50	Sí	Isla Genómica	Sistema Toxina-Antitoxina
	V	16.45	92	Sí	Profago	Infección y Replicación de Bacteriófago
	I	59.73	19	Sí	Isla Genómica	Ensamblaje de Pilus
phe1A	II	3.21	86	No	Transposón	Resistencia a drogas
	III	13.59	18	Sí	Isla Genómica	Estructura y Ensamblaje de Fimbrias
	IV	13.26	16	Sí	Isla Genómica	Resistencia a drogas
pro2A	I	5.78	-	Sí	Isla Genómica	Transporte de Aminoácidos
	II	12.14	21	Sí	Isla Genómica	Desconocida
	I	3.12	18	Sí	Profago (vestigio)	Desconocida
sec1A	II	7.81	20	Sí	Isla Genómica	Desconocida
	III	11.15	21	Sí	Isla Genómica	Desconocida
	IV	4.60	-	Sí	Profago (vestigio)	Desconocida
	I	19.58	-	No	Isla Genómica	Transporte de azúcares
ser1A	II	28.55	15	Sí	Isla Genómica	Ensamblaje de Pilus
	III	42.94	102	Sí	Profago	Infección y Replicación de Bacteriófago
ser2A	I	8.60	23	Sí	Isla Genómica	Desconocida
ser3A	I	53.31	69	Sí	Profago	Infección y Replicación de Bacteriófago
	II	42.17	68	Sí	Profago	Infección y Replicación de Bacteriófago
	I	36.61	-	Sí	Profago	Infección y Replicación de Bacteriófago
	II	9.92	67	Sí	Profago (vestigio)	Desconocida
	III	12.69	-	No	Isla Genómica	Desconocida
	IV	34.84	-	Sí	Isla Genómica	Resistencia a drogas
	V	8.92	97	Sí	Profago (vestigio)	Desconocida
	VI	11.64	36	Sí	Isla Genómica	Desconocida
ssrA	VII	31.06	46	Sí	Isla Genómica	Desconocida
	VIII	21.45	-	Sí	Isla Genómica	Desconocida
	IX	27.70	15	Sí	Isla Genómica	Desconocida
	X	18.26	-	Sí	Profago (vestigio)	Desconocida
	XI	32.63	-	Sí	Isla Genómica	Desconocida
	XII	40.23	36	Sí	Profago	Infección y Replicación de Bacteriófago
	XIII	37047	129	Sí	Isla Genómica	Sistema Toxina-Antitoxina
	XIV	12.83	53	Sí	Profago (vestigio)	Desconocida
	I	10.77	23	Sí	Profago (vestigio)	Desconocida
thr2A	II	20.57	23	Sí	Isla Genómica	Estructura y Ensamblaje de Fimbrias
	III	15.22	-	No	Isla Genómica	Transporte de Aminoácidos
	IV	5.91	-	No	Transposón	Desconocido
	V	10.05	69	No	Transposón	Desconocido

	VI	36.91	65	No	Transposón	Desconocida
	VII	23.17	-	No	Isla Genómica	Estructura y Ensamblaje de Fimbrias
	VIII	11.37	121	Sí	Profago (vestigio)	Desconocida
	IX	12.04	17	Sí	Profago (vestigio)	Desconocida
	X	13.73	50	Sí	ICE	Desconocida
thr2A	XI	17.41	41	Sí	Isla Genómica	Desconocida
	XII	22.64	-	Sí	Isla Genómica	Infección y Replicación de Bacteriófago
	XIII	14.82	-	Sí	Isla Genómica	Estructura y Ensamblaje de Fimbrias
	XV	17.86	127	Sí	Isla Genómica	Desconocida
	XVI	39.47	40	Sí	Isla Genómica	Desconocida

¹DR: repetidos directos. La ausencia de un tamaño implica que no fueron identificados DR en el MGE.

²Fue considerada la función que pudiese determinar la virulencia de la cepa portadora, otras funciones pueden estar codificadas en los MGEs.

³Basado en reporte anterior de las islas que ocupan estos sitios (Marcoleta, et al., 2016).

ANEXO IV: Islotipos de las cepas analizadas.

Cepa	Clase de GI insertada en cada <i>locus</i>																				
	<i>arg2A</i>	<i>arg3A</i>	<i>asn1A</i>	<i>asn1B</i>	<i>asn1C</i>	<i>asn1D</i>	<i>gly1A</i>	<i>leu2A</i>	<i>leu5A</i>	<i>met1C</i>	<i>met1E</i>	<i>phe1A</i>	<i>pro2A</i>	<i>sec1A</i>	<i>ser1A</i>	<i>ser2A</i>	<i>ser3A</i>	<i>ssrA</i>	<i>thr2A</i>		
1084			II		VII	II			II					I	I					IV	
342		IV	X	XI			I		VII	I			I		I					III	XVI
1158	IV		III			I	I		III				III							VII	III
30660		I				II			I				I							I	I
30684						II			I				I							I	I
32192		I				II			I				I		III						I
34618		I				II			I				I								I
234-12	III						I		XI		V				I					VIII	X
43816 KPPR1						IV			VIII				III		I	I	I				XIII
500_1420		I				II			I				I		III						I
BAA-2146		I							I				I		II					I	
blaNDM									V				I	II							I
CAV1193		II							I	I	II									II	II
CAV1344		II							I	I	II									II	II
CAV1392		IV							I				I								I
CAV1596		I			VI	II			I				I							I	I
CG43							IX	III	IV				III								II
DMC1097						II			I				I							I	I
HK787		VII		VI			I		IV				III								II
HKUOPLC									XIV						I					IV	VIII
HS11286		IV		VI					I				I								I
J1									XIII					III						XIV	XV
JM45		IV							I				I							I	I

KCTC 2242		VI			I		III		III			V	III
KP13	VI		V				III						II
KP5-1	V			XII			XII	III		I	I	VI	XIV
Kp52.145		V		III	I	III	III	II		II			III
KP617					V		VI	I		III		I	V
KpN01	I	III			VIII	I	VI		I				II
KpN06	I	III			VIII	I	VI		I				II
KPNIH1					II		I			I			I
KPNIH10		I			II		I			I			I I
KPNIH24		I			II		I			I		III	I
KPNIH27							I			IV			X XI
KPNIH29		VIII			I		X				III		XI XII
KPNIH30		I			II		I			I		III	
KPNIH31	II						VI		III		II		II
KPNIH32		I		III	II		I			I			I I
KPNIH33		I	III		II		I			I			I I
KPR0928		I			II		I			I		III	
MGH 78578							II	II	III	III			IX VII
MS6671	II						IX		III		II	II	II
NTUH-K2044					IV		II			II	I	I	IV
PittNDM01					V		VI	I		III		I	V
PMK1					III		VI	I	IV	III	II	I	VI
RJF999				VII	III		II				I	I	XII IV
RYC492				VII			VII			II		I	
U25					V		VI	I	III	III		I	V
UHKPC07		I			II		I			I		III	I

UHKPC33	I	II	I	I	III	I
XH209		II	V		IV	XIII IX
