

Tesis Doctoral

Estudios genómicos y funcionales en *Anaplasma marginale* y en bacterias intracelulares de la clase alfa-proteobacteria y su relación con la reducción del genoma

Nuñez, Pablo Alfredo

2012

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Nuñez, Pablo Alfredo. (2012). Estudios genómicos y funcionales en *Anaplasma marginale* y en bacterias intracelulares de la clase alfa-proteobacteria y su relación con la reducción del genoma. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Nuñez, Pablo Alfredo. "Estudios genómicos y funcionales en *Anaplasma marginale* y en bacterias intracelulares de la clase alfa-proteobacteria y su relación con la reducción del genoma". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2012.

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

**Estudios genómicos y funcionales en *Anaplasma marginale* y
en bacterias intracelulares de la clase α -proteobacteria y su
relación con la reducción del genoma.**

Tesis presentada para optar por el título de Doctor de la Universidad de Buenos
Aires en el área de Ciencias Biológicas

LIC. PABLO ALFREDO NUÑEZ

Director de tesis: Dr. Marisa Farber

Consejero de Estudios: Dr. Omar Coso

Lugar de trabajo:

Instituto de Biotecnología

Centro de Investigaciones Veterinarias y Agronómicas (CICVyA)

Centro Nacional de Investigaciones Agropecuarias (CNIA)

Instituto Nacional de Tecnología Agropecuaria (INTA)

Buenos aires

Argentina

Buenos aires, marzo de 2012

Agradecimientos,

A mi madre, padre y hermano.

A nati, familia completa y amigos.

A los compañeros del laboratorio, del Instituto y los que colaboraron con la tesis.

Soy parte de una generación que nació en los últimos años de la dictadura militar. No la registré personalmente; pero la sufro desde que tengo uso de razón. Cursé el secundario en un colegio de la Universidad de Buenos Aires, el Carlos Pellegrini, entre los años 1993-1998 y empecé a estudiar Biología en el año 2000 en la Facultad de Ciencias Exactas y Naturales de la UBA. El 20 de diciembre del 2001, con un país en crisis, con represión y muertos, rendía un final de física I. Me formé pensando que tendría que irme afuera al graduarme. Me formé en décadas tristes. Décadas que, por solo hablar de nuestra área, el Sistema Educativo y el Sistema Científico Nacional sufrieron la desfinanciación y desarticulación completa. Con un Estado vaciado y gobernado por las corporaciones.

Mucho cambió desde el 2003 en la Argentina.

Como tantos otros graduados, me quedé a realizar mi doctorado con una beca del CONICET. En una Facultad y un INTA que mejoraron sustancialmente su presupuesto, sus recursos humanos, salarios e infraestructura. En un país que tiene un Ministerio nuevo, el MINCYT, para seguir fortaleciendo el Sistema Científico Nacional y contribuir a industrializar el país generando empleos de mayor calidad e impulsando la distribución de la riqueza. Con un Estado gobernado por la política, en democracia y que piensa en los millones de argentinos.

Recibo el título de Doctor, en un país que ya es distinto, que quiere y tiene que ser mejor. En una Argentina que nos necesita.

Un país donde se reencontraron y revalorizaron dos de las herramientas más importantes de una sociedad para construir un país distinto, la ciencia y la política.

Por todo lo que falta.

INDICE

Resumen	5
Abstract.....	7
Diversidad genómica y biológica en la clase α -proteobacteria.	9
<i>Anaplasma marginale</i> y su impacto en la producción bovina de la Argentina.....	13
Impacto de los patógenos del orden Rickettsiales como agentes etiológicos de enfermedades emergentes en la Argentina y en el mundo.	14
Genomas bacterianos completamente secuenciados y oportunidades.	17
Objetivo general.....	19
Objetivos específicos.	19
Capítulo I.	20
1.1. Antecedentes	21
1.1.1. Control de la Anaplasmosis.....	21
1.1.2. Infección, inmunidad y proteínas de superficie.	23
1.2. Resultados	26
1.3. Discusión.....	36
Capítulo II.	38
2.1. Antecedentes	39
2.1.1. Sistemas de secreción de proteínas.	39
2.1.2. El sistema de secreción “Twin Arginine Translocation Pathway (Tat)”	40
2.2. Resultados.	42
2.3. Discusión.....	58
Capítulo III.	66
3.1. Antecedentes	67
3.1.1 El cromosoma bacteriano y los modelos para explicar la existencia de operones.	67
3.1.2 La reducción del genoma y el destino de los operones.	69
3.2. Resultados.	74
3.3. Discusión.....	95
Conclusiones generales.	102
Materiales y métodos.....	105
Capítulo I.	105
Selección de antígenos recombinantes.	105
Análisis <i>in silico</i> y genómica comparativa.	105
Expresión de las proteínas recombinantes en <i>E. coli</i>	105
Amplificación, clonado y secuenciación.....	106
Expresión y purificación de proteínas recombinantes.....	106
Animales utilizados en el estudio.	107
Linfoproliferación de células mononucleares periféricas.	108
Extracción de ARN y RT-PCR.	108
Immunoblotting.	109
Ensayos de <i>Real time</i> q-PCR.	109
Análisis estadístico.	109
Capítulo II.	111
Análisis filogenético y organización de los genes <i>tatABC</i>	111
Alineamientos múltiples.	111
Extracción de ARN y RT-PCR	111
Construcciones y plásmidos utilizados	112
Bacterias utilizadas y condiciones de crecimiento	112
Microscopía.....	113

RT-PCR y <i>Real-time</i> q-PCR.....	113
Predicción <i>in silico</i> de sustratos Tat.....	114
Capítulo III.....	115
Genomas y operones.....	115
Identificación de Ortólogos.....	115
Análisis de Persistencia.....	116
Análisis de expresión.....	119
Análisis de niveles de proteínas.....	119
Análisis de esencialidad.....	120
Cálculos y análisis de matrices.....	120
Análisis estadísticos.....	122
Anexo.....	123
Capítulo I.....	123
Capítulo II.....	125
Capítulo III.....	129
Publicaciones durante el período de la tesis de Doctorado.....	150
Bibliografía.....	151

ESTUDIOS GENÓMICOS Y FUNCIONALES EN *ANAPLASMA MARGINALE* Y EN BACTERIAS INTRACELULARES DE LA CLASE α -PROTEOBACTERIA Y SU RELACIÓN CON LA REDUCCIÓN DEL GENOMA.

Resumen

La subdivisión α - de la clase proteobacteria representa un grupo sumamente diverso y heterogéneo de bacterias con una gran variabilidad en sus características genómicas, biológicas y ecológicas. Dentro de esta subdivisión se encuentra el orden Rickettsiales que ha cobrado gran importancia durante los últimos años por agrupar a diversos patógenos emergentes que resultan relevantes para la salud pública y salud animal. Entre ellos, se destaca *Anaplasma marginale*, causante de la Anaplasmosis bovina enfermedad transmitida por garrapatas que tiene gran impacto sobre la actividad ganadera en nuestro país. Es importante concentrar esfuerzos en comprender mejor las preguntas: ¿Qué tienen en común y en qué difieren estos organismos en términos de información en el genoma y en su biología? ¿Podemos identificar y explicar las diferencias en sus capacidades de adaptación en términos del contenido génico y la conservación de la estructura y función de determinados mecanismos moleculares? Este trabajo se centra en aspectos inmunológicos, funcionales y evolutivos tomando como modelo a *A. marginale* y analizando comparativamente organismos relacionados del orden Rickettsiales en el contexto de la clase α -proteobacteria.

La caracterización del conjunto de proteínas de membrana externa y de los sistemas de secreción representa un desafío importante por ser los principales blancos de interacción de las bacterias con las células hospedadoras y con las células del sistema inmune. Este estudio permitió la identificación de dos proteínas en *A. marginale* con capacidad antigénica que fueron capaces de estimular una respuesta inmune específica. Estudios futuros permitirán verificar experimentalmente la localización superficial, como también corroborar su potencialidad como antígenos protectivos en nuevas formulaciones de vacunas racionales. La caracterización del sistema de secreción de proteínas "Twin Arginine translocación pathway (Tat)" aporta nueva información acerca de la funcionalidad del sistema en dos patógenos de gran relevancia: *A. marginale* y *B. abortus*. A pesar de que sistema Tat en *A. marginale* presentan una organización genómica fragmentada, se pudo demostrar la conservación de su funcionalidad en este organismo. Esta organización se opone a la estructura clásica en operon bajo una única secuencia regulatoria característica de la mayoría de los sistemas proteicos multi-componentes. Asimismo, en *A. marginale* la diferencia de los niveles de abundancia de ARNm de cada uno de los genes correlaciona con los niveles de proteínas descriptos para complejos funcionales, sugiriendo la existencia de una adaptación a mecanismos de regulación alternativos gen-específicos. Las bacterias intracelulares estrictas tendrían un uso limitado del sistema Tat, mientras que organismos de vida libre y bacterias del suelo presentarían un uso moderado a alto. Sin embargo, tanto en *A. marginale* como en *B. abortus* el sistema Tat podría estar involucrado en procesos relevantes de adaptación y patogénesis.

Los procariontes coordinan la transcripción de genes involucrados en complejos metabólicos o de proteínas mediante la organización de operones. En la actualidad,

existen diferentes hipótesis acerca de las fuerzas que rigen estos procesos de organización, tanto para determinar su origen como los mecanismos de mantenimiento de este tipo de estructura. A partir de la fragmentación del operon *tat* y la adaptación a mecanismos de regulación gen-específicos, como así también de otros casos descritos recientemente en la literatura, surgió la hipótesis de la existencia de un proceso generalizado de fragmentación de operones en organismos asociados a ambientes intracelulares con genomas reducidos en tamaño. En tal sentido, fue posible comprobar que en las α -proteobacterias de genomas pequeños tanto el número como el largo de los operones era menor en comparación con bacterias de genomas también pequeños pero pertenecientes a otros grupos. Este comportamiento observado en las α -proteobacterias sería compatible con un proceso generalizado de fragmentación. Estos resultados se explicarían teniendo en cuenta que los procesos de reducción genómica se describieron como procesos evolutivos independientes en los diferentes linajes. Del análisis de las diferentes propiedades de los genes presentes en operones en diversos grupos filogenéticos se pudo detectar tanto resultados que reflejan diversidad como patrones consistentes y conservados. La mayor frecuencia de fragmentación supone el incremento de regiones de regulación gen-específicas en genomas de tamaño reducido, y/o en genes bajo escasa presión de selección, como se observa para los genes menos persistentes, de baja expresión y no-esenciales. La aparición de mecanismos de control gen-específico versus un control concertado en operón estaría fuertemente influenciado por los procesos de mutación, selección y deriva. Dada la diversidad de la clase y la complejidad de procesos que afectan y determinan la organización de los cromosomas, el conocimiento acerca de los mismos ayudará a conocer los procesos celulares, moleculares que subyacen a la diversidad.

Aspiramos a que los estudios realizados en la tesis aporten al conocimiento general de los organismos del orden Rickettsiales para avanzar en la caracterización de un grupo de bacterias que requiere más atención y estudio como patógenos emergentes de amplia distribución.

Palabras clave: α - proteobacteria, Rickettsiales, *Anaplasma marginale*, proteínas de superficie, inmunidad, vacunas, sistemas de secreción, Tat, intracelulares obligadas, reducción del genoma, fragmentación de operones, persistencia, expresión, esencialidad, evolución de operones.

GENOMIC AND FUNCTIONAL STUDIES IN *ANAPLASMA MARGINALE* AND INTRACELLULAR α -PROTEOBACTERIA AND ITS RELATION WITH GENOME REDUCTIVE EVOLUTION.

Abstract

The α -proteobacteria class represents an extremely diverse and heterogeneous group of bacteria with a high variability in their genomic, biological and ecological features. The order Rickettsiales has become more important over the last years, since it groups together emerging pathogens that are relevant to human and animal health. Among them, *Anaplasma marginale* is the etiological agent of bovine Anaplasmosis, an infectious tick-borne disease that highly impacts on livestock farming activity in our country. It is important to focus efforts to better understand the following questions: what do these organisms have in common and in what do they differ, taking into account their genome and biology? Can we identify and explain the differences in their adaptive capacities in terms of genetic content, structure and functionality conservation of molecular mechanisms? This work focus on immunological, functional and evolutionary topics based on *A. marginale* in comparison with related organisms from the order Rickettsiales in the context of the α -proteobacteria class.

The characterization of the set of outer membrane proteins and secretion systems represents a significant challenge, since they are the main targets of interaction of bacteria with host cells and the immune system, and also can be responsible of the export of virulence factors. This study allowed the identification of two novel proteins with antigenic capacities in *A. marginale* which were able to stimulate a specific immune response. Future studies will allow to verify their surface location, as well as their potential as protective antigens in new vaccine formulations. The characterization of protein secretion system "Twin Arginine translocation pathway (Tat)" provides new information on the functionality of the system in two relevant pathogens: *A. marginale* and *B. abortus*. The functionality of the Tat system was demonstrated in *A. marginale*, irrespective of the genomic and regulatory divergence relative to the classical operon organization under unique regulatory sequence. Difference in the abundance of mRNA levels which correlated with the protein levels described for functional complexes suggests the existence of alternative adaptive mechanisms by gene-specific regulation. Pathogenic intracellular bacteria would have a limited usage of the Tat system, while free-living organisms and soil bacteria could have a moderate to high usage. However, in both pathogens analyzed in this study, the Tat system could be involved in several adaptive or pathogenic mechanisms. Prokaryotes usually coordinate transcription of genes involved in metabolism or protein complexes through the organization of operons. Nowadays, different hypotheses are under discussion about the forces that govern these organizations, both regarding the origin and maintenance of this type of structure. Considering the *tat* operon fragmentation as well as other cases recently described in the literature, we hypothesized the existence of a more generalized operon fragmentation event in small genome organisms associated with intracellular lifestyles. In this regard, the α -proteobacteria showed less usage and length of operons supporting a generalized fragmentation process, while the opposite pattern was observed in small genomes from other groups. These results could be explained considering that genome reduction events were described as independent evolutionary processes in different lineages.

The analysis of the different genes features and phylogenetic groups let us detect abundant diversity as well as consistent and conserved patterns among groups. The highest operon fragmentation frequency involves the increase of gene-regulatory regions in small genomes and/or in operons of genes with lower adaptive value as it is observed for non-persistent, lowly-expressed and non-essential genes. This suggests that gene-specific mechanisms versus operon organization are strongly shaped by mutation, selection and genetic drift. Considering the diversity and complexity of the events that affect the organization of bacterial chromosomes, a better understanding of chromosome organization will help to understand the cellular and molecular mechanisms underlying diversity.

We believe this thesis contribute to the overall knowledge of organisms of the Rickettsiales order, making it possible to improve the characterization of a bacterial group that requires more attention and study as emerging pathogens widely distributed.

Keywords: α -proteobacteria, Rickettsiales, *Anaplasma marginale*, surface proteins, immunity, vaccines, secretion systems, Tat, obligate intracellular, genome reduction, operon fragmentation, persistence, expression, essentiality, evolution of operons.

Antecedentes generales.

Diversidad genómica y biológica en la clase α -proteobacteria.

La subdivisión α - de la clase proteobacteria representa un grupo sumamente diverso y heterogéneo de bacterias. Está compuesta por diferentes grupos filogenéticos representados por diversos ordenes que agrupan y representan sus propiedades biológicas [1]. Las bacterias pertenecientes a este grupo presentan una gran variabilidad en sus capacidades metabólicas, morfología y ciclos de vida. Habitan diferentes nichos ecológicos; organismos de vida libre en el agua y el suelo, o en asociaciones, obligadas o facultativas, intra- y extra-celulares con eucariotas, incluyendo organismos intracelulares, nematodos, artrópodos, plantas y animales.

Es importante concentrar esfuerzos en comprender mejor las preguntas ¿Qué tienen en común, y en qué difieren, estos organismos en términos de información en el genoma y de su biología? ¿Comparten estrategias y rasgos genéticos para permitir la interacción con las células hospedadora? ¿Podemos identificar y explicar las diferencias en sus capacidades de adaptación en términos de contenido génico y conservación de la estructura y función de determinados mecanismos moleculares?

Hasta la fecha, 172 genomas completos de bacterias de la clase α -proteobacteria se encuentran completamente secuenciados y anotados. Los genomas muestran una variación notable en el tamaño, desde 0,86 a 9,1 Mb (**Figura 1**). Algunas bacterias atípicas, secuenciadas recientemente como *Candidatus Hodgkinia cicadicola Dsem* presenta un genoma de 140.000 nucleótidos [2].

Los miembros de este grupo presentan prácticamente, el espectro completo de ciclos de vida; desde organismos de vida libre como *Caulobacter crescentus* hasta parásitos pericelulares de plantas como *Agrobacterium tumefaciens str. C58*, patógenos intracelulares facultativos como *Brucella* spp. y *Bartonella* spp., simbioses de plantas, conocidos como *Rhizobia*, hasta patógenos intracelulares obligados de mamíferos o artrópodos como los géneros de *Rickettsia*, *Anaplasma*, *Ehrlichia* y *Wolbachia*.

El orden Rhizobiales, compuesto por al menos 8 familias, es un grupo muy diverso filogenéticamente (**Figura 1**). Se compone principalmente de las bacterias del suelo que presentan asociaciones simbióticas con plantas. Las bacterias colonizan intracelularmente los nódulos de las raíces y participan activamente en el proceso de fijación de nitrógeno generando el 50% del nitrógeno total fijado anualmente.

Presentan también una amplia diversidad en relación al tamaño del genoma desde 1,46 a 9,23 Mb (**Figura 1**). En particular, *Agrobacterium* spp. son organismos fitopatógenos que a través de infecciones pericelulares provocan un crecimiento *tumor-like* e inducen un crecimiento no controlado de las raíces en plantas susceptibles. Dentro de este orden también se encuentran las especies intracelulares facultativas que corresponden a los clados *Brucella* spp. y *Bartonella* spp. Las especies de *Bartonella* poseen genomas pequeños desde 1,4 a 2,6 Mb, están asociados a diferentes artrópodos vectores, son capaces de infectar una amplia gama de animales y, algunas en particular, provocan importantes enfermedades en humanos. Las especies de *Brucella* spp. también establecen asociaciones intracelulares facultativas, son las causantes de Brucelosis, la zoonosis más extendida en el mundo, infectando rango muy amplio de mamíferos.

El orden Rickettsiales agrupa a diferentes familias (*Anaplasmataceae*, *Rickettsiaceae*) de bacterias intracelulares obligadas. Dentro de la familia *Anaplasmataceae*, las especies de *Anaplasma* spp., *Ehrlichia* spp. y *Neorickettsia* spp. son patógenos de impacto tanto en salud humana como veterinaria, mientras que *Wolbachia* spp. son especies simbioses de artrópodos y nematodos [3]. En particular, los miembros del género *Anaplasma* y *Ehrlichia*, replican en vacuolas citoplasmáticas de las células hematopoyéticas del hospedador mamífero y a su vez tienen como hospedador-vector a las garrapatas (**Figura 2**). *Anaplasma marginale* infecta eritrocitos y es responsable de la Anaplasmosis bovina, una de las enfermedades transmitidas por garrapatas de mayor impacto en la producción bovina de las regiones tropicales y subtropicales. *Anaplasma phagocytophilum*, es transmitida por garrapatas del grupo *Ixodes* spp., siendo el agente causante de la Anaplasmosis Granulocítica Humana (HGA) [4, 5]. *Ehrlichia chaffeensis* es un patógeno emergente en los Estados Unidos, agente etiológico de la Ehrlichiosis Monocítica Humana (HME) [6]. Las especies de *Wolbachia* spp. son conocidas por su capacidad de inducir procesos de distorsión en las proporciones de sexos en las poblaciones, siendo transmitidas transovarialmente por hembras de invertebrados a su progenie. Dentro de la familia *Rickettsiaceae*, el género *Rickettsia* se subdivide principalmente en tres grupos: **i)** el *Typhous Group* (TG) que incluye *Rickettsia prowazekii* y *Rickettsia typhi*, que son los agentes etiológicos del tifus epidémico y endémico, respectivamente; **ii)** el grupo de la fiebre manchada (*spotted fever group*, SFG) compuesto por las especies causantes de la fiebre de las montañas rocallosas (*Rocky Mountain Fever*) *Rickettsia rickettsii*, de la fiebre manchada del Mediterráneo, *Rickettsia conorii* [7] y *Rickettsia massillae* y el **iii)** que incluye especies más divergentes, como *Rickettsia canadensis* y *Rickettsia bellii* [8-10] (**Figura 2**). La mayor cantidad de las especies conocidas de las Rickettsias se

clasifican en el grupo de la fiebre manchada (SFG), siendo en su gran mayoría transmitidas por garrapatas [11]. Hasta hace algunos años, *Rickettsia rickettsii*, el agente etiológico de la *Spotted Fever* en Brasil y de *Rocky Mountain Fever* en los Estados Unidos, era la única especie de *Rickettsia* conocida capaz de infectar garrapatas en América Latina [12]; en la actualidad recientes trabajos indican una distribución mas amplia que será detallada mas adelante.

A pesar de su diversidad fenotípica, las relaciones dentro de la clase α -proteobacteria tiene un fuerte soporte filogenético utilizando, diferentes métodos y algoritmos para estimar las relaciones filogenéticas como también diferentes grupos de genes tanto ribosomales (rRNA) como un conjunto de genes altamente conservados (denominado *core genome*) [13]. El orden Rickettsiales representa un linaje divergente en la filogenia indicando la transición a ambientes intracelulares, acoplado a un proceso evolutivo de reducción genómica [14]. Por otro lado, el orden Rhizobiales (excluyendo a *Bartonella* spp.), con genomas más grandes y variables, pertenecen a un grupo filogenético con cierta divergencia asociado a un proceso de expansión genómica. Estudios computacionales a través análisis de deleciones, duplicación o nuevos genes permitieron estimar el tamaño de genoma del ancestro común en 3000-5000 genes [14] y proponen la existencia de dos principales tendencias para arribar a las especies actuales: expansión y reducción genómica. Por un lado, un proceso de reducción del genoma para las bacterias intracelulares obligadas que asociadas con invertebrados, mamíferos incluyendo a humanos, evolucionaron a través de procesos de pérdidas de genes; y un proceso donde las especies de vida libre y asociadas a plantas evolucionaron por expansión genómica.

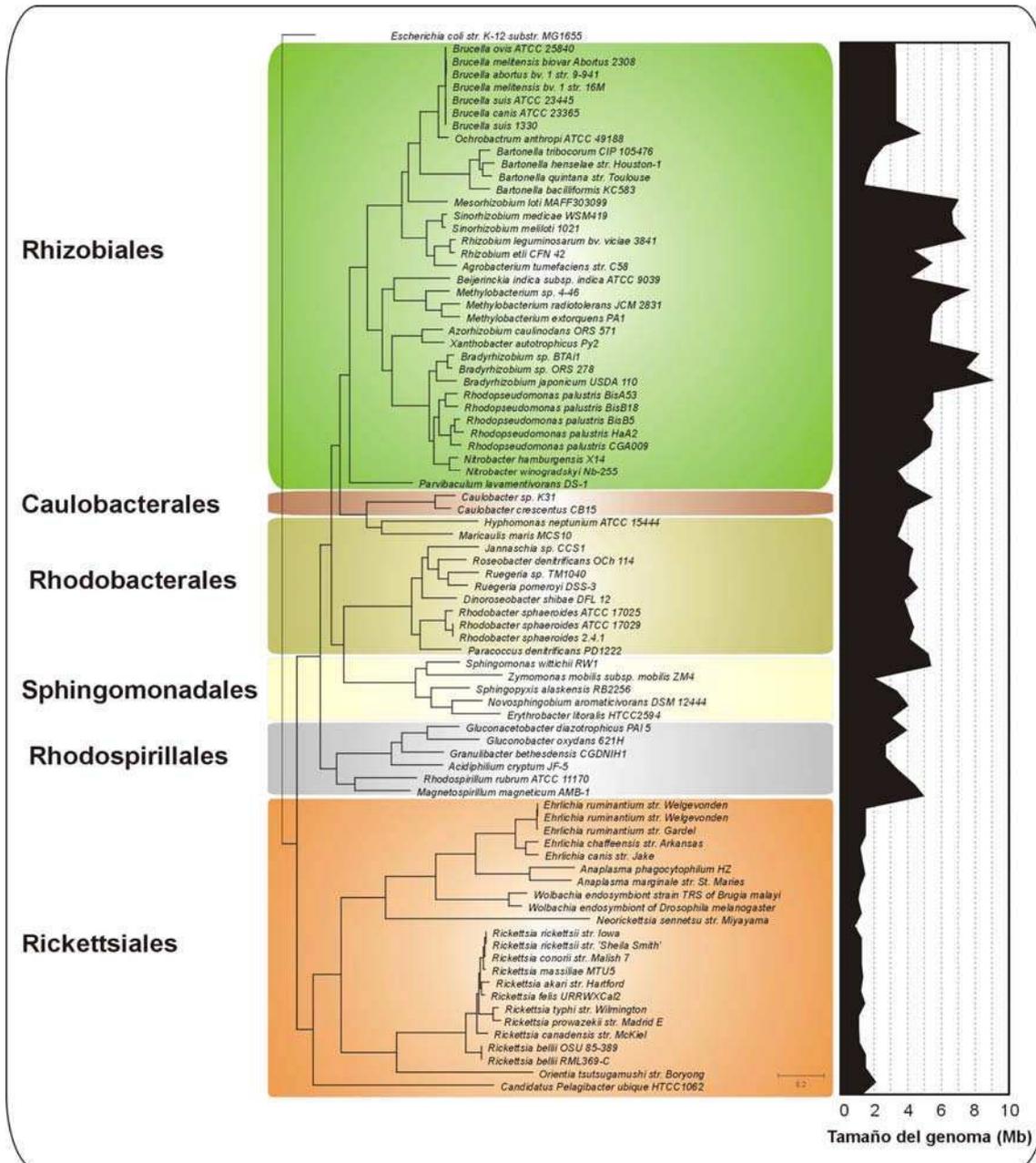


Figura 1. Filogenia de la clase α -proteobacteria.

Análisis filogenético para 82 organismos de la clase α -proteobacteria (Neighbor Joining, NJ) utilizando las matrices de distancia estimadas por el modelo Jones–Thornton–Taylor (JTT), MEGA 4.0, bootstrap=1000. En árbol se realizó con el conjunto de proteínas presentes en >80 genomas (*core genome* \approx 90 proteínas). Las proteínas ortólogas de los diferentes organismos se alinearon individualmente utilizando el ClustalW y luego se concatenaron los alineamientos para la construcción de la filogenia. Los resultados concuerdan con la descripción y agrupamiento en los órdenes establecidos. Como grupo de referencia externo (*outgroup*) se utilizó *Escherichia coli* str. K-12 substr. MG1655.

***Anaplasma marginale* y su impacto en la producción bovina de la Argentina.**

El género *Anaplasma* fue descrito en 1910 por el Dr. Arnold Theiler, que reconoció por primera vez mediante microscopía unos puntos en los márgenes de los eritrocitos de los bovinos afectados por una enfermedad específica [15-17]. Estas inclusiones que fueron identificados luego como *Anaplasma marginale*, habían sido referidas con frecuencia en los glóbulos rojos del ganado anémico, asociadas erróneamente al ciclo de vida del protozoo *Babesia* spp., agente causal de la babesiosis bovina, enfermedad transmitida por garrapatas que se presenta con un cuadro inmunopatológico característico de anemia y desmejoramiento, presentándose con frecuencia en forma de brotes con tasas variables de mortalidad. Sin embargo Theiler determinó correctamente que Babesiosis y Anaplasmosis eran dos enfermedades diferentes que a menudo coexisten en el mismo animal. Tras estos hallazgos, la Anaplasmosis como enfermedad de entidad propia, no tardó en ser reconocida en regiones tropicales y subtropicales donde las garrapatas son enzoóticas.

La anaplasmosis bovina a menudo resulta en el desarrollo de anemia de leve a grave e ictericia, sin hemoglobinemia y hemoglobinuria. La enfermedad clínica, causada por la replicación de la bacteria dentro de los eritrocitos maduros se caracteriza por anemia, pérdida marcada de peso y muerte [18, 19]. Los síntomas clínicos pueden incluir fiebre, pérdida de peso, el aborto, letargia, ictericia, y con frecuencia la muerte en los animales mayores de 2 años. Los bovinos que se recuperan de la infección aguda permanecen infectados en forma persistente desarrollando ciclos recurrentes de bacteremia [20, 21].

Las garrapatas, y particularmente *Rhipicephalus (Boophilus) microplus* (Acari: *Ixodidae*) en América Latina, es el vector biológico de las especies de *Anaplasma* spp. [22], participando asimismo en la transmisión los insectos hematófagos [23-25] y los fomites a través de la transmisión mecánica de glóbulos rojos infectados a animales susceptibles.

Esta enfermedad infecciosa causa importantes pérdidas económicas debido a la alta morbilidad y mortalidad en el ganado, con \$800 millones de pérdidas anuales solamente en América latina [26] y también produce un impacto directo en la productividad afectando la posibilidad de mejoramiento genético en las regiones donde esta enfermedad es enzoótica, a través de la introducción de especies importadas

como *Bos taurus*, altamente susceptibles a enfermedades transmitidas por garrapatas [27].

Las infecciones ocurren en regiones tropicales y subtropicales donde la garrapata es enzootica, detectándose también brotes de la enfermedad en regiones por fuera de esta área debido fundamentalmente al traslado de animales infectados a áreas no enzooticas [28-30]. Esta patología es la principal restricción para el desarrollo de la producción ganadera en muchos países o regiones. En la Argentina, la distribución correlaciona con la distribución del vector y afecta principalmente a las regiones del NOA y NEA. El movimiento de animales entre diferentes regiones, conjuntamente con la posterior transmisión mecánica contribuye a la dispersión de la bacteria en áreas no endémicas, produciendo frecuentemente brotes de la enfermedad. La enfermedad es enzootica en los países de la región latinoamericana con la excepción de regiones templadas, frías, áreas desérticas y montañosas [31]. La distribución se espera que continúe cambiando, en parte, debido a los efectos del cambio climático global que influye en la distribución de vectores y hospedadores [32].

Impacto de los patógenos del orden Rickettsiales como agentes etiológicos de enfermedades emergentes en la Argentina y en el mundo.

Los patógenos del género *Anaplasma* y *Ehrlichia* (familia *Anaplasmataceae*) dentro del orden Rickettsiales constituyen la mayor causa de enfermedades transmitidas por garrapatas en animales y humanos. Como se describió anteriormente *A. marginale* es sumamente importante en nuestra región por el impacto en la producción ganadera, sin embargo, las demás bacterias de la familia, aunque aún menos estudiadas en nuestro país, o algunas sin distribución aparente en la actualidad, representan un grupo que requiere atención por sus implicancias para la salud pública humana y veterinaria como patógenos emergentes (**Figura 2**).

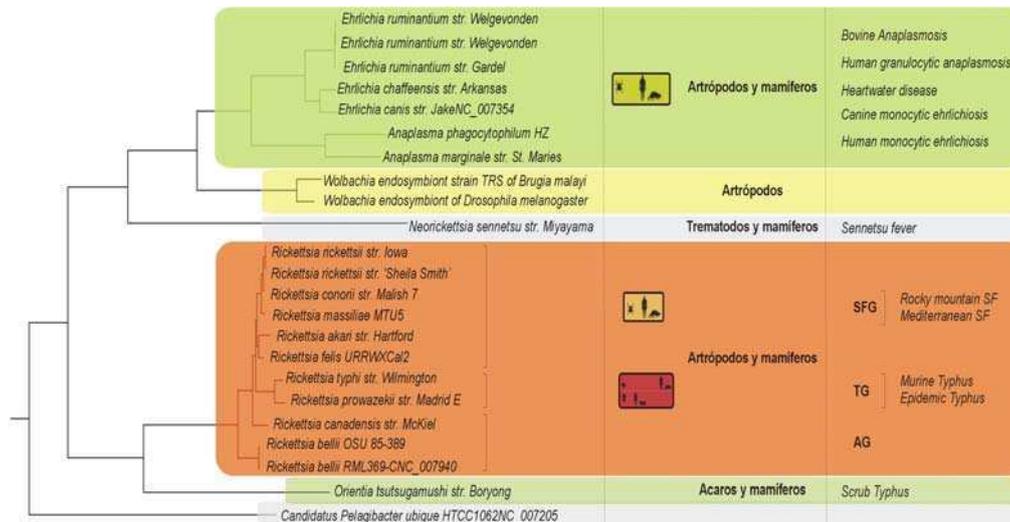


Figura 2. Filogenia del orden Rickettsiales.

Árbol Filogenético (Neighbor Joining, NJ) para 23 organismos del orden Rickettsiales utilizando las matrices de distancia estimadas por el modelo Jones–Thornton–Taylor (JTT), MEGA 4.0, bootstrap=1000. En árbol se realizó con el conjunto de proteínas presentes en >80 genomas (*core genome* ≈ 90 proteínas). Las proteínas ortólogas de los diferentes organismos se alinearon individualmente utilizando el ClustalW y luego se concatenaron los alineamientos para la construcción de la filogenia. Los resultados concuerdan con la descripción y agrupamiento en los órdenes establecidos. Como *outgroup* se utilizó *Escherichia coli* str. K-12 *substr.* MG1655. Las bacterias del orden Rickettsiales tienen complejos ciclos de transmisión involucrando diversos vectores y reservorios y pudiendo ocasionar distintas enfermedades. SFG (Spotted fever group), TG (Typhus group), AG (transition group). Se ilustra el organismo, el vector-hospedador y la enfermedad que produce.

Dentro de la familia *Anaplasmataceae*, *Anaplasma phagocytophilum* es el agente causante de la anaplasmosis granulocítica humana (HGA). La infección por *A. phagocytophilum* causa fiebre, cefalea, mialgias, anorexia, y escalofríos [33]. *A. phagocytophilum* es transmitida por *Ixodes* spp. Los casos de HGA corresponden a la distribución de *Ixodes* spp. y fueron identificados en Nueva Inglaterra, la región del Atlántico medio y el norte de California en los Estados Unidos, como así como en diversas partes de Europa. *A. phagocytophilum* es una de los principales agentes causantes de la Ehrlichiosis en el mundo. También en la familia *Anaplasmataceae*, las diferentes especies de *Ehrlichia* spp. son patógenos de relevancia. En América Central y del Sur, se reportaron casos humanos de Ehrlichiosis con evidencia serológica compatible en Venezuela, Brasil, México y Chile, aunque la bacteria no ha sido aislada [34]. Recientemente, se reportó evidencia molecular de una infección con *E. chaffeensis* en un niño de 9 años de edad en Venezuela [35]. En Argentina,

anticuerpos reactivos para *E. chaffeensis*, o para especies antigénicamente relacionadas, se detectaron en muestras de sueros humano durante un estudio serológico en la Provincia de Jujuy, donde se informaron casos fatales de enfermedades febriles [36]. Aun así, poco se conoce de la epidemiología en América del Sur y particularmente en nuestro país. En Brasil, el ciervo *Blastocerus dichotomus* se propone como el reservorio natural [37]. Recientemente, en nuestro laboratorio realizamos la identificación molecular de *Ehrlichia chaffeensis* en la garrapata *Amblyomma parvum* en nuestro país, en el 9,2% de los especímenes colectados. De las 16 garrapatas *A. parvum* que resultaron positivas, 5 se encontraban infestando humanos, lo que constituye un alerta de potencial riesgo para la salud pública. A su vez, se identificaron variantes genéticas de las *E. chaffeensis* indicando la necesidad de estudiar la distribución y su posible correlación es con hospedadores, reservorios y/o patogenicidad [38].

Durante los últimos años, diversas especies de *Rickettsia* spp. se identificaron en garrapatas de América Latina. Entre ellas, *Rickettsia bellii* en las especies de garrapatas *Amblyomma* spp. y en *Ixodes loricatus* de Brasil [39-41], *Rickettsia parkeri* en garrapatas *Amblyomma* spp. de Uruguay [42], *Candidatus "Rickettsia andeanae"* en *Amblyomma maculatum* y *Ixodes boliviensis* en Perú (Blair 2004), *Rickettsia rhipicephali* en garrapatas *Haemaphysalis juxtakochi* en Brasil [37], *Rickettsia prowazekii* en *Amblyomma* spp. en México [43], y *Rickettsia* spp. cepa ARANHA en *Amblyomma longirostre* de Brasil [12]. Además, se identificó una cepa de rickettsia llamada transitoriamente *Candidatus "Rickettsia amblyommii"* aislada de *Amblyomma cajennense* y *Amblyomma coelebs* de garrapatas en Brasil [40]. Por otro lado, seis casos de rickettsiosis de fiebre manchada (*spotted fever*), dos de ellos mortales, se describieron en la provincia de Jujuy, aunque no fue determinada la especie causal de la enfermedad [36]. La especie de garrapata *Amblyomma neumanni*, es la garrapata que más frecuentemente se identificó infectando humanos en las regiones del noroeste de nuestro país [44]. En un reciente estudio se identificaron *Rickettsia bellii* y "*Candidatus Rickettsia amblyommii*" en *A. neumanni* en garrapatas colectadas en el norte de la provincia de Córdoba, en lo que significó el primer registro de Rickettsias en garrapatas de la Argentina y en esa especie en particular [45]. Aunque no se haya estudiado la patogenicidad de estas bacterias identificadas, conocer su distribución y su potencial patogenicidad, representan importantes desafíos para contribuir al estudio de las zoonosis y enfermedades de importancia en salud pública. Frente a un caso de potencial Rickettsiosis en la región del Delta del Paraná, provincia de Buenos Aires, se realizó un estudio que identificó la presencia de *R. parkeri* en nuestro país en garrapatas *A. Triste*. Por otro lado, dos estudios identificaron una nueva especie de

Rickettsia perteneciente al SFG, "*Candidatus Rickettsia* spp. strain Argentina", en la región del noroeste de nuestro país [46, 47]. En uno de los estudios realizado por nuestro laboratorio, identificamos que el 82,8% de las *Amblyomma* spp. fueron positivas para esta bacteria y la prevalencia de infecciones por rickettsia fue del 93,9% en *A. parvum*, del 54,5% en *A. pseudoconcolor* y el 24% en *A. tigrinum*. Además en este trabajo identificamos la presencia de *R. bellii* en garrapatas *A. tigrinum* por primera vez [47]. Se ha reportado el hallazgo de *R. felis* en pulgas del gato en Brasil y Perú [48, 49] agente recientemente reportado en *R. sanguineus* [50]. Finalmente, también se comunicó el hallazgo de *R. massiliae* en *R. sanguineus* en la Ciudad de Buenos Aires [51].

Los datos anteriores dan cuenta de una amplia distribución aun poco conocida de las bacterias de este orden en nuestra región, resultando muy necesario generar mayor conocimiento de la epidemiología, de sus características inmunológicas, como también de aspectos relacionados con sus mecanismos biológicos de adaptación y patogénesis y evolución genómica que permitan conocer más acerca de este grupo de bacterias.

Genomas bacterianos completamente secuenciados y oportunidades.

Los avances recientes en metodologías de secuenciación han permitido incrementar notablemente la cantidad de genomas bacterianos completos disponibles en las bases de datos públicas para análisis funcionales y evolutivos. Por esto, la disponibilidad de un número creciente de genomas bacterianos secuenciados junto con las nuevas herramientas desarrolladas para el análisis *in silico* a través de un enfoque de genómica comparativa permiten abordar el estudio de los patógenos a través del análisis exhaustivo de sus genomas. De esta manera es posible la identificación de mecanismos conservados de patogénesis y de adaptación a vectores u hospedadores específicos comunes entre patógenos relacionados filogenéticamente. En este sentido, la utilización de herramientas bioinformáticas permiten la predicción de función y/o localización celular de un número determinado de secuencias codificantes o complejos de proteínas. A su vez, las estrategias posteriores, tales como, la caracterización funcional e inmunológica, permiten investigar los procesos complejos que se dan entre el huésped, los vectores y el patógeno. Estos enfoques contribuyen a la comprensión de los mecanismos de virulencia y a la identificación los genes potencialmente responsables de esos mecanismos, abriendo nuevos horizontes

en el campo del diagnóstico, del desarrollo de vacunas de nueva generación y del estudio de la interacción y adaptación de patógenos y simbioses.

El objetivo de la tesis es explorar aspectos inmunológicos, funcionales y evolutivos tomando como modelo a *Anaplasma marginale* y organismos relacionados en el contexto del orden Rickettsiales y de la clase α -proteobacteria. De esta forma aspiramos a contribuir al conocimiento general de los organismos del orden Rickettsiales para avanzar en la caracterización de un grupo de bacterias que requiere más atención y estudio como patógenos emergentes de amplia distribución.

Objetivo general.

Comprender los mecanismos tanto generales como específicos de la patogénesis, la adaptación al medio y la interacción con el vector de *A. marginale* y organismos relacionados de la clase α -proteobacteria. Contribuir a la identificación de genes potencialmente candidatos en el campo del desarrollo de vacunas y a la comprensión de los mecanismos moleculares responsables de la interacción y adaptación de patógenos y simbioses.

Objetivos específicos.

1. Estudiar *in silico* y mediante ensayos de inmunogenicidad la potencialidad de genes seleccionados como antígenos candidatos a fin de corroborar la capacidad para estimular una respuesta inmunológica.
2. Identificar y caracterizar el sistema de secreción de proteínas "Twin arginine translocation pathway (Tat)" con el objetivo de inferir la función y el rol en la capacidad de adaptación y patogénesis.
3. Estudiar de la organización y distribución de operones en genomas de la clase α -proteobacteria para comprender la relación entre la reducción del genoma en bacterias del orden Rickettsiales y su impacto en la organización de operones.

Capítulo I.

Inmunogenicidad de proteínas de membrana de *A. marginale*.

Objetivo específico.

Estudiar *in silico* y mediante ensayos de inmunogenicidad de la potencialidad de genes seleccionados como antígenos candidatos a fin de corroborar la capacidad para estimular una respuesta inmunológica.

1.1 Antecedentes

1.1.1 Control de la Anaplasmosis.

1.1.2. Infección, inmunidad y proteínas de superficie.

1.2. Resultados

1.3 Discusión

1.1. Antecedentes

1.1.1. Control de la Anaplasmosis.

Los mecanismos de control de la anaplasmosis no han cambiado sustancialmente en los últimos 60 años. Las medidas de control varían de acuerdo a la ubicación geográfica e incluyen: el control de artrópodos por aplicación de acaricidas, la administración profiláctica de los antibióticos y la vacunación. El control de los artrópodos no se practica en muchas áreas por los requerimientos de logística y gastos asociados, además de que puede proteger parcialmente contra la transmisión de *A. marginale*. A menudo la transmisión de la bacteria se produce por transmisión mecánica de sangre infectada por fomites. La quimioterapia, probablemente utilizada con más frecuencia para la prevención de la enfermedad, fundamentalmente en los Estados Unidos, es un procedimiento costoso y muchas veces no se aplica al ganado en amplia escala. Por otra parte, el uso intensivo de antibióticos lleva el riesgo de emergencia asociado al surgimiento de cepas resistentes de microorganismos. La vacunación ha sido una forma económica y eficaz para controlar parcialmente en todo el mundo la Anaplasmosis bovina [26].

Las vacunas para el control se dividen en dos grupos principales: vivas y muertas. Ambos tipos de vacunas inducen inmunidad protectora que silencia o previene la enfermedad clínica, pero ninguno impide la persistencia de la infección de *A. marginale*. Las infecciones persistentes en el ganado contribuyen también a la mayor propagación, ya que son una fuente de organismos para la transmisión mecánica y biológica. Aunque las investigaciones llevadas a cabo en las últimas dos décadas han contribuido sustancialmente a un mayor conocimiento de la composición de antígenos de *A. marginale*, del papel de las principales proteínas de superficie y de su potencialidad en conferir inmunidad protectora, actualmente no se desarrollan ni comercializan vacunas basadas en tecnologías moleculares que representarían métodos de inmunizaciones más seguro que las vacunas vivas.

Las vacunas vivas se basan fundamentalmente en la exposición del ganado a través de la inoculación con eritrocitos infectados con *Anaplasma centrale*. *A. centrale*, aislada por el Dr. Arnold Theiler a principios de los años 1900, es en la actualidad la variedad más ampliamente utilizada como vacuna viva para el control de la anaplasmosis bovina [17]. Se utiliza en varias zonas del mundo, incluyendo África, Australia, Israel y varios países de América Latina; aunque no se encuentra habilitada la licencia para su uso en los Estados Unidos. Theiler observó que *A. centrale*, organismo estrechamente relacionado filogenéticamente, fue menos patógeno para el

ganado que *A. marginale* e incluso los bovinos infectados con *A. centrale* desarrollaron una inmunidad protectora contra la infección por *A. marginale*. La reciente publicación del genoma completo de *A. centrale* [52] provee información muy valiosa para una mejor comprensión de las bases de esta capacidad de protección que no ha sido demasiado estudiada aún.

Otra estrategia ha sido infectar a los terneros a inmunizar con *A. marginale* y luego tratarlos con tetraciclina en el inicio del aumento de la temperatura corporal o de la detección de parasitemia. Este método requiere un estrecho control del ganado y no es aplicable para su uso en grandes rodeos. Incluso un tratamiento oportuno con tetraciclina de la reacción post-inoculación, resulta ineficiente en prevenir la infección aguda [53].

Para la producción de vacuna vivas, los terneros esplenectomizados se inoculan experimentalmente con cepas seleccionadas y son utilizados como fuente de sangre infectada. Este procedimiento, sin embargo, es riesgoso teniendo en cuenta la posibilidad de transmisión de otros hemoparásitos. Los bovinos que desarrollan infecciones persistentes, en general adquieren protección a lo largo de la vida y no requieren re-vacunación [54]. Sin embargo, pueden ser susceptibles a la exposición con otros genotipos genéticamente diferentes de *A. marginale*, lo que adquiere relevancia ya que representa un obstáculo frente al traslado de animales de una región a otra.

Las vacunas atenuadas *A. marginale* se producen por irradiación o por pasajes de los organismos por otros hospedadores (ovejas o ciervos). Sin embargo, la protección proporcionada por este tipo de vacunas atenuadas ha sido menos confiable y eficiente [26, 53]

El uso de las vacunas muertas desarrolladas en USA durante los años 1960 es escaso debido a que dejaron de comercializarse a partir de 1999. Representaron algunas ventajas frente a las vacunas vivas, como un menor riesgo de contaminación con otras enfermedades infecciosas y reacciones luego de la inoculación. Sin embargo, usualmente requerían la necesidad de refuerzos anuales, un alto costo de la purificación de eritrocitos infectados por *A. marginale*, la falta de protección cruzada contra otras variantes genéticas y el riesgo de inducir respuestas autoinmunes contra los glóbulos rojos, además de conferir una protección menor que las vacunas vivas.

Por lo tanto, actualmente están planteados importantes desafíos para el diseño y desarrollo de nuevas vacunas más seguras y efectivas para lograr una inmunidad protectora, que prevenga la infección del ganado y también de las garrapatas, de manera de disminuir su capacidad vectorial. Uno de los principales desafíos es que

sean también capaces de conferir protección cruzada frente a diferentes cepas y variantes genéticas.

1.1.2. Infección, inmunidad y proteínas de superficie.

Durante los últimos años, aunque solamente en cierta medida, se avanzó en comprender la naturaleza de la respuesta inmunitaria del ganado a la infección por *A. marginale*, como así también en la identificación de nuevas herramientas y desarrollos de inmunizaciones con capacidad de generar respuestas protectoras y su caracterización [55].

A. marginale invade y replica en glóbulos rojos y durante la infección aguda puede alcanzar niveles muy altos de infección de eritrocitos (10^9 bacterias por ml de sangre) [56]. La anemia se produce por la remoción de los eritrocitos infectados por las células fagocíticas del timo. La infección aguda en animales sanos generalmente se resuelve, pero no se alcanza a eliminar completamente la infección generando infecciones persistentes que pueden mantenerse a lo largo de la vida completa de los animales con rangos de bacteremia entre 10^2 a $> 10^7$ bacterias por ml de sangre [56]. Ambas fases, infección aguda y persistente, se caracterizan por una alta carga de bacterias. Como la bacteria no replica en células que expresen las moléculas MHC, se hipotetiza, y demostró en ensayos experimentales, que la inmunidad protectora se correlaciona con una respuesta de linfocitos T CD4+ y la producción de INF γ para la activación de los macrófagos y para colaborar con células B para potenciar la producción de anticuerpos IgG2 [55]. Durante el curso de la infección se identificaron altos títulos (3000-100.000) fluctuantes de anticuerpos IgG1 e IgG2 [56], principalmente dirigidos a los antígenos inmunodominantes de superficie de la familia de proteínas MSP (*Membrane surface proteins*). La función de los anticuerpos se postula relacionada con la neutralización de las bacterias extracelulares durante el proceso de invasión a eritrocitos, como también con el proceso de opsonización. Durante las infecciones persistentes, se postuló y describió un mecanismo muy relevante de recombinación y variación antigénica que permite evadir la respuesta inmune que consiste en la expresión de variantes de epitopes B y T de las proteínas de superficie MSP-2 y MSP-3 [21, 57-59]. Durante la infección si bien no se reportó que *A. marginale* sea capaz de suprimir la respuesta inmune, como si se describió para patógenos relacionados como *A. phagocytophilum*, algunos trabajos recientes describen cierta capacidad de regulación inducida por la infección de la respuesta de

células T CD4+ [60, 61], aunque los experimentos se realizaron con protocolos específicos de vacunación y desafío con proteínas de superficies MSP-2 y MSP1.

En una primera instancia, el foco para la identificación de proteínas inmunogénicas estuvo orientado a la identificación y caracterización estas pocas y bien definidas proteínas MSP. Se caracterizaron seis proteínas MSP en *A. marginale*, y los ensayos vacunales con estas proteínas purificadas confirieron diferentes grados de respuesta protectora [62, 63]. Sin embargo, aunque algunas actúan como proteínas inmunodominantes, confiriendo altas respuestas de anticuerpos IgG, no se identificó una asociación entre la magnitud de la respuesta de anticuerpos con un control total de la infección o de los niveles de bacteriemia [64] y se postuló que la respuesta estaría dirigida hacia proteínas de membrana conservadas sub-dominantes [65-67] hacia donde se han concentrado esfuerzos durante los últimos años.

Inmunizaciones con fracciones enriquecidas en membrana externa de *A. marginale* indujeron protección eficiente contra la bacteremia y la enfermedad después de los desafíos experimentales [68-70]. Por esto surgió la pregunta de qué proteínas o conjunto de proteínas de superficie son responsables o candidatos de conferir altos niveles de protección. En esto consistió la estrategia de los últimos 5 años para la identificación de nuevas proteínas que puedan caracterizarse como potenciales vacunas recombinantes, capaces de estimular respuestas y conservadas durante la infección como también en diferentes aislamientos geográficos.

En relación a las proteínas de superficie, estas pueden ser relevantes en la interacción con las células hospedadoras, como también pueden tener un rol relevante en conferir inmunidad. La secuenciación y anotación del genoma permitió confirmar esta familia de proteínas y ampliarla, prediciendo que de las 949 secuencias codificantes del genoma completo, 62 pertenecen a dos superfamilias de proteínas de superficie inmunodominantes, MSP-1 y MSP-2 [71]. El análisis genómico evidenció 163 secuencias codificantes (CDSs) con péptido señal, de las cuales 160 contienen dominios transmembrana; generando valiosa información para el estudio de nuevos antígenos de superficie, adhesinas o involucrados en sistemas de secreción de proteínas. A su vez, la anotación del genoma estableció funciones para un 59% de los CDSs permaneciendo un alto porcentaje de proteínas con función desconocida.

Algunos trabajos realizados que combinaron enfoques genómicos, proteómicos, inmunológicos y de espectrometría de masa identificaron inmunógenos de membrana externa. Extractos de membrana externa de la bacteria se separaron en geles de 2 dimensiones, y luego de la incubación con sueros de animales inmunizados con los mismos extractos de membrana externa se identificaron por espectrometría de masa las proteínas principalmente reconocidas. En teoría, este tipo de estudio permite

la identificación de los antígenos inmunodominantes como también los subdominantes [66]. En este estudio se identificaron 20 proteínas inmunogénicas, incluyendo algunos componentes del sistema de secreción de proteínas *typeIV secretion system* (T4SS; VirB9-1, VirB9-2 y VirB10) entre otras proteínas hipotéticas. Otros estudios han intentado describir la composición del proteoma de la superficie de la bacteria, y comparar los estadios intraeritrocíticos con el estadio dentro de la garrapata [68]. A su vez, otros autores se han concentrado en algunos candidatos como sistemas de secreción de proteínas o proteínas integrales de membrana. Investigaciones recientes, incorporan caracterizaciones *in silico* y experimental a través de ensayos de linfoproliferación, reconocimiento de sueros de animales infectados y han abierto un nuevo espectro de potenciales candidatos o combinación de antígenos vacunales a partir de estudios de identificación masiva [67, 72].

Estrategias experimentales realizadas anteriormente en nuestro laboratorio contribuyeron a profundizar este análisis; a través de la biblioteca de proteínas de fusión a fosfatasa alcalina se identificaron proteínas potencialmente localizadas en la membrana externa (datos no publicados). Por otra parte, utilizando la estrategia de hibridización sustractiva basada en la técnica de PCR (datos no publicados), se analizaron las diferencias genéticas entre *A. marginale* y *A. centrale*, pudiéndose identificar 7 secuencias específicas de *A. marginale*. Estas estrategias permitieron identificar candidatos para ser evaluados como potenciales antígenos sub-dominantes capaces de estimular una respuesta inmune para contribuir a la identificación del conjunto de proteínas de superficie de la bacteria a ser ensayados en nuevas formulaciones vacunales.

1.2. Resultados

El objetivo de este capítulo fue la selección, caracterización y estudio de la inmunogenicidad de nuevos antígenos, proteínas de membrana de *A. marginale* para contribuir a la identificación de proteínas que puedan actuar como antígenos subdominantes capaces de estimular una respuesta inmune y contribuir a la identificación del conjunto de proteínas de superficie de la bacteria a ser ensayados en nuevas formulaciones vacunales.

A través de análisis *in silico* para la identificación de motivos conservados, dominios, localización subcelular y del análisis de la distribución de los genes en organismos del orden Rickettsiales (mediante BLASTp) se seleccionaron tres genes candidatos: AM127 y AM1108 (obtenidos a partir de las estrategias experimentales) que presentan dos segmentos transmembrana y pertenecen a un clúster conservado entre patógenos intracelulares y endosimbiontes (**Figura 1.4**). Por otro lado a partir del análisis bioinformático se seleccionó el gen AM216 que posee un segmento transmembrana y pertenece a un clúster conservado únicamente entre *Anaplasma* spp. y *Ehrlichia* spp. (**Figura1.4**).

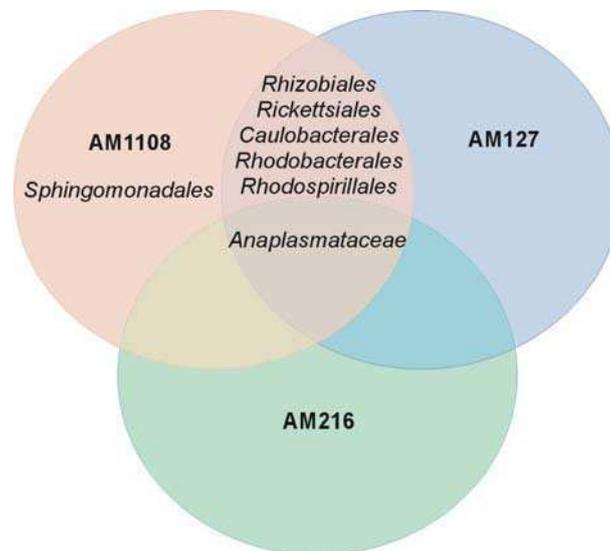


Figura 1.4. Ortólogos en la clase α -proteobacteria.

Distribución de los genes involucrados en el estudio en los diferentes órdenes de la clase α -proteobacteria. Los ortólogos se identificaron en los miembros de cada orden mediante BLASTp con criterios de e-value $< 10^{-5}$, % cobertura del 70%.

Las proteínas se analizaron con programas disponibles en servidores web, para predecir su estructura y localización subcelular. Si bien, la capacidad de

predicción es variable en función de los programas disponibles y sus métodos, y no representan análisis concluyentes en sí mismos, son una herramienta para aportar información para una descripción más completa de las proteínas en estudio, basándose en la secuencia de nucleótidos o proteínas.

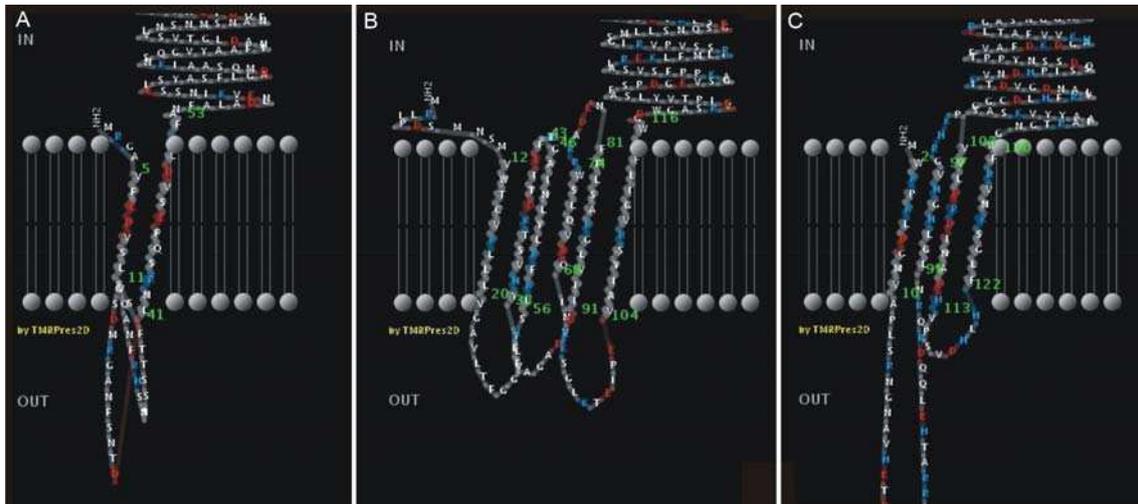


Figura 1.5. Predicción de la estructura 2-D.

Predicción de la estructura en 2-D dimensiones para a las proteínas de *A. marginale* **(A)** AM1108, **(B)** AM127 y **(C)** AM216, con respecto a la membrana externa utilizando el programa “*PRED-TMBB: graphical representations of the predicted topology*” [73]. Múltiples pasos transmembrana β se observaron en las tres proteínas. El valor obtenido para las tres secuencias aminoacídicas fue de 2.941, 2.955 y 2.959 respectivamente. Valores menores al valor umbral 2.965 indican la posibilidad de que las proteínas sean de localización en membrana externa.

La predicción de estructura en 2-D utilizando el programa PRED-TMBB caracterizó a las tres proteínas como integrales de membrana, con varios pasos transmembrana (TM), dominios intracelulares o periplásmicos, y dominios extracelulares (**Figura 1.5**).

Tabla 1.1. Análisis in silico de las proteínas candidatas seleccionadas.

ID	Anotación en el genoma	ORF predicho	ORF	Péptido señal	Segmentos TM		Localización subcelular	Dominios y anotación		
		Tamaño (pb)	Tamaño (aa)	SignalP	TM-PRED	TMBB	PSORT	Conserved domain (Pfam)	Blast2GO	Uniprot
AM1108	<i>Hypothetical protein</i>	2094	698	Si	2	2	<i>Non-Cytoplasmic</i>	<i>Pentapeptide repeat</i>	<i>Pentapeptide repeat</i>	<i>Pentapeptide repeat</i>
AM127	<i>Hypothetical protein</i>	3000	999	S	2	6	<i>OMP</i>	NA	<i>Acetamidase formamidase</i>	NA
AM216	<i>Hypothetical protein</i>	2530	842	No	1	4	<i>Unknown</i>	NA	NA	NA

Tabla 1.1. Predicción *in silico* de las proteínas candidatas.

Síntesis de la información disponible de análisis predictivos para diferentes características predichas de acuerdo a la secuencias para las proteínas utilizadas en el estudio. TM: Transmembrana; NA: Sin anotación; OMP: proteína de membrana externa (*outer membrane protein*).

Los programas de predicción de segmentos transmembrana detectaron diversos pasos para las tres proteínas analizadas. Utilizando la anotación del BLAST2GO, la proteína AM1108 presento una anotación con un dominio conservado de *pentapeptide repeats* y la proteína AM127 de *acetamidase formamidase*, mientras que no se identificaron anotaciones para ortólogos de la proteína AM216 siendo específica de la familia *Anaplasmataceae* (Tabla 1.1).

Para estudiar la expresión de los genes involucrados, a través de RT-PCR se verificó la transcripción de los tres genes en *A. marginale* S1A aislada de una infección natural (**Figura 1.6**).

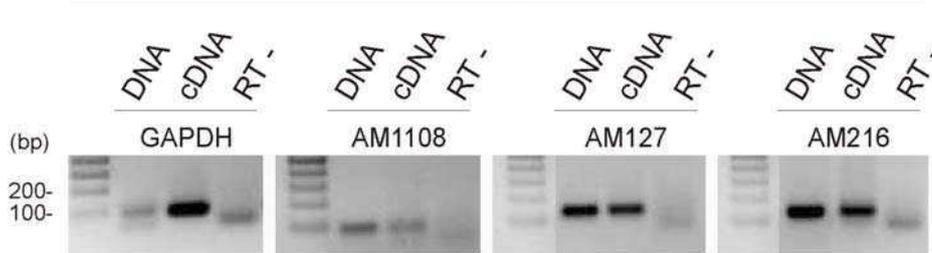


Figura 1.6. Transcripción de los genes candidatos.

Para confirmar la expresión de los genes AM1108, AM127 y AM216 en *A. marginale* se realizaron ensayos de RT-PCR. Como control positivo se realizó la amplificación de GAPDH y la amplificación a partir de ADN total; y como control negativo la amplificación sin agregado de la enzima RT. Los productos de amplificación por PCR utilizando primers específicos (**Tabla S 1.1, Anexo**) a partir del cDNA se visualizaron en geles de agarosa 2% teñidos con bromuro de etidio. Los resultados demostraron la transcripción de los 3 genes en estudio.

Posteriormente, los genes seleccionados fueron amplificados por PCR con primers diseñados a partir del ADN de *A. marginale* S1A (**Tabla S 1.1, Anexo**) y clonados para su expresión como proteínas recombinantes de fusión a *tag*-histidina para corroborar la expresión y permitir la purificación. Las proteínas fueron divididas en dos segmentos (N-terminal y C-terminal), excluyendo la secuencia del péptido señal potencialmente tóxicos para *E. coli*, para evitar mutaciones durante la amplificación por PCR de fragmentos largos, para optimizar la expresión y disminuir la degradación de la proteína (**Figura 1.12, materiales y métodos**). Se verificó la purificación e integridad de acuerdo al peso molecular esperado a través de electroforesis en geles de poliacrilamida (SDS-PAGE) teñidos con *Coomassie blue* y *Western blot* con anticuerpos anti-histidina (**Figura 1.7**).

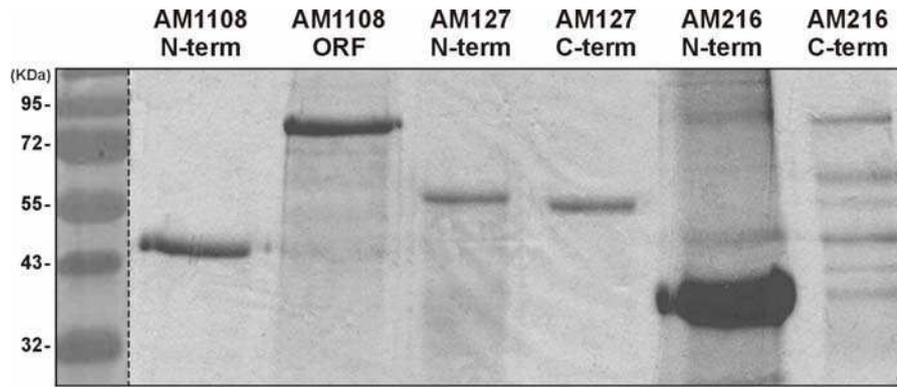


Figura 1.7. Western blot de proteínas recombinantes.

Se verificó la integridad y purificación de las proteínas recombinantes. SDS-PAGE Poli(acrilamida Tris/glicina 10% (30% acrilamida/bisacrilamida, 1,5M Tris [pH 8,8], 10% SDS, 10% APS, TEMED, H₂O) teñido con *coomasie blue* para visualizar las proteínas recombinantes purificadas.

Para evaluar la inmunogenicidad de los potenciales antígenos se utilizaron tres animales naturalmente infectados y un animal control (*naive*, no infectado). Los animales infectados provenían de una infección natural y se recolectó sangre en la estación experimental del EEA-INTA Mercedes (Corrientes, Argentina). Se corroboró que los animales utilizados (Nº 99, 282 y 285) estén infectados ya sea en infecciones agudas o persistentes con *A. marginale*, a través de PCR del gen *msp5* (*membrane surface protein-5*) (**Figura 1.8**).

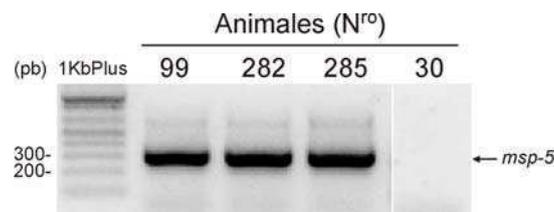


Figura 1.8. PCR diagnóstica de infección.

Los animales utilizados en el estudio (infectados y control) se analizaron por PCR para confirmar la infección por *A. marginale*. A partir de sangre infectada se realizó una PCR para amplificar el gen *msp-5* según protocolos establecidos. Los resultados de la PCR se visualizaron en geles de agarosa 1,5% teñidos con bromuro de etidio.

Se realizaron ensayos de linfoproliferación de células mononucleares periféricas (PBMC) que se incubaron con las proteínas recombinantes (AM1108-N-term, AM1108 ORF, AM127 N-term, AM127 C-term, AM216 N-term y AM216 C-term) para estudiar la respuesta de linfoproliferación específica.

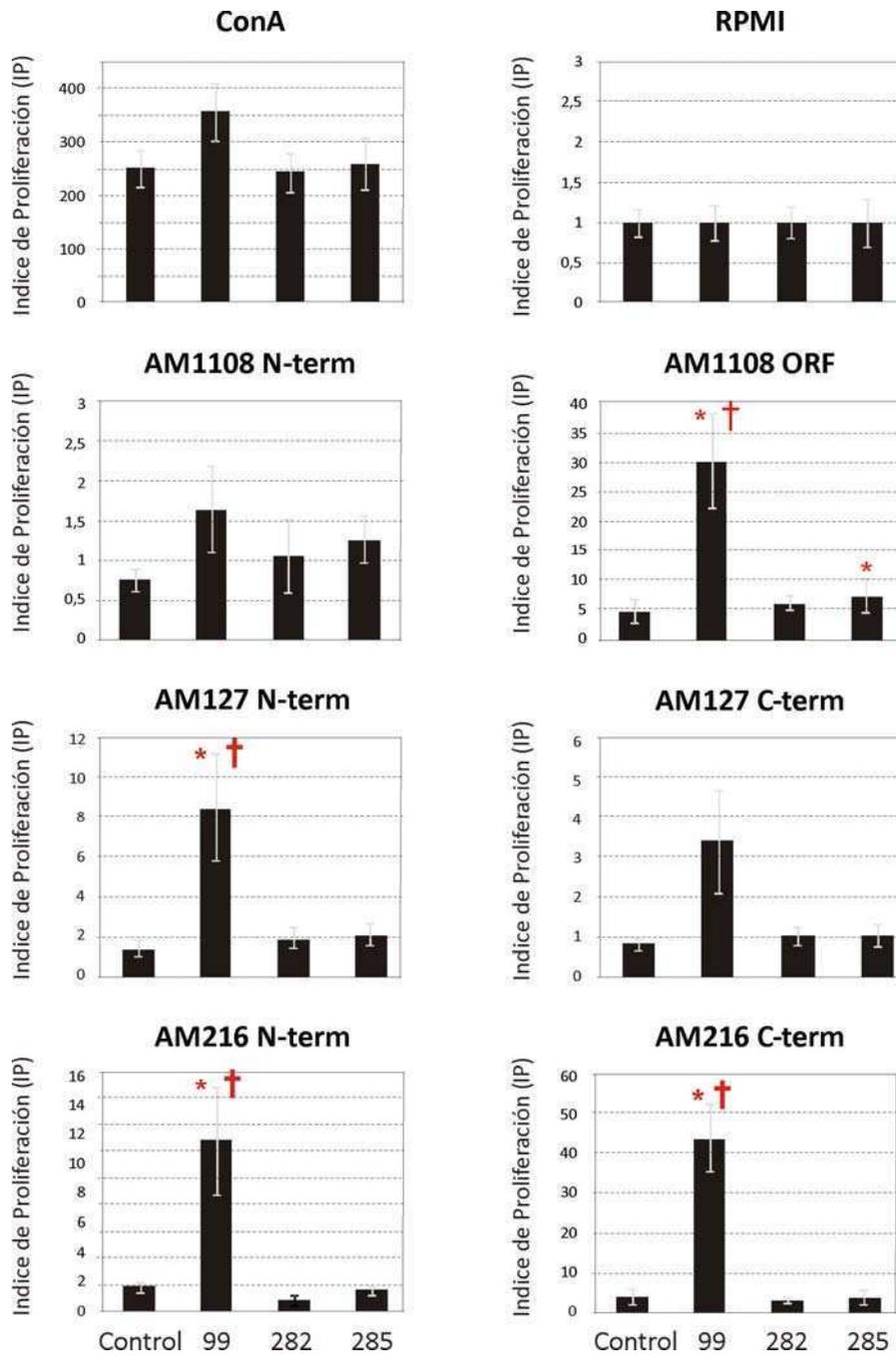


Figura 1.9. Ensayos de linfoproliferación.

Linfoproliferación de células PBMC incubados durante 4 días con las proteínas recombinantes. Los resultados se presentan como índice de proliferación (IP) calculado como el valor de cpm medio de PBMC incubadas con proteínas recombinantes (1ng/ul)/ valor cpm medio de PBMC incubadas con RPMI completo (Control). Resultados son representativos de

tres experimentos independientes (ANOVA; † $p < 0.001$ significativo respecto al animal control, * $p < 0.05$ significativo respecto al medio RPMI. Los ensayos se repitieron con incubaciones de 6 días con muy similares resultados.

Mediante la incubación con el estimulador inespecífico (Concanavalina A) se verificó la integridad y viabilidad de las células PBMC purificadas observando altos índices de proliferación (**Figura 1.9**).

Se comprobó una estimulación significativa de proliferación cuando los PBMC provenientes del animal infectado N° 99 fueron estimulados con los antígenos AM1108 ORF y AM216 C-term; y se observó proliferación específica, aunque de menor intensidad, para los antígenos AM216 N-term y AM127 N-term. PBMC del animal N° 285 respondieron con el antígeno AM1108 ORF con un índice de proliferación significativamente mayor que el valor umbral (Índice Proliferación > 3) (**Figura 1.9, Tabla 1.2**). Los segmentos de las proteínas AM1108 N-term y AM127 C-term, no estimularon proliferación de PBMC.

En resumen en este estudio las proteínas AM1108 ORF y AM216 C-term estimularon respuestas fuertes y significativas de proliferación celular de linfocitos; mientras que las proteínas AM216 N-term y AM127 N-term también estimularon respuestas significativas aunque moderadas.

Tabla 1.2. Respuesta específicas de la linfoproliferación de las células PBMC a los diferentes antígenos recombinantes de *A. marginale*.

Animales	Índice de proliferación (IP)							
	ConA	RPMI	AM1108 N-term	AM1108 ORF	AM127 C-term	AM127 N-term	AM216 C-term	AM216 N-term
Naive	250,85 ±34,71 *	1 ±0,17	0,74 ±0,13	4,61 ±1,97	0,79 ±0,15	1,29 ±0,39	3,6 ±2,08	1,58 ±0,35
99	355,12 ±54,14 *	1 ± 0,219	1,65 ±0,55	30,38 ±7,95 a * †	3,42 ±1,32	8,53 ±2,73 a * †	44,05 ±8,57 a * †	11,53 ±3,64 a * †
282	243,78 ±36,99 *	1 ± 0,19	1,05 ±0,47	6,10 ±1,22 a	0,99 ±0,23	1,82 ±0,53	2,66 ±0,84	0,67 ±0,37
285	259,25 ±47,49 *	1 ± 0,30	1,26 ±0,30	7,31 ±2,87 a *	1,01 ±0,27	2 ±0,54	3,28 ±1,81	1,36 ±0,33

± Error estándar (ES)

Los datos se presentan como el Índice de Proliferación (IP). Representa el promedio de 3 resultados de experimentos independientes. IP = Promedio (cpm) de cultivos de PBMC incubados con proteínas recombinantes / Promedio (cpm) de PBMC incubados con medio RPMI. **IP >3** se consideran respuestas proliferativas.

Diferencias significativas [p < 0.001] de las respuestas IP: **IP > 3 (ª)**, significativo respecto al medio RPMI (*****) significativo respecto al animal control (**†**).

Para evaluar la respuesta serológica de diferentes sueros se realizaron ensayos de *western blot*. Se analizó la presencia de anticuerpos específicos frente a las proteínas recombinantes AM1108 ORF y AM216 C-term que estimularon los mayores índices de linfoproliferación. Los sueros fueron preadsorbidos con extractos de *E. coli* para disminuir las reacciones inespecíficas.

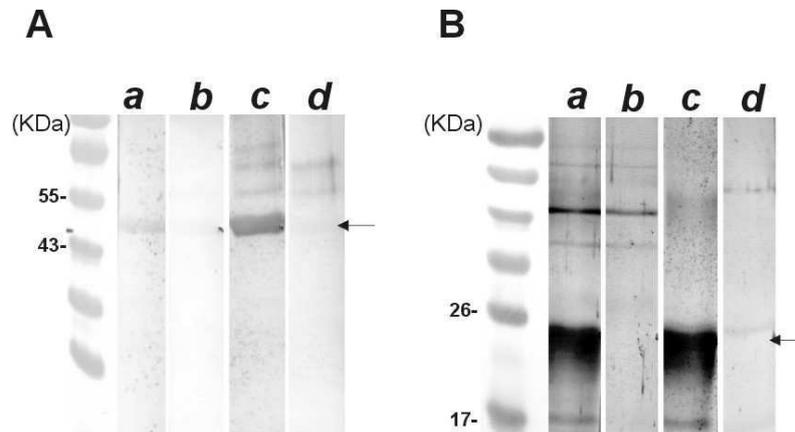


Figura 1.10. Western blot para las proteínas recombinantes.

Se realizaron ensayos de *Western blot* con diferentes sueros experimentales y control para las proteínas AM1108 ORF (A) y MSP-5 (B) como control positivo. Los sueros utilizados fueron: (a) suero de animal inmunizado con fracciones de membrana externa de *A. marginale*; (b) suero de animal previamente a la inmunización con fracciones de membrana externa de *A. marginale* (Provistos por Dr. G. Palmer, USA); (c) suero de animal infectado con *A. marginale* N° 640 infección experimental; (d) suero pre-inmune de animal N° 640 (extraído previamente a la infección). Los western blot se revelaron con un anticuerpo secundario anti-bovino conjugado con fosfatasa alcalina y según se describe en la sección de materiales y métodos.

En los ensayos de *western blot* se observó reactividad específica frente al antígeno AM1108 ORF con sueros inmunizado con fracciones de membrana externa de *A. marginale* como también del suero del animal infectado experimentalmente (Figura 1.10A; calle a, c), indicando la presencia de anticuerpos específicos. No se detectó reactividad de los sueros preinmunes frente a los antígenos analizados (Figura 1.10A, calle b, d). Resultados similares se obtuvieron para el control (Figura 1.10B). La reactividad de los sueros del animal inmunizado con extractos de membrana externa sugiere la localización en membrana externa de la proteína AM1108, coincidiendo con las predicciones *in silico*. No observamos un reconocimiento específico para el antígeno AM216 C-term (resultados no mostrados) con los sueros analizados.

Por último, se evaluó para células PBMC incubadas con los antígenos AM1108 ORF y AM216 C-term, el perfil de citoquinas que estimulaban mediante RT-qPCR.

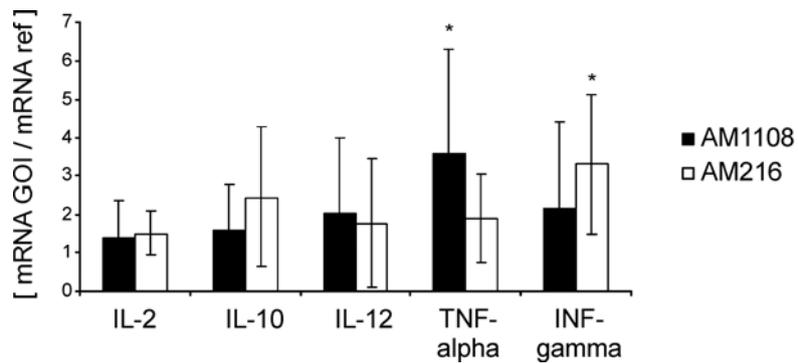


Figura 1.11. Perfil de estimulación de citoquinas.

Estimulación de citoquinas por los antígenos recombinantes AM1108 ORF y AM216. Se grafican resultados de tres experimentos independientes (* $p < 0.05$). Detalle de los genes de referencia en materiales y métodos.

Los ensayos indicaron que TNF- α e INF- γ se encontraron significativamente elevados para los antígenos AM1108 y AM216 C-terminal, respectivamente. Para el animal control no infectado, no se observaron resultados significativos cuando fueron estimulados con los antígenos recombinantes (**Figura 1.11**).

1.3. Discusión

En este capítulo se estudió *in silico* y mediante ensayos de inmunogenicidad de la capacidad para estimular una respuesta inmunológica por parte de proteínas seleccionados como antígenos candidatos.

Diversos trabajos experimentales acoplados a predicciones bioinformáticas, se realizaron en organismos de las familias *Anaplasmataceae* y *Rickettsiae*, para estudiar el proteoma de las superficies de las bacterias utilizando ensayos de biotinilado (sulfo-NHS-SS-biotina) o cross-linking de proteínas de la superficie externa de las bacterias, estudios proteómicos de extractos de membrana externa y posterior identificación por MALDI-TOFF, con el objetivo de contribuir a la descripción de los proteomas de superficie [66, 68, 74, 75]. En un estudio realizado en *A. marginale* que combinó la separación de proteínas provenientes de fracciones de membrana externa en geles de dos dimensiones (2D) e identificación utilizando los sueros de los animales inmunizados con las fracciones de membrana externa, se identificaron una serie de proteínas candidatas con capacidad protectora. De las 24 proteínas identificadas, MSP-2, MSP-3 y MSP-5 fueron ampliamente descritas como inmunogénicas, aunque como se describió anteriormente, están relacionadas con mecanismos de variación antigénica reduciendo su capacidad como potenciales vacunas. Otras 15 proteínas fueron identificadas, entre las que se encuentra una de las proteínas utilizadas en este estudio, AM127, y fueron reconocidas por los sueros de los tres animales infectados [66]. A su vez, la proteína AM127 fue predicha por diferentes programas de predicción como una proteína membrana externa y con ortólogos presentes en organismos de la familia *Anaplasmataceae*. En este estudio, la proteína AM127 fue identificada en diferentes órdenes de la clase con la excepción del orden Sphingomonadales. También fue predicha como potencial antígeno de membrana confirmando los resultados e identificación experimental previa. Se confirmó además la expresión del gen, al menos en el estadio intraeritrocítico. La proteína carece de información asociada en la anotación en el genoma (proteína hipotética) [71] aunque la anotación del Blast2GO realizada en este trabajo identificó una anotación asociada como *Acetamidase formamidase*, al igual que el ortólogo identificado en *A. phagocytophilum* (APH_0110). Los ensayos de estimulación de células PBMC dieron resultados significativos para el segmento AM127 N-term y con valores moderados, aunque mayores que el valor umbral.

La proteína AM216 que pertenece a un clúster conservado únicamente en *Anaplasma* spp. y *Ehrlichia* spp. no tiene ortólogos en otras familias de la clase. Tampoco fue identificada por otros estudios ni reviste dominios o estructuras conservadas. Sin embargo, los programas de predicción de localización subcelular la describen como potencial proteína de membrana externa, aunque no se evidenció reconocimiento específico de los sueros de los animales infectados o inmunizados con extractos de membrana externa. Los resultados de la estimulación de PBMC dieron resultados significativos, y también demostró estimular la expresión de IFN-gamma, siendo esta una de las citoquinas descritas como relevante en el curso de la infección.

En relación a la proteína AM1108 se identificaron ortólogos en diversos organismos del orden Rickettsiales. En diversos estudios realizados con organismos del orden Rickettsiales, *E. chaffeensis*, *A. phagocytophilum* y en *N. sennetsu*, se identificaron un conjunto de proteínas expuestas en la superficie entre las que se identificó el ortólogo de la proteína AM1108 anotada como proteína hipotética con un dominio de *pentapeptide repeat domain protein* en *A. marginale* (NSE_0725 en *N. sennetsu*; APH_1170 en *A. phagocytophilum*; ECH_1083 en *E. chaffeensis*). AM1108 pertenece a un clúster conservado en la clase, y con predicción de localización en membrana, aparte de los sucesivos trabajos experimentales donde se demostró en organismos relacionados su localización en la superficie. Para esta proteína se obtuvieron resultados positivos en la estimulación de PBMC, en el perfil de citoquinas estimuló la producción de TNF-alpha. A su vez, la proteína recombinante fue reconocida por los sueros de animales infectados con extractos de membrana externa como una nueva evidencia de su localización de superficie.

Este trabajo permitió la identificación de dos proteínas de AM1108 como AM216 con capacidad antigénica, que fueron capaces de estimular una respuesta inmunológica específica. Estudios futuros permitirán corroborar la potencialidad como antígenos eficientes en nuevas formulaciones vacunales.

Capítulo II.

El sistema Tat en *A. marginale* y en la clase α -proteobacteria

Objetivo específico.

Identificar y caracterizar el sistema de secreción de proteínas “Twin arginine tranlocation pathway (Tat)” con el objetivo de inferir la función y el rol en la capacidad de adaptación y patogénesis.

2.1. Antecedentes

2.1.1 Sistemas de secreción de proteínas.

2.1.2. El sistema de secreción “Twin Arginine Translocation Pathway (Tat)”

2.2. Resultados

2.3. Discusión

2.1. Antecedentes

2.1.1. Sistemas de secreción de proteínas.

En bacterias, los sistemas de secreción y translocación de proteínas adquieren gran relevancia en los mecanismos de patogénesis e interacción con las células hospedadoras, ya que permiten la secreción y exportación de proteínas bacterianas que pueden actuar como factores de virulencia o localizarse en la membrana externa.

La translocación de proteínas a través de la membrana citoplasmática y la inserción en membrana se realiza por dos mecanismos principalmente: el sistema **Sec**, que es la principal ruta de exportación de proteínas desplegadas [76], y el sistema *twin-arginine translocation pathway* (**Tat**), responsable de exportar proteínas plegadas [77].

La diversidad genómica y biológica de la subdivisión α -proteobacteria se ve representada en la multiplicidad de estrategias que establecen en la asociación con múltiples hospedadores como se describió anteriormente. Muchas bacterias simbiotes o patógenos de plantas y animales utilizan los sistemas de secreción especializados como mecanismos moleculares específicos para establecer interacciones con células hospedadoras [1].

La ubicuidad de los sistemas de secreción correlaciona con la alta variabilidad de la composición y organización de los genomas, lo que podría implicar un compromiso para la funcionalidad de los complejos proteicos. Los genomas reducidos del orden Rickettsiales usualmente muestran la ausencia de genes ortólogos u organizaciones anómalas de los genes en los clústers involucrados en el mismo proceso metabólico o en los sistemas complejos de secreción de proteínas. En estos casos, la conservación de la funcionalidad no es evidente, o bien es posible que se vea afectada y es necesario estudiarla experimentalmente [78-81].

El sistema Tat permite la traslocación de proteínas a través de la membrana interna en bacterias gram-negativas. Las proteínas integrales de membrana, TatA, TatB y TatC son los componentes principales de este sistema [77]. Mientras que los componentes fueron identificados en muchos organismos de la clase utilizando los genomas disponibles, no se avanzó demasiado en el conocimiento del funcionamiento, rol e importancia del sistema Tat en la adaptación y virulencia en diversos grupos de bacterias.

Como se describió previamente en la Capítulo I, en nuestro laboratorio una biblioteca de proteínas de fusión a fosfatasa alcalina permitió la identificación de

secuencias codificantes para proteínas de exportación o de membrana. Entre los genes identificados, algunos de ellos codificaban para proteínas de membrana externa, antígenos de superficie con dominios conservados de adhesinas o para las proteínas que constituyen los componentes de los sistemas de secreción *secretory pathway type I* y *twin arginine traslocation pathway* (Tat). Si bien la anotación del genoma de *A. marginale* permitió corroborar la integridad del sistema de secreción *Sec, type I* y *type IV*, esto no se verificaba respecto del sistema Tat en *A. marginale* [71] donde se describió como ausente o incompleto. En tal sentido, los datos experimentales y la posterior búsqueda en el genoma permitieron corregir la mencionada omisión, siendo posible identificar en el genoma los tres ORFs que codifican para las proteínas que conforman el complejo proteico del sistema Tat. *TatA* al ser un pequeño ORF resultó omitida en la anotación del genoma, posiblemente por ser un ORFs de tamaño menor a 200 pares de bases. A su vez, el gen que codifica para la proteína TatB se encontraba anotado como proteína hipotética.

2.1.2. El sistema de secreción “Twin Arginine Translocation Pathway (Tat)”

El sistema de secreción de proteínas Tat se encuentra presente en la gran mayoría de las bacterias donde fue estudiado, algunas arqueas y en la membrana tilacoidal de plástidos en plantas [82]. Tres proteínas denominadas TatA, TatB y TatC componen el complejo proteico; sin embargo la estructura proteica y organización en diferentes organismos, es diversa.

Las bacterias gram-negativas usualmente presentan los tres componentes [83-85] formando un complejo proteico heteromultimérico localizado en la membrana interna [77]. En cambio, TatB se encuentra ausente en la mayoría de las especies gram-positivas y arqueas, formando un complejo mínimo compuesto por TatAC.

Los genes *tat* usualmente se encuentran organizados en un operón (*tatABC*) en la mayoría de los organismos estudiados, mientras que solo en algunos casos se describieron organismos que tenían los genes como unidades transcripcionales simples [86]. La estequiometría de la transcripción de las subunidades es crítica para la actividad de complejo Tat [87]. La proteína TatA es la más abundante, presente en aproximadamente 20-23 veces más que las proteínas TatB y TatC que se encuentran representadas en cantidades equimolares en el complejo proteico [77]; requiriendo mayores niveles de transcripción que los otros genes involucrados [88].

Las proteínas TatA- y TatB-*type*, se encuentran altamente relacionadas con un posible ancestro común [77]. Ambas están compuestas por un dominio transmembrana α -helice N-terminal, conectados a través de un dominio “bisagra” a un dominio anfipático, y seguido por un dominio C-terminal de largo variable [77, 89]. En organismos que carecen de TatB, se describió que TatA es capaz de retener ambas funciones biológicas [90-93]. En efecto se demostró a través de experimentos *in vivo* que TatA de *E. coli* es capaz de prescindir de los requerimientos de TatB para realizar un transporte eficiente [94]. La proteína TatC es la más conservada de las proteínas del complejo, donde el grado de conservación es mayor en los 6 pasos transmembrana [77, 83, 95].

Los péptidos señales de las proteínas exportadas por esta vía mantienen una estructura común en relación a las proteínas secretadas vía *Sec*, pero fundamentalmente se caracterizan por un dominio de dímeros del aminoácido Arginina (RR) en la región N-terminal, seguido por una región hidrofóbica (h-región) y una región cargada positivamente antes del sitio de clivaje, que sería la responsable de evitar el transporte vía *Sec* [77]. Estudios recientes demostraron que variantes al característico dímero de RR, como motivos R-N-R, o las variantes KR, RK y KK en la región N-terminal preservan la capacidad de ser exportados vía Tat [96-98]. La función principal descrita para TatC es el reconocimiento del péptido señal y sus principales propiedades [99].

Aunque se demostró que el sistema Tat esencial para la virulencia y simbiosis en diversas bacterias que interactúan con plantas y animales [100-112], pocos estudios se realizaron para comprender el rol del mecanismo en la clase α -proteobacteria [100, 113, 114].

El objetivo de este capítulo de la tesis fue explorar el rol del sistema Tat en la clase, estudiando la funcionalidad en relación a diferentes estructuras de organización de los genes. Particularmente nos focalizamos en *Anaplasma marginale* y *Brucella abortus*, patógenos ampliamente conocidos de la clase, y representantes de diferentes estilos de vida: intracelular facultativo y obligado, respectivamente.

2.2. Resultados.

El objetivo fue estudiar la presencia, función y utilización del sistema en *A. marginale*, contrastar las características con respecto a las bacterias de la clase α -proteobacteria para comprender la relevancia del sistema en la biología bacteriana y posibles implicancias durante la infección. Considerando organismos con una estructura diferente para el sistema, también nos propusimos analizar los mismos aspectos en *B. abortus* para comparar ambas organizaciones y contribuir con la caracterización de este sistema en ambos patógenos. Para esto nos propusimos: *i)* identificar los componentes del sistema y la organización en el genoma en la clase; *ii)* estudiar la funcionalidad en organismos modelos; *iii)* identificar y caracterizar los posibles sustratos del sistema en los organismos estudiados y grupo filogenético relacionado.

Con el objetivo de explorar la arquitectura de los genes *tat* en la clase α -proteobacteria, realizamos un estudio comparativo seleccionando 53 genomas y representando los diferentes órdenes de la clase. Confirmamos la presencia de los tres genes ortólogos *tatA*, *tatB* y *tatC* en 42 de los 53 organismos estudiados (**Figura 2.1A**); excepto para *A. marginale str. St. Maries* y *Ehrlichia ruminantium str. Gardel* que no presentaban el gen *tatA* en la secuenciación y anotación del genoma. A partir del resultado previo utilizando la biblioteca genómica de fosfatasa alcalina se identificó un marco de lectura abierto (ORF) de 171 nucleótidos, omitido durante la secuenciación y anotación del genomas completo de *A. marginale str. St. Maries* [71], y flanqueado por 2 loci, AM392 y AM394, (localizado 21 nucleótidos hacia el extremo 5' de una proteína hipotética y 18 nucleótidos hacia el extremo 3' de la proteína recR). La traducción del ORF predicho fue una proteína de 53 aminoácidos con una alta conservación con el gen identificado como *tatA* (**Figura 2.2**). La secuencia para el gen *tatA* de *A. marginale str. Salta* se incorporó a la base de datos del GenBank bajo el número de acceso númeroGI JQ409478 (100% identidad de secuencia con el gen *tatA* de *A. marginale str. St. Maries*). Este ORF identificado como el gen *tatA* fue posiblemente omitido durante la anotación del genoma debido a su pequeño tamaño, probablemente al igual que ocurrió en *E. ruminantium str. Gardel*. A través de Tblastn confirmamos que las 11 especies restantes analizadas de los géneros *Rickettsia*, *Neorickettsia*, *Orientia* y *Wolbachia* carecen del gen *tatB*. La estructura de los genes *tatABC* mostró una gran diversidad en la clase; los genomas analizados se organizaron en 3 grupos de acuerdo a la conservación de la organización en operon y aspectos generales de la arquitectura de los genes del complejo (**Figura 2.1A**). El

primer grupo comprende los genomas de los ordenes Rhizobiales (incluyendo *B. abortus*), Caulobacterales, Rhodobacterales, Sphingomonadales y dos especies del orden Rhodospirillales, *Rhodospirillum rubrum* ATCC 11170 y *Magnetospirillum magneticum* AMB-1. Estos organismos presentaron la organización característica con los tres genes en organización de operon descrita para la mayoría de los organismos utilizados. El segundo grupo presentó una estructura parcialmente dispersa donde el locus *tatA* se ubica por fuera del operon *tatBC*, aunque en la misma hebra. Este grupo se conformó por dos organismos, *Gluconobacter oxydans* y *Acidiphilium cryptum* JF-5, ambos de la familia Acetobacteraceae. El otro grupo incluyó los géneros incluidos en el orden Rickettsiales: *Anaplasma*, *Ehrlichia*, *Neorickettsia* (familia Anaplasmataceae), *Wolbachia* y *Rickettsia* (familia Rickettsiaceae), que presentaron una organización completamente fragmentada para los genes *tat* (**Figura 2.1A**, **Figura 2.1B**). Los organismos que carecen de *tatB*, con la excepción de *Neorickettsia* spp. también codifican para *tatA* y *tatC* en diferentes hebras del genoma.

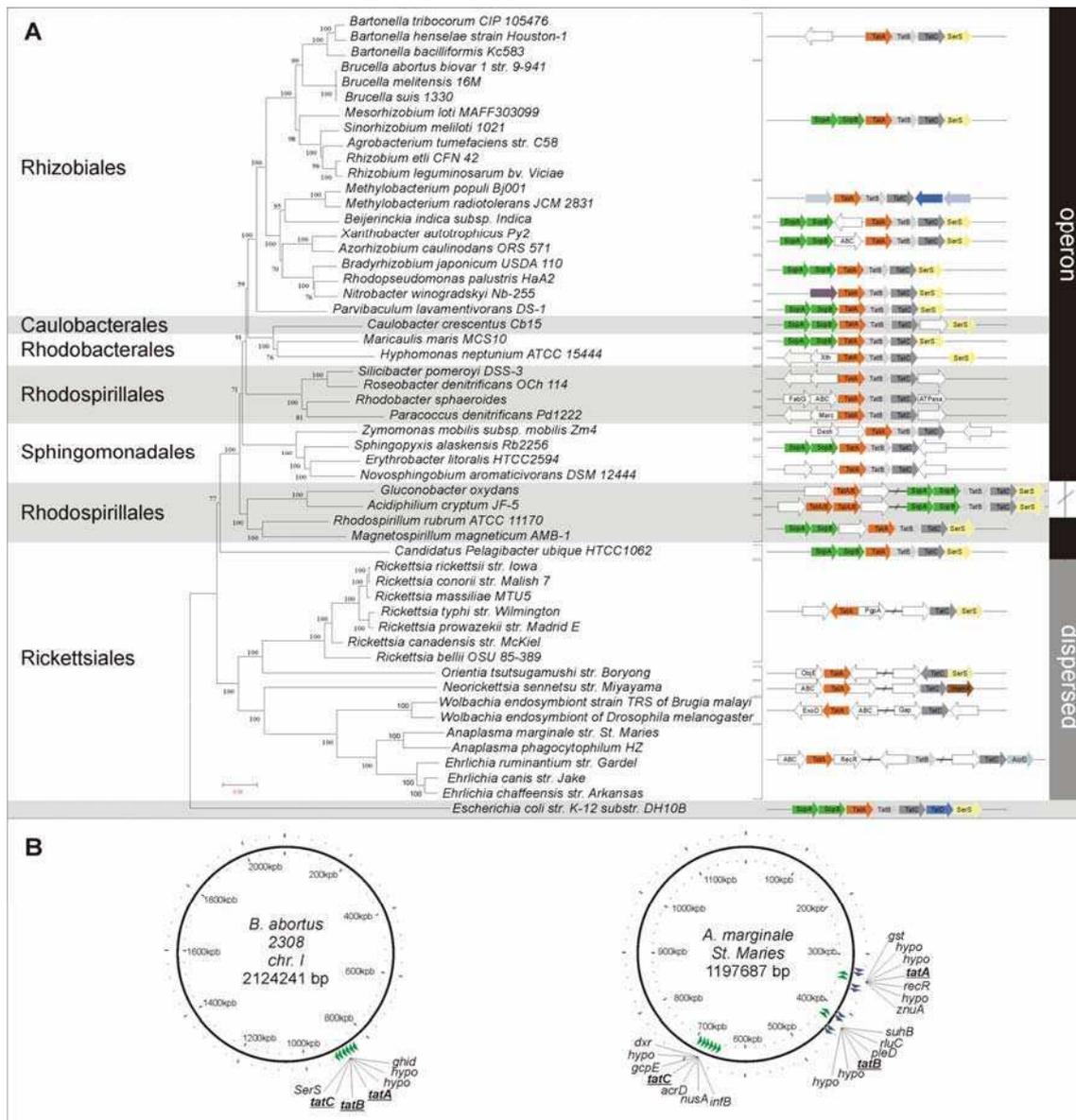


Figura 2.1. Filogenia y organización de los genes *tatABC* en las α -proteobacteria.

(A) Panel izquierdo: Análisis filogenético utilizando Neighbor-joining, (NJ-JTT, 1000 bootstrap, MEGA4) para los 53 organismos analizados basados en el alineamiento de 11 proteínas concatenadas (RNA Pol $\beta\gamma\beta'$, alanyl-tRNA synthetase, phenylalanyl-tRNA synthetase, arginyl-tRNA synthetase, EF-Tu, EF-G, RecA, GyrA, Gyr β , Hsp70). γ -proteobacteria *E. coli* K12 fue utilizado como grupo externo de referencia para la inferencia filogenética. **Panel Derecho:** organización genómica de los genes *tatA*, *B* y *C* en los organismos analizados. Los corchetes indican los organismos que comparten la organización de los genes. **(B)** Organización de los genes *tatA*, *tatB* y *tatC* en el genoma de *A. marginale* str. St. Maries y *B. abortus* 2308 (Cromosoma I) utilizando el programa Circular genome View (CGView).

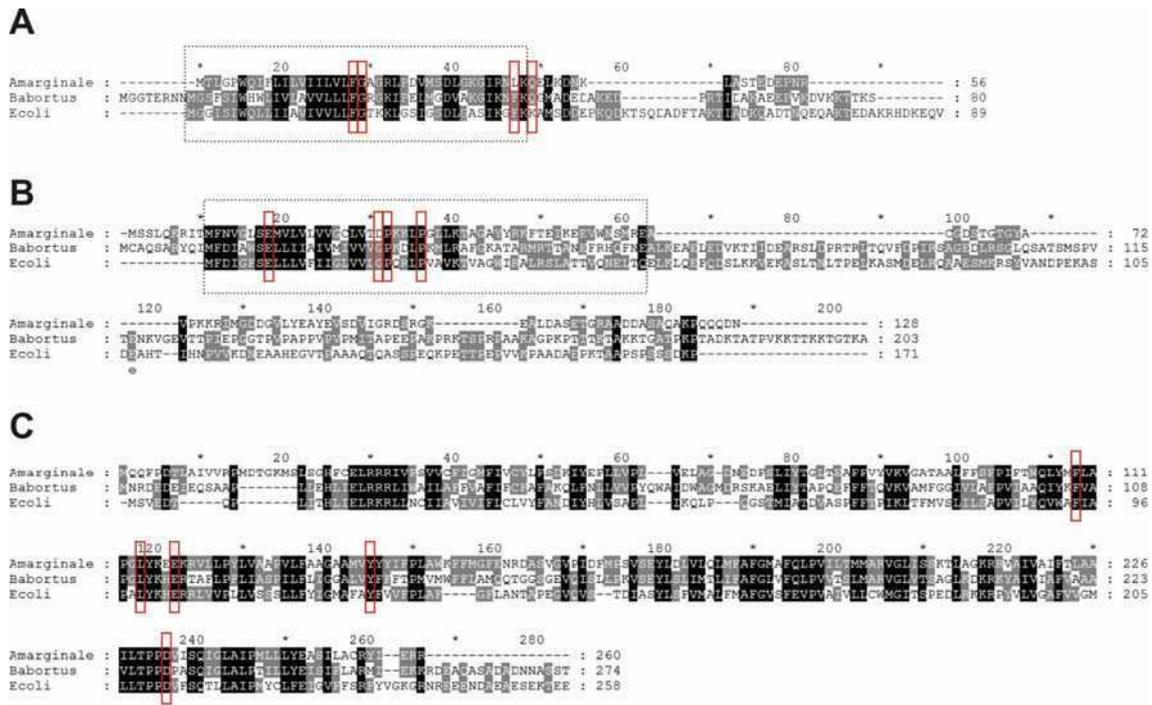


Figura 2.2. Alineamiento de las secuencias de proteínas TatA, B y C.
 Alineamiento múltiple de la secuencia de aminoácidos de (A) TatA, (B) TatB y (C) TatC para *A. marginale* str. *St. Maries*, *B. abortus* 2308 y *E. coli* K12 utilizando el CLUSTAL W [115].

Las principales diferencias entre las proteínas TatA y TatB están asociadas con el largo global de la proteína, de la región bisagra y la conservación de la helice anfipática. Ambos genes *tatA* y *tatB* están truncados en el orden Rickettsiales generando proteínas más cortas que las presentes en la mayoría de los organismos y en *E. coli*, organismo donde la funcionalidad y mecanismo del sistema fue más estudiado. Sin embargo, aunque las proteínas sean más cortas, en los alineamientos se observa que conservan los dominios funcionales descritos como los primeros 40 aminoácidos, y los principales residuos que se consideran críticos para la funcionalidad estudiados mediante experimentos de mutagénesis dirigida.

Para TatA, los residuos F20 y G21, descritos como parte de región bisagra, se encuentran estrictamente conservados. Sin embargo, la F39 postulada esencial para la actividad, se encuentra reemplazada por otro residuo no polar, leucina, en *A. marginale*. La K41 en *E. coli* esta reemplazada en ambos organismos por una glutamina.

La secuencia global de TatB parece conservar los dominios esenciales aunque existan importantes sustituciones aminoacídicas. La región transmembrana se encuentra altamente conservada; incluyendo el invariante Acido glutámico (E8) del

dominio, para el que se propuso un rol en la traslocación de los protones como fuente de energía para la traslocación de las proteínas [85]. Ambas Prolinas propuestas como parte de la región bisagra están conservadas [116]. La Glycina21, que fue descrita como necesaria para mantener una translocación eficiente [116], está reemplazada por una Isoleucina in *A. marginale*. La región anfipática está menos conservada que el extremo N-terminal, pero el patrón de los tres Ácidos Glutámicos parece conservarse.

Para el caso de TatC, el alineamiento múltiple indica una conservación de los seis pasos transmembrana indicando una conservación global de la estructura. Los residuos F94, E103 y D211 descritos como determinantes para la funcionalidad, están conservados [117] y los residuos L99A y Y126 considerados importantes pero no esenciales para la actividad de TatC, también se conservan [117].

Posteriormente, bajo la hipótesis de que la pérdida de estructura de operon pudiese modificar la regulación de la transcripción de manera que afecte la funcionalidad, estudiamos en primer lugar, la transcripción de los genes *tat* con diferentes organizaciones. Utilizando RT-PCR en *A. marginale* (dispersa) y, comparativamente, en *B. abortus* (operon), demostramos la transcripción de los genes en ambos organismos (**Figura 2.3A, Figura 2.3B**). Para *B. abortus*, utilizando primers que reconocen la región 3' de uno de los genes y el 5' del gen contiguo (**Figura 2.3D**), utilizando RT-PCR se amplificó la secuencia esperada a partir del ADNc, demostrando que ambos pares de genes se encuentran en la misma molécula de ARNm. Se confirmó la presencia de un mensajero policistrónico para los genes *tatABC* en *B. abortus* que excluye al gen *serS* (**Figura 2.3C, Figura 2.3D**).

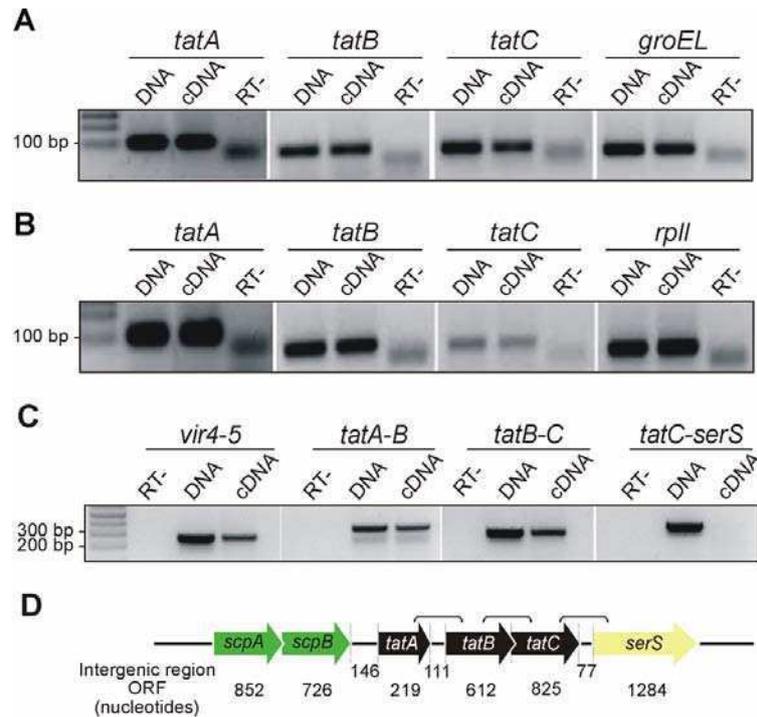


Figura 2.3. Transcripción de los genes *tatA*, *tatB* y *tatC*.

(A) RT-PCR para *A. marginale* (B) RT-PCR para *B. abortus*. Control positivo: amplificación de DNA; Control negativo: amplificación sin agregado de la enzima RT. (C) Mensajero policistrónico para *B. abortus*; el operon *Vir4-5* se utilizó como control positivo. (D) Esquema de la organización de los genes *tatABC* y regiones intergénicas donde se diseñaron los *primers* para el análisis de transcripción policistrónica en *B. abortus*. Las bandas que se observan en las calles RT- en las Figuras A y B corresponden a dímeros de primers que no amplificaron.

Luego de corroborar la presencia y transcripción de los genes *tat* en ambos organismos, se estudió la conservación de la funcionalidad del sistema Tat a través de ensayos de complementación heteróloga utilizando las cepas de *E. coli* mutantes con una delección en fase para los genes *tat* que da lugar a diferentes cepas con el sistema Tat inactivo (provistas por Dr. Tracy Palmer, Molecular Microbiology, University of Dundee, Scotland). Las mutantes utilizadas MC4100-P (parental; control), JARVIG (Δ *tatA*), BOD (Δ *tatB*), BILKO (Δ *tatC*) y DADE (Δ *tatABC*) son incapaces de translocar sustratos específicos que permiten crecer en condiciones selectivas. Las cepas mutantes se complementaron individualmente con los respectivos genes ortólogos de los diferentes organismos. Para esto, se clonaron los genes *tatA*, *tatB* y *tatC* de *A. marginale*, *B. abortus* y *E. coli* (Control) en el plásmido pUNIPROM bajo el promotor del operon *tat* de *E. coli*. Estos plásmidos se utilizaron en ensayos de complementación heteróloga para estudiar si eran capaces de restituir la función del sistema bajo diferentes condiciones selectivas [118-120].

Cultivos de cepas mutantes (Δtat -) transformados con los genes experimentales y controles, fueron crecidos durante la noche, se diluyeron 1:100 en medio LB completo e incubados a 37 °C hasta DO de 0.4-0.6 para ser analizadas por microscopía (DIC, contraste de fases). Por microscopia fue posible identificar un fenotipo de células correspondiente al sistema Tat inactivo que se visualiza como largas cadenas de bacterias que se mantienen unidas luego de la división celular, debido a la imposibilidad de exportar un sustrato al periplasma que sería el encargado de finalizar la unión entre células luego de la división. El control positivo, células complementadas con genes de *E. coli*, recuperaron la funcionalidad, lo que se refleja en células completamente divididas. Por el contrario, las células complementadas con el vector sin inserto, mostraron largas cadenas de células unidas. Para células complementadas con genes de *A. marginale* solo se observó complementación parcial por TatA, observando ambos fenotipos. Para *B. abortus*, también se observó una complementación parcial para TatA, mientras que TatB y TatC demostraron restaurar completamente la función (**Figura 2.4**).

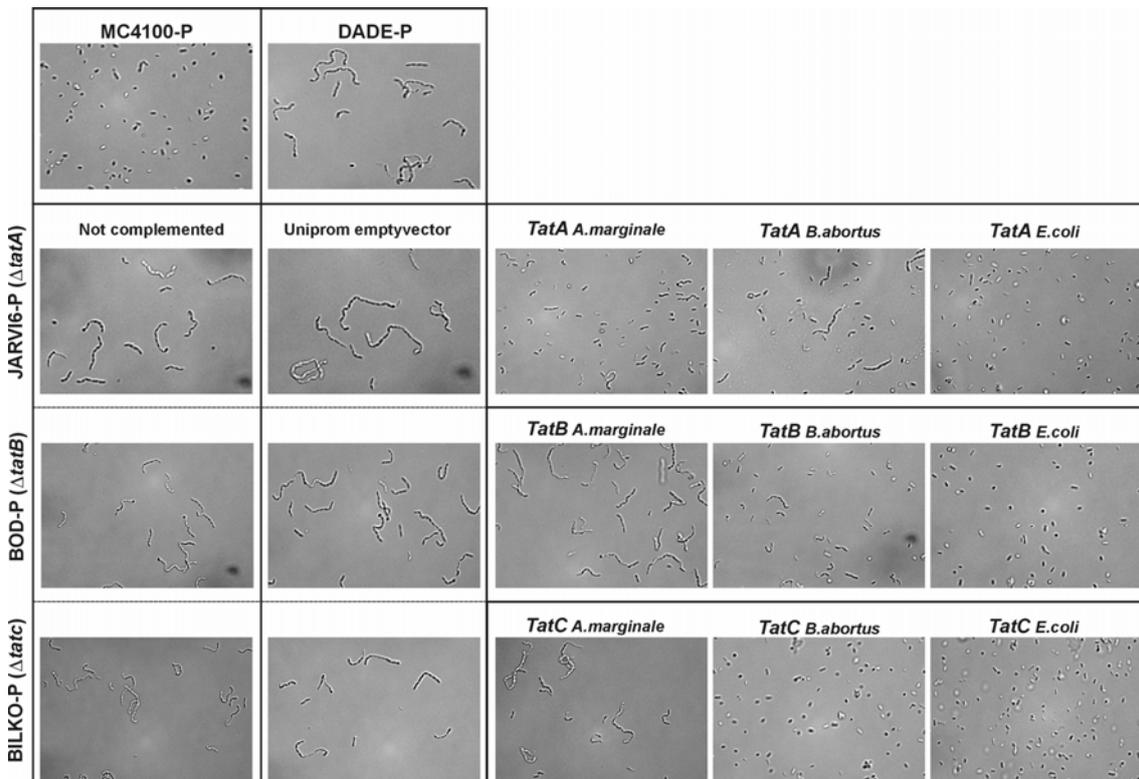


Figura 2.4. Fenotipo cell-chain.

La funcionalidad de TatA, TatB o TatC se analizó por microscopía de contraste de fases para estudiar la habilidad para completar la división celular. Las mutantes Tat de *E. coli* (JARVI6-P, BOD-P y BILKO-P) se complementaron con los genes heterólogos experimentales de *A.*

marginale, *B. abortus* o *E. coli* (control). Como control positivo se utilizó además M4100-P y como control negativo la triple mutante (para todo el operon *tatABC*, DADE-P) y las mutantes sin complementar, o complementadas con el vector pUNIPROM sin inserto. Las células se examinaron con microscopio confocal Leica TCS-SP5 Laser Confocal Microscope 63X.

Utilizando también otros ensayos se evaluó la capacidad de los genes heterólogos de complementar *E. coli* mutantes. La exportación por el sistema Tat de las proteínas AmiA y AmiC, involucradas en remodelar la pared celular, permite crecer en un medio con 2% de SDS. Células con un sistema Tat inactivo son incapaces de crecer en un medio con 2% de SDS y por ende la habilidad de crecer bajo esta condición es un indicador de funcionalidad del sistema. El crecimiento de cultivos complementados se evaluó cualitativamente en placas de LB-agar con 2% SDS y se cuantificó en las mismas condiciones de crecimiento pero en cultivos líquidos. Por otro lado, analizando la funcionalidad bajo otras condiciones selectivas, *E. coli* es capaz de crecer en condiciones anaeróbicas utilizando TMAO como aceptor final de electrones debido a la exportación a través del sistema Tat de la enzima TMAO reductasa (TorA) soluble en el periplasma. Entonces, la habilidad de crecer en condiciones anaeróbicas con glicerol como única fuente de carbono y TMAO como aceptor de electrones es un indicador de funcionalidad del sistema.

Para *A. marginale*, TatA fue capaz de complementar la función del gen nativo de *E. coli* demostrando que conserva la función y la capacidad de interactuar con las proteínas TatBC de *E. coli* (**Figura 2.5**). TatB fue capaz de complementar la función pero únicamente bajo condiciones selectivas de anaerobiosis y TMAO, y no en el medio con 2%SDS (**Figura 2.6**); mientras que TatC no fue capaz de complementar la funcionalidad (**Figura 2.7**).

Para *B. abortus*, TatA de no complementó la función de su homólogo bajo la condición de SDS pero si de TMAO en anaerobiosis (**Figura 2.5**). TatB y TatC restauraron la función del sistema (**Figura 2.6, Figura 2.7**).

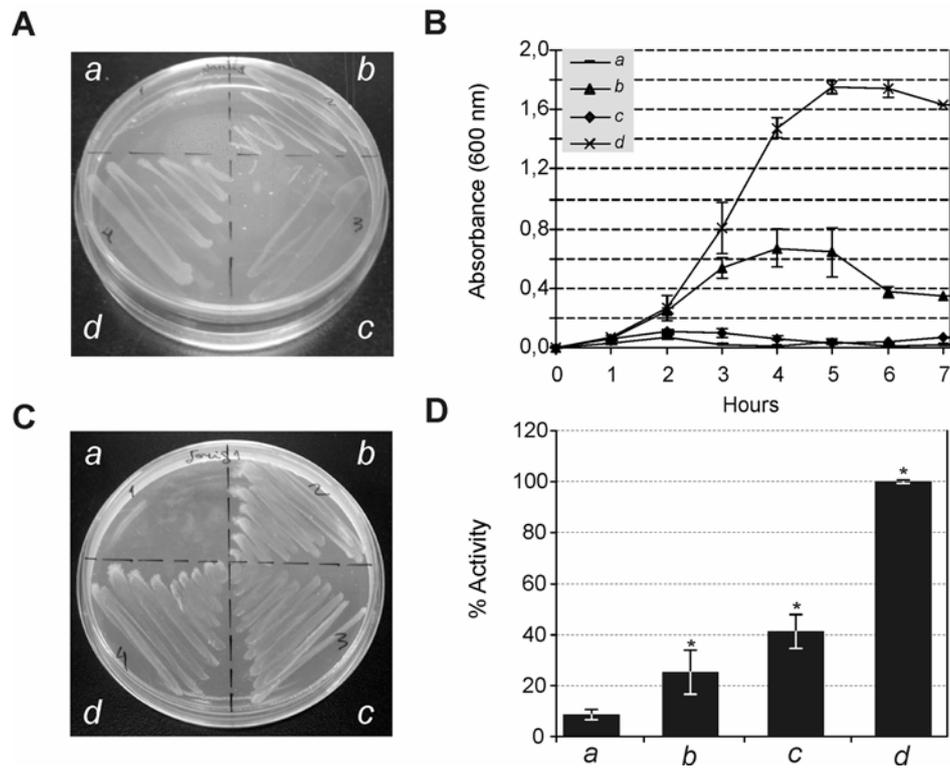


Figura 2.5. Funcionalidad de TatA.

Complementación de *E. coli* $\Delta tatA$ - $\Delta tatE$ (JARVIG-P) con los genes heterólogos *tatA* de *A. marginale* (b), *B. abortus* (c) y *E. coli* (d). *Uniprom empty vector* (a) se utilizó como control negativo. **(A)** Bacterias crecidas en un medio sólido agar LB 2% SDS. **(B)** Bacterias crecidas en medio LB líquido 2% SDS. **(C)** Bacterias crecidas bajo condiciones de anaerobiosis en medio mínimo con glicerol como fuente de carbono y TMAO como aceptor de electrones. **(D)** La actividad de la enzima TMAO se cuantificó a partir de las fracciones periplásmicas. Un 100% de la actividad se determinó a partir de la fracción periplásmica de las cepas mutantes complementadas con los genes nativos de *E. coli*. (*) $p < 0,05$, ANOVA.

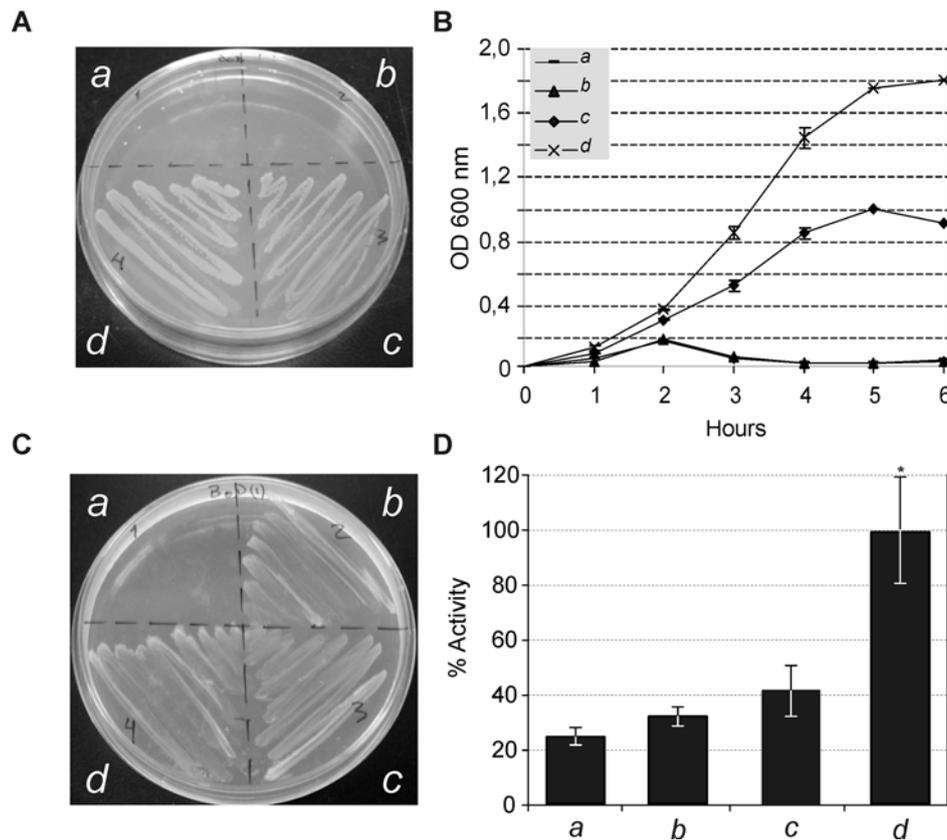


Figura 2.6. Funcionalidad de TatB.

Complementación de *E. coli* Δ *tatB* (BOD-P) con los genes heterólogos *tatB* de *A. marginale* (b), *B. abortus* (c) y *E. coli* (d). *Uniprom empty vector* (a) se utilizó como control negativo. **(A)** Bacterias crecidas en un medio sólido agar LB 2% SDS. **(B)** Bacterias crecidas en medio LB líquido 2% SDS. **(C)** Bacterias crecidas bajo condiciones de anaerobiosis en medio mínimo con glicerol como fuente de carbono y TMAO como aceptor de electrones. **(D)** La actividad de la enzima TMAO se cuantificó a partir de las fracciones periplásmicas. Un 100% de la actividad se determinó a partir de la fracción periplásmica de las cepas mutantes complementadas con los genes nativos de *E. coli*. (*) $p < 0,05$, ANOVA.

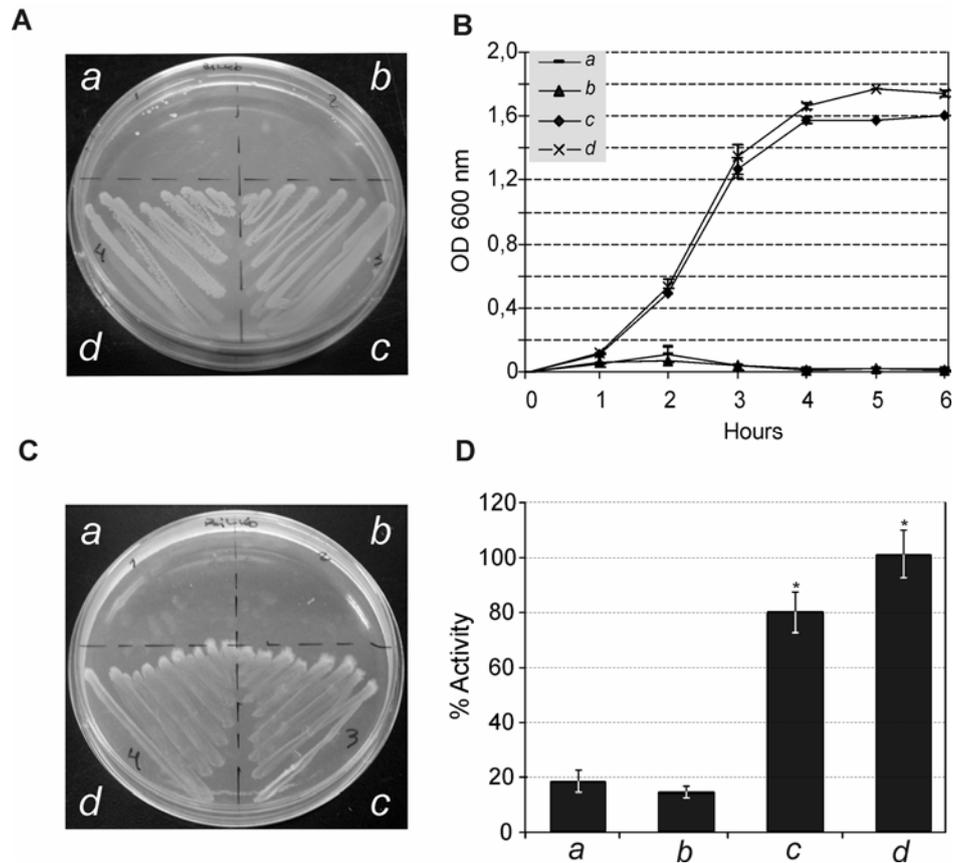


Figura 2.7. Funcionalidad de TatC.

Complementación de *E. coli* Δ *tatC* (BILKO-P) con los genes heterólogos *tatC* de *A. marginale* (b), *B. abortus* (c) y *E. coli* (d). *Uniprom empty vector* (a) se utilizó como control negativo. **(A)** Bacterias crecidas en un medio sólido agar LB 2% SDS. **(B)** Bacterias crecidas en medio LB líquido 2% SDS. **(C)** Bacterias crecidas bajo condiciones de anaerobiosis en medio mínimo con glicerol como fuente de carbono y TMAO como aceptor de electrones. **(D)** La actividad de la enzima TMAO se cuantificó a partir de las fracciones periplásmicas. Un 100% de la actividad se determinó a partir de la fracción periplásmica de las cepas mutantes complementadas con los genes nativos de *E. coli*. (*) $p < 0,05$, ANOVA.

Los resultados indicarían que la funcionalidad de las diferentes subunidades es dependiente del sustrato específico que se transporte observando diferentes resultados en los distintos ensayos. Estos resultados en conjunto sugieren que *A. marginale* TatA y TatB son capaces de restablecer la funcionalidad en el sistema heterólogo de *E. coli*, mientras que TatC no fue capaz de restaurar la función en ninguna condición. TatB solo complementó la función en la condición de selección anaeróbica, sin restaurar la función que permite el crecimiento en un medio con 2% SDS; esto podría deberse a la incapacidad de TatB de *A. marginale* de reconocer el péptido señal específico de los sustratos de *E. coli*, como también a que la condición

de crecimiento con SDS 2% resulta más estricta, como fuera observado por otros autores [118]. Esto es consistente con el rol que poseen en el ensamblado de las subunidades del complejo proteico donde se propone que TatA forma un complejo oligomérico generando un poro, mientras que TatB y TatC son las responsables de interactuar con el péptido señal específico de los sustratos, para luego interactuar con TatA en membrana. En el caso de *B. abortus* las tres subunidades demostraron ser capaces de restaurar la función en este sistema heterólogo. Es importante destacar que TatA no fue capaz de complementar la función cuando se evaluó en la condición de crecimiento con 2% SDS lo que podría deberse, como se mencionó, a una condición de crecimiento más estricta o también a la incapacidad de reconocer al péptido señal específico.

Considerando que la estequiometría del complejo proteico es relevante para la funcionalidad, y por ende, la transcripción juega un rol importante conjuntamente con las diferentes organizaciones de los genes en el genoma, utilizando *real time* RT-qPCR estudiamos los niveles de transcripción de los diferentes genes *tat* en ambos organismos. Los niveles de transcripción se calcularon como concentración de cDNA (ng/μl) del gen *tat* (A, B o C) / (ng/μl) del gen *housekeeping* GroEL (*A. marginale*) o RPLL (*B. abortus*). Para cada par de primers se estimó la concentración óptima que no genere dímeros y se calibró la eficiencia de cada uno. Para estimar la concentración de cDNA se realizaron diluciones seriadas del cDNA para construir curvas de concentración estándar (*R value* > 0.99). Los resultados se analizaron utilizando el *software* REST utilizando métodos estadísticos no paramétricos (*Pair Wise Fixed Reallocation Randomisation*).

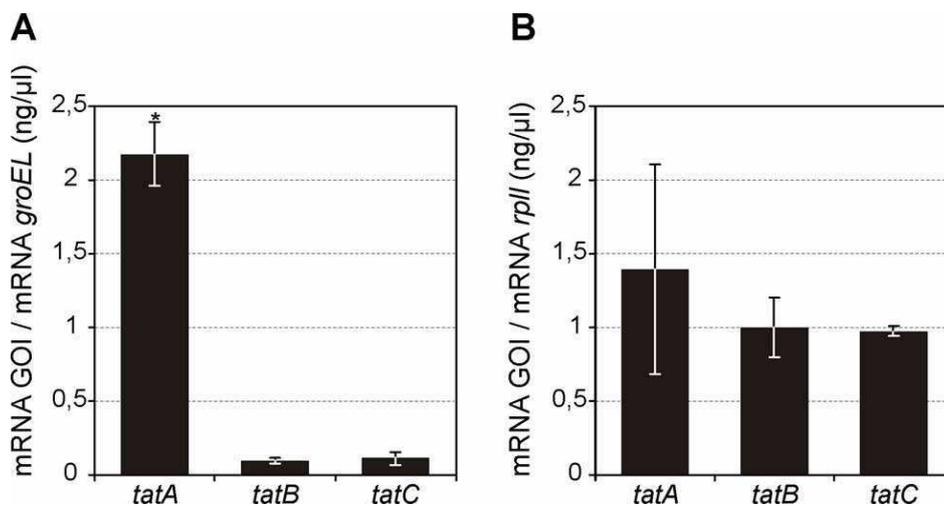


Figura 2.8. Niveles de abundancia de transcritos para los genes *tat*.

Análisis de los niveles de abundancia de los ARNm de los genes *tatA*, *tatB* y *tatC* en **(A)** *A. marginale* str. *St. Maries* y **(B)** *B. abortus* 2308. (*) $p < 0.001$, *Pair Wise Fixed Reallocation Randomization Test*. **GOI**: *gene-of-interest*.

Los resultados demostraron una expresión equivalente de los tres genes para *B. abortus*, como se esperaría para un operon (**Figura 2.8B**), sin embargo para *A. marginale* se determinó una expresión de *tatA*, 23 y 20 veces mayor que *tatB* y *tatC*, respectivamente (**Figura 2.8A**). Es interesante destacar que la relación identificada para los niveles de transcripción de los genes coincide con la estequiometría descrita por diferentes autores para el complejo proteico funcional en la membrana plasmática. Tomando en consideración que las tasas obtenidas para la transcripción son equivalentes a la estequiometría del complejo proteico, este resultado podría ser una evidencia indirecta no solamente de la funcionalidad del sistema pese a perder la estructura de operon, sino también de un mecanismo de regulación transcripcional alternativo a la transcripción policistónica.

Posteriormente a demostrar la funcionalidad del sistema Tat en el sistema heterólogo para los organismos seleccionados, estudiamos *in silico*, utilizando los diferentes programas de predicción, los potenciales sustratos translocados por el sistema. Analizamos las familias de *Anaplasmataceae* y *Brucellaceae* utilizando los tres algoritmos diseñados y publicados para la detección de los péptidos señales N-terminal del sistema. Identificamos los potenciales sustratos en 12 especies de la familia *Anaplasmataceae* y 12 especies de la familia *Brucellaceae*. Consideramos positivas las proteínas predichas por al menos 2 programas. Los potenciales sustratos se agruparon de acuerdo a las categorías asignadas para Clúster of Orthologs Groups (COGs). Aunque confirmar los sustratos requeriría una aproximación experimental, este análisis es una primera aproximación para comprender el número de potenciales proteínas que utilizarían el sistema permitiendo conocer un poco más acerca de rol que podría tener el sistema Tat en los organismos en estudio, como también posibles modificaciones en los péptidos señales.

Se identificaron tres sustratos positivos con COGs asociados en la familia *Anaplasmataceae* (**Tabla 2.1**). Identificamos la proteína *Rieske Fe-S protein* (COG0723) presente en 8 de los 12 genomas analizados. En *Wolbachia endo.of Culex quinquefasciatus*, detectamos la proteína *FtsI/penicillin-binding 2* (COG0768). Finalmente, identificamos al COG *Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases, COG1028)* en *Anaplasma centrale* str. *Israel*. Los sustratos positivos se identificaron en los restantes organismos de la familia para

identificar las proteínas ortólogas (cuando estén presentes) y analizar posibles modificaciones en los péptidos señales por lo que no fueron identificados por los programas predictivos. Se encontraron ortólogos para la proteína perteneciente al COG1028 en todos los organismos de la familia. *A. centrale* con el característico péptido señal conformado por las RR y los aminoácidos consenso. Los restantes organismos presentan modificaciones en una o ambas argininas, presentando reemplazos por el aminoácido Lisina (RK, KR o KK) (**Figura S2.1, Anexo**). Sumado a esto, el codón de iniciación del respectivo gen en *A. marginale str. St. Maries* y *A. marginale str. Florida* fue anotado “río arriba” (hacia extremo 5') lo que afecta la predicción del programa, que en general el consenso determina el péptido señal en los primeros 30 aminoácidos. La proteína *penicillin-binding protein* (COG0768) fue identificada en *Wolbachia* endosymbiont of *Culex quinquefasciatus* por dos programas (TatP and TATFIND), pero solamente por uno (TATFIND) en las restantes especies del genero *Wolbachia* spp. Se identificaron ortólogos de la proteína únicamente en *A. marginale (str. St. Maries y str. Florida)* y en *A. centrale Israel*, que mostraron bloques conservados a lo largo de la proteína y aminoácidos conservados en el péptido señal Tat. Sin embargo, el péptido señal presenta una sustitución próxima a las RR (una Isoleucina en vez de una Serina o Alanina) que probablemente afecte la capacidad del algoritmo TATFIND de predicción (**Figura S2.2, Anexo**).

Tabla 2.1. Definición COG para los sustratos positivos (a).

Anaplasmataceae (12 genomes)			
COG	Definición GOG	Presente en <i>Anaplasmataceae</i>	Presente en <i>Brucellaceae</i>
COG0723	<i>Rieske Fe-S protein</i>	8 de 12	si
COG0768	<i>Cell division protein FtsI/penicillin-binding</i>	1 de 12	no
COG1028	<i>Dehydrogenases with different specificities</i>	1 de 12	no

(a) Las proteínas se consideraron positivas cuando fueron predichas por al menos 2 de los programas.

El análisis en la familia *Brucellaceae* se realizó en 10 organismos y se identificaron 250 proteínas positivas como sustratos del sistema que pudieron ser agrupadas en 22 categorías COGs (**Tabla 2.2**) con diferentes niveles de representación dentro de la familia, y 7 COGs exclusivos para *Ochrobactrum anthropi*.

Dentro de los potenciales sustratos, 14 tienen una anotación como proteínas hipotéticas sin COGs relacionados. Entre las categorías COGs mayormente representadas se encontraban proteínas denominados “periplasmic components of solute-binding proteins”. La presencia de péptidos señal del sistema fue previamente descrito en componentes periplásmicos como los transportadores ABC en especies como *Rhizobium leguminosarum* *bv. viciae* [100] y Halobacteraceae [121, 122]. Los sistemas de importación solo se encuentran en procariotas y contienen dominios ABC y dominios en la membrana interna, con proteínas de unión (BPs) extracitoplasmáticas para unir específicamente a los compuestos del respectivo sistema ABC. En las bacterias gram-negativas, las BPs se ubican en el periplasma [123]. Los sistemas ABC importan una gran cantidad de compuestos en la célula procariota incluyendo péptidos, poliaminas, iones, aminoácidos, hierro y sulfatos [123, 124]. También identificamos COGs relacionados a las “Nitrous oxide reductasa” (*Nos*; COG4263). Estudios previos describieron el rol del sistema Tat en la translocación de las *Nos* (reductasas de ON) en *Pseudomonas stutzeri* [125], donde el sistema Tat demostró ser necesario para establecer el proceso anaeróbico de la denitrificación. La enzima *Nos* es una de las cuatro reductasas de *Brucella* spp. involucradas en la isla de denitrificación (*denitrification island*) que le permite a la bacteria crecer bajo condiciones de baja concentración de oxígeno dentro de los macrófagos por medio de la respiración utilizando nitratos [126]. Finalmente, el COG0723 (Rieske Fe-S protein) fue el único sustrato compartido entre el secretoma tat identificado en ambas familias *Anaplasmataceae* y *Brucellaceae*.

Tabla 2.2. Definición COG para los sustratos positivos (a).

<i>Brucellaceae</i> (10 genomas)			
COG	Definición GOG	Presente en <i>Brucellaceae</i>	Presente en <i>Rickettsiales</i>
COG0723	<i>Rieske Fe-S protein</i>	10 de 10	si
COG0747	<i>ABC-type dipeptide transport system, PC</i>	10 de 10	no
COG1651	<i>DSBA oxidoreductase; thiol:disulfide interchange</i>	10 de 10	si

COG2132	<i>Putative multicopper oxidases</i>	10 de 10	no
COG2041	<i>Sulfite oxidase and related enzymes</i>	10 de 10	no
COG1376	<i>Uncharacterized protein conserved in bacteria</i>	10 de 10	no
COG4263	<i>Nitrous oxide reductase</i>	9 of 10	no
COG3206	<i>Uncharacterized protein involved in exopolysaccharide biosynthesis</i>	9 of 10	no
COG3319	<i>Thioesterase domains of type I polyketide synthases or non-ribosomal peptide synthetases</i>	9 of 10	no
COG3683	<i>ABC-type uncharacterized transport system, PC</i>	9 of 10	no
COG4134	<i>ABC-type uncharacterized transport system, PC</i>	9 of 10	no
COG4213	<i>ABC-type xylose transport system, PC</i>	8 of 10	no
COG1477	<i>Membrane-associated lipoprotein involved in thiamine biosynthesis</i>	8 of 10	no
COG0715	<i>ABC-type nitrate/sulfonate/bicarbonate transport, PC</i>	8 of 10	no
COG1464	<i>ABC-type metal ion transport system, PC /surface antigen</i>	8 of 10	no
COG4663	<i>TRAP-type mannitol/chloroaromatic compound transport system, PC</i>	8 of 10	no
COG2989	<i>Uncharacterized protein conserved in bacteria</i>	8 of 10	no
COG1574	<i>Predicted metal-dependent hydrolase with the TIM-barrel</i>	7 of 10	no
COG4166	<i>ABC-type oligopeptide transport system, PC</i>	5 of 10	no
COG2340	<i>Uncharacterized protein with SCP/PR1 domains</i>	5 of 10	no
COG0612	<i>Predicted Zn-dependent peptidases</i>	4 of 10	Si
COG0741	<i>Soluble lytic murein transglycosylase and related regulatory proteins</i>	2 of 10	si
COG0243	<i>Anaerobic dehydrogenases, selenocysteine-containing</i>	Ochrobactrum	no
COG1879	<i>ABC-type sugar transport system, PC</i>	Ochrobactrum	no
COG2837	<i>Predicted iron-dependent peroxidase</i>	Ochrobactrum	no
COG0683	<i>ABC-type branched-chain amino acid transport, PC</i>	Ochrobactrum	no
COG0687	<i>Spermidine/putrescine-binding PC</i>	Ochrobactrum	no
COG3019	<i>Predicted metal-binding protein</i>	Ochrobactrum	no
COG3246	<i>Uncharacterized conserved protein</i>	Ochrobactrum	no

Abreviaciones: PC: proteína periplásmica.

(a) Las proteínas se consideraron positivas cuando fueron predichas por al menos 2 de los programas.

2.3. Discusión

En este capítulo de la tesis, abordamos el estudio y caracterización del sistema Tat en dos importantes patógenos, *A. marginale str. St. Maries* y *B. abortus 2308*. Identificamos los componentes del sistema y estudiamos la organización genómica, la conservación estructural y funcional de los genes *tatA*, *tatB* y *tatC* comparativamente con la clase α -proteobacteria y el conjunto de potenciales sustratos del sistema en cada una de las familias.

Se realizó la identificación y análisis de la organización en el genoma de los genes, como también la conservación de la secuencia y de los dominios estructurales. Por otro lado se analizó la transcripción, la conservación funcional y los niveles de expresión contrastando una organización dispersa y una organización clásica en operon, para estudiar el impacto de la conservación de la organización genómica y regulatoria en la funcionalidad. También se realizaron análisis del uso del sistema Tat para la exportación de potenciales sustratos, utilizando diversos programas disponibles para analizar el posible rol del sistema en la exportación de proteínas para conferir capacidades de adaptación o patogénesis en los diferentes grupos.

En los últimos años se realizaron muchos avances para desentrañar los mecanismos moleculares y bioquímicos del sistema Tat como un mecanismo alternativo al sistema Sec para la secreción de proteínas. Sin embargo, escasos avances se realizaron para conocer el rol del sistema en la clase α -proteobacteria [100, 113, 127-129]. Como se mencionó anteriormente, este grupo presenta una gran variabilidad genómica y biológica (1-10 Mb) asociada con procesos extremos de reducción y expansión genómica con la eliminación e incorporación de nuevos genes [14], gran diversidad en estilos de vida y nichos ecológicos (desde intracelulares obligados a organismos de vida libre [1]), que podrían ser parcialmente explicados a través de una mejor y mayor comprensión y conocimiento acerca del rol de los sistemas de secreción, y específicamente del sistema Tat, como actores claves para la exportación de sustratos, la interacción con la célula hospedadora y con el ambiente.

Se analizaron 53 genomas completamente anotados de la clase α -proteobacteria. Se confirmó la presencia de los tres genes *tatA*, *tatB* y *tatC* para el ensamble de la maquinaria de traslocación en casi todos los miembros de la clase. Identificamos los genes *tatA* en *A. marginale str. St. Maries* y *Ehrlichia ruminantium str. Gardel*, que fueron significativamente más cortos que los ortólogos en los otros órdenes perteneciente a la clase, revelando que los ORFs pequeños suelen ser omitidos en los procesos de anotación [80]. Las bacterias intracelulares obligadas que

sufrieron procesos de reducción genómica asociados a una alta frecuencia de rearrreglos cromosómicos [1], aunque también podría ser el resultado de procesos de recombinación intracromosómica, la organización de los genes que forman parte de complejos con multicomponentes, podrían verse afectados cuando la selección no favorece el agrupamiento de los genes en el mismo operon o en regiones próximas en el genoma. En este sentido, la organización de los genes *tat* reveló una gran variabilidad y diversidad dentro de la clase. En la mayoría de los géneros, los genes *tat* presentaron una estructura canónica, conformada por los tres genes organizados en el mismo operon (*tatABC*). Por el contrario, el orden Rickettsiales, que sufrió procesos evolutivos extremos de reducción genómica, mostró una organización absolutamente dispersa en el genoma de los genes *tatA*, *tatB* y *tatC* (**Figura 2.1B**). Una organización dispersa para los genes *tat*, fue previamente descrita para *Rickettsia prowazekii* [86]. Hipotetizamos que la reducción genómica asociada a los procesos de rearrreglos cromosómicos podría tener un efecto en la organización anómala, debido a que los genomas pertenecientes al orden Rickettsiales son el resultado de fuerzas de selección hacia un proceso de reducción del genoma como una adaptación a un estilo de vida intracelular. Esta observación parece no corresponderse con la observación para las especies de *Bartonella* spp., dado que presentan una organización clásica en operon *tatABC*, pese al pequeño genoma que también es producto de un proceso de reducción; sin embargo, aunque también son organismos con genomas pequeños y hábitats intracelulares, diferentes estudios apoyan la idea de que los procesos de reducción ocurrieron independientemente en los linajes de *Rickettsia* y *Bartonella* [14].

Recientemente, en un estudio se describió la organización y patrones de transcripción del operon de la succinato deshidrogenasa (*sdh*) en *Anaplasma phagocytophilum*, otra bacteria intracelular obligada con genoma reducido [130]. Los autores observaron que los genes *sdh* involucrados se conservaban en el genoma y también conservaban residuos críticos, sugiriendo que estas subunidades conservan la funcionalidad, pero describieron también una atípica reorganización genómica, con una fragmentación del operón y un inusual patrón de expresión. Los autores observaron que algunos genes fragmentados presentaron, alternativamente, ATG o GTG codones de inicio, con o sin secuencias Shine-Dalgarno (SD). Estudios anteriores han demostrado que el actual sitio de SD o la utilización de ATG como codón de inicio generalmente se traducen en un aumento del nivel de proteínas. Por lo tanto, estas diferencias entre los codones de inicio y los sitios de SD podrían representar mecanismos de regulación alternativos para el control de la expresión en operones fragmentados. Otros organismos de la clase también muestran

duplicaciones de genes y se han propuesto como un mecanismo adicional para aumentar el nivel de transcripción [130].

Observaciones similares de una asociación entre una fragmentación de operones y reducción del genoma fueron descritas para el sistema de secreción tipo IV o T4SS (*Vir*) [79, 80]. El sistema de secreción *Vir* es un factor de virulencia esencial que presenta una organización en operón de 11 genes, en la mayoría de los organismos, como es el caso del *virB/virD* T4SS de *Agrobacterium tumefaciens*, donde fue más estudiado. Los autores observaron en el organismo *Ehrlichia chaffeensis* perteneciente al orden Rickettsiales, el operón *virBD* se divide en dos operones de cinco y seis genes (*virB3-virB6* y *virB8-virD4*). Además, entre estos dos operones se identificaron duplicaciones de genes que componen al sistema. Utilizando ensayos de movilidad electroforética (EMSA) identificaron una proteína que se une específicamente a las regiones promotoras río arriba de ambos fragmentos del operón. Esta proteína, anotada como un factor de transcripción hipotético, se ha propuesto como regulador del operón fragmentado. Interesantemente, este hipotético factor de transcripción no se encuentra en los organismos de la clase α -proteobacteria como un único operón de 11 genes, pero si está presente en el orden Rickettsiales y *Gluconobacter oxydans* que presentan una organización fragmentada del operón [79]. Otros estudios anteriores han demostrado que, al igual que los genes rRNA de las especies de Rickettsia, los genomas de *A. phagocytophilum*, *A. marginale*, y *E. chaffeensis* contienen fragmentado el operón rRNA, lo que sugiere que un evento evolutivo distante dio lugar a la fragmentación de este operón en un ancestro del orden Rickettsiales [130-132].

Esto apoya la hipótesis que en los organismos de genomas reducidos que sufrieron intensos eventos de reordenamientos genómicos, estos podrían producir la fragmentación, y la consecuente pérdida de una expresión coordinada para garantizar cantidades necesarias de cada proteína para la funcionalidad. En este caso, mecanismos alternativos de regulación del tipo gen-específicos, como la presencia de ATG, GTG secuencias de SD, el uso de codones, y factores de transcripción, podrían actuar como mecanismos alternativos de regulación para co-regular los genes dispersos (**Figura 2.9**). Teniendo en cuenta el proceso de pérdida masiva de genes por reducción del genoma, la ausencia del gen *tatB* en las especies de *Rickettsia*, *Neorickettsia*, *Orientia* y *Wolbachia* podría llevar a la pérdida de función del sistema Tat en estos organismos. Sin embargo, en organismos gram-positivos y arqueas, se describieron sistemas que carecen de TatB [90, 91] y permanecen funcionales como un complejo TatAC [91, 92]. A su vez, un estudio reciente donde sustituyeron algunos aminoácidos de la proteína TatA se demostró que la actividad biológica de TatA y TatB

se condensa en una sola proteína en aquellos sistemas que carecen de TatB identificable [94]. Por esto, la proteína no pudo ser identificada en *Rickettsia* spp, *Neorickettsia* spp, *Wolbachia* spp y *Orientia* spp, organismos donde la funcionalidad no fue demostrada aun y que aunque se conserven los genes *tatA* y *tatC*, requerirá futuros estudios.

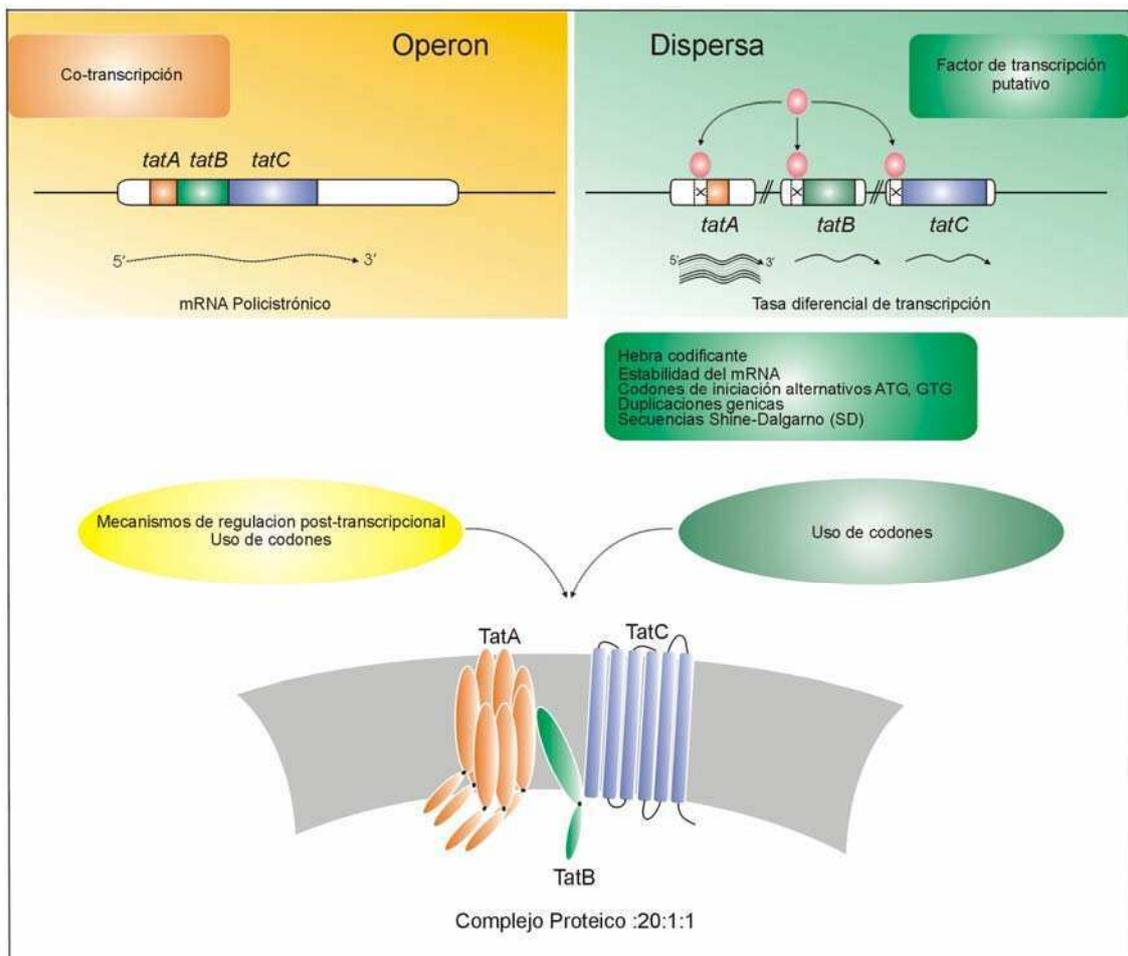


Figura 2.9. Esquema hipótesis de regulación alternativos.

Pese a la fragmentación del operon *tatABC* en *A. marginale* y de una proteína de pequeño tamaño, el análisis de la secuencia indicó una conservación de los aminoácidos esenciales, las características estructurales y segmentos críticos de proteínas, lo que sugiere la conservación de la funcionalidad (Figura 2.2).

Los resultados experimentales de *A. marginale* demostraron que sólo la subunidad TatA puede restaurar la funcionalidad del complejo Tat en el sistema heterólogo de *E. coli*. Es importante tener en cuenta que siempre que se ensayó la complementación entre especies, las proteínas TatA conservan un alto nivel de la

función en el huésped heterólogo lo que sugiere que las limitaciones en la función de TatA son menos severas que las de TatB o TatC [114, 118]. Esto es consistente con el papel de la subunidad de TatA en el complejo de proteínas que se auto-ensambla en la formación de canales formado homo-multímeros. Por el contrario, los requerimientos en la complementación heteróloga con las proteínas TatB o TatC, es probable sean más estrictos ya que estos dependerían del reconocimiento de los péptidos señal no nativos de *E. coli*, que pueden mostrar diferencias en las secuencias. Dado que TatB interactúa con cada uno de los componentes de Tat y con péptidos señal específicos, se podría esperar que la complementación cruzada con esta subunidad sea menos eficiente que con otras proteínas Tat. TatB de *A. marginale* permitió el crecimiento de la mutante BOD-P en un medio selectivo que contiene TMAO indicando la función Tat (**Figura 2.6**), pero no permitió el crecimiento en los medios conteniendo 2% SDS (**Figura 2.6A, Figura 2.6B**) probablemente debido a un efecto específico del sustrato involucrado en el test. TatC de *A. marginale* no complementó la mutante BILK0-P en diferentes medios selectivos (**Figura 2.7**). Aunque la expresión de la proteína TatC de *A. marginale* no se pudo demostrar en el sistema heterólogo porque carecemos de un anticuerpo contra la proteína nativa, se corroboró la expresión de los genes *tatA*, *tatB* y *tatC* de *A. marginale* en las cepas de *E. coli* por RT-PCR luego de la complementación (resultados no mostrados).

En conjunto, nuestros resultados sugieren que *A. marginale* conserva un sistema Tat funcional, ya que TatA y tatB son capaces de restaurar la función. La estructura de TatC está conservada y también los aminoácidos esenciales, aunque no sea capaz de complementar la funcionalidad en el sistema heterólogo (**Tabla 2.3**). Considerando que TatC ha sido implicado como un factor determinante de la especificidad de sustratos Tat-dependientes de la secreción a través del reconocimiento de los péptidos señal Tat [77, 133], y teniendo en cuenta que el genoma de *A. marginale* no codifica para ningún sustrato del sistema involucrados en estos ensayos experimentales, probablemente la falta de complementación pueda explicarse por la incapacidad de reconocer péptidos señal Tat de *E. coli*. Otra posible explicación podría ser la incapacidad total o parcial de interactuar con el complejo heterólogo. Sin embargo una potencial co-evolución entre los péptidos señales de *A. marginale* (y organismos relacionados) y la maquinaria específica de reconocimiento representa una pregunta y desafío interesante a abordar.

Las tres subunidades de *B. abortus* fueron capaces de restaurar la función de Tat sugiriendo no solo la conservación de la funcionalidad, sino también el reconocimiento de sustrato (**Figura 2.5C, Figura 2.5D, Figura 2.6 y Figura 2.7**), excepto para el medio selectivo de 2%SDS (**Figura 2.5A, Figura 2.5B y Tabla 2.3**).

Tabla 2.3. Resultados de la funcionalidad de las subunidades Tat.

<i>Anaplasma marginale</i> str. <i>St. Maries</i>				
	<i>Cell-chain phenotype</i>	<i>SDS-resistant phenotype</i>	Crecimiento Anaeróbico-TMAO	Actividad TMAO reductasa (%)^a
TatA	si	si	si	25,48 *
TatB	no	no	si	32,33
TatC	no	no	no	13,03
<i>Brucella melitensis</i> biovar <i>abortus</i> 2308				
TatA	no	no	si	41,37 *
TatB	si	si	si	41,67
TatC	si	si	si	78,82 *
(a) Actividad relativizada al control positivo				
(*) p<0.05				

Los resultados del análisis de los niveles de expresión de los ARNm de los genes *tat* para *A. marginale* son interesantes ya que diferentes autores describen la estequiometría del complejo de proteínas TatABC (20-30:1:1) [77, 133] con un gran exceso molar de la subunidad TatA. Este resultado podría revelar una evidencia indirecta de funcionalidad, como también la existencia de un mecanismo alternativo de regulación. Para *B. abortus*, se demostró que los genes *tatABC* se transcribe en un ARN policistrónico, como se espera de un sistema de expresión en operón, excluyendo al gen *serS*. La exclusión de este cuarto gen se estudio con el objetivo de diseñar mutantes específicas que todavía no se han realizado en este organismos para estudiar su relación con la viabilidad y capacidad de infección de la bacteria. En este caso, de un mensajero policistrónico, se requerirían mecanismos de regulación post-transcripcionales para lograr cantidades relativas acordes a la estequiometria correcta para este sistema de multi-componentes (**Figura 2.9**).

Aunque las familias *Brucellaceae* y *Anaplasmataceae* están relacionadas filogenéticamente, se encontraron importantes diferencias en el conjunto de proteínas potencialmente secretadas por el sistema Tat (secretoma Tat). Nuestros datos son consistentes con algunos obtenidos previamente en la clase α -proteobacteria, donde independientemente de la filogenia, las bacterias patogénicas tendrían un uso limitado del sistema Tat, mientras que organismos de vida libre y bacterias del suelo tienen un uso moderado a alto [134]. Esta característica que relaciona el uso del sistema Tat con un estilo de vida mostró que en la familia *Brucellaceae*, *Ochrobactrum antropi* evidenció un número más alto de potenciales sustratos que las demás especies de la

familia (**Tabla 2.2**). En este sentido, *Brucella* spp., como bacterias intracelulares facultativas, parecen tener un uso moderado del sistema Tat, intermedio entre patógenos y organismos de vida libre. Esta hipótesis está sustentada por la relativa gran cantidad de transportadores ABC que fueron predichos como potenciales sustratos del sistema Tat; transportadores de alta afinidad de unión a diferentes sustratos para importar los nutrientes de hábitats competitivos y variables, aunque la mayoría de las veces *Brucella* spp. puede adquirir la nutrición de un nicho estable. Notablemente, ambas familias *Anaplasmataceae* y *Brucellaceae*, exhibieron como único sustrato plenamente compartido a la proteína *ubiquinol-cytochrome c reductase iron-sulfur subunit* (*Rieske iron-sulfur domain-containing protein*). La proteína Rieske Fe/S es una subunidad esencial de las mitocondrias y de los complejos bc1 de bacterias, como actores fundamentales en la transferencia de funciones *redox* en el transporte de electrones en la respiración oxidativa. Pertenece a una clase de sustratos Tat que son proteínas integrales de membrana con un péptido señal no clivado que funciona como un anclaje a membrana y un dominio largo periplásmico. Recientemente se demostró que el sistema Tat es indispensable para una correcta integración del péptido señal de esta proteína en la membrana en la bacteria gram-negativa y patógeno intracelular *Legionella pneumophila* [135]. Aunque las bacterias patógenas de la familia *Anaplasmataceae* tendrían pocos sustratos del sistema Tat, no podemos descartar que la identificación esté subestimada debido a las frecuentes anotaciones incorrectas de los sitios de inicio de la traducción durante los procesos de anotación automática de los genomas bacterianos (Figura 1 y Figura 2, Capítulo III, Anexo).

Dos proteínas que fueron detectadas como positivas por solamente uno de los algoritmos (TATFIND) en la familia *Anaplasmataceae* son, la proteína de unión a penicilina 2 (*Penicillin-binding protein 2*, PBP2-COG0768) y la proteína *Rod shape-determining protein MreB* (MreB - COG1077) (Figura 3 y Figura 4, Capítulo III, Anexo). Las proteínas de unión a penicilina son una clase bien caracterizada de enzimas necesarias para el ensamble del peptidoglicano a partir de precursores sintetizados intracelularmente. En particular, la función de la proteína PBP2 se relaciona con los enlaces peptídicos cruzados y estaría implicada en la fase de alargamiento del crecimiento celular [136]. Además, se demostró que MreB se pliega en una estructura similar a la actina eucariota [137]. Teniendo en cuenta que la forma celular de las bacterias se piensa que está determinado por la capa de la pared celular de peptidoglicano, se ha propuesto que los filamentos de MreB podrían funcionar como un citoesqueleto que sirve de organizador o un dispositivo de rastreo para el complejo de la biosíntesis de PBP-2 peptidoglicano en *Caulobacter crescentus* [138]. Por otro

lado, *A. marginale* reúne un haz de filamentos de actina durante la infección intracelular, y las observaciones estructurales han demostrado que la organización de filamentos de actina es una característica de esta bacteria que infecta eritrocitos. En *A. marginale* una proteína asociada al apéndice (AAP), sin ortólogos conocidos en otros patógenos [139], ha sido implicada en la formación del apéndice. Sin embargo, parece poco probable que una sola proteína pueda ser responsable del montaje completo de los apéndices de actina. Se detectó un péptido señal Tat sólo en MreB de *A. marginale* spp. y en PBP2 sólo en *A. marginale* y *Wolbachia*. Se verificó la ausencia de ortólogos de PBP2 en *Ehrlichia* spp, *Neorickettsia* y *Anaplasma phagocytophilum* por búsquedas BlastP y TblastN. En conjunto, aunque es necesario realizar diversos experimentos, MreB y PBP2 a través de la vía Tat podrían ser candidatos potenciales a ser considerados como parte del mecanismo de invasión mediado por el apéndice.

Estos resultados y observaciones presentados proveen nuevos resultados, evidencias y enfoques para la caracterización del sistema Tat y de nuevas proteínas potencialmente secretadas por esta vía para desentrañar y comprender el rol del sistema en otorgar capacidades de adaptación y estrategias de infección.

Capítulo III.

Estudio del impacto de la reducción genómica en la organización de operones.

Objetivo específico.

Estudiar de la organización y distribución de operones en genomas de la clase α -proteobacteria para comprender la relación entre la reducción del genoma en bacterias del orden Rickettsiales y su impacto en la organización de operones.

3.1. Antecedentes

3.1.1 El cromosoma bacteriano y los modelos para explicar la existencia de operones.

3.1.2 La reducción del genoma y el destino de los operones.

3.2. Resultados.

3.3. Discusión.

3.1. Antecedentes

3.1.1 El cromosoma bacteriano y los modelos para explicar la existencia de operones.

Un patrón conspicuo en los cromosomas bacterianos es la existencia de operones multigenicos descritos hace algunas décadas [140, 141], sin embargo aún hoy se encuentran en estudio y discusión diferentes hipótesis para explicar los mecanismos que influyen en ese tipo de patrón de organización. Muchos procesos celulares y biológicos interactúan con el cromosoma bacteriano y esa interacción es considerada por diferentes autores como una de las principales fuerzas que determinan la estructura del genoma y su organización [142]. La interacción ocurre en diferentes escalas: **i)** local, influyendo sobre la expresión génica, o **ii)** global, afectando la replicación; ambas determinan diferentes grados o niveles de organización como son los motivos de secuencia, repeticiones, operones, dominios, “islas”, macrodominios, plásmidos o, el recientemente definido, como *chromid*.

El cromosoma bacteriano, a diferencia de los eucariotas, es homogéneo en cuanto a la distribución de los genes en el genoma y usualmente los genes se cotranscriben en operones [142]. La publicación y disponibilidad de genomas bacterianos completos representan una fuente muy importante de información para comprender el rol de los genes en la organización del cromosoma. Una aproximación de genómica comparativa provee nuevas herramientas para estudiar la dinámica de organización de los genomas, el rol de los genes y de los procesos celulares, biológicos, ecológicos y su valor adaptativo en relación a los procesos evolutivos que determinan la organización de los genes en el cromosoma.

Inicialmente, los primeros trabajos describieron lo que resultaba aparente; que genes que codificaban para enzimas relacionadas tendían a estar próximos o adyacentes en el cromosoma bacteriano, y que más aún, el orden de los genes en el sentido transcripcional se correspondía con el orden de las enzimas en la ruta metabólica. Así fue descrito por Jacob y Monod en el año 1961 en la descripción y caracterización del operon *lac* [140]. Sin embargo, dentro de lo paradigmático del ejemplo, el operon *lac*, representa una rareza: por un lado el sistema está codificado por dos unidades transcripcionales (para el regulador y el operón *lacZYA*); por otro lado, codifica para un transportador y dos enzimas, poniendo en evidencia que la relación funcional no está limitada a la participación en una misma ruta metabólica. A su vez, el operon *lac* está presente solamente en algunas especies de *E. coli*, mientras

que no se encuentra en la gran mayoría de los genomas bacterianos secuenciados. Esto ejemplifica la multiplicidad de factores y preguntas que subyacen a las causas, existencia y mantenimiento de los operones como mecanismo de regulación.

Diferentes modelos plantean hipótesis para explicar las causas de la existencia de los operones. El **modelo regulatorio** (*regulatory model*), que posee mayores evidencias, se basa en que genes funcionalmente relacionados se acercan adaptativamente en el cromosoma por las ventajas que ofrecería la co-regulación [143]. La concatenación de los genes río debajo de una misma secuencia regulatoria, como ocurre en los operones, es una estrategia de *todo-o-nada*, ya que si ocurre una mutación en la secuencia regulatoria o el promotor falla, deja de expresarse el conjunto de los genes, y sólo la expresión del conjunto tiene valor adaptativo [142]. Por otro lado, la transferencia exitosa del operon a otro organismo mediante mecanismos de transferencia lateral del operon (HGT; *horizontal gene transfer*), también depende de la transferencia de las secuencias regulatorias. En bacterias, la transcripción está acoplada a la traducción, y la co-transcripción también permite, o favorece, la posibilidad de co-localizar la traducción lo que se ha demostrado que resulta ser adaptativo, por ejemplo, para complejos que pueden comenzar a plegarse durante el proceso de co-traducción [144] o facilitando la colocalización conjunta en la célula [145]. Para genes que tienen que ser expresados en cantidades equivalentes, la co-transcripción disminuiría las diferencias de niveles de expresión y favoreciendo estequiometrías más precisas. La mayoría de los operones más conservados codifican para proteínas que interactúan entre sí [146]. Otro aspecto que sustenta el modelo regulatorio sería que la degradación conjunta de ARNm policistrónico facilitaría el proceso de regulación y haría más eficiente la expresión génica [142]. Un modelo alternativo, denominado “operon egoísta” (***operon selfish model***) [147] es más controversial y propone que los genes se agrupan en estructura de operon como una forma de hacer más eficiente la capacidad para transferirse como genes funcionalmente relacionados. En bacterias, los genes se transfieren como un mecanismo habitual mediante HGT, y un clúster de genes funcionalmente relacionado aumenta las probabilidades de transferir un módulo de funciones completa. Recientemente ha surgido un nuevo, denominado ***persistent model*** [148] que propone la formación de operones para genes esenciales y con alto valor adaptativo, de manera tal que al ocupar un menor espacio en su distribución en el cromosoma, disminuyen la probabilidad de verse afectados por mutaciones de inserción y deleciones en el cromosoma [148]. En este caso, organismos con los genes más compactos, y organizados en clústers adquieren mayor resiliencia a los efectos letales o deletéreos de las mutaciones. Este último modelo se enmarca en un estudio que

propone que genes que comparten secuencias regulatorias (al estar organizados en un operon) disminuyen las probabilidades de mutaciones en relación a genes dispersos con secuencias individuales de control (gen-específicas) que aumentarían de esta manera la secuencias blanco y aumentando entonces las probabilidades de mutaciones y cambios deletéreos que podrían conllevar a la pérdida de la funcionalidad [149]. Este modelo propone que los mecanismos generales y básicos de la genética de poblaciones son las fuerzas predominantes que moldean la estructura y organización de los genomas, en contraposición con los que ponen el énfasis en la importancia de los mecanismos y procesos funcionales y metabólicos.

3.1.2 La reducción del genoma y el destino de los operones.

Los organismos con genomas reducidos en asociación con una transición a ambientes intracelulares, se caracterizan por haber sufrido procesos de rápida evolución, desvíos en la composición nucleotídica y re- anotación de codones [150]. A su vez los genomas pequeños, sufrieron, y algunos actualmente sufren, procesos de pérdidas masivas de genes; mientras que otros genes esenciales para la interacción con las células hospedadores se mantienen. Las mutaciones y la deriva génica, en un contexto de poblaciones de bajo tamaño efectivo y reproducción asexual estricta en algunos casos, son los principales motores de esa rápida evolución, donde las deleciones o mutaciones deletéreas aumentan la probabilidad de fijarse en un contexto de una selección natural menos eficiente [151] (**Figura 3.1**).

Inicialmente se había descrito que las bacterias de genomas pequeños eran primitivas, pero posteriormente basados en análisis filogenéticos se comprobó posteriormente que las bacterias con genomas reducidos (por ejemplo miembros de la familia *Rickettsiaceae*) derivan en realidad de bacterias con genomas grandes [150]. Con la secuenciación masiva de genomas, se corroboró esto y también se describieron diferentes procesos independientes de reducción genómica convergente en diferentes clados diversos, incluyendo: Mollicutes, α -proteobacteria, γ -Proteobacteria y Spirochaetas [14, 152].

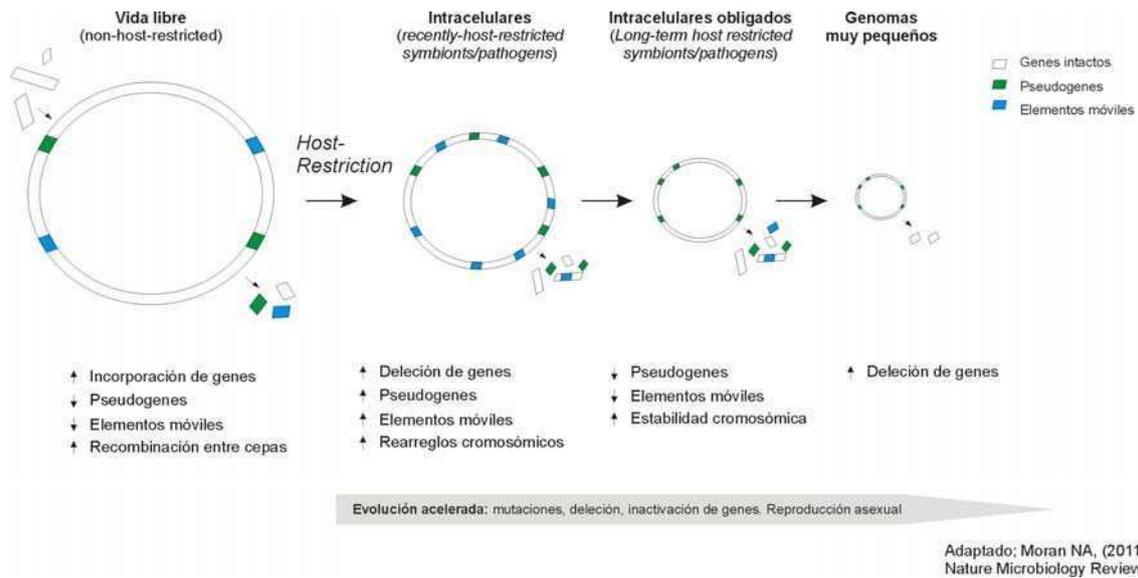


Figura 3.1. Etapas de la reducción genómica en bacterias.

Esquema adaptado del trabajo de Mc Cutcheon y Moran NA [150] describiendo las etapas del proceso de reducción genómica en organismos intracelulares, obligados o facultativos, con ambientes limitados, bajos tamaños efectivos poblacionales y reproducción asexual.

Durante los últimos años diversos trabajos abordaron por un lado, el estudio de los mecanismos que influyen en los procesos de formación de operones en el genoma bacteriano analizando la organización y distribución de los genes en el genoma, como también el rol de la persistencia, de la esencialidad y de la co-regulación de vías metabólicas en mantener a los genes en clúster. En particular, escasos trabajos estudiaron la relación entre el tamaño del genoma y las estrategias regulatorias, incluyendo a los operones. Fundamentalmente, y considerando que la reducción de los genomas está íntimamente asociada a procesos de pérdida masiva de genes, a la recombinación y rearrreglos genómicos, estos procesos podrían tener una influencia directa sobre la organización de los genes, y por ende, en sus estrategias de regulación. En este contexto, diferentes factores y variables podrían tener un valor adaptativo para influir en la conservación o no de los operones como estrategia de regulación; mientras que también podrían existir o preponderar otros mecanismos de regulación alternativa a los operones. Por otro lado, para complejos de proteínas o mecanismos biológicos donde se ha descrito que la organización en operon contribuye a garantizar la regulación conjunta o es relevante para la funcionalidad, es muy importante conocer el impacto que podría tener un proceso de potencial fragmentación en la funcionalidad de sistemas o complejos proteicos que presentan organizaciones

dispersas, que podrían o bien perder la funcionalidad, o presentar mecanismos alternativos de regulación.

Los casos particulares estudiados en la Capítulo II como la fragmentación del operon *tat*, conjuntamente con los ejemplos descritos recientemente en la bibliografía asociados al sistema T4SS [79] y al operon *sdh* [130] y su fragmentación en organismos con genomas pequeños del orden Rickettsiales (**Figura 3.2**), plantean una serie de preguntas acerca de cómo evolucionan los operones en bacterias intracelulares que sufrieron procesos de reducción genómica y nos permiten plantear la hipótesis de la existencia de un proceso generalizados de fragmentación de operones en estos organismos con genomas chicos. La conservación de la funcionalidad del sistema Tat como se demostró en el capítulo anterior, pese a su completa desorganización en relación a la organización característica, sugiere en este caso la adaptación a mecanismos de regulación alternativos gen-específicos.

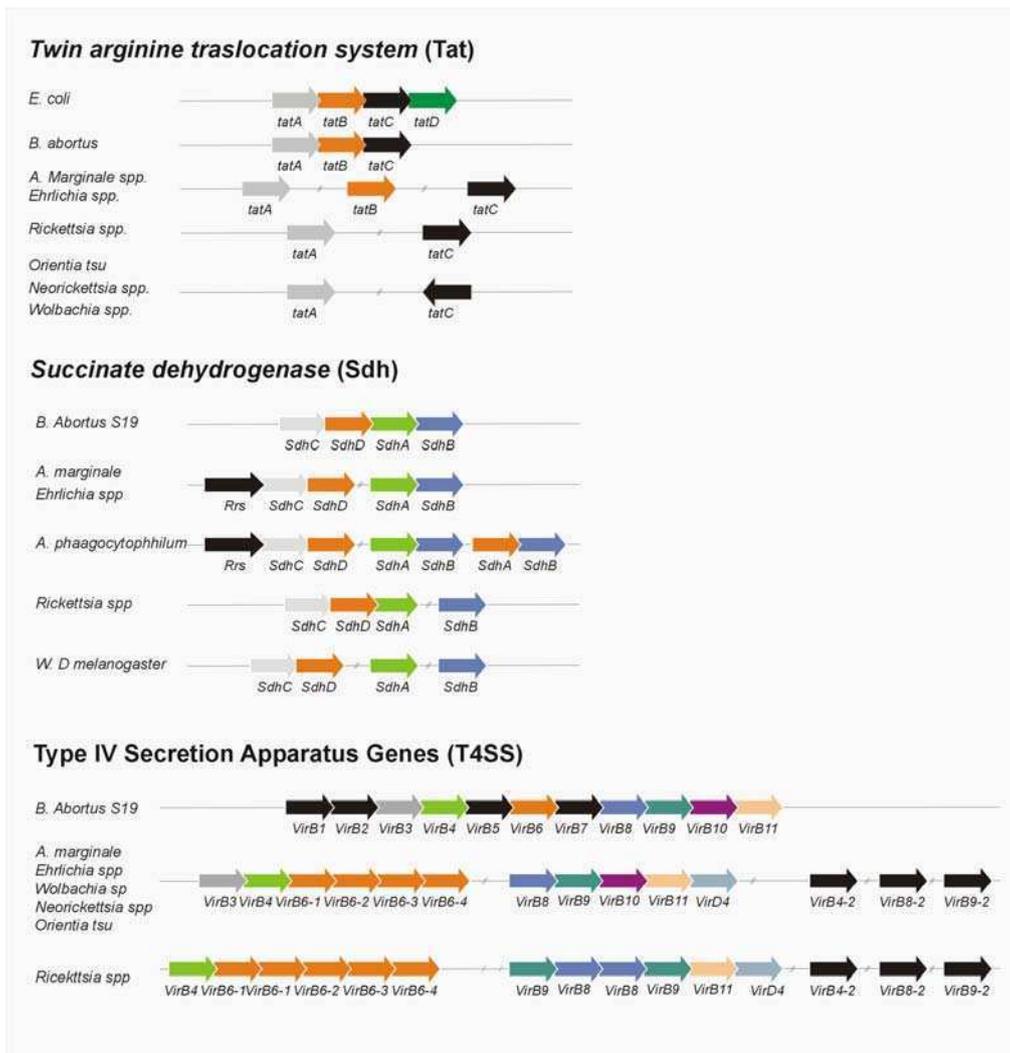


Figura 3.2. Esquema de casos de fragmentación de operones en el orden Rickettsiales.

Esquema de organización de los operones Tat, sdh y T4SS en diferentes organismos del orden Rickettsiales y de la clase α -proteobacteria. Se observa que en los tres casos los organismos intracelulares obligados, simbiotes o patógenos, del orden Rickettsiales presentan estructuras fragmentadas para los operones. En el caso del operon Tat se observa en algunos organismos que los genes se ubican en diferentes hebras del ADN; en otros casos, además de operones fragmentados se observan duplicaciones o deleciones en los genes de los complejos.

Escasos trabajos abordaron durante los últimos años la relación entre el tamaño del genoma y los operones. En un estudio realizado en año 2003 [153] en el que se analizaron 26 genomas bacterianos disponibles pertenecientes a diferentes clases de bacterias observaron una correlación positiva entre el tamaño del genoma y el número de unidades transcripcionales simples (monocistrónicas), donde los genomas grandes tendrían menos genes organizados en operones. El modelo denominado **selfish operon**, propone que la organización en operones se relacionaría con una facilidad a ser transferidos mediante HGT. A su vez los genomas grandes sufrieron una expansión genómica, en parte, por transferencia de genes mediante HGT [147]. Entonces se podría esperar que genomas grandes contengan una mayor proporción de genes en operon, sin embargo, los autores encontraron lo opuesto. Por el contrario, si los operones contribuyeran a facilitar la regulación conjunta en los genomas grandes que codificarían para una red más compleja de circuitos regulatorios, se podría esperar una mayor frecuencia de operones. Sin embargo, al observar lo contrario bajo las condiciones y métodos de este estudio, los autores aluden a que en los genomas grandes la organización predominante podría ser mediante elementos de control individuales, o gen-específica. En otro trabajo [154] que estudió la complejidad transcripcional en genomas pequeños, para *Mycoplasma pneumoniae*, uno de los genomas más pequeños secuenciados hasta el momento, los autores identificaron una inusual y no descripta complejidad (alto porcentaje de transcritos alternativos, ARNm antisentidos y la heterogeneidad en los niveles de expresión entre los genes de un operon), que no puede ser explicado por los 8 factores de transcripción codificados en el genoma, sugiriendo la existencia de mecanismos y procesos regulatorios complejos, aún en organismos aparentemente sencillos en términos de tamaño del genoma. Únicamente, estudiando la media del largo de los operones (promedio de número de genes por operon) para organismos de diferentes grupos filogenéticos, los autores describieron una tendencia global donde la reducción del genoma conlleva a operones más largos acoplando genes con funciones diferentes. Sin embargo, las evidencias para esta afirmación no se encuentran

desarrolladas y analizando las distribuciones para los diferentes grupos filogenéticos se observan importantes diferencias que no fueron estudiadas [155].

Un estudio que analizó la relación entre el contenido génico (analizando los COGs, Clúster of Orthologs Groups) y el tamaño del genoma, identificó distintas categorías que describen la evolución de los principales procesos biológicos y su relación con los procesos de expansión y reducción genómica [156]. Describieron que los genomas grandes presentan una alta proporción de genes pertenecientes a categorías de regulación y metabolitos secundarios, en comparación con genomas medianos y chicos. Este trabajo confirmó estudios previos [157] que indicaban que los genomas chicos preferencialmente perdieron elementos regulatorios (factores σ) presuntamente relacionados con un ambiente intracelular más estable y homogéneo, que disminuye el valor adaptativo de una red de regulación compleja para los genes.

Para esto nos propusimos estudiar si el fenómeno de fragmentación de operones es un proceso común en organismos que sufrieron reducción genómica, y evaluar diferentes hipótesis asociadas a las características de los genes y genomas y su influencia en la conservación. Con el fin de hacer un análisis comparativo de la conservación de la estructura de operon se analizaron 81 organismos de la α -proteobacteria (**Tabla S3.1, Anexo**) por considerarlo un grupo diverso de bacterias gram-negativas en sus tamaños del genomas como también en sus propiedades biológicas, ofreciendo un excelente sistema modelo para estudiar las fuerzas, los mecanismos y las tasas por el cual los genomas de bacterias evolucionan.

Conocemos como los reguladores evolucionaron en estos organismos por trabajos previos, por lo que nos planteamos estudiar como otro factor fundamental de la regulación, los operones, evolucionaron. En particular, nos propusimos poner a prueba la hipótesis de un patrón de fragmentación de operones en los genomas reducidos para aproximarnos a las respuestas para las preguntas de las siguientes preguntas, ¿Cómo evolucionan los operones en bacterias intracelulares? ¿Cuán generalizable es el fenómeno de dispersión de operones en bacterias con genomas reducidos? ¿En qué medida esto puede ser consecuencia del proceso de reducción del genoma? ¿Operones “fragmentados” se mantienen como sistemas funcionales? ¿Existen mecanismos alternativos de regulación que permitan una regulación conjunta?

3.2. Resultados.

Análisis exploratorios de la distribución de operones en la clase α -proteobacteria.

Con el objetivo de analizar la organización y distribución de operones en los organismos seleccionados de acuerdo a su representación en los diferentes ordenes de la clase y su tamaño del genoma (**Figura 3.3, Tabla S3.1**) se analizó en relación al genoma, el número de genes, de proteínas, de operones y el porcentaje de genes en operon o como unidades transcripcionales simples.

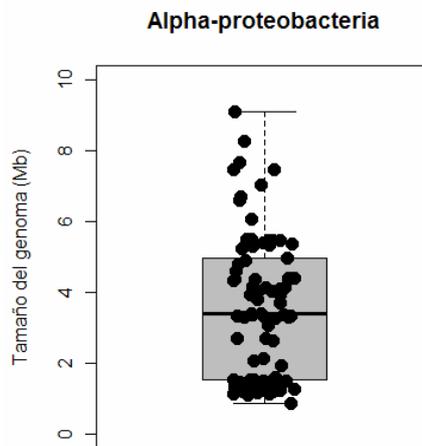


Figura 3.3. Boxplot tamaño de genomas de organismos.

Distribución de los tamaños de genomas de los 81 organismos de la clase α -proteobacteria analizados en el estudio. Las barras representan los valores mínimos y máximos; la caja gris los valores entre los cuartiles 25-75%, y la línea, la mediana. La identificación de los organismos se encuentra en la **Tabla S3.1** (Anexo).

Se observó una correlación positiva entre el tamaño del genoma y el número de genes, proteínas, operones y con el porcentaje de genes en operon. Estas relaciones eran las esperadas de acuerdo a genomas reducidos con un menor número de genes (**Figura 3.4, Tabla S 3.4**).

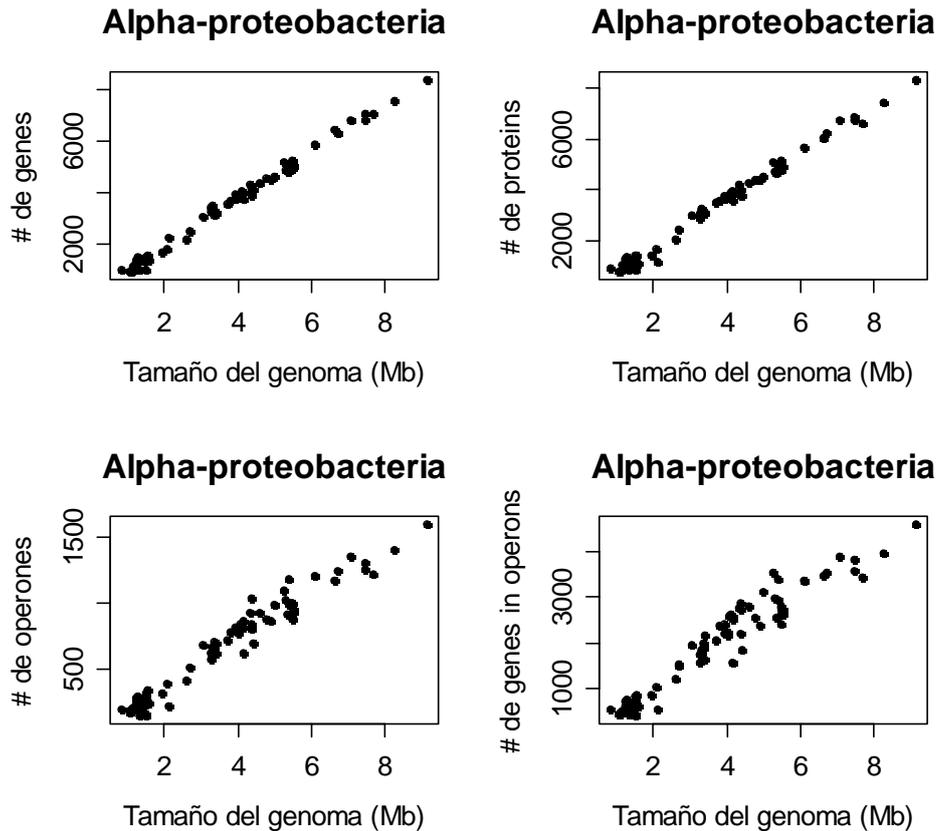


Figura 3.4. Análisis exploratorios de características de los genomas.

Para cada uno de los 81 genomas de la clase α -proteobacteria se estudió en relación al tamaño del genoma: **(A)** número de genes; **(B)** número de proteínas; **(C)** número de operones; **(D)** número de genes en organización de operon. La información de cada genoma se descargó de la base de datos del NCBI (archivos ".rtp") y la información de los operones predichos de la base de datos DOOR. Para obtener la información de los archivos para el conjunto de organismos se diseñaron scripts en bash para procesar la información. número: Número.

Para estudiar un uso diferencial de la organización de los operones como estrategia regulatoria en los organismos de la clase, calculamos la proporción de genes en organización en operon y dispuestos en unidades transcripcionales simples y su relación con el tamaño del genoma. Al número total de genes codificados en el genoma, se le restó el número de genes predichos en organización de operon.

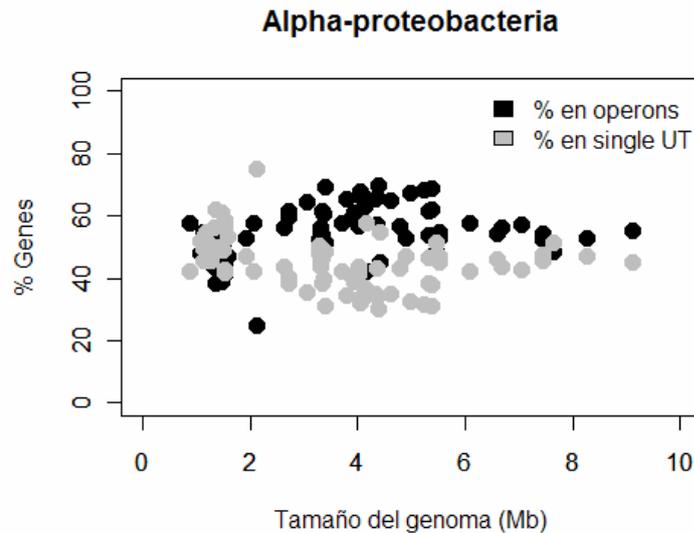


Figura 3.5. Genes en operones y en unidades monocistrónicas.

Porcentaje de genes en organización de operon y en unidades transcripcionales simples (single UT) en los 81 organismos de clase α -proteobacteria. Se observó una correlación positiva ($\rho = 0.36$; $P = 0.0007424$) para el % de genes en operon en relación con el tamaño del genoma.

Se observó una correlación significativa donde los genomas pequeños presentan una mayor proporción de genes en unidades monocistrónicas, y una menor proporción de genes en organización de operon en relación a los genomas grandes (**Figura 3.5, Tabla S 3.4**). Este resultado sugiere que los genomas chicos de esta clase presentan un menor uso de los operones como estrategia de organización, en relación a los genomas más grandes.

Posteriormente analizamos el largo promedio de los operones en los organismos de la clase abarcados en el estudio. Para esto se calcularon las medidas de estadística descriptiva (media, mediana, moda) a partir de los archivos completos de la información de los operones. La información completa puede observarse en la **Tabla S 3.4**.

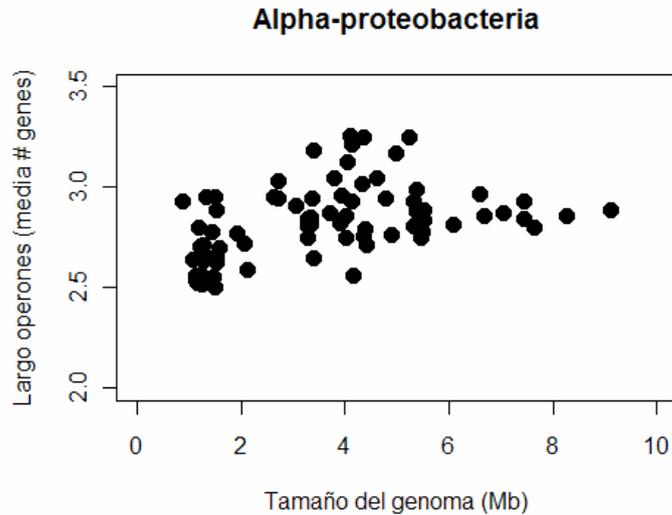


Figura 3.6. Largo de los operones.

Promedio de genes por operon en los 81 organismos de clase α -proteobacteria. Se observó una correlación positiva ($\rho = 0.43$; $P = 6.197 \cdot 10^{-5}$) para el largo de los operones en relación con el tamaño del genoma.

Como se desprende de la **Figura 3.6** los genomas pequeños presentan operones más cortos, mostrando una correlación positiva entre el número promedio de genes por operon y el tamaño del genoma.

Análisis de la conservación de pares de genes en operon.

Para estudiar la hipótesis de fragmentación de operones en genomas pequeños, analizamos el porcentaje de conservación de pares de genes en operon analizando el conjunto total de genes ortólogos para cada una de las comparaciones de a pares entre *E. coli* y los organismos estudiados (ver detalle de los procedimientos para los cálculos en materiales y métodos; tabla de resultados en Anexo **Tabla S 3.5**).

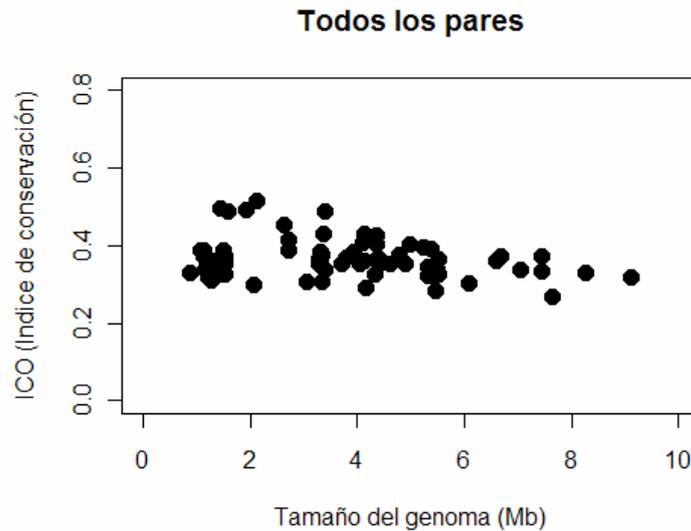


Figura 3.7. Porcentaje de conservación de pares de genes en operon.

Se analizó el % de conservación de operones entre cada organismo y *E. coli*. Se analizó en cada una de las comparaciones el conjunto completo de ortólogos de acuerdo a las matrices generadas y detalladas en materiales y métodos. Se estudio la correlación de las variables *Spearman's rank correlation* ($\rho = -0.13229$, $P = 0.2391$) que no mostró ser significativa.

No se observó una correlación significativa entre el tamaño del genoma y la conservación de pares de genes en operon. Sin embargo, se observó una tendencia que muestra que los genomas pequeños presentarían un mayor porcentaje de conservación al analizar el conjunto de genes ortólogos en cada organismo y *E. coli* (**Figura 3.7**). Esta tendencia, iría en contra de la hipótesis planteada, donde esperaríamos encontrar, de acuerdo a las observaciones anteriores, un porcentaje mayor de pares de genes fragmentados.

Para profundizar e intentar responder algunas preguntas entre la relación entre el contenido génico de los operones y la presión de selección para mantenerlos o fragmentarse, e hipotetizar acerca de los procesos biológicos y evolutivos subyacentes, estudiamos como afectan diferentes propiedades de los genes como la persistencia, los niveles de proteínas en las células, los niveles de expresión y la característica de esencialidad. Por lo tanto, para comprender mejor el proceso, en primer lugar, desglosamos el conjunto total de genes ortólogos para cada comparación, de acuerdo a sus niveles de persistencia. Se analizó la conservación diferenciando los pares de genes persistentes (presentes en >90% de los genomas) y No-persistentes (<50% de los genomas).

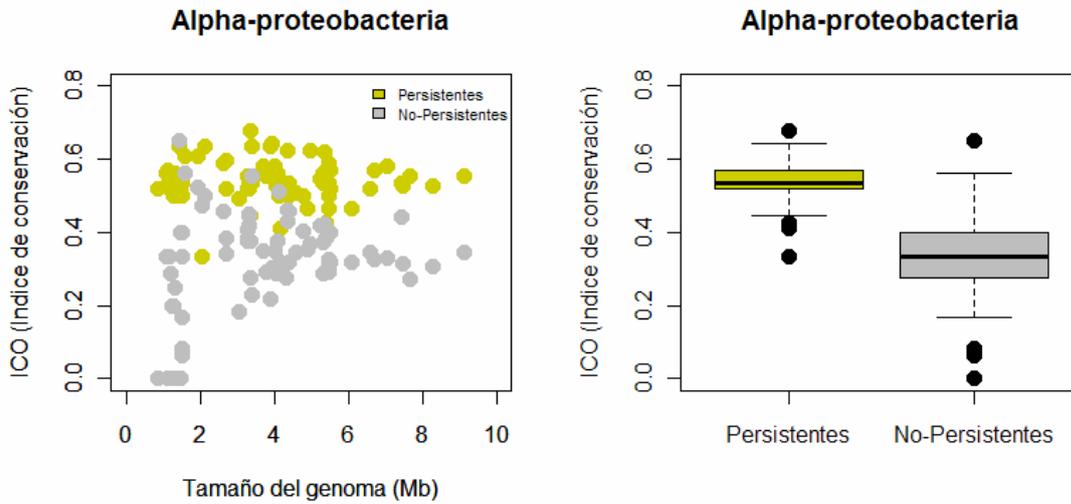


Figura 3.8. Influencia de la persistencia en la conservación de operones.

Se analizó el % de conservación de operones entre cada organismo y *E. coli* para los diferentes conjuntos de ortólogos (Persistentes vs. No-persistentes). Las matrices generadas para el conjunto de ortólogos en cada comparación se dividieron de acuerdo a los niveles de persistencia de los genes. Se estudió la correlación de las variables (*Spearman's rank correlation*): Tamaño del genoma vs. Persistentes $\rho = -0.0542899$, $P = 0.6303$, Tamaño del genoma vs. No-Persistentes $\rho = 0.2613815$, $P = 0.01842$. El análisis de Paired t-test (Persistent vs. Non-persistent, $P < 2.2e-16$) mostró diferencias significativas para el comportamiento de los diferentes grupos de genes en relación con el tamaño del genoma.

Se observó un patrón de conservación diferente para los diferentes sets de ortólogos. Los genomas pequeños mostraron una mayor conservación para los genes persistentes, pero una menor conservación para los operones compuestos por genes menos persistentes sugiriendo que el rol de los genes es relevante. En relación al tamaño del genoma, se observó una correlación positiva con el porcentaje de conservación de operones compuestos por genes No-persistentes. Este resultado sugiere que en los genes persistentes que forman parte del *core genome* están bajo presión de selección para mantenerse en operon, pero las funciones más accesorias no lo están. Por otro lado, en conjunto, los operones compuestos por genes persistentes presentaron mayores porcentajes de conservación que los No-persistentes (Figura 3.8).

Influencia de los niveles de proteínas.

Posteriormente estudiamos si los niveles de proteínas para los cuales codifican los genes involucrados en el operon podrían tener un impacto, teniendo en cuenta la hipótesis planteada por diferentes autores denominada **balance hypothesis** donde postulan que el desbalance de la concentración de los niveles de proteínas de los componentes de un complejo proteína-proteína pueden ser deletéreos para las células [158], donde la sub- o sobre expresión disminuye el fitness.

La hipótesis de estudio en este punto particular fue que los operones que codifican para niveles de proteínas desbalanceados (y que pueden requerir regulación más fina) se fragmentan más que operones que codifican para similares niveles de proteínas.

La distribución de los niveles de proteínas para los pares de genes de *E. coli* se muestra en la **Figura 3.9**. Como se observa, la frecuencia de pares de genes para los diferentes niveles esperables de acuerdo a los experimentos de proteómica analizados son relativamente constantes. Para considerar niveles de proteínas similares o diferentes consideramos como punto de corte la mediana de la distribución.

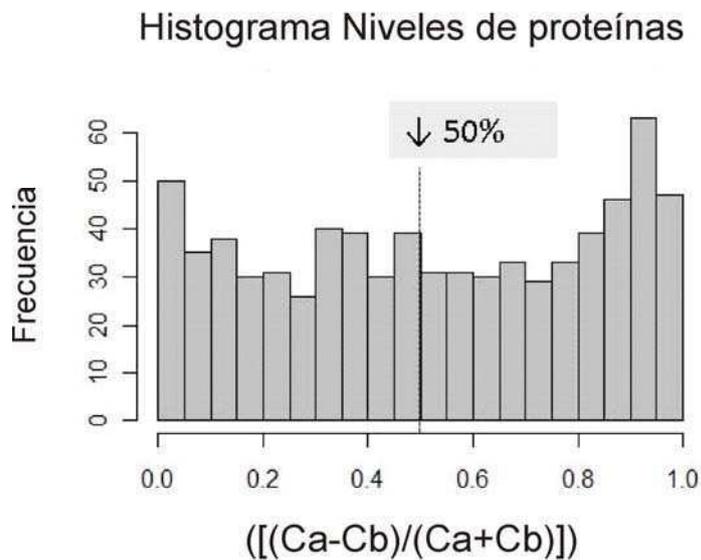


Figura 3.9. Niveles de Proteínas.

Distribución de los niveles de proteínas para cada par de genes en operon para *E. coli*. Con la información de la concentración (**C**) determinada para el par de genes **a** y **b**, se calculó la relación $\left[\frac{Ca-Cb}{Ca+Cb}\right]$ para estimar la diferencia de niveles considerando la magnitud absoluta.

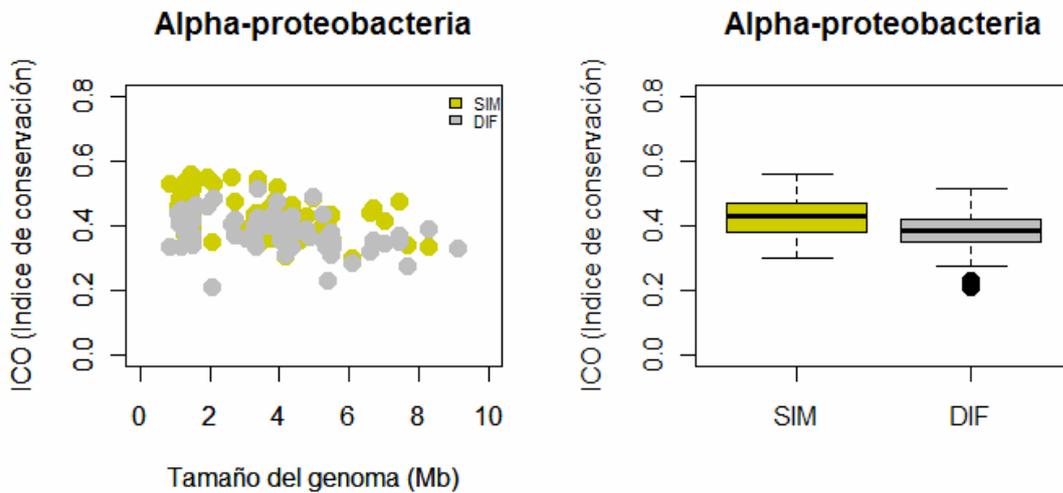


Figura 3.10. Influencia de los niveles de proteínas celulares en la conservación de operones.

Se analizó el % de conservación de operones entre cada organismo y *E. coli* para los diferentes conjuntos de ortólogos (con niveles de proteínas similares (SIM) vs. diferentes (DIF). Las matrices generadas para el conjunto de ortólogos en cada comparación se dividieron de acuerdo a los niveles de proteínas codificadas por los genes. Se estudió la correlación de las variables (*Spearman's rank correlation*): Tamaño del genoma vs. SIM: $\rho = -0.4759469$, $P = 7.097e-06$, Tamaño del genoma vs. DIF: $\rho = -0.454259$, $P = 2.047e-05$, donde ambas distribuciones tuvieron una correlación significativa con el tamaño del genoma. El análisis de *Paired t-test* (SIM vs. DIF, $P < 1.721e-11$) demostró diferencias significativas para el comportamiento de los diferentes grupos de genes en relación con el tamaño del genoma.

Si bien ambos conjuntos presentaron la misma tendencia, no pudiendo asignar una relación diferencial para el tamaño del genoma, los pares de genes que presentan niveles de proteínas similares en la célula, demostraron presentar mayores porcentajes de conservación que los pares de genes que tienen niveles proteicos desbalanceados, en acuerdo con la hipótesis planteada (**Figura 3.10**).

Influencia de los niveles de expresión.

Teniendo en cuenta que los operones están directamente relacionados con los niveles de expresión de los genes, estudiamos si los niveles de expresión tenían una influencia sobre la conservación de los genes en operon. De acuerdo a nuestra hipótesis esperamos una mayor conservación de operones que contengan genes de alta expresión.

La distribución de los niveles de expresión para los pares de genes de *E. coli* se muestra en la **Figura 3.11**. Como se observa la frecuencia de pares de genes con niveles de expresión alto es menor que los de baja expresión. Para considerar niveles de expresión alta (HE) o niveles de expresión baja (LE) consideramos como punto de corte la mediana de la distribución.

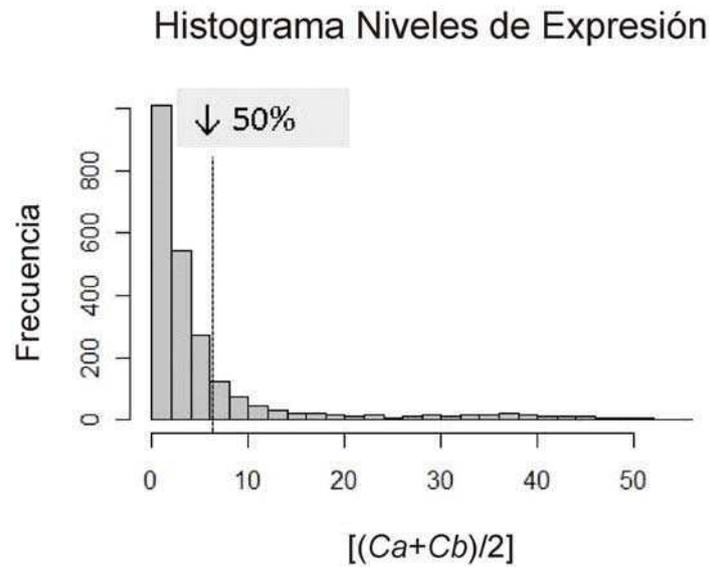


Figura 3.11. Niveles de expresión.

Distribución de los niveles de expresión de los pares de genes en operon para *E. coli*. Con la información de la cantidad (**C**) estimada para los niveles de ARNm del par de genes **a** y **b**, se calculo la relación el promedio $[(Ca+Cb) / 2]$ para determinar el nivel de expresión del par de genes a analizar.

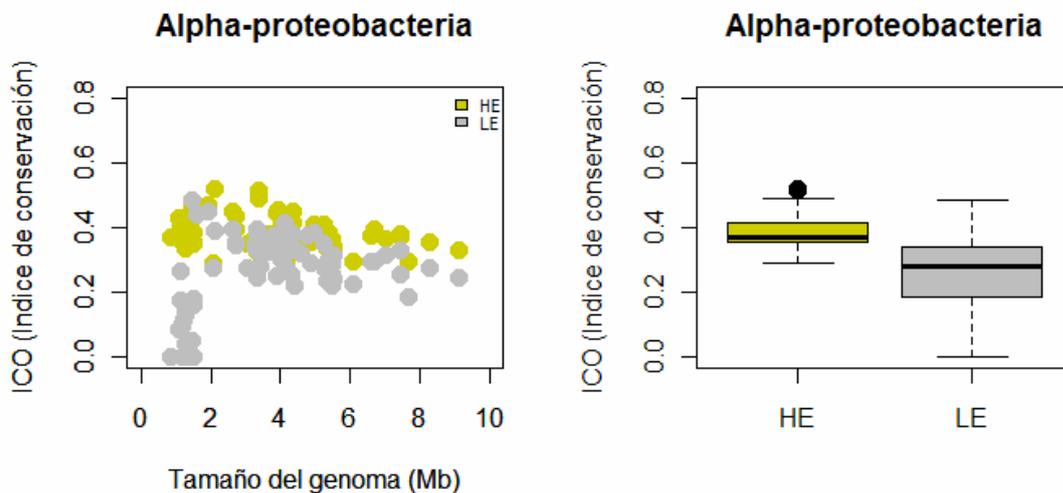


Figura 3.12. Influencia de los niveles de expresión de los genes en la conservación de operones.

Se analizó el % de conservación de operones entre cada organismo y *E. coli* para los diferentes conjuntos de ortólogos con niveles de expresión altos [HE] vs. bajos [LE]. Las matrices generadas para el conjunto de ortólogos en cada comparación se dividieron de acuerdo a los niveles de expresión del par de genes. Se estudió la correlación de las variables (*Spearman's rank correlation*): Tamaño del genoma vs. HE: $\rho = -0.3519388$, $P = 0.001273$, Tamaño del genoma vs. LE: $\rho = 0.3049113$, $P = 0.005644$, donde ambas distribuciones tuvieron una correlación significativa con el tamaño del genoma. El análisis de Paired t-test (HE vs. LE, $P < 1.581e-14$) mostró diferencias significativas para el comportamiento de los diferentes grupos de genes en relación con el tamaño del genoma.

Los resultados mostraron diferentes patrones de conservación para los genomas pequeños; mostraron un mayor porcentaje de conservación para los pares de genes con niveles de expresión alto, y menor porcentaje de conservación para los bajos niveles de expresión. Al igual que para los resultados de genes No-persistentes, en relación al tamaño del genoma se observó una correlación positiva con el porcentaje de conservación de operones compuestos por genes de bajos niveles de expresión. Esto indicaría que este grupo de genes de con bajos niveles de expresión (LE-LE) tienen una alta frecuencia de fragmentación en genomas chicos. En conjunto, todos los pares de genes de alta expresión mostraron un mayor porcentaje de conservación en relación a los de baja expresión (**Figura 3.12**).

Influencia de los niveles de esencialidad.

Otra variable para estudiar si el contenido génico es relevante para la conservación de los operones fue la condición de genes esenciales (**Tabla 3.1**). De acuerdo a nuestra hipótesis esperábamos una mayor conservación de operones que contengan genes esenciales que no esenciales.

Tabla 3.1. Genes esenciales en <i>E. Coli</i>.	
Genes Esenciales	642
Genes no esenciales	3269
EE, NN y NE/EN pares de genes en operon en <i>E. coli</i>	
EE	368
NE/EN	734
NN	2507

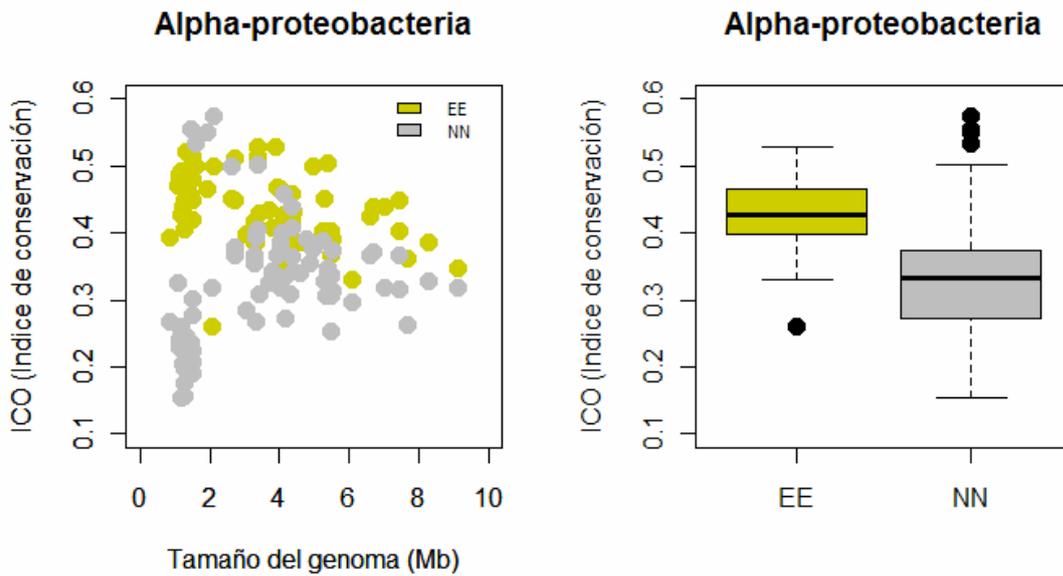


Figura 3.13. Influencia de la esencialidad de los genes en la conservación de operones.

Se analizó el % de conservación de operones (analizando de a pares de genes) entre cada organismo y *E. coli* para los diferentes conjuntos de ortólogos esenciales (EE) vs. No esenciales (NN). Las matrices generadas para el conjunto de ortólogos en cada comparación se dividieron de acuerdo a la característica de los genes. Se estudió la correlación de las variables (*Spearman's rank correlation*): Tamaño del genoma vs. EE: $\rho = -0.4862902$, $P = 4.173e-06$, Tamaño del genoma vs. NN: $\rho = 0.3429597$, $P = 0.001723$, donde ambas distribuciones tuvieron una correlación significativa con el tamaño del genoma. El análisis de Paired t-test (EE vs. NN, $P < 3.471e-14$) mostró diferencias significativas para el comportamiento de los diferentes grupos de genes en relación con el tamaño del genoma.

Al igual que para el análisis de los niveles de expresión, los genomas pequeños evidenciaron un comportamiento diferente en la conservación de los operones compuestos por genes esenciales o no esenciales. Se observó que en los genomas pequeños, los pares de genes conformados por genes “esenciales” se conservan más que los pares de genes conformados por genes “no esenciales”. A su vez, para ambos grupos de genes, se observó una correlación significativa con el tamaño del genoma y de signos opuestos para ambos grupos de genes (**Figura 3.13**).

Análisis de conservación en diferentes clados de bacterias.

Utilizando los mismos programas y scripts desarrollados para estudiar la distribución y comportamiento de los operones en las α -proteobacterias, realizamos el mismo análisis en otros grupos filogenéticos compuesto por organismos con

variaciones en el tamaño de los genomas. Se realizó este análisis para analizar las mismas variables en otros clados, y evaluar la consistencia de los resultados.

Para esto realizamos el mismo análisis en las clases β -proteobacteria y Firmicutes (excluyendo a los organismos pertenecientes a Tenericutes: *Mycoplasma* spp. que tienen genomas extremadamente pequeños y presentan muchas particularidades que podrían afectar la generalidad del análisis). Los organismos se seleccionaron de modo de cubrir la amplitud de los diferentes tamaños de genomas, representando las diferentes especies y evitando sobre representar las especies con genomas de diferentes cepas.

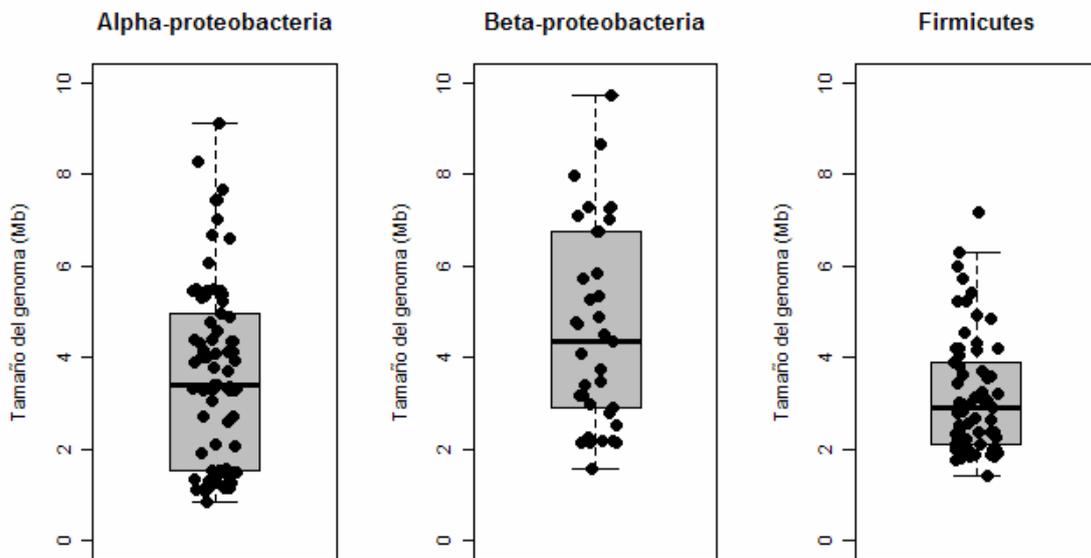
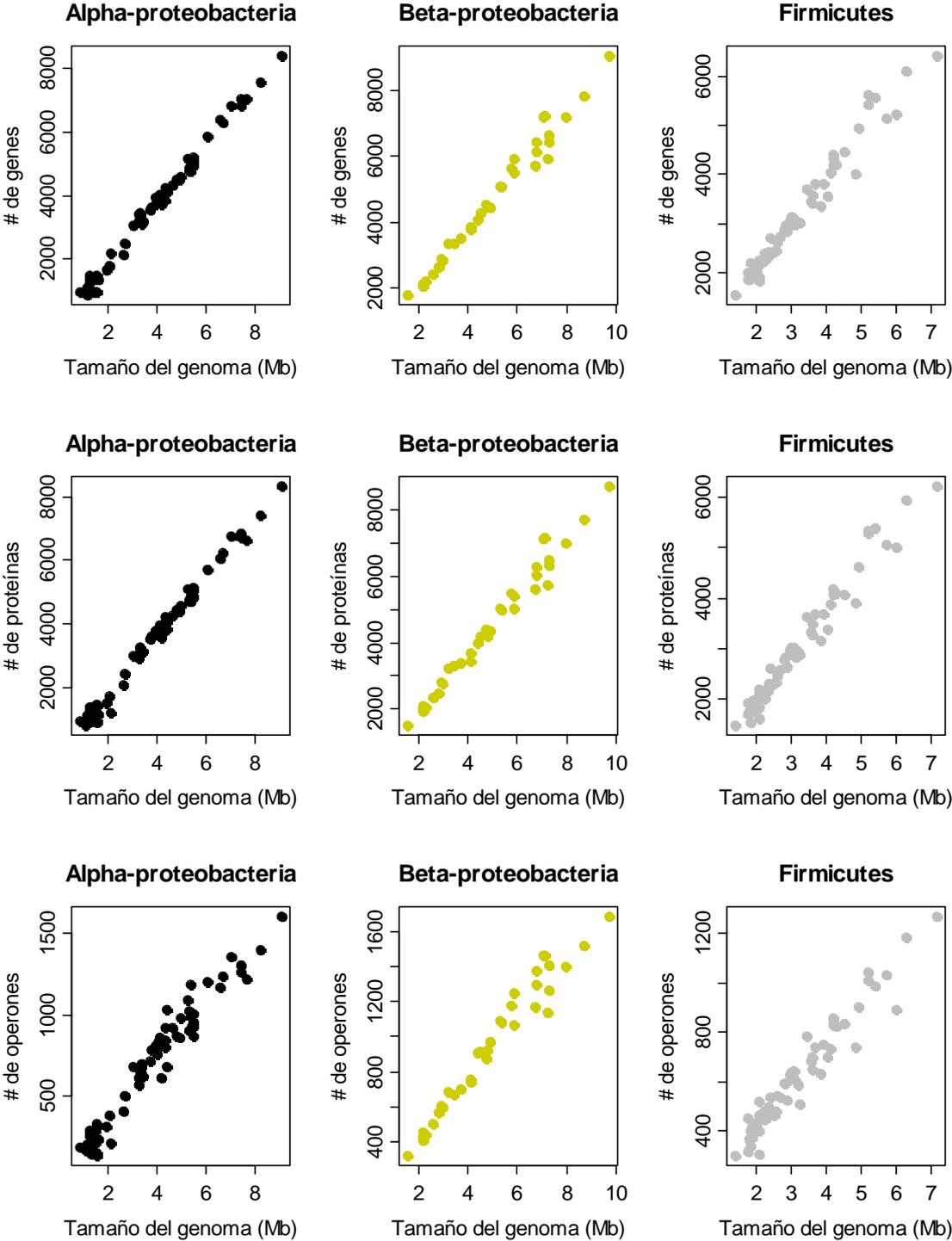


Figura 3.14. Tamaño de genomas.

Distribución de los tamaños de genomas de los 81 organismos de la clase α -proteobacteria, 37 organismos de la clase β -proteobacteria y 65 de la clase Firmicutes analizados en el estudio (Anexo, Capítulo IV). Las barras representan los valores mínimos y máximos; la caja gris los valores entre los cuartiles correspondientes al 25-75%, y la línea, corresponde a la mediana. La identificación de los organismos de la clase β -proteobacteria se detalla en la **Tabla S3.2**. La identificación de los organismos del phylum Firmicutes se detalla en la **Tabla S3.3**. Los análisis filogenéticos para los diferentes grupos se ilustran en las **Figuras S3.1 y Figuras S3.2**, respectivamente (Anexo).



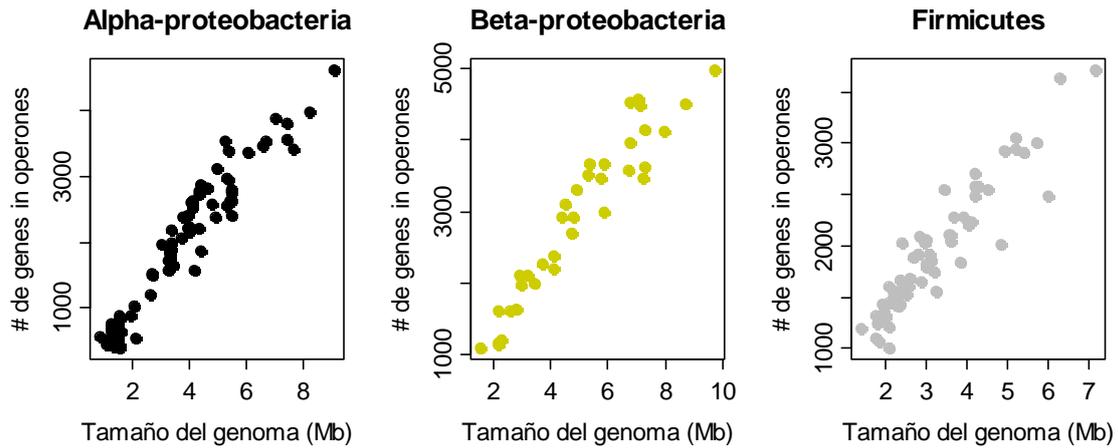


Figura 3.15. Análisis exploratorios de características de los genomas.

Para cada uno de los 81 organismos de la clase α -proteobacteria, 37 organismos de la clase β -proteobacteria y 65 de la clase Firmicutes se analizó: **(A)** número de genes; **(B)** número de proteínas; **(C)** número de operones; **(D)** número de genes en organización de operon. La información de cada genoma se descargó de la base de datos del NCBI (archivos “.rtp”) y la información de los operones predichos de la base de datos DOOR.

En los tres grupos analizados, se observaron las correlaciones esperadas para el número de genes, proteínas, operones y genes en operon en relación al tamaño del genoma (**Figura 3.15**).

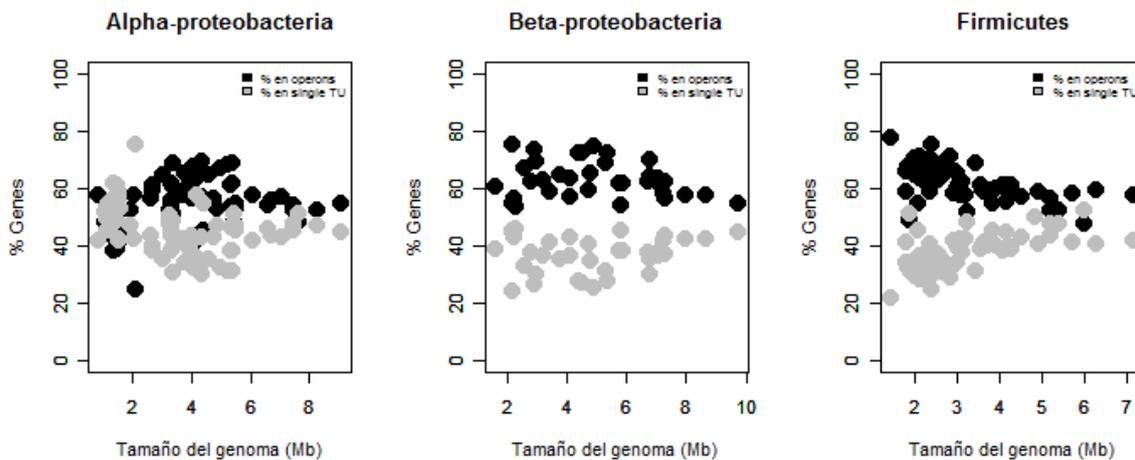


Figura 3.16. Genes en operones y en unidades monocistrónicas.

Porcentaje de genes en organización de operon y en unidades transcripcionales simples (single UT) en los 81 organismos de clase α -proteobacteria, 37 organismos de la clase β -proteobacteria y 65 de la clase Firmicutes. Para las α -proteobacteria se observó una correlación positiva ($\rho = 0.36$; $P = 0.0007424$) para el porcentaje de genes en operon en

relación con el tamaño del genoma. Para las β -proteobacteria no se observó una correlación significativa ($\rho = -0.01422475$; $P = 0.9336$) para el porcentaje de genes en operon en relación con el tamaño del genoma. Para las Firmicutes se observó una correlación negativa ($\rho = -0.6216346$; $P = 7.19e-08$) para el porcentaje de genes en operon en relación con el tamaño del genoma.

Al estudiar el porcentaje de genes en organización de operon teniendo en cuenta en número total de genes; o sea analizando el “uso” de operones, se observaron importantes diferencias entre los grupos. Como se describió para las α -proteobacteria se observó una correlación significativa donde los genomas pequeños presentan una mayor proporción de genes en unidades monocistrónicas, y una menor proporción de genes en organización de operon en relación a los genomas grandes (Figura 3.5). Para las β -proteobacteria, no se observó correlación con el tamaño del genoma, y los organismos analizados presentaron una mayor proporción de genes en operon, independientemente del tamaño del genoma. Por el contrario para las Firmicutes, las bacterias de genomas chicos presentaron un mayor porcentaje de genes en organización de operon, contrastando con los resultados de la clase α -proteobacteria (Figura 3.16).

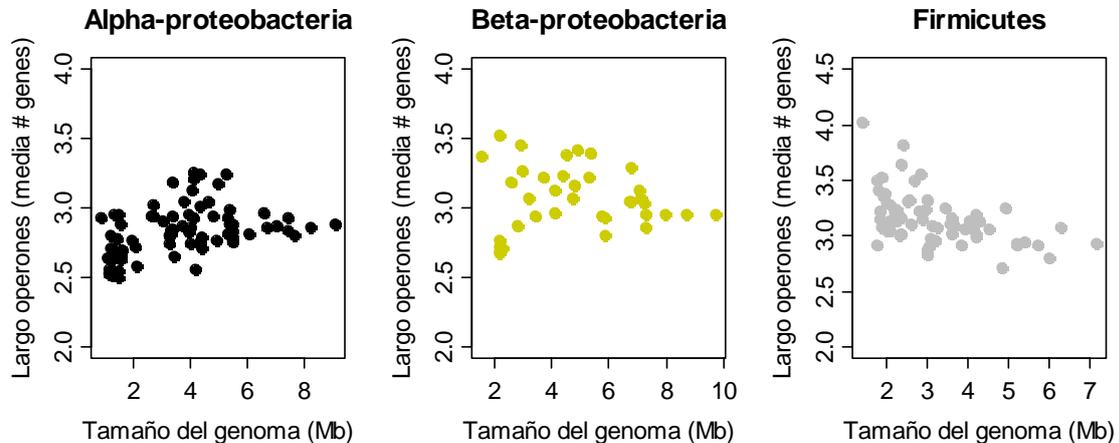


Figura 3.17. Largo de los operones.

Número promedio de genes por operon para cada uno de los 81 organismos de la clase α -proteobacteria, 37 organismos de la clase β -proteobacteria, 65 de la clase Firmicutes (Anexo, Capítulo IV). Para las α -proteobacteria, se observó una correlación positiva significativa ($\rho = 0.43$; $P = 6.197 \times 10^{-5}$) para el largo de los operones en relación con el tamaño del genoma. Para las β -proteobacteria, no se observó correlación significativa ($\rho = -0.01469891$; $P = 0.9314$) para el largo de los operones en relación con el tamaño del genoma. Para las Firmicutes, se observó correlación negativa significativa ($\rho = -0.5188811$; $P = 1.287e-05$) para el largo de los operones en relación con el tamaño del genoma; opuesta a la observada en las α -proteobacteria.

El análisis del largo promedio de los operones describe diferencias interesantes entre la clase α -proteobacteria y Firmicutes, observando correlaciones opuestas en relación al tamaño del genoma. Mientras las bacterias de genomas chicos de la clase α -proteobacteria tienen operones más cortos que las de genomas grandes, se observa lo contrario con las bacterias pertenecientes a Firmicutes. Las β -proteobacteria, no mostraron una correlación significativa con el tamaño del genoma (Figura 3.17)

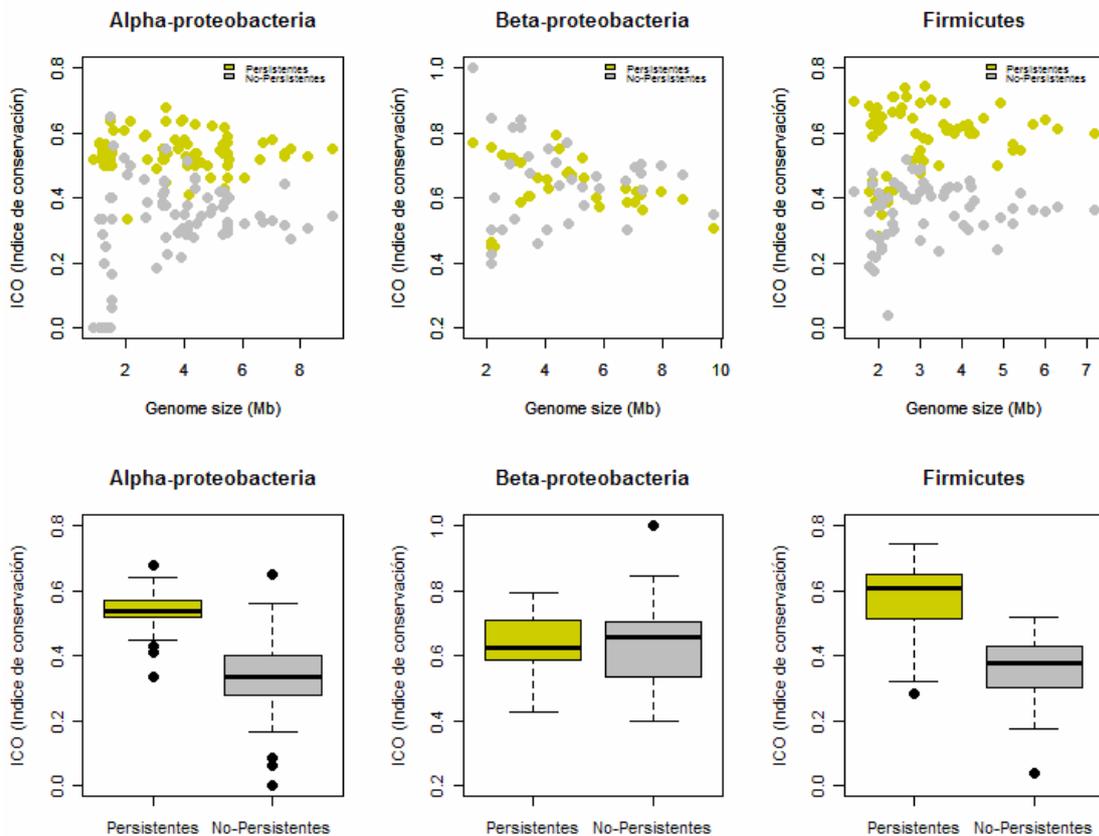


Figura 3.18. Influencia de la persistencia en la conservación de operones.

Se analizó el % de conservación de operones (analizando de a pares de genes) entre cada organismo y *E. coli* para los diferentes conjuntos de ortólogos (Persistentes vs. No persistentes). Las matrices generadas para el conjunto de ortólogos en cada comparación se dividieron de acuerdo a los niveles de persistencia de los genes. **β -proteobacteria:** Spearman's rank correlation: Tamaño del genoma vs. Persistentes $\rho = -0.1436025$, $P = 0.3965$; Tamaño del genoma vs. No-Persistentes $\rho = 0.04795256$, $P = 0.7781$. Paired t-test: Persistent vs. Non-persistent, $P = 0.3476$). No se encontraron correlaciones significativas ni diferencias significativas para el comportamiento de los diferentes grupos de genes en relación con el tamaño del genoma. **Firmicutes:** Se estudio la correlación de las variables Spearman's rank correlation: Tamaño del genoma vs. Persistentes $\rho = 0.06530346$, $P = 0.6053$, Tamaño del genoma vs. No-Persistentes $\rho = 0.1634687$, $P = 0.1932$. Paired t-test: Persistent vs. Non-persistent, $P < 2.2e-16$) mostró diferencias significativas para el comportamiento de los diferentes grupos de genes en relación con el tamaño del genoma.

En el análisis de la influencia de la persistencia, las β -proteobacteria no presentaron relación con el tamaño del genoma ni tampoco diferencias entre los dos grupos de genes. Por el contrario, para las Firmicutes, se observa un patrón similar a las α -proteobacteria, y aunque las correlaciones no son significativas para el tamaño del genoma, se observaron diferencias significativas para los diferentes conjuntos de genes, apoyando los resultados anteriores de una mayor conservación de los pares de genes en operon compuestos por genes persistentes-persistentes (**Figura 3.18**).

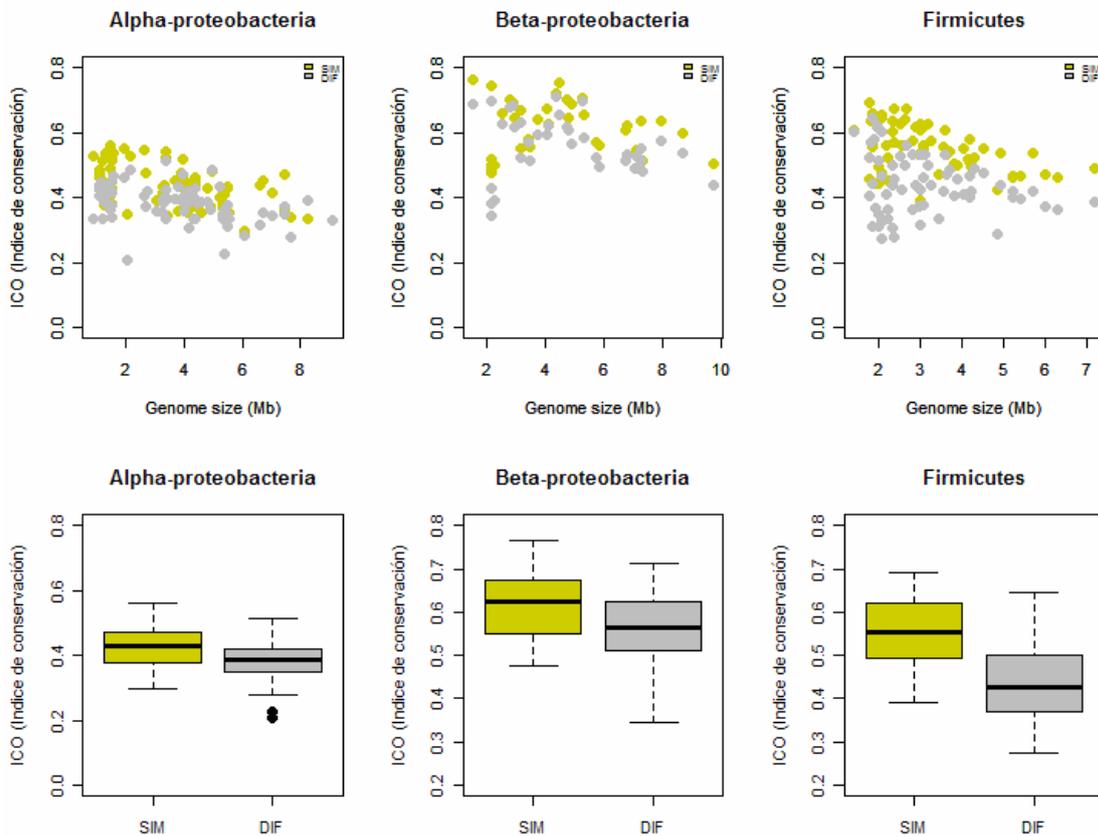


Figura 3.19. Influencia de los niveles de proteínas en la conservación de operones.

Se analizó el % de conservación de operones (analizando de a pares de genes) entre cada organismo y *E. coli* para los diferentes conjuntos de ortólogos SIM vs. DIF. Las matrices generadas para el conjunto de ortólogos en cada comparación se dividieron de acuerdo a los niveles de expresión de los genes. β -proteobacteria: *Spearman's rank correlation*: Tamaño del genoma vs. SIM: $\rho = -0.1132187$, $P = 0.5046$; Tamaño del genoma vs. DIF $\rho = -0.1464051$, $P = 0.3872$. Paired t-test: SIM - DIF, $P = 5.365e-12$). No se encontraron correlaciones significativas, ni diferencias significativas para el comportamiento de los diferentes grupos de genes en relación con el tamaño del genoma, pero sí entre los grupos de genes. **Firmicutes**: Se estudio la correlación de las variables *Spearman's rank correlation*: Tamaño del genoma vs.

SIM $\rho = -0.3990297$, $P = 0.0009928$, Tamaño del genoma vs. DIF $\rho = -0.2024697$, $P = 0.1058$. Paired t-test: SIM - DIF, $P < 2.2e-16$) mostró diferencias significativas para el comportamiento de los genes SIM en relación con el tamaño del genoma y entre los diferentes grupos de genes.

Para las α -proteobacteria, se observaron correlaciones negativas significativas para los dos grupos de genes (SIM, DIF). En las β -proteobacteria, no se observaron correlaciones significativas; mientras que para los Firmicutes se observaron las mismas tendencias que en la clase α -proteobacteria, aunque la correlación únicamente fue significativa para el grupo de pares SIM. Sin embargo al comparar los conjuntos de genes SIM Vs. DIF, en los tres clados se identificaron diferencias significativas para los conjuntos de genes, demostrando una mayor conservación para los pares de genes que codifican para niveles de proteínas balanceados en las células (Figura 3.19).

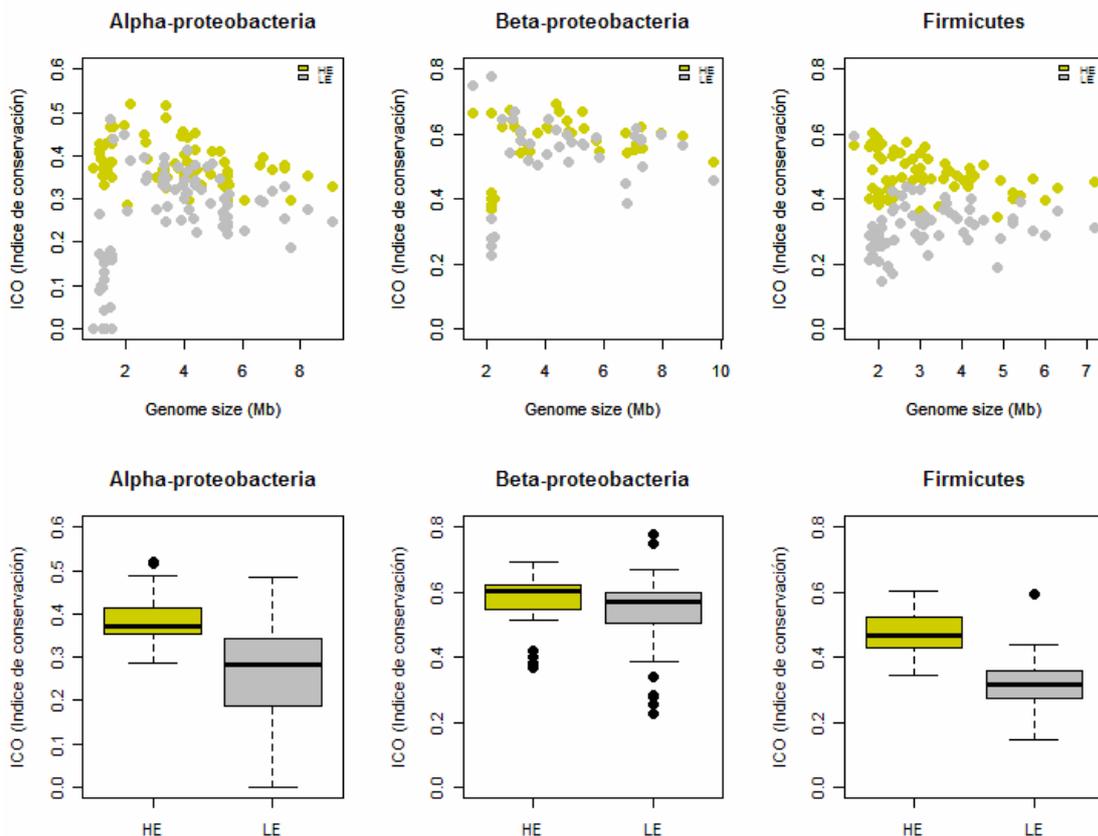


Figura 3.20. Influencia de los niveles de expresión en la conservación de operones. Se analizó el % de conservación de operones (analizando de a pares de genes) entre cada organismo y *E. coli* para los diferentes conjuntos de ortólogos HE vs. LE. Las matrices

generadas para el conjunto de ortólogos en cada comparación se dividieron de acuerdo a los niveles de expresión de los genes. **β -proteobacteria:** *Spearman's rank correlation:* Tamaño del genoma vs. HE: $\rho = -0.03354869$, $P = 0.8437$; Tamaño del genoma vs. LE $\rho = 0.01114272$, $P = 0.9481$. Paired t-test: HE - LE, **$P = 0.001544$**). No se encontraron correlaciones significativas ni diferencias significativas para el comportamiento de los diferentes grupos de genes en relación con el tamaño del genoma, pero sí entre los grupos de genes. **Firmicutes:** Se estudio la correlación de las variables *Spearman's rank correlation:* Tamaño del genoma vs. HE $\rho = -0.2963287$, **$P = 0.01685$** , Tamaño del genoma vs. LE $\rho = 0.339908$, **$P = 0.005601$** . Paired t-test: HE - LE, **$P < 2.2e-16$**) demostró diferencias significativas para el comportamiento de los diferentes grupos de genes en relación con el tamaño del genoma y entre los grupos de genes.

Los resultados de la influencia de expresión en las **α -proteobacterias** mostraron diferentes patrones de conservación para los genomas pequeños; donde los pares de genes de alta expresión (HE) correlacionaron negativamente con el tamaño del genoma, y los pares de baja expresión (LE) tuvieron una correlación positiva. Esto indicaría que los genes LE-LE tienen una mayor frecuencia de fragmentación en genomas chicos. Si bien las **β -proteobacterias**, como en los análisis anteriores, no mostraron una tendencia clara, las Firmicutes presentaron el mismo patrón que las **α -proteobacterias**, con correlaciones significativas y opuestas para ambos grupos de genes en relación al tamaño del genoma. En conjunto, todos los pares de genes de alta expresión mostraron un mayor porcentaje de conservación en relación a los LE (**Figura 3.20**).

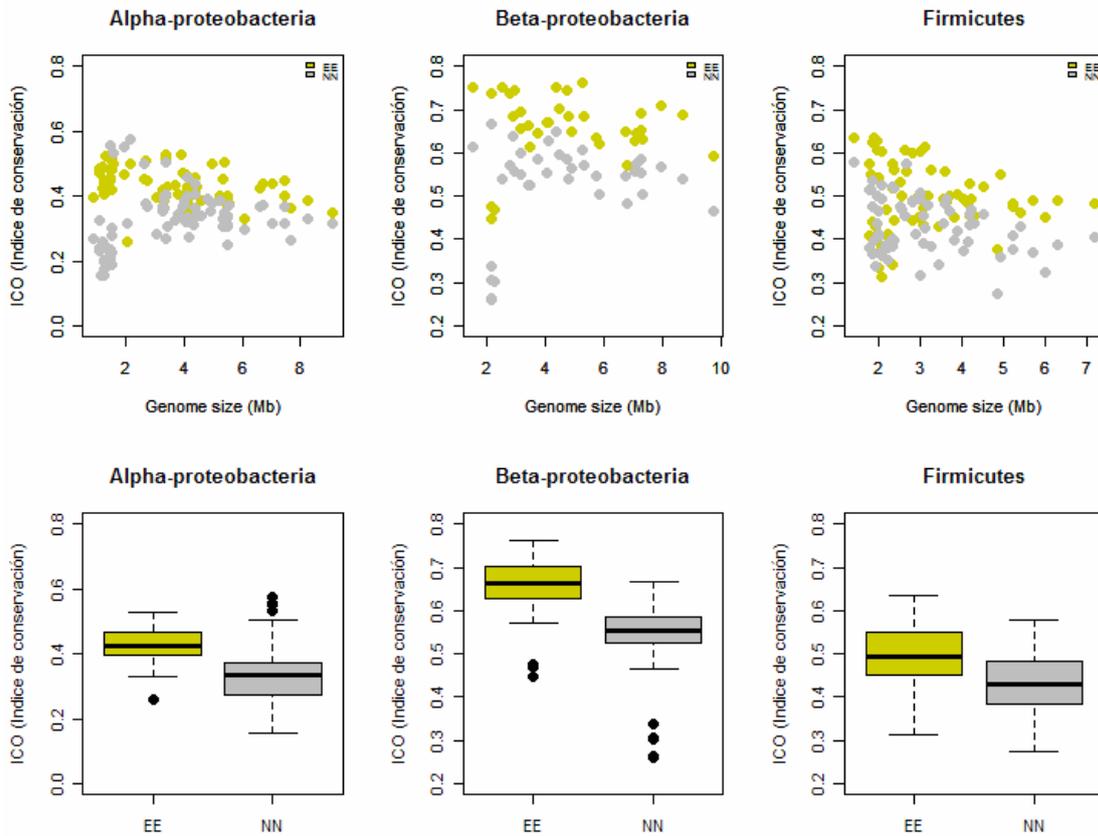


Figura 3.21. Influencia de la esencialidad en la conservación de operones.

Se analizó el % de conservación de operones (analizando de a pares de genes) entre cada organismo y *E. coli* para los diferentes conjuntos de ortólogos (EE vs. NN). Las matrices generadas para el conjunto de ortólogos en cada comparación se dividieron de acuerdo a los niveles de expresión de los genes. **β -proteobacteria:** Spearman's rank correlation: Tamaño del genoma vs. EE: $\rho = -0.01470588$, $P = 0.9312$; Tamaño del genoma vs. NN $\rho = 0.02607871$, $P = 0.8782$. Paired t-test: EE - NN, **$P < 2.2e-16$**). No se encontraron correlaciones significativas ni diferencias significativas para el comportamiento de los diferentes grupos de genes en relación con el tamaño del genoma, pero sí entre los grupos de genes. **Firmicutes:** Se estudio la correlación de las variables Spearman's rank correlation: Tamaño del genoma vs. EE $\rho = -0.1079953$, $P = 0.3918$, Tamaño del genoma vs. NN $\rho = -0.2614102$, **$P = 0.03543$** . Paired t-test: EE - NN, **$P = 5.302e-11$**) mostró diferencias significativas para el comportamiento de los genes NN en relación con el tamaño del genoma y entre los grupos de genes.

Los resultados de la influencia la esencialidad en las α -proteobacterias mostraron diferentes patrones de conservación para los genomas pequeños; donde los pares de genes esenciales (EE) correlacionaron negativamente con el tamaño del genoma, y los pares de genes no-esenciales (NN) tuvieron una correlación positiva. Esto indicaría, que los genes no esenciales tienen una mayor frecuencia de fragmentación en genomas chicos. Las β -proteobacterias, como en los análisis anteriores, no mostraron correlaciones significativas. A diferencia de las α -

proteobacterias, las Firmicutes mostraron correlaciones negativas con la conservación de los genes no-esenciales. En conjunto, todos los pares de genes compuestos por genes esenciales (EE) presentaron un mayor porcentaje de conservación en relación a los pares compuestos por genes no esenciales (NN) (**Figura 3.21**).

3.3. Discusión

El objetivo de este capítulo fue contribuir a un mayor conocimiento de cómo evolucionan las estrategias regulatorias con la reducción del genoma, y específicamente como evolucionó la estrategia de organización y regulación de la transcripción mediante los operones.

Considerando la fuerte reducción genómica en las α -proteobacterias, involucrando procesos de delección masiva de genes y rearrreglos cromosómicos, nos propusimos poner a prueba la hipótesis de un patrón de fragmentación de operones en los genomas reducidos. Para esto, analizamos en primer lugar 81 organismos de la clase α -proteobacteria describiendo las características y propiedades de los genomas (tamaño, número de genes, número operones, número genes en operon, y el largo de los operones, **Figuras 3.3 a 3.6**) y posteriormente diseñamos un método, basándonos en análisis de ortología e información de operones predichos y mediante operaciones de matrices para estimar comparativamente el grado de conservación de genes en operon. Considerando los antecedentes e hipótesis relacionadas, analizamos diferentes características y propiedades de los genes y su potencial influencia sobre la conservación (**Figuras 3.7 a 3.13**). Posteriormente, una vez diseñado el proceso, incorporamos otros grupos filogenéticos para evaluar la consistencia de los resultados, que podría sugerir hipótesis más generales para los factores que influyen sobre la conservación de los operones en el contexto de los diferentes modelos propuestos en la literatura (**Figuras 3.14 a 3.21**).

En el análisis de características de los genomas y distribución de operones, teniendo en cuenta la pérdida masiva de genes que sufrieron las bacterias intracelulares obligadas de la clase α -proteobacteria, era esperable encontrar que tengan también un menor número de operones, y un menor número absoluto de genes en organización de operon (**Figura 3.4**). La proporción de genes organizados en operon mostró una correlación significativa con el tamaño del genoma, donde los genomas chicos presentaron un menor porcentaje respecto a los genomas grandes en esta clase de bacterias. En los genomas chicos se observó que el porcentaje de genes en operon es similar al porcentaje de genes en unidades monocistrónicas; mientras que a medida que aumenta el tamaño del genoma, la tendencia parece invertirse, y los genomas grandes tienen un mayor “uso” de los operones (**Figura 3.5**). *Candidatus Pelabibacter Ubique*, mostró un uso alto de la organización en operon (85%) y fue excluido del análisis, ya que requería un estudio particular. Por otro lado, los genomas pequeños mostraron operones más cortos, observándose una correlación positiva del

largo promedio de los operones (número promedio de genes) con el tamaño del genoma (**Figura 3.6**). En un trabajo reciente [155] realizado en la bacteria de genoma muy pequeño *Mycoplasma pneumoniae* (Tenericutes, Mollicutes), los autores describieron que la reducción del genoma conlleva a operones más largos con genes con diferentes funciones acopladas. Al analizar el largo promedio de los operones, para diferentes grupos de bacterias, describieron en el conjunto que los genomas chicos tenían operones más largos, pero al desglosar entre grupos filogenéticos (α -proteobacteria, γ -proteobacteria y Firmicutes), observaron una tendencia diferencial para las α -proteobacteria, que no fue abordada en el estudio.

La fragmentación de operones no necesariamente implica una menor proporción de genes en operon ya que también puede producir una mayor frecuencia de operones más cortos (por ej. un operon de 4 genes evoluciona a 2 operones de 2 genes cada uno), o reorganizaciones más complejas de operones (ver ejemplos en **Figura 4.2**). Sin embargo, en la clase α -proteobacteria existe una relación entre el tamaño del genoma y los operones, donde los genomas pequeños presentan una menor proporción de los genes en operones, y a su vez, presentan operones más cortos; características compatibles con un proceso generalizado de fragmentación de operones.

De acuerdo a las observaciones particulares (operon *tat*, operon *sdh* o operon **T4SS**), conjuntamente con los procesos de delección masiva de genes (por ej. disminución en la cantidad de factores de transcripción [156]) y rearrreglos genómicos asociados a la reducción del genoma, esperábamos encontrar un menor número de operones u operones más cortos, en organismos de genoma pequeño. Estos organismos están asociados a ambientes intracelulares obligados y condiciones ambientales estables, y pérdida de factores de transcripción. Estas características podrían haber disminuido la presión de selección para mantener juntos a los genes, que se habrían adaptado a otro de regulación gen-específica. Sumado a esto, la adaptación a ambientes intracelulares se asocia con una disminución de los tamaños efectivos poblacionales [151] donde se espera una mayor tasa de fijación de mutaciones levemente deletéreas, incluyendo inserciones y delecciones en regiones regulatorias [159]. Por esto esperábamos también encontrar un patrón generalizado de fragmentación de operones en relación al tamaño del genoma.

Al estudiar el porcentaje de conservación (**Figura 3.7**), no se observó una correlación significativa entre el tamaño del genoma y la conservación de pares. Sin embargo, se obtuvo una tendencia que muestra que los genomas pequeños presentan un mayor porcentaje de conservación al analizar el conjunto de genes ortólogos en cada organismo y *E. coli*. Dicha tendencia difiere según se trate de genes persistentes

(presente en > del 90% de los genomas) o no-persistentes (presente en < del 50% de los genomas). Los pares compuestos por genes persistentes no mostraron una correlación significativa, demostrando que al comparar genes ortólogos con fuertes restricciones funcionales las organizaciones de operon no varían sustancialmente. Por el contrario, los pares de genes compuestos por genes no persistentes presentaron una fuerte correlación con el tamaño del genoma, evidenciando una mayor frecuencia de fragmentación (**Figura 3.8**). A su vez, independientemente de la relación con el tamaño del genoma, los pares no-persistentes presentaron un porcentaje de conservación significativamente menor que los pares persistentes, indicando que el rol de los genes es relevante. Estos resultados sugieren que los genes persistentes están bajo una fuerte selección para mantenerse juntos pese a la reducción del genoma, pero los genes con funciones más accesorias, no lo están.

En relación al estudio de cómo podrían afectar los niveles de proteínas balanceados (SIM) o desbalanceados (DIF), la hipótesis denominada *balance hypothesis* [158], postula que el desbalance de la concentración de los niveles de proteínas de los componentes de un complejo proteína-proteína pueden ser deletéreos para las células donde la sub- o sobre expresión disminuye el valor adaptativo. Como se describió anteriormente, en condiciones ambientales menos variables, las necesidades de regulación se vuelven menos relevantes, y diferentes autores asocian estos procesos con una pérdida en los factores de transcripción y reguladores transcripcionales. Hipotetizamos que esta ausencia o disminución de mecanismos de regulación podría afectar en mayor medida a este tipo de operones, y por lo tanto podrían tener una menor presión de selección para mantenerse juntos, haciendo más propenso un proceso de fragmentación. No obtuvimos diferencias para el comportamiento de ambos grupos (SIM y DIF) en relación al tamaño del genoma (ambos correlacionan negativamente); sin embargo los resultados indicaron que los pares de genes en operon que codifican para niveles de proteínas similares (balanceados) presentan mayores porcentajes de conservación que los pares de genes en operon que codifican para niveles desbalanceados (**Figura 3.10**).

Con respecto a la hipótesis de la influencia de los niveles de expresión en la conservación de operones, hipotetizamos que mantener la regulación de genes de alta expresión adquiere mayor importancia y relevancia, ya que para los genes altamente expresados y/o complejos abundantes mal regulados son más costosos para la economía energética que los complejos menos abundantes y con menores tasas de expresión. Los resultados para la clase α -proteobacteria demostraron que en los genomas chicos, los pares de genes con alta expresión se conservan más que los de baja expresión, apoyando la hipótesis planteada. Como se observa en la **Figura 3.12**,

los pares HE mostraron una correlación positiva significativa, mientras que los pares LE una correlación de signo opuesto también significativa. A su vez, en conjunto, los pares analizados demostraron que los pares LE-LE se conservan menos que los pares HE-HE.

En relación a la hipótesis de la influencia de la esencialidad de los genes, al igual que para los niveles de expresión, los pares de genes compuestos por genes Esenciales-Esenciales (E-E) y los No Esenciales-No Esenciales (NE-NE) tuvieron un patrón diferente en los genomas pequeños. Mostraron una correlación con diferente signo y ambas significativas. A su vez, los pares E-E presentaron altos índices de conservación, mientras que los NE-NE presentaron índices significativamente más bajos (**Figura 3.13**).

En conclusión, los resultados demostraron que los genomas chicos poseen un menor porcentaje de genes en operon y además tienen operones más cortos (mayor frecuencia de operones de dos genes), compatibles con un proceso de fragmentación. Los resultados sugieren que aunque no se trataría de un proceso generalizado de fragmentación de operones, en los genomas chicos, los pares de genes con menores niveles de persistencia, expresión y esencialidad presentan mayores frecuencias de fragmentación. Estos resultados demuestran que el rol de los genes y de sus propiedades influye en los niveles de conservación de los operones, lo que reflejaría diferentes restricciones evolutivas sobre la estructura/arquitectura organizativa de dichos genes.

Para demostrar que la influencia de estos factores en la organización de los operones es parte de un proceso adaptativo asociado a la reducción del genoma durante la transición a un estilo de vida intracelular se requieren estudios comparados más profundos como el análisis de Contrastes Filogenéticos Independientes [160].

Los porcentajes de genes organizados en operon y en unidades transcripcionales simples varían entre los diferentes linajes filogenéticos. En las β -proteobacteria no se observa una relación con el tamaño del genoma, y todas las bacterias analizadas muestran un porcentaje equivalente de genes en organización de operon (y mayor que el porcentaje de genes en unidades transcripcionales simples). Las α -proteobacterias presentan una correlación positiva para el porcentaje de genes en operon y el tamaño del genoma, mientras que para las Firmicutes, la relación es opuesta, y presentan una correlación negativa con el tamaño del genoma, donde los genomas chicos evidencian un mayor porcentaje de genes en operones que los genomas grandes. A su vez, en Firmicutes, para los genomas chicos, la diferencia entre el porcentaje de genes en cada organización se hace más extrema y evidente,

para organismos con genomas chicos, con porcentajes entre el 60-80% de genes en operon, y entre el 20-40% en unidades monocistrónicas (**Figura 3.16**).

En un principio, se pensaba que los genomas pequeños correspondían a organismos ancestrales y primitivos. Sin embargo, con los avances en los análisis filogenéticos basados en el gen 16S durante los años 1980, se propuso que estos organismos derivaban de genomas más grandes. A partir de los años 2000, con el desarrollo de numerosos genomas completamente secuenciados, se comprobaron y describieron procesos evolutivos de reducción genómica que ocurrieron de forma independiente (convergencia evolutiva) en diferentes clados como α -proteobacteria, γ -proteobacteria Firmicutes y Spirochaetes, sugiriendo la acción de mecanismos evolutivos generalizados [14, 150].

En las β -proteobacterias, no se han descrito procesos clásicos de reducción del genoma asociados a organismos con estilo de vida intracelulares. Las especies pertenecientes al clúster *Polynucleobacter necessarius subsp. asymbioticus* QLW-P1DMWA-1 (2,15Mb) y *Polynucleobacter necessarius subsp. necessarius* STIR1 (1,56Mb) representan linajes muy cercanos filogenéticamente y con estilos de vida contrastantes; mientras el primero es un organismo obligado de vida libre estricta, el otro en un endosimbionte obligado. Por otra parte son organismos con porcentajes altos de genes organizados en operon (75 y 70%, respectivamente) y operones largos (> 3,5 genes/operon), lo que evidencia un grupo particular que debería analizarse por separado. Además, estos genomas chicos comparten muy pocos genes ortólogos con *E. coli*, lo cual hace más complicados los análisis. Si bien las bacterias de la clase β -proteobacterias mostraron diferencias significativas en los niveles de conservación para los diferentes grupos de pares de genes (HE vs. LE, SIM vs. DIF y EE vs. NN; no así para persistencia), no mostraron correlaciones significativas en relación al tamaño del genoma.

El phylum Firmicutes mostró resultados similares a los obtenidos para la clase α -proteobacteria. Se identificó una correlación positiva (aunque no significativa) entre los porcentajes de conservación de los pares de genes no-persistentes y el tamaño del genoma (**Figura 3.18**). A su vez, los pares de genes persistentes tienen mayores índices de conservación que los pares no-persistentes. Al igual que las α -proteobacteria, los resultados indicaron que los pares de genes en operon que codifican para niveles de proteínas balanceados presentan mayores porcentajes de conservación que los pares de genes en operon que codifican para niveles desbalanceados (**Figura 3.19**). Los resultados obtenidos para los niveles de expresión fortalecen la hipótesis, ya que al igual que para las α -proteobacteria, se obtuvieron correlaciones opuestas y diferencias significativas de los niveles de conservación para

ambos grupos de genes (HE vs. LE) sumando evidencia de que los pares de baja expresión tienen una mayor frecuencia de fragmentación que los pares con altos niveles de expresión. Para los resultados de esencialidad, a diferencia de las α -proteobacteria para los pares de genes no-esenciales, el índice de conservación mostró una correlación negativa. Por otro lado, los dos grupos de genes (EE vs. NN) presentaron diferencias significativas concordantes con los resultados anteriores.

En conclusión, las propiedades de los genes como la persistencia, niveles de proteínas, expresión y esencialidad, resultaron ser aspectos relevantes para el origen y mantenimiento de los operones.

Según se ha propuesto, los operones que típicamente poseen una única secuencia regulatoria, minimizan las regiones susceptibles a mutaciones en complejos metabólicos u otros sistemas de genes. En muchos casos dichas regiones corresponden a secuencias de regulación gen-específica, que representan potenciales blancos de mutaciones perjudiciales para el conjunto del sistema [149]. Así, un mayor número de regiones blanco de mutaciones, representaría una desventaja adaptativa (que depende del tamaño de la región blanco y de la tasa de mutación) [149].

En todos los organismos el genoma adquiere mutaciones constantemente en forma de inserciones o deleciones, substituciones de bases o rearrreglos cromosómicos. La mayoría de estos cambios suelen ser deletéreos, y en tal caso, eliminados por selección purificadora, o bien neutrales quedando entonces sujetos a los efectos estocásticos de la deriva genética. La predominancia de una u otra fuerza evolutiva dependerá en gran medida del tamaño poblacional efectivo del organismo en cuestión: cuanto menor sea el tamaño efectivo, mayores serán los efectos de la deriva en detrimento de la selección. En las bacterias intracelulares, de tamaños efectivos poblacionales bajos, o que usualmente experimentan “cuellos de botellas” (durante la transmisión), la selección purificadora es menos eficiente para eliminar la mutación, permitiendo entonces la fijación de cambios levemente deletéreos por deriva. La mayor frecuencia de fragmentación, que supone el incremento de regiones de regulación gen-específicas en genomas chicos, y/o en genes de menor presión de selección como observamos para los genes menos persistentes, de baja expresión y no-esenciales, podrían ser entonces consecuencias de los determinantes ambientales y mecanismos y principios básicos en el contexto de la genética de poblaciones [149]. De esta forma, la aparición de mecanismos de control individual de genes versus control concertado de un conjunto de genes en operón, estaría fuertemente sujeto y moldeado por los procesos de mutación, selección y deriva.

El debate del origen y mantenimiento de los operones, y los procesos evolutivos subyacentes continúa con evidencias a favor y en contra de diferentes

modelos, y requerirá más estudios para dilucidar los factores preponderantes que influyen en el surgimiento y el destino de los operones en los distintos linajes procariontes.

Conclusiones generales.

En el marco de los objetivos propuestos, el trabajo desarrollado se orientó a descifrar los mecanismos tanto generales como específicos de la patogénesis, la adaptación al medio e interacción con el vector de *A. marginale* y organismos relacionados de la clase α -proteobacteria. El trabajo avanzó sobre aspectos inmunológicos, funcionales y evolutivos tomando como modelo a *Anaplasma marginale* y analizando comparativamente organismos relacionados del orden Rickettsiales y en el contexto de la clase α -proteobacteria en general.

La caracterización del conjunto de proteínas de membrana externa o superficie de las bacterias representa un desafío importante por ser los principales blancos de interacción de las bacterias con las células hospedadoras y con las células del sistema inmune. Este estudio permitió la identificación de dos proteínas en *A. marginale* con capacidad antigénica que fueron capaces de estimular una respuesta inmune específica. Algunas de estas proteínas presentaron una localización superficial en organismos relacionados del orden Rickettsiales y demostraron capacidad antigénica. Estudios futuros permitirán verificar experimentalmente la localización superficial, como también corroborar su potencialidad como antígenos protectivos en nuevas formulaciones de vacunas racionales a través de ensayos en el hospedador mamífero. El estudio contribuyó a incrementar el conocimiento no solo de las proteínas de membrana, sino también de los sistemas de secreción, que en muchos casos son responsables de la interacción con la célula hospedadora mediante la exportación de factores de virulencia y proteínas responsables de la interacción. La caracterización del sistema de secreción de proteínas "Twin Arginine translocación pathway (Tat)" brinda nueva información de la funcionalidad del sistema en dos patógenos de gran relevancia: *A. marginale* y *B. abortus*. En tal sentido, demostramos la conservación de la funcionalidad del sistema Tat en *A. marginale*, pese a una organización genómica fragmentada a diferencia de las estructuras canónicas descritas de un operon bajo una única secuencia regulatoria. La diferencia de los niveles de abundancia de ARNm que correlacionan con niveles proteicos descritos para complejos funcionales sugiere la existencia de una adaptación a mecanismos de regulación alternativos gen-específicos. Las bacterias patogénicas tendrían un uso limitado del sistema Tat, mientras que organismos de vida libre y bacterias del suelo tienen un uso moderado a alto. En particular, en *Brucella* spp. el sistema Tat podría tener un rol en la traslocación de los transportadores ABC participando de los sistemas de importación relacionados con la incorporación de una gran cantidad de compuestos incluyendo péptidos,

poliaminas, iones, aminoácidos, hierro y sulfatos. Por otro lado, la potencial translocación de la enzima Nos, una de las cuatro reductasas de *Brucella* spp. involucradas en la isla de denitrificación que permite, por medio de la respiración utilizando a los nitratos, sobrevivir a la bacteria en condiciones de baja concentración de oxígeno dentro de los macrófagos, podría ser necesaria como mecanismo de adaptación frente a la respuesta inmune. Al estudiar el mensajero policistrónico del operon Tat, y comprobar que excluye al gen *serS* será necesario generar y evaluar en *Brucella* spp. mutantes para el sistema (Δtat) para evaluar el rol del sistema Tat durante la infección de macrófagos. En el caso de *A. marginale*, la presencia del péptido señal Tat detectado por un programa de predicción para MreB y PBP2 y su posible participación en el mecanismo de invasión mediado por el apéndice, representa un importante desafío experimental que requerirá futuros estudios. Aunque las bacterias patógenas de la familia *Anaplasmataceae* tendrían pocos sustratos del sistema Tat, no podemos descartar que la identificación esté subestimada debido a las frecuentes anotaciones incorrectas de los sitios de inicio de la traducción durante los procesos de anotación automática de los genomas bacterianos. Por otro lado, considerando que no fue posible demostrar la funcionalidad de TatC para *A. marginale*, y que podría deberse a la incapacidad de esta subunidad a reconocer los péptidos señal de *E. coli*, una potencial co-evolución de los péptido señal Tat y de la maquinaria de reconocimiento específico medida por el complejo TatBC, representa otro interesante desafío.

Los procariotas coordinan la transcripción de genes involucrados en complejos metabólicos o de proteínas mediante la organización de operones. Diferentes hipótesis están en discusión en la actualidad sobre las fuerzas que rigen estos procesos de organización, tanto el origen como el mantenimiento. Considerando la gran variabilidad de los tamaños de los genomas, y específicamente la fuerte reducción en organismos asociados a hábitats intracelulares, nos propusimos estudiar la posible relación y consecuencias de estos procesos sobre la organización de los genes y la evolución de los operones como estrategia regulatoria. Considerando la fragmentación del operon *tat* y la adaptación a mecanismos de regulación gen-específicos, como también de otros casos descritos recientemente en la literatura, nos propusimos estudiar cuan generalizable podría ser el proceso de fragmentación de operones y su posible asociación con el proceso de reducción del genoma. Del análisis de diferentes propiedades de los genes y diversos grupos filogenéticos, fue posible detectar abundante diversidad como también patrones consistentes y conservados. Al analizar el uso de operones en los diferentes grupos, las α -proteobacterias evidenciaron un menor uso y largo de los operones, mientras que observamos un patrón opuesto en

los genomas pequeños de las Firmicutes. Considerando que los procesos evolutivos de reducción del genoma fueron descritos como procesos independientes en diferentes linajes, es probable que las consecuencias en la organización y estructura de los genomas tengan también patrones diferentes en función de las numerosas variables involucradas y procesos evolutivos subyacentes a cada contexto y grupo filogenético. La mayor frecuencia de fragmentación supone el incremento de regiones de regulación gen-específicas en genomas chicos, y/o en genes de menor presión de selección como observamos para los genes menos persistentes, de baja expresión y no-esenciales; donde la aparición de mecanismos de control gen-específico versus un control concertado en operón estaría fuertemente sujeto y moldeado por los procesos de mutación, selección y deriva. La subdivisión α -proteobacteria representa un grupo sumamente diverso y heterogéneo en sus capacidades, biológicas y ecológicas. Considerando la complejidad de procesos que afectan y determinan la organización de los cromosomas, una mejor comprensión de la organización de los cromosomas ayudará a conocer los procesos celulares, moleculares que subyacen a la diversidad.

Aspiramos a que los hallazgos y caracterizaciones realizadas en la tesis aporten al conocimiento general de los organismos del orden Rickettsiales para avanzar en la caracterización de un grupo de bacterias que requiere más atención y estudio como patógenos emergentes de amplia distribución.

Materiales y métodos.

Capítulo I.

Selección de antígenos recombinantes.

Los antígenos a evaluar fueron seleccionados a través de estrategias experimentales o de predicción *in silico*. Proteínas de exportación o de membrana fueron identificadas por un trabajo anterior en el laboratorio utilizando el sistema de *library alkaline phosphatase gene (PhoA) fusion system* (resultados no publicados) que nos permitió identificar un número de proteínas potencialmente localizadas en membrana o con destino de exportación. La proteína *hypothetical protein* AM1108 identificada por este método fue seleccionada para posteriores análisis. Incluimos también la proteína *hypothetical protein* AM127 identificada utilizando la técnica de *PCR-based subtractive hybridization* (resultados no publicados). La proteína *hypothetical protein* AM216 fue incorporada al estudio.

Análisis *in silico* y genómica comparativa.

Utilizamos las bases de datos de Blast2GO (<http://www.blast2go.de/>) [161] para la anotación funcional de las secuencias. Para predecir la presencia de péptido señal y sitio de clivaje utilizamos el algoritmo SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>) [162]. Para predecir potenciales sitios transmembrana utilizamos los programas TMpred, *transmembrane prediction algorithm* (http://www.ch.embnet.org/software/TMPRED_form.html) y TMHMM. Para predecir la localización subcelular utilizamos el programa PSORT (www.psort.org). Para la búsqueda de ortólogos y similitud utilizamos el BLASTp (www.ncbi.nlm.nih.gov/BLAST) y para realizar los alineamientos el ClustalX (www.ebi.ac.uk/clustalw) o t_coffee (www.tcoffee.org).

Expresión de las proteínas recombinantes en *E. coli*.

Las proteínas seleccionadas (AM1108, AM127 y AM216) fueron expresadas como proteínas recombinantes de fusión a 6xHis-Tag, en *Escherichia coli* en vectores de expresión pBAD/TOPO o pRSET *expression vector* (Invitrogen). Las proteínas AM1108, AM127 and AM216 fueron clonadas dividiendo el ORF en dos fragmentos (N-terminal y C-terminal) (Figura 3) para optimizar la expresión y disminuir los niveles de degradación. Los péptidos señales se excluyeron por ser potencialmente tóxicos para *E. coli*.

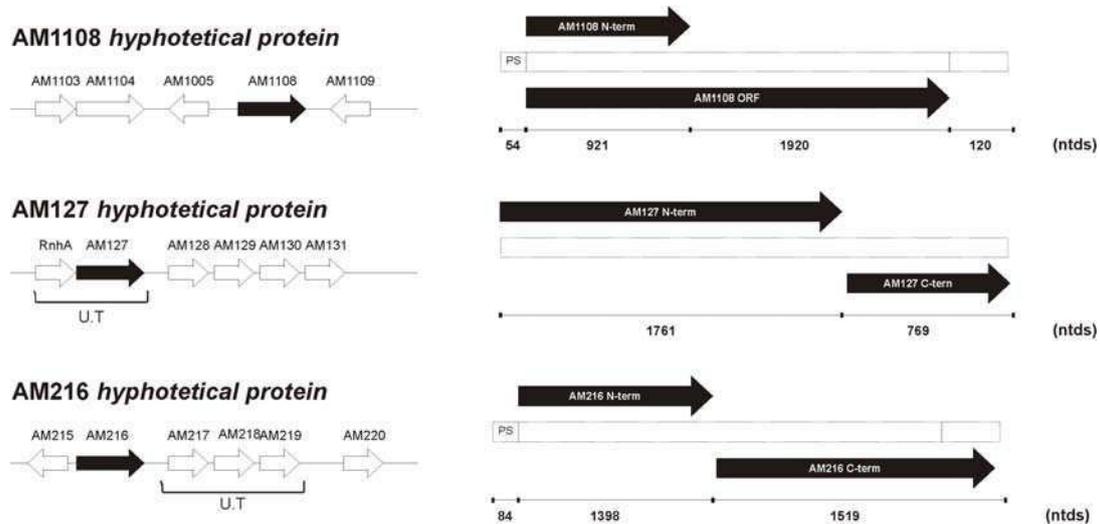


Figura 1.10. Esquema de la estrategia de clonado para antígenos candidatos.

Las proteínas AM1108, AM216 y AM127 se dividieron en dos fragmentos (N-terminal y C-terminal) para el clonado, expresión y análisis inmunológicos. Las flechas negras indican las proteínas recombinantes y los números indican la longitud en nucleótidos de los correspondientes genes o fragmentos.

Amplificación, clonado y secuenciación.

ADN de *A. marginale str. St Maries* fue provisto por Dr. Guy Palmer (Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, Washington) y se utilizó para amplificar las secuencias genómicas de las tres proteínas candidatas. Se diseñaron primers específicos (**Tabla S1.1, Anexo**) y mediante PCR utilizando protocolos estándar se amplificaron las secuencias de interés. Los fragmentos se clonaron en los plásmidos pTOPO2.1 o pGEMT (Invitrogen) y se secuenciaron para confirmar la integridad de la secuencia y el marco de lectura correcto (Unidad de Secuenciación, Instituto de Biotecnología, INTA). Las secuencias se analizaron con el programa Vector NTI (Vector NTI® Software). Las secuencias clonadas en plásmidos pRSET se utilizaron para transformar bacterias *E. coli* BL21 pLys y los vectores pBAD para transformar bacterias *E. coli* XLBlue, ambas competentes (métodos químicos) mediante protocolos estándar.

Expresión y purificación de proteínas recombinantes.

Para la expresión de las proteínas recombinantes, 500 ml de medio de cultivo LB (Luria-Bertani) completo fue inoculado con 5 ml de de cultivo (*overnight*) de las

bacterias complementadas con los plásmidos de expresión. Las cepas *E. coli* BL21 y XLBlue fueron inducidas con 1mM de solución de IPTG o solución final de 0.2% de arabinosa, respectivamente. Las células fueron centrifugadas después de la incubación a 37°C, congeladas y sonicadas bajo condiciones desnaturalizantes (8M urea). Se agregaron inhibidores de proteasas (Roche) para inhibir la acción de proteasas. Las proteínas recombinantes se purificaron utilizando columnas de cromatografía de inmunoafinidad anti-HIS (ProBond Resin, Invitrogen) utilizando columnas de Níquel (Ni^{2+}) bajo condiciones desnaturalizantes. La elución se realizó con un gradiente de imidazol de 125mM a 1M. Las proteínas recombinantes se dializaron durante la noche a 4°C contra soluciones con concentraciones decrecientes de urea hasta buffer PBS (estéril, pH = 7.0). Las proteínas purificadas se evaluaron en geles de poliacrilamida (SDS-page) para evaluar el nivel de purificación y comparar el peso molecular esperado y observado. Los geles se tiñeron con Coomassie Blue y se realizaron alternativamente transferencias a membranas de nitrocelulosa (Bio-Rad, Laboratories) para *Westerns blot* con anticuerpos anti-HIS. Las membranas luego de la transferencia se bloquearon con una solución de 5% de leche, e incubaron por 1 hora con un anticuerpo monoclonal anti-Histidina (dilución = 1:3000). Utilizando un anticuerpo secundario anti-ratón conjugado con fosfatasa alcalina y posterior revelado con NBT-DCIP (Promega) se identificaron las proteínas recombinantes y se comparó el peso molecular esperado con el observado.

Animales utilizados en el estudio.

Se evaluaron tres animales naturalmente infectados (Aberdeen Angus) N° 99, 282 y 285 de un año de edad provenientes de un establecimiento agropecuario de la localidad de Mercedes (Corrientes, Argentina) y un animal control (no infectado) de un año de edad proveniente del campo experimental del INTA (Castelar). La infección con *A. marginale* fue confirmada utilizando una *heminested* PCR [163] para la detección de la secuencia de *msp5*. Se realizó una extracción de sangre de los animales que fue colectada en tubos con anticoagulante heparina y se purificó el ADN total [164] para realizar la PCR. Brevemente, a 400ul de sangre, se agregaron 800ul de buffer de lisis (0.32 M Sucrosa, 10mM Tris-HCl (pH 7.5), 5 mM MgCl_2 , 1% v/v Triton X-100). Se centrifugó a máxima velocidad por 90 segundos y se removió el sobrenadante, repitiendo el paso de 2-4 veces, hasta remover completamente la hemoglobina. El precipitado se resuspendió en 100 μl de PBNB (PCR Buffer con Nonionic Detergents: 50 mM KCl, 10 mM Tris-HCl [pH 8.3], 2.5 mM MgCl_2 , 0.1 mg/ml gelatin, 0.45% (v/v) Nonidet P40, 0.45% (v/v) Tween 20), con 60 $\mu\text{g}/\text{ml}$ proteinasa K y se incubó *overnight* a 56°C. Posteriormente, las muestras se calentaron 10 minutos a 97°C para inactivar

la proteinasa K. El ADN se visualizó en geles de agarosa 0.8% teñido con bromuro de etidio y se midió la concentración utilizando el equipo Nanodrop ND-1000 (NanoDrop Technologies). Se utilizaron primers específicos para *msp-5* de *A. marginale strain* Florida (GenBank *accession* M93392) de acuerdo a: *external forward*, GCATAGCCTCCCCCTCTTTC (*msp5* posiciones 254 a 273); *external reverse*, TCCTCGCCTTGCCCCTCAGA (*msp5* posiciones 710 a 692); *internal forward*, TACACGTGCCCTACCGACTTA (*msp5* posiciones 367 a 387). Los productos de la amplificación se visualizaron en geles de agarosa 2% teñido con bromuro de etidio.

Linfoproliferación de células mononucleares periféricas.

A partir de las muestras de sangre de los animales involucrados en el estudio se purificaron las células mononucleares periféricas (PBMC) utilizando el kit de separación Histopaq (Sigma). Se cultivaron $2,5 \times 10^5$ PBMC/*well* en placas de 96-*wells* en un volumen final de 200 μ l de medio RPMI (10% suero fetal bovino, SFB). Para estudiar la respuesta de linfoproliferación de las células PBMC específica a las proteínas recombinantes (AM1108 N-term, AM1108 ORF, AM127 N-term, AM127 C-term, AM216 N-term y AM216 C-term) se agregaron concentraciones de 1ng/ μ l por cuadruplicado. Se utilizó como positivo inespecífico concanavalina A para confirmar la viabilidad de las células PBMC. Se utilizó como control negativo medio RPMI (10% SFB) y una proteína no relacionada de *Brucella melitensis*. Después de 4 o 6 días de incubación, los PBMC se marcaron con 1 μ Ci de methyl-3H, se colectaron en papel de fibra de vidrio y se determinó la proliferación celular a través de incorporación de Timidina [3H] en un contador de centelleo (Wallac, Winspectral). La respuesta de proliferación se expresa como [índice de Proliferación = cpm en presencia de antígeno recombinante / cpm RPMI].

Extracción de ARN y RT-PCR.

A partir de las muestras de sangre se extrajo ARN total utilizando RNeasy miniRNA extraction kit (QIAGEN). La concentración y grado de pureza se determinó midiendo la absorbancia a A260 nm y el cociente A260/A280 nm con el equipo Nanodrop ND-1000 (NanoDrop Technologies). Un (1 μ g) de ARN se trató con 1 μ l de DNase I *amplification grade* (Invitrogen) y se incubó a temperatura ambiente por 30 min. La enzima DNase I se inactivó por el agregado de 1 μ l de 25mM EDTA y posterior incubación a 65°C por 10 min. Luego para la reacción de RT se agregaron 1 μ l de random primers, 1 μ l DNTPs 10mM y agua Milliq hasta volumen final de 13 μ l. La *mix* se calentó por 5 min a 65°C y luego se incubó 5 min en hielo y se centrifugó para colectar. Luego del agregado de 4 μ l First strand Buffer, 1 μ l DTT 0,1M, 1 μ l Super script III (Invitrogen), la

mix se incubó 5 min a temperatura ambiente, seguido por 60 min a 50°C y la reacción se terminó por incubación a 70°C por 15 min. Como control de contaminación con DNA, una *mix* idéntica se preparó sin el agregado de la enzima retrotranscriptasa Super Script III.

Immunoblotting.

Las proteínas recombinantes AM1108 y AM216 C-term se analizaron mediante Western blots para estudiar el reconocimiento de anticuerpos específicos utilizando los sueros de los animales en estudio (Nº 30 -control-, 99, 282, 285 –infectados-), sueros de referencia positivos (de animales infectados) y negativos (de animales no infectados) y sueros de animales inmunizados con proteínas de membrana externa (Provistos por Dr. G. Palmer, USA).

Ensayos de *Real time* q-PCR.

Los primers se diseñaron utilizando el programa Primer Express Version 2.0 (Applied Biosystems) a partir de las secuencias de bovinos depositadas en el GenBank: IL-2 (AF535144), IL10 (gi), IL-12p35 (AJ271034), TNF- α (AF011926), IFN- γ (Z54144). Como gen *housekeeping* de referencia se utilizó el gen de la *Polymerase II (Pol II)*. Se utilizaron diluciones seriadas de cDNA para construir curvas de calibración y calcular la eficiencia de cada par de primers. Se consideraron ensayos con un regresión lineal de $R > 0.99$. El ensayo de q-PCR se realizó con el equipo ABI PRISM®, 7000, Sequence Detection System, Applied Biosystems) utilizando *Quantitect SYBR green* (QIAGEN).

Análisis estadístico.

Para los ensayos de linfoproliferación se calculó el Índice de Proliferación (IP) como el cociente:

$$IP = [cpm \text{ (células incubadas con proteínas recombinantes)} / cpm \text{ (células incubadas con medio RPMI)}]$$

Los resultados para los ensayos de IP son representativos de 3 experimentos independientes. Se analizaron con métodos de Análisis de la Varianza (ANOVA) utilizando el programa Infostat y Statistica 6.0. Para los ensayos de realtime PCR se utilizó un método de cuantificación relativa [165]. El método calcula la expresión relativa (R) de un gen *target* en base de la eficiencia de los primers y en análisis de los ciclos comparativamente entre los diferentes tratamientos, control y relativo a los genes de referencia. Para estos cálculos se utilizó el programa Relative Expression

Software Tool (REST) [165] que utiliza métodos no paramétricos *Pair Wise Fixed Reallocation Randomisation Test*.

Capítulo II.

Análisis filogenético y organización de los genes *tatABC*.

Se descargaron las secuencias de aminoácidos de 11 proteínas conservadas (RNA Pol $\beta\gamma\beta'$, alanyl-tRNA synthetase, phenylalanyl-tRNA synthetase, arginyl-tRNA synthetase, EF-Tu, EF-G, RecA, GyrA, Gyr β yHsp70) de 53 α -proteobacterias de la base de datos Refseq del NCBI y su identidad se confirmó utilizando el BLASTp (*E-value* = 10^{-5} ; *Query coverage* > 70%). Los ortólogos identificados mediante BLASTp en los organismos estudiados, se alinearon con el ClustalW [115] y cada alineamiento se procesó con el programa Gblocks [166] para eliminar regiones pobremente alineadas y posteriormente los alineamientos se concatenaron para realizar el análisis filogenético. Se realizó un árbol filogenético utilizando las secuencias concatenadas para las 53 bacterias mediante el método de neighbor-joining (JTT Model, 1000 bootstrap, con el MEGA 4 [167]). Los genes *tatA*, *B* y *C* de los diferentes organismos se localizaron en los genomas de los organismos estudiados y fueron visualizados en Mapviewer del NCBI (<http://www.ncbi.nlm.nih.gov/projects/mapview/>) o alternativamente en KEGG (<http://www.genome.jp/kegg/>). TatABC de *A. marginale*, y *B. abortus* 2308 se visualizaron en el programa Circular genome Viewer (CGView) [168].

Alineamientos múltiples.

Las secuencias de aminoácidos se descargaron de la base de datos del NCBI. Los alineamientos múltiples de las secuencias de las proteínas TatA, TatB y TatC de *A. marginale* str. *St. Maries*, *B. abortus* 2308 y *E. coli* K12 se realizaron utilizando el programa CLUSTAL W [115].

Extracción de ARN y RT-PCR

El ARN total de *A. marginale* str. *Salta* [169] se obtuvo de estabilatos de sangre congelada infectada conservados a -80°C de un bovino infectado, mientras que para *B. abortus* 2308 se obtuvo a partir de cultivos de 3 días en medio TBS (BD, USA) cultivado a 37°C y 200 rpm. Se utilizó *RNAeasy mRNA extraction kit* (QIAGEN) para la extracción de RNA. La concentración y grado de pureza se estimó con la absorbancia a 260 nm (A_{260}) y relación A_{260}/A_{280} con el equipo Nanodrop ND-1000 (NanoDrop Technologies). Un microgramo ($1\mu\text{g}$) de ARN se trató con $1\mu\text{l}$ de *DNase I amplification grade* (Invitrogen) y se incubó a temperatura ambiente por 30 min. *DNase I* se inactivó por el agregado de $1\mu\text{l}$ de 25mM EDTA y posterior incubación a 65°C por 10 min. Se agregaron $1\mu\text{l}$ de ($3\mu\text{g}/\mu\text{l}$) *Random primers* (Invitrogen, USA), $1\mu\text{l}$ DNTPs 10mM

(Promega, USA) y Milliq agua (Millipore) hasta llegar a 13µl de volumen final. La mezcla se calentó por 5 min. a 65 °C y luego se incubó 5 min. en hielo y centrifugó para colectar. Luego del agregado de 4µl *First strand Buffer*, 1µl DTT 0,1M, 1µl Super script III (Invitrogen), la *mix* se incubó 5 min. a temperatura ambiente, seguido por 60 min. a 50°C y la reacción se terminó por incubación a 70 °C por 15 min. Como control de contaminación con ADN, una *mix* idéntica se preparó sin el agregado de *RT Super Script III*.

Construcciones y plásmidos utilizados

El ADN genómico de *A. marginale str. St. Maries* fue provisto por Dr. Guy Palmer (Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, Washington). El ADN de *B. abortus 2308* se obtuvo a partir de cultivos de 3 días en medio TBS (BD, USA) cultivado a 37°C y 200 rpm inactivados por calor (a 99°C durante 10 min. y centrifugados por 2 minutos a 13,000 g). Se procesaron con 3 ciclos consecutivos de congelado-descongelado de acuerdo a protocolos establecidos [170]. Posteriormente 5 µl del sobrenadante se utilizaron para ensayos de PCR. Se diseñaron primers específicos utilizando el programa *Primer Express* (**Tabla S 2.1, Anexo**), y se realizaron las PCR mediante protocolos estándar para amplificar en ambos organismos los genes *tatA*, *tatB* y *tatC*. Los fragmentos amplificados se clonaron en plásmidos pTOPO 2.1 (Invitrogen) para después subclonarlos en el plásmido pUNIPROM [171] bajo el promotor *tat* de *E. coli* [118] y posteriormente se secuenciaron para comprobar la integridad. Una lista de los plásmidos pUNIPROM utilizados se muestra en la **Tabla S 2.2, Anexo**. Durante todos los pasos de clonado, *E. coli* se cultivó aeróbicamente en medio LB en condiciones y utilizando concentraciones de antibióticos estándar.

Bacterias utilizadas y condiciones de crecimiento

Los plásmidos y construcciones realizadas se utilizaron para complementar *E. coli* mutantes para el sistema Tat (MC4100-P [*wild-type*], JARVI6-P, BOD-P, BILK0-P y DADE-P; **Tabla S 2.2, Anexo**) [85, 118] para estudiar la funcionalidad en un sistema heterólogo mediante ensayos de complementación. Las cepas mutantes de *E. coli* JARVI6-P (Δ *tatA*), BOD-P (Δ *tatB*) y BILK0-P (Δ *tatC*) [118] se complementaron con plásmidos pUNIPROM conteniendo *tatA*, *tatB* y *tatC* de los organismos *A. marginale str. St. Maries* y *B. abortus 2308*, utilizando protocolos estándar de transformación. Las cepas mutantes complementadas con el vector pUNIPROM vacío y la cepa DADE-P (Δ *tatABC*) se utilizaron como controles negativos. La cepa *wild type* MC4100-P y las mutantes *E. coli tat* complementadas con pUNIPROM conteniendo los genes nativos de *E. coli tatA*, *tatB* y *tatC*, pFAT415, pFAT 416 y pFAT417 [85], respectivamente, se

utilizaron como controles positivos. Para los ensayos de funcionalidad de los genes heterólogos, luego de la transformación las cepas se incubaron en diferentes condiciones selectivas:

(i) 2% SDS: las bacterias se incubaron en medio LB líquido a 37°C durante la noche y posteriormente en placas con LB agar suplementadas con 2% SDS [117], o en medio LB líquido 2% SDS, donde se midió el crecimiento registrando la densidad óptica del cultivo a 600 nm durante 8 horas [118];

(ii) condiciones de anaerobiosis: las bacterias se incubaron a 37°C durante la noche y se evaluó la capacidad de crecimiento en placas de agar conteniendo medio mínimo M9 suplementado con 0.5% glicerol y 0.4% trimethylamine-N-oxide (TMAO) e incubado en una jarra anaeróbica bajo atmósfera de hidrógeno y dióxido de carbono [118];

(iii) Ensayos de TMAO reductasa: para los ensayos de actividad de la enzima TMAO reductase se utilizaron fracciones subcelulares preparadas a partir de cultivos de 30 ml incubados durante la noche sin agitación a 37°C en medio líquido LB TMAO/glicerol suplementado con 50% glicerol y 20% TMAO bajo condiciones anaeróbicas. Las fracciones periplásmicas se obtuvieron con el método de *shock* osmótico en frío [118, 172]. Sobre las fracciones periplásmicas luego del ensayo enzimático se midió la concentración total de proteínas (Pierce, Thermo Scientific, USA) para relativizar la medición de la actividad de la enzima TMAO reductasa. La actividad de la enzima TMAO oxidoreductasa se registro de acuerdo a [173] (Thermo MultiSkan Spectrum, Thermo Scientific).

Microscopía.

Las bacterias *E. coli* mutantes complementadas con genes heterólogos y controles se incubaron durante la noche a 37 °C. Posteriormente se realizaron diluciones 1:100 en medio LB y se incubaron a 37°C hasta densidad óptica de 0.6–0.8 medido a 600 nm [119]. Las bacterias se visualizaron con microscopía de contraste de fases utilizando un microscopio Leica TCS-SP5 confocal (Leica Microsystems GmbH, Wetzlar, Germany) a 63x y utilizando un objetivo HCX PL APO CS 63.0 x 1.20 WATER UV.

RT-PCR y Real-time q-PCR.

Se diseñaron primers específicos utilizando el programa Primer Express Versión 2.0 (Applied Biosystems) (**Tabla S 2.1, Anexo**). Como genes housekeeping se utilizaron *groEl* para *A. marginale* y *rplI* para *B. abortus* 2308. Diluciones seriadas al décimo de cDNA se utilizaron para construir una curva de calibración y para calcular la eficiencia para cada par de primers en los ensayos de real-time PCR. Ensayos con un coeficiente de correlación *r value* >0.99 se consideraron válidos. Los ensayos de

realtime PCR se realizaron con un equipo de ABI PRISM®, 7000, Sequence Detection System, Applied Biosystems, utilizando el kit comercial de Quantitect SYBR green (QIAGEN). Los resultados se analizaron con el software disponible *Relative expression software tool (REST)* para evaluar los niveles de expresión de los genes *tatA*, *tatB* y *tatC* [165].

Predicción *in silico* de sustratos Tat.

Se identificaron potenciales sustratos del sistema Tat del set completo de proteínas codificadas por los genomas completos en diferentes genomas anotados en 12 especies de la familia *Anaplasmataceae* y 10 especies de la familia *Brucellaceae*. Para la identificación de potenciales sustratos se utilizaron 3 programas disponibles: TatP (<http://www.cbs.dtu.dk/services/TatP/>; [174]), TATFIND (<http://signalfind.org/tatfind.html>; [175]) y PRED-TAT (<http://www.compgen.org/tools/PRED-TAT/> [176]). El programa TatP combina la búsqueda de patrones en las secuencias de aminoácidos con redes neuronales entrenadas para detectar sitios de clivaje y otra para determinar los aminoácidos de los consensos establecidos para la secuencias del péptidos señales Tat. TATFIND predice péptidos señales Tat mediante la búsqueda e identificación de patrones de aminoácidos e hidrofobicidad. PRED-TAT aplica *Hidden Markov models* para predecir y discriminar entre péptidos señales Tat y Sec. TatP y TATFIND se corrieron en sus respectivos servidores vía web. Los autores de PRED-TAT proveen un repositorio de genomas bacterianos pre-procesados para los cuales se extrajeron los sustratos predichos para los genomas de las familias en el analizadas (<http://www.compgen.org/tools/PRED-TAT/supplement/genomes>). Consideramos como sustratos positivos proteínas predichas como sustratos por al menos 2 de los 3 programas.

Capítulo III.

Genomas y operones.

Los genomas de *E. coli* K-12 MG1655 y las secuencias completas de los genomas de los organismos de la clase α -proteobacterias, β -proteobacterias, y Firmicutes fueron descargados de <ftp://ftp.ncbi.nih.gov>. Se descargaron los archivos “.faa” con la información de las proteínas codificadas por el genoma para los análisis de BlastP e identificación de ortólogos. Se descargaron los archivos “.rpt” para el análisis del tamaño del genoma, número de genes y número de proteínas. Por otro lado, la información de la predicción de los operones para los genomas se descargaron de la base de datos provista por DOOR “Database of prokaryotic Operons” [177] (<http://csbl1.bmb.uga.edu/OperonDB>). La precisión de la predicción de DOOR resultó consistente con los datos experimentales y operones recuperados por RegulonDB [178] y es de las base de datos más analizadas y consultadas en la bibliografía.

Identificación de Ortólogos.

Los ortólogos se identificaron con el criterio de mejores *hits* recíprocos (RBH) utilizando el BLASTP con un límite de *e-value* 10^{-5} y un porcentaje de cobertura del 70%. Se realizó a través de una estrategia de “estrella” entre cada uno de los 81 organismos de la clase α -proteobacteria a analizar *vs.* *E. coli* K-12 MG1655, identificando de esta manera para cada comparación de a pares el conjunto de genes compartidos (genes ortólogos). A su vez, para los genes de *E. coli* se calculó el % de genomas donde se identificó el ortólogo pudiendo clasificar la persistencia de cada uno de los genes como el % de genomas donde estaba presente.

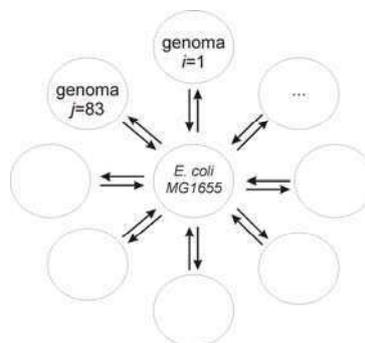


Figura 3.22. Estrategia de estrella para la identificación de ortólogos.

Las flechas indican los BlastP (*e-value* 10^{-5} , 70% de cobertura) recíprocos para identificar los mejores *hits* recíprocos (RBH) entre ambos genomas. De esta manera, se estableció para cada

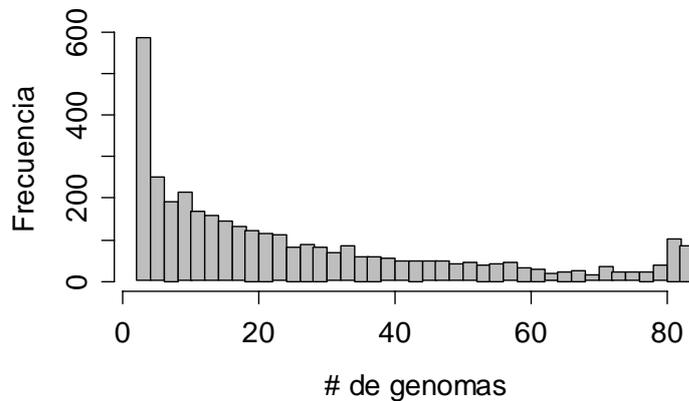
par de comparación (organismo X vs. *E. coli*) el conjunto de genes ortólogos a comparar. Por otro lado para cada gen de *E. coli* se estableció el % de genomas donde se identificó un ortólogo y por lo tanto, el grado de persistencia del gen.

Análisis de Persistencia.

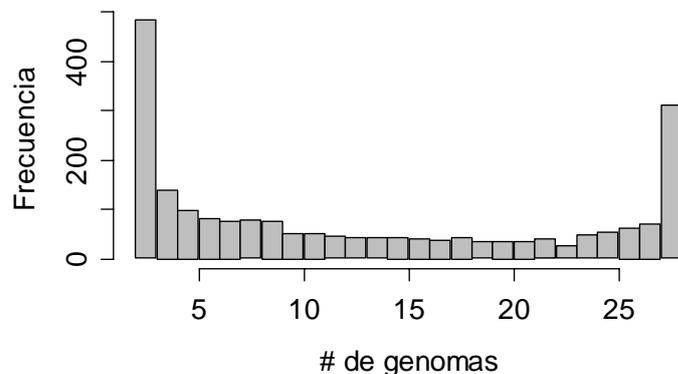
Luego de realizar los análisis de BlastP entre *E. coli K-12 MG1655* y cada uno de los 81 organismos de la clase α -proteobacteria, se identificaron los ortólogos de cada gen de *E. coli* en los restantes genomas. Con este procedimiento para cada uno de los genes de *E. coli* se identificó en cuantos de los 81 genomas se identificó un ortólogo. Analizando en cuantos genomas (% de ocurrencia) se identificó el ortólogo, se establecieron los criterios para determinar genes persistentes y no persistentes. Los criterios establecidos fueron:

- i) Genes en $>$ del **90 %** de los genomas = genes persistentes;
- ii) Genes entre el **50-90 %** de los genomas = nivel intermedio de persistencia; iii)
- Genes en $<$ del **50 %** de los genomas = genes no persistentes.

Alpha-proteobacteria - Persistencia



Beta-proteobacteria - Persistencia



Firmicutes - Persistencia

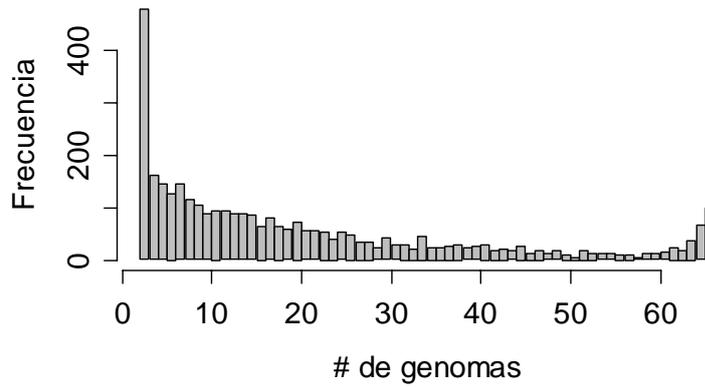


Figura 3.23. Histogramas de persistencia de los genes de *E. coli*.

Para cada gen de *E. coli* se calculó en los diferentes grupos filogenéticos (α -proteobacteria, β -proteobacteria y Firmicutes) el porcentaje de genomas donde se identificó el ortólogo, y se graficó la distribución. De acuerdo a los puntos de corte establecidos se seleccionaron los grupos de persistencia para posteriores análisis.

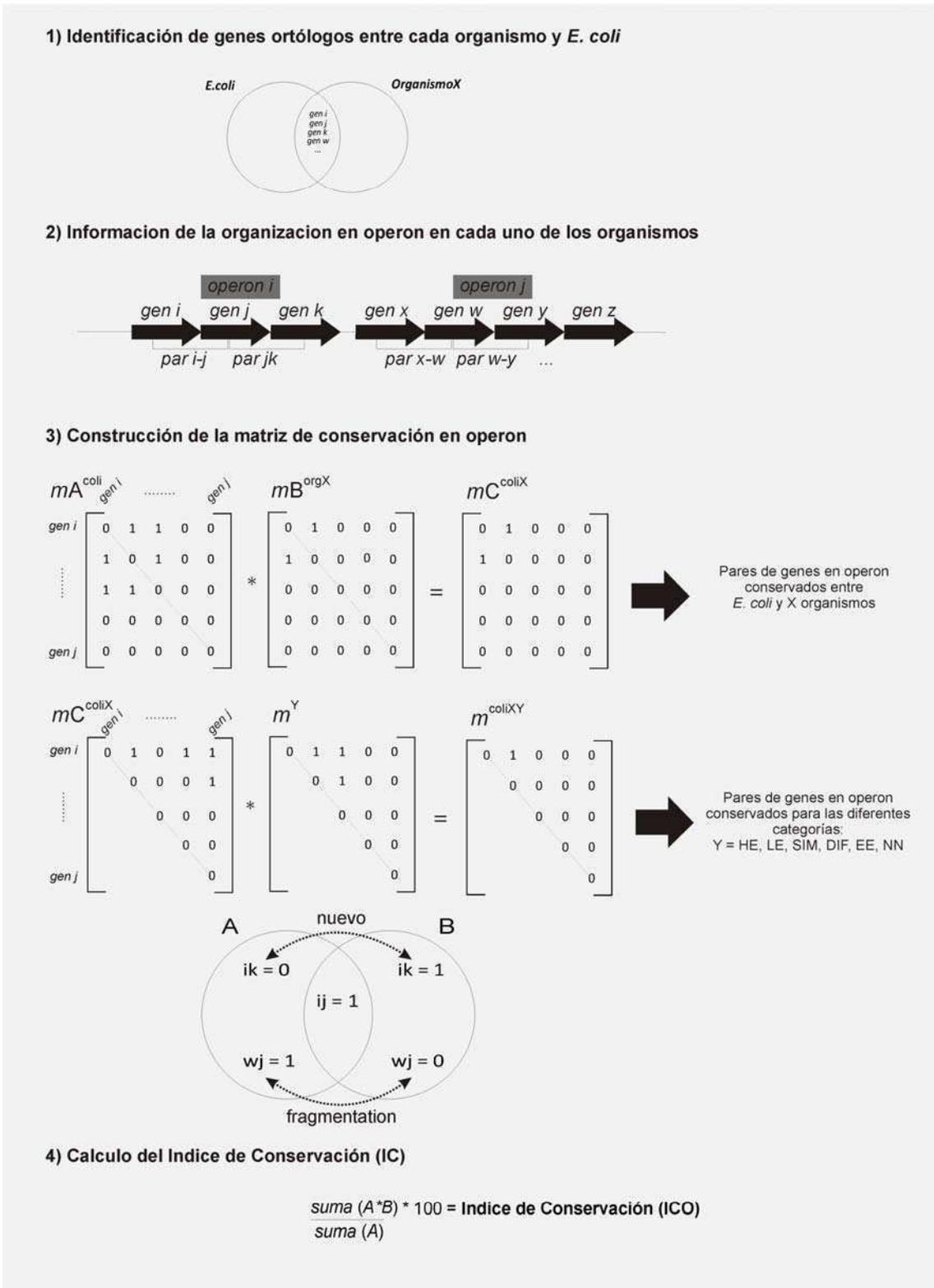


Figura 3.24. Esquema para el análisis de conservación de operones en los diferentes organismos.

(1) Como ilustra la Figura 3.22 el primer paso fue realizar los BLASTp de a pares entre *E. coli* y los restantes 81 organismos; de esta manera identificamos para cada comparación, el conjunto de genes ortólogos. (2) Como se describió anteriormente la información completa de los

operones predichos en cada uno de los organismos se descargó de la base de datos DOOR a partir de la cual se armaron mediante scripts en *bash* y *perl* los pares de genes en operon. Para esto consideramos los pares de genes contiguos dentro de un operon, y no todos los pares posibles que podrían generarse. **(3)** Para cada comparación construimos 2 matrices (una para *E. coli* y otra para el organismo X; teniendo en cuenta solamente el conjunto de genes ortólogos para cada una de las comparaciones de a pares) donde de acuerdo a la información de operones, se le asignó en cada organismo, el valor “1” si el par de genes pertenecía al mismo operon, y “0” si no lo estaban. Del producto de matrices, se calculó el número de pares de genes ortólogos que se conservan en operon. Para el análisis de la relación entre las propiedades de los genes (expresión, esencialidad, niveles de proteínas) y la conservación, primero se construyó una matriz con la información respecto a *E. coli*. Para el caso de alta expresión, se construyó una matriz donde se asignó “1” si el par de genes tenía alta expresión o “0” si era de baja expresión. Y lo mismo se realizó para las restantes categorías (SIM-DIF; EE-NN). De la multiplicación de la matriz de propiedad, con la matriz de conservación en operones, se obtienen los valores de conservación de pares de genes para cada categoría. **(4)** Para calcular el % de conservación de genes en operon para cada comparación, se multiplicaron las matrices, de modo que al conservarse, la matriz resultante conserva un “1”, mientras que si no se conserva, tendrá un “0”. Al sumar la matriz resultante, obtenemos el valor correspondiente al número total de pares de genes conservados, que luego se relativizó al número total de genes en operon en *E. coli*, respectivamente para cada comparación, ya que el número de ortólogos y por lo tanto las dimensiones de cada una de las matrices, es específica para cada comparación de a pares.

Análisis de expresión.

Para estimar los niveles de ARNm se utilizó un conjunto de datos de acceso público de experimentos de transcriptómica realizados en *E. coli K-12 MG1655* (**Figura 3.25**). Los datos obtenidos de diferentes experimentos se normalizaron respecto a sus propios valores de referencia para normalizar los experimentos con el fin de hacer el promedio para estimar los niveles de transcripción. Para estimar los niveles de ARNm para cada par de genes en operon (*a*, *b*) en *E. coli* consideramos el promedio de ambos valores dentro de un par para obtener la tasa de expresión del par de genes en operón. Para definir los pares de genes en operón con baja o alta expresión, se realizó un histograma, se analizó la distribución y se utilizó la mediana como corte de 50% de los datos.

Análisis de niveles de proteínas.

Para estimar las concentraciones de proteínas por célula se utilizó un conjunto de datos de acceso público de experimentos proteómica realizados en *E. coli K-12 MG1655* (**Figura 3.25**). Luego de la normalización de los experimentos, se cuantificó la relación $[(Ca-Cb)/(Ca+Cb)]$ para cada par de genes de genes en operón. De esta manera nos independizamos de la magnitud de la concentración de la proteína en la diferencias de las concentraciones. Para definir los pares de genes en operón con

niveles de proteínas similares o diferentes, se realizó un histograma para ver la distribución de los casos, y se utilizó un corte de 50% de los datos.

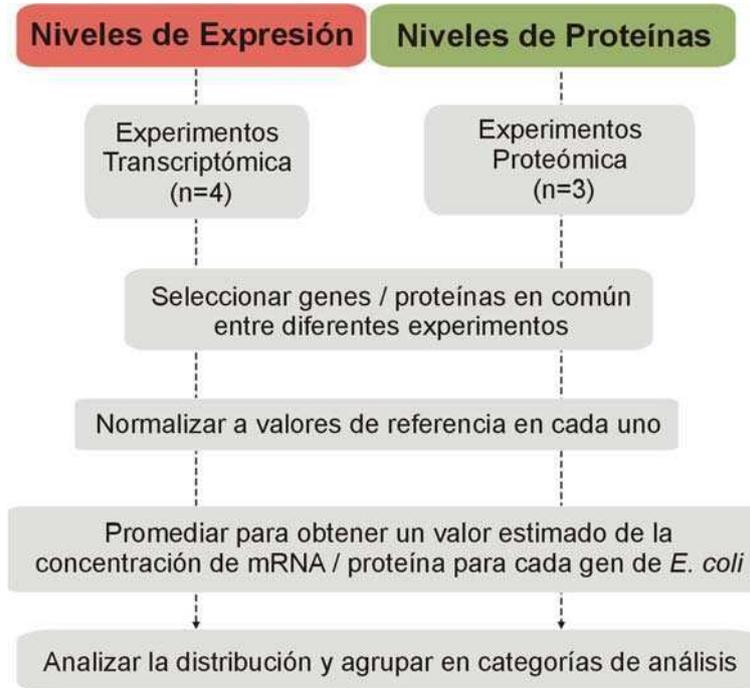


Figura 3.25. Esquema de información de ARNm y proteínas.

Para estimar los niveles de ARNm se utilizó un conjunto de datos de acceso público de experimentos de transcriptómica realizados en *E. coli* K-12 MG1655. Los datos obtenidos de diferentes experimentos [179-182] se normalizaron respecto a sus propios valores de referencia para normalizar los experimentos con el fin de hacer el promedio para estimar los niveles de transcripción. El mismo procedimiento se realizó para estimar los niveles de proteínas por celular de acuerdo a experimentos publicados [182-184].

Análisis de esencialidad.

De acuerdo a la información provista por la base de datos RegulonDB (<http://regulondb.ccg.unam.mx/>) y al trabajo [185] se definió para cada gen de *E. coli* si representa un gen esencial o no-esencial.

Cálculos y análisis de matrices.

Para los cálculos de conservación de pares de operon para analizar las hipótesis de la influencia de los niveles de expresión, los niveles proteicos y de la característica de esencialidad, realizamos análisis similares al de matrices de conservación para analizar el rol de la persistencia (Explicado en la **Figura 3.24**).

Generamos matrices simétricas de coocurrencia (con “0” en la diagonal) con el set de genes ortólogos para cada comparación de pares con *E. coli*.

i) Para cada organismo generamos una matriz ($m\mathbf{B}$), con un gen $(i,j) = 1$ si ambos se encuentran en el mismo operon, o $\text{gen}(i,j) = 0$ si no (considerando la diagonal $\text{diag}(m\mathbf{B}) = 0$). Por lo tanto, generamos 81 pares de matrices para el set de ortólogos entre *E. coli* y cada organismo.

Por lo tanto, multiplicando:

$$m\mathbf{A}^{\text{coli}} * m\mathbf{B}^{\text{orgX}} = m\mathbf{C}^{\text{coliX}}$$

Donde $m\mathbf{C}^{\text{coliX}}$ es la matriz con 1 en los pares genes que se conservan en el mismo operon en ambos organismos:

$$m\mathbf{C}^{\text{coliX}}(i,j) = 1 \rightarrow m\mathbf{A}^{\text{coli}}(i,j) = 1 \text{ y } m\mathbf{B}^{\text{orgX}}(i,j) = 1$$

De igual manera, obtuvimos una matriz $m\mathbf{D}^{\text{coliX}}$ con “1” si el par de genes se conservan en un genoma pero no en ambos.

$$m\mathbf{D}^{\text{coliX}}(i,j) = 1 \rightarrow m\mathbf{A}^{\text{coli}}(i,j) = 1 \text{ y } m\mathbf{B}^{\text{orgX}}(i,j) = 0; m\mathbf{A}^{\text{coli}}(i,j) = 0 \text{ y } m\mathbf{B}^{\text{orgX}}(i,j) = 1$$

ii) Para el análisis de las características de expresión, esencialidad y niveles proteicos también construimos matrices simétricas para abordar el análisis.

Para esto nos basamos en la información provista por diferentes experimentos disponibles de transcriptómica, proteómica de *E. coli*, según **Figura 3.25**.

- a. *Expresión*, dos matrices: $\mathbf{HE}(i,j) = 1$ si ambos genes que participan en el mismo operon son considerados de alta expresión, y 0 si no lo están. Y otra matriz donde, $\mathbf{LE}(i,j) = 1$ si ambos genes dentro de un operon son de baja expresión, y 0 si no lo están.
- b. *Niveles de proteínas*, dos matrices: $\mathbf{SIM}(i,j) = 1$ si ambos genes dentro de un operon expresan niveles de proteínas similares, y 0 si no lo están. Y otra matriz donde, $\mathbf{DIF}(i,j) = 1$ si ambos genes codifican para niveles de proteínas diferentes en la células, y 0 si no lo están.

- c. *Esencialidad* (EE, EN [o NE], o NN), en este caso, generamos tres matrices siguiendo el mismo criterio explicado arriba. Siendo E = esencial, N no-esencial.

Con este procedimiento, luego realizamos la multiplicación para calcular los diferentes índices, por ej:

$mC^{colix} * HE$ = matriz con 1 para los pares de genes en operon con alta expresión que se conservan en ambos organismos; o “0” en caso contrario.

$mC^{colix} * LE$ = matriz con 1 para los pares de genes en operon con baja expresión que se conservan en ambos organismos; o “0” en caso contrario.

Realizamos el mismo análisis para: SIM, DIF, E-E y NE-NE. Y también para cada uno de los 81 organismos de la clase α -proteobacteria, teniendo en cuenta en cada comparación de a pares los genes ortólogos.

Análisis estadísticos.

Las correlaciones fueron analizadas con métodos de Correlación *Spearman-rank*. Las distribuciones pareadas mediante *t-paired test*. Los análisis fueron realizados utilizando los paquetes disponibles en el programa R (*The R Project for Statistical Computing*, <http://www.r-project.org/>)

Anexo.

Capítulo I.

Tabla S 1.1 Primers utilizados para amplificación y clonados.

ID gen	Primer	Target	Tamaño Target (pb)	
	Dirección*	Secuencia (5' - 3')		
AM1108	F	GCCTGGGTCGCAATTGTAT	N-term	921
	R Interno	ATTCACCCCCACCATAACTG		
	F	GCCTGGGTCGCAATTGTAT	ORF	1920
	R	TCTAAACCCGACGAAAGCACC		
AM127	F	GGGGATCCGGGGTGCTGTACTATCACG	N-term	1397
	R Interno	GGGGATCCCTTCGTGTTGGCAACACC		
	F Interno	GGGGATCCGGTGTGGCCAACACGAAG	C-term	1521
	R	GGGGATCCCTTGCTGCTTCACGTTACCG		
AM216	F	GGGGATCCATGTGGAAACCGAAGTTGG	N-term	1755
	R	GGGGATCCACTTGCACGGGTGCCACTG	C-term	764

*F, forward; R, reverse.

Tabla S 1.2 Primers utilizados para los ensayos de RT-PCR

ID gen	Primer	Target	Tamaño Target (pb)
	Dirección*	Secuencia (5' - 3')	
AM1108	F	AATATCGCGGACTGGCATGT	101
	R	CCTCGCTCAATACCGCATTT	
AM127	F	CATGAAGAACAACGTGCCCA	108
	R	GGCGCATTATCATTGGCAA	
AM216	F	GCGTTGCGAGAGATGACAAA	102
	R	CGGAAAGTCTCCACATGACGTA	

*F, forward; R, reverse.

Tabla S 1.3. Niveles de transcripción de las citoquinas en PBMC como respuesta a la incubación con proteínas recombinantes.

Proteína recombinante	Citoquina	Expression ratio (R)	Error estándar (ES)	p-value
AM1108	IL2	1,39	0,99	p > 0,05
	IL10	1,6	1,16	p > 0,05
	IL12	2,03	1,95	p > 0,05
	TNF-alpha	3,59	2,69	p = 0,051
	INF-gamma	2,18	2,21	p > 0,05
AM216 C-term	IL2	1,51	0,57	p > 0,05
	IL10	2,45	1,83	p > 0,05
	IL12	1,78	1,67	p > 0,05
	TNF-alpha	1,9	1,15	p > 0,05
	INF-gamma	3,31	1,8	p = 0,043

Capítulo II.

Tabla S 2.1 Primers utilizados para las RT-PCR y Real-time qPCR.

Primers	Secuencia (5' - 3')
RTTatAAMF	GGTCCGTGGCAGATTTTTCTT
RTTatAAMR	CTCAGACTACGGATGCCCTTTC
RTTatBAMF	ACGAGGCTTACGAAGTTAGCGA
RTTatBAMR	TTGGCTTGTGCAGAGGCAT
RTTatCAMF	CTCCGTCTGTGTTTTATCGG
RTTatCAMR	TTATCGCCTGCAAGCTCCA
RT16SF	TGATCCTGGCTCAGAACGAAC
RT16SR	ACCCGTCTGCCACTAACCATAC
RTGroELF	GGCAAGATAGCACAGTGCGTAA
RTGroELR	TACCGTCAGTCCTTTTCGACCTC
RTTatABruF	TGATCGTTCTGGCGGTTGT
RTTatABruR	CCATGCCCTGCTTGAAATTC
RTTatBBruF	GTCATGATTGTGGTGGTTCGGT
RTTatBBruR	TGATGGCGAAACTCGTTGG
RTTatCBruF	TTTGCTTTTGGCCTGGTCTT
RTTatCBruR	GCATATTTGCGCTTGTCTTC
IF1F	ACTGCTGCCCAATGCAATG
IF2R	CCAGAACACGGATACGGTTCTT
RPLLF	TGAGCTGTCCAAGCTTCTCGA
RPLLR	GAATTCGGTCTTTTCTTCTGCG

Tabla S 2.2 Cepas y plasmidos utilizados en el estudio.

Bacterias	Genotipo	Fuente
MC-4100-P	Km ^r	T. Palmer Lab
JARVIG-P	$\Delta tatA\Delta tatE$; Km ^r	T. Palmer Lab
BOD-P	$\Delta tatB$; Km ^r	T. Palmer Lab
BILKO-P	$\Delta tatC$; Km ^r	T. Palmer Lab
DADE-P	$\Delta tatABC$; Km ^r	T. Palmer Lab
DH5 α		Promega
Plasmidos		
pUNIPROM	Amp ^R	T. Palmer Lab
pUNIPROM_AmTatA	Amp ^R	Este trabajo
pUNIPROM_AmTatB (<i>AM476</i>)	Amp ^R	Este trabajo
pUNIPROM_AmTatC (<i>AM740</i>)	Amp ^R	Este trabajo
pUNIPROM_BaTatA (<i>BAB1_0901</i>)	Amp ^R	Este trabajo
pUNIPROM_BaTatB (<i>BAB1_0902</i>)	Amp ^R	Este trabajo
pUNIPROM_BaTatC (<i>BAB1_0903</i>)	Amp ^R	Este trabajo
pFAT415	Amp ^R	T. Palmer Lab
pFAT416	Amp ^R	T. Palmer Lab
pFAT417	Amp ^R	T. Palmer Lab

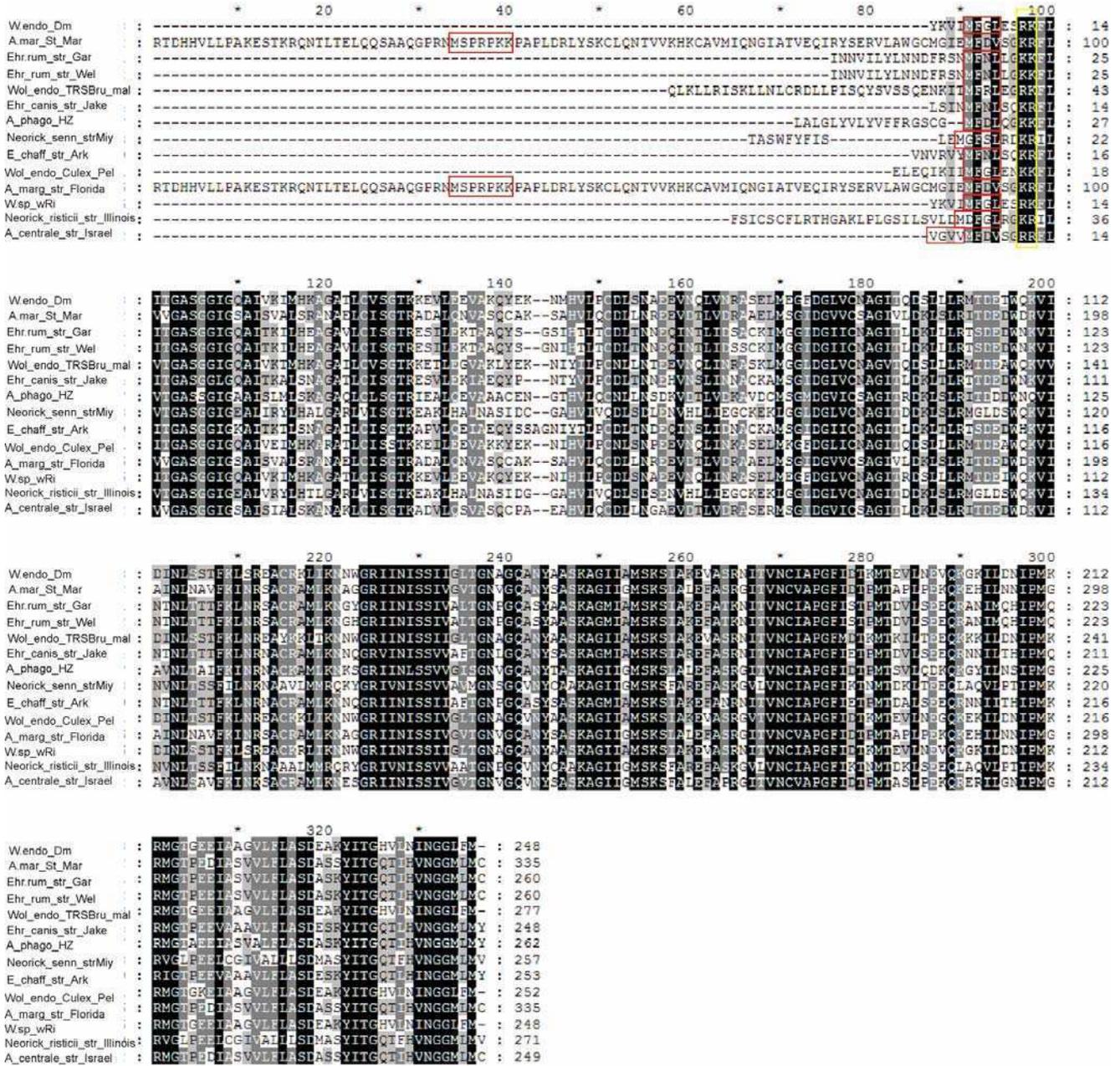


Figura S 2.1. COG1028, Anaplasmataceae protein sequence alignment

Multiple alignment of orthologs amino acid sequences (Dehydrogenases with different specificities) using CLUSTAL W [115] from *Anaplasma centrale str. Israel*, *Anaplasma marginale str. Florida*, *Anaplasma marginale str. St. Maries*, *Anaplasma phagocytophilum HZ*, *Ehrlichia canis str. Jake*, *Ehrlichia chaffeensis str. Arkansas*, *Ehrlichia ruminantium str. Gardel*, *Ehrlichia ruminantium str. Welgevonden*, *Neorickettsia risticii str. Illinois*, *Neorickettsia sennetsu str. Miyayama*, *Wolbachia endosymbiont of Culex quinquefasciatus*, *Wolbachia endosymbiont of Drosophila mel*, *Wolbachia endosymbiont str. TRS Brugia malayi*, *Wolbachia sp. wRi*. Annotated start sites were highlighted in red, RR motif and its variants were highlighted in yellow.

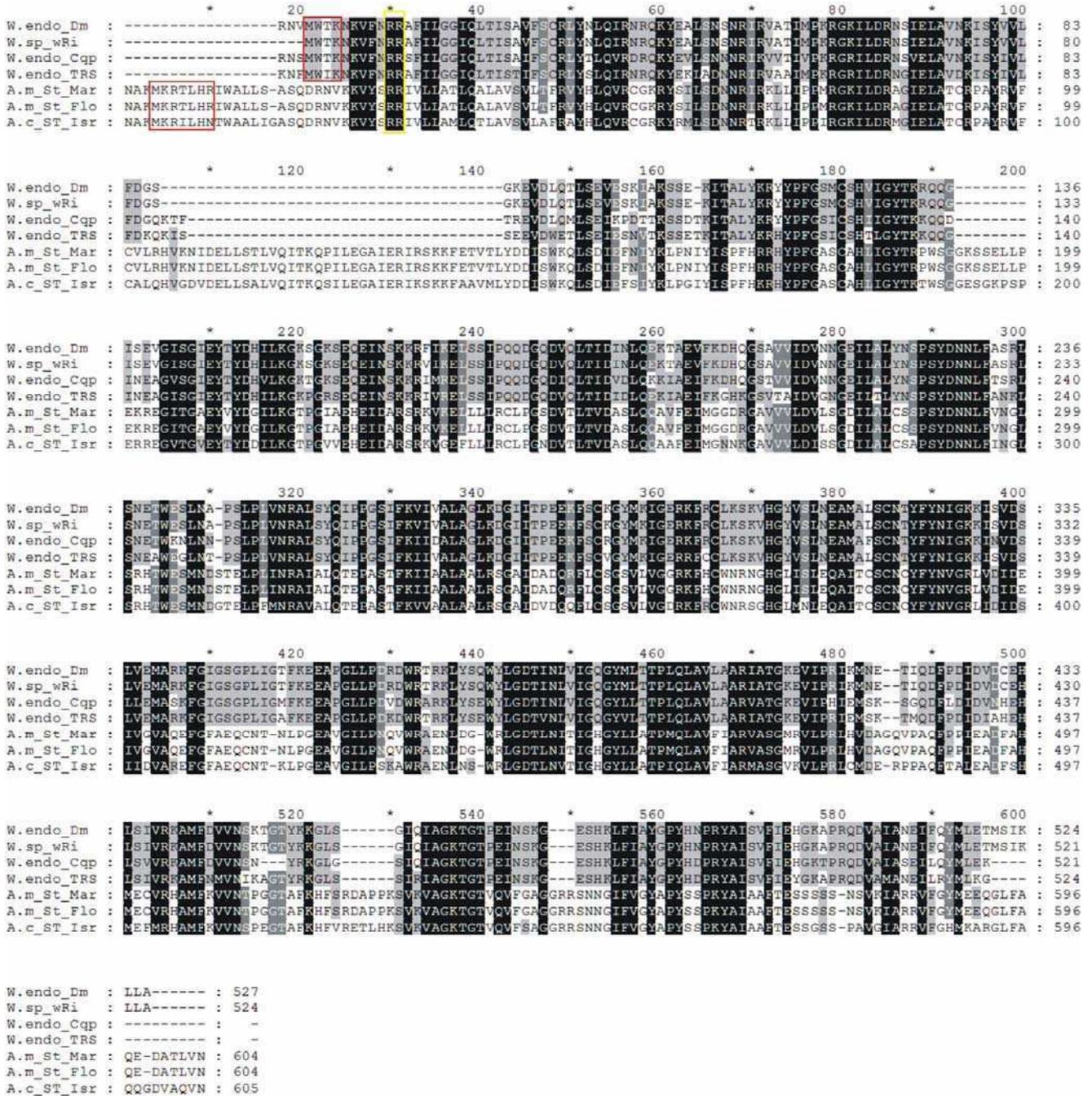


Figura S 2.2 COG0768, Anaplasmataceae protein sequence alignment

Multiple alignment of orthologous amino acid sequences (Cell division protein FtsI/penicillin-binding protein 2) using CLUSTAL W [115] from *Anaplasma centrale str. Israel*, *Anaplasma marginale str. Florida*, *Anaplasma marginale str. St. Maries*, *Wolbachia endosymbiont of Culex quinquefasciatus*, *Wolbachia endosymbiont of Drosophila mel*, *Wolbachia endosymbiont str. TRS Brugia malayi*, *Wolbachia sp. wRi*. Annotated start sites were highlighted in red, RR motif and its variants were highlighted in yellow.

Capítulo III.

Tabla S 3.1 Organismos de la clase α -proteobacteria.

Genoma ID	Organismo
NC_000963	<i>Rickettsia prowazekii</i> str. Madrid E
NC_002678	<i>Mesorhizobium loti</i> MAFF303099
NC_002696	<i>Caulobacter crescentus</i> CB15
NC_002978	<i>Wolbachia</i> endosymbiont of <i>Drosophila melanogaster</i>
NC_003047	<i>Sinorhizobium meliloti</i> 1021
NC_003062	<i>Agrobacterium tumefaciens</i> str. C58
NC_003103	<i>Rickettsia conorii</i> str. Malish 7
NC_003317	<i>Brucella melitensis</i> bv. 1 str. 16M
NC_003911	<i>Ruegeria pomeroyi</i> DSS-3
NC_004310	<i>Brucella suis</i> 1330
NC_004463	<i>Bradyrhizobium japonicum</i> USDA 110
NC_004842	<i>Anaplasma marginale</i> str. Str. St. Maries
NC_005295	<i>Ehrlichia ruminantium</i> str. Welgevonden
NC_005296	<i>Rhodopseudomonas palustris</i> CGA009
NC_005955	<i>Bartonella quintana</i> str. Toulouse
NC_005956	<i>Bartonella henselae</i> str. Houston-1
NC_006142	<i>Rickettsia typhi</i> str. Wilmington
NC_006526	<i>Zymomonas mobilis</i> subsp. <i>mobilis</i> ZM4
NC_006677	<i>Gluconobacter oxydans</i> 621H
NC_006831	<i>Ehrlichia ruminantium</i> str. Gardel
NC_006832	<i>Ehrlichia ruminantium</i> str. Welgevonden
NC_006833	<i>Wolbachia</i> endosymbiont strain TRS of <i>Brugia malayi</i>
NC_006932	<i>Brucella abortus</i> bv. 1 str. 9-941
NC_007109	<i>Rickettsia felis</i> URRWXCa2
NC_007205	<i>Candidatus Pelagibacter ubique</i> HTCC1062
NC_007354	<i>Ehrlichia canis</i> str Jake
NC_007406	<i>Nitrobacter winogradskyi</i> Nb-255
NC_007493	<i>Rhodobacter sphaeroides</i> 2.4.1
NC_007618	<i>Brucella melitensis</i> biovar Abortus 2308
NC_007626	<i>Magnetospirillum magneticum</i> AMB-1
NC_007643	<i>Rhodospirillum rubrum</i> ATCC 11170
NC_007722	<i>Erythrobacter litoralis</i> HTCC2594
NC_007761	<i>Rhizobium etli</i> CFN 42
NC_007778	<i>Rhodopseudomonas palustris</i> HaA2
NC_007794	<i>Novosphingobium aromaticivorans</i> DSM 12444
NC_007797	<i>Anaplasma phagocytophilum</i> HZ
NC_007798	<i>Neorickettsia sennetsu</i> str. Miyayama
NC_007799	<i>Ehrlichia chaffeensis</i> str. Arkansas
NC_007802	<i>Jannaschia</i> sp. CCS1

NC_007925	<i>Rhodopseudomonas palustris</i> BisB18
NC_007940	<i>Rickettsia bellii</i> RML369-C
NC_007958	<i>Rhodopseudomonas palustris</i> BisB5
NC_007964	<i>Nitrobacter hamburgensis</i> X14
NC_008044	<i>Ruegeria</i> sp. TM1040
NC_008048	<i>Sphingopyxis alaskensis</i> RB2256
NC_008209	<i>Roseobacter denitrificans</i> OCh 114
NC_008343	<i>Granulibacter bethesdensis</i> CGDNIH1
NC_008347	<i>Maricaulis maris</i> MCS10
NC_008358	<i>Hyphomonas neptunium</i> ATCC 15444
NC_008380	<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> 3841
NC_008435	<i>Rhodopseudomonas palustris</i> BisA53
NC_008686	<i>Paracoccus denitrificans</i> PD1222
NC_008783	<i>Bartonella bacilliformis</i> KC583
NC_009049	<i>Rhodobacter sphaeroides</i> ATCC 17029
NC_009428	<i>Rhodobacter sphaeroides</i> ATCC 17025
NC_009445	<i>Bradyrhizobium</i> sp. ORS 278
NC_009484	<i>Acidiphilium cryptum</i> JF-5
NC_009485	<i>Bradyrhizobium</i> sp. BTAi1
NC_009488	<i>Orientia tsutsugamushi</i> str. Boryong
NC_009505	<i>Brucella ovis</i> ATCC 25840
NC_009511	<i>Sphingomonas wittichii</i> RW1
NC_009636	<i>Sinorhizobium medicae</i> WSM419
NC_009667	<i>Ochrobactrum anthropi</i> ATCC 49188
NC_009719	<i>Parvibaculum lavamentivorans</i> DS-1
NC_009720	<i>Xanthobacter autotrophicus</i> Py2
NC_009879	<i>Rickettsia canadensis</i> str. McKiel
NC_009881	<i>Rickettsia akari</i> str. Hartford
NC_009882	<i>Rickettsia rickettsii</i> str. 'Sheila Smith'
NC_009883	<i>Rickettsia bellii</i> OSU 85-389
NC_009900	<i>Rickettsia massiliae</i> MTU5
NC_009937	<i>Azorhizobium caulinodans</i> ORS 571
NC_009952	<i>Dinoroseobacter shibae</i> DFL 12
NC_010103	<i>Brucella canis</i> ATCC 23365
NC_010125	<i>Gluconacetobacter diazotrophicus</i> PAI 5
NC_010161	<i>Bartonella tribocorum</i> CIP 105476
NC_010169	<i>Brucella suis</i> ATCC 23445
NC_010172	<i>Methylobacterium extorquens</i> PA1
NC_010263	<i>Rickettsia rickettsii</i> str. Iowa
NC_010338	<i>Caulobacter</i> sp. K31
NC_010505	<i>Methylobacterium radiotolerans</i> JCM 2831
NC_010511	<i>Methylobacterium</i> sp. 4-46
NC_010581	<i>Beijerinckia indica</i> subsp. <i>indica</i> ATCC 9039

Tabla S 3.2 Organismos de la clase β -proteobacteria.

Genoma ID	Organismo
NC_002927	<i>Bordetella bronchiseptica</i> RB50
NC_002928	<i>Bordetella parapertussis</i> 12822
NC_002929	<i>Bordetella pertussis</i> Tohama I
NC_002946	<i>Neisseria gonorrhoeae</i> FA 1090
NC_003112	<i>Neisseria meningitidis</i> MC58
NC_003116	<i>Neisseria meningitidis</i> Z2491
NC_004757	<i>Nitrosomonas europaea</i> ATCC 19718
NC_005085	<i>Chromobacterium violaceum</i> ATCC 12472
NC_006348	<i>Burkholderia mallei</i> ATCC 23344
NC_006350	<i>Burkholderia pseudomallei</i> K96243
NC_007298	<i>Dechloromonas aromatica</i> RCB
NC_007404	<i>Thiobacillus denitrificans</i> ATCC 25259
NC_007434	<i>Burkholderia pseudomallei</i> 1710b
NC_007509	<i>Burkholderia</i> sp, 383
NC_007650	<i>Burkholderia thailandensis</i> E264
NC_007947	<i>Methylobacillus flagellatus</i> KT
NC_007951	<i>Burkholderia xenovorans</i> LB400
NC_008060	<i>Burkholderia cenocepacia</i> AU 1054
NC_008702	<i>Azoarcus</i> sp, BH72
NC_008767	<i>Neisseria meningitidis</i> FAM18
NC_008835	<i>Burkholderia mallei</i> NCTC 10229
NC_009074	<i>Burkholderia pseudomallei</i> 668
NC_009076	<i>Burkholderia pseudomallei</i> 1106a
NC_009079	<i>Burkholderia mallei</i> NCTC 10247
NC_009138	<i>Hermiimonas arsenicoxydans</i>
NC_009379	<i>Polynucleobacter necessarius</i> subsp, <i>asymbioticus</i> QLW-P1DMWA-1
NC_009659	<i>Janthinobacterium</i> sp, Marseille
NC_010002	<i>Delftia acidovorans</i> SPH-1
NC_010120	<i>Neisseria meningitidis</i> 053442
NC_010170	<i>Bordetella petrii</i> DSM 12804
NC_010508	<i>Burkholderia cenocepacia</i> MC0-3
NC_010524	<i>Leptothrix cholodnii</i> SP-6
NC_010531	<i>Polynucleobacter necessarius</i> subsp, <i>necessarius</i> STIR1
NC_010645	<i>Bordetella avium</i> 197N
NC_012559	<i>Laribacter hongkongensis</i> HLHK9
NC_012791	<i>Variovorax paradoxus</i> S110
NC_012968	<i>Methylotenera mobilis</i> JLW8

Tabla S 3.3 Organismos del phylum Firmicutes.

Genoma ID	Organismo
NC_000964	<i>Bacillus subtilis</i> subsp <i>subtilis</i> str 168
NC_002570	<i>Bacillus halodurans</i> C-125
NC_002662	<i>Lactococcus lactis</i> subsp <i>lactis</i> II1403
NC_002952	<i>Staphylococcus aureus</i> subsp <i>aureus</i> MRSA252
NC_003869	<i>Thermoanaerobacter tengcongensis</i> MB4
NC_003997	<i>Bacillus anthracis</i> str Ames
NC_004193	<i>Oceanobacillus iheyensis</i> HTE831
NC_004350	<i>Streptococcus mutans</i> UA159
NC_004368	<i>Streptococcus agalactiae</i> NEM316
NC_005362	<i>Lactobacillus johnsonii</i> NCC 533
NC_005945	<i>Bacillus anthracis</i> str Sterne
NC_006177	<i>Symbiobacterium thermophilum</i> IAM 14863
NC_006270	<i>Bacillus licheniformis</i> ATCC 14580
NC_006449	<i>Streptococcus thermophilus</i> CNRZ1066
NC_006582	<i>Bacillus clausii</i> KSM-K16
NC_006814	<i>Lactobacillus acidophilus</i> NCFM
NC_007503	<i>Carboxydotherrmus hydrogenoformans</i> Z-2901
NC_007576	<i>Lactobacillus sakei</i> subsp <i>sakei</i> 23K
NC_007644	<i>Moorella thermoacetica</i> ATCC 39073
NC_007907	<i>Desulfitobacterium hafniense</i> Y51
NC_008024	<i>Streptococcus pyogenes</i> MGAS10750
NC_008054	<i>Lactobacillus delbrueckii</i> subsp <i>bulgaricus</i> ATCC 11842
NC_008261	<i>Clostridium perfringens</i> ATCC 13124
NC_008525	<i>Pediococcus pentosaceus</i> ATCC 25745
NC_008528	<i>Oenococcus oeni</i> PSU-1
NC_008530	<i>Lactobacillus gasseri</i> ATCC 33323
NC_008555	<i>Listeria welshimeri</i> serovar 6b str SLCC5334
NC_008593	<i>Clostridium novyi</i> NT
NC_009009	<i>Streptococcus sanguinis</i> SK36
NC_009012	<i>Clostridium thermocellum</i> ATCC 27405
NC_009253	<i>Desulfotomaculum reducens</i> MI-1
NC_009442	<i>Streptococcus suis</i> 05ZYH33
NC_009454	<i>Pelotomaculum thermopropionicum</i> SI
NC_009513	<i>Lactobacillus reuteri</i> DSM 20016
NC_009617	<i>Clostridium beijerinckii</i> NCIMB 8052
NC_009633	<i>Alkaliphilus metalliredigens</i> QYMF
NC_009725	<i>Bacillus amyloliquefaciens</i> FZB42

NC_009848	Bacillus pumilus SAFR-032
NC_009922	Alkaliphilus oremlandii OhILAs
NC_010001	Clostridium phytofermentans ISDg
NC_010080	Lactobacillus helveticus DPC 4571
NC_010321	Thermoanaerobacter pseudethanolicus ATCC 33223
NC_010337	Heliobacterium modesticaldum Ice1
NC_010380	Streptococcus pneumoniae Hungary19A-6
NC_010424	Candidatus Desulforudis audaxviator MP104C
NC_010610	Lactobacillus fermentum IFO 3956
NC_010999	Lactobacillus casei BL23
NC_011295	Coprothermobacter proteolyticus DSM 5265
NC_011567	Anoxybacillus flavithermus WK1
NC_011660	Listeria monocytogenes HCC23
NC_011725	Bacillus cereus B4264
NC_011898	Clostridium cellulolyticum H10
NC_011899	Halothermothrix orenii H 168
NC_012004	Streptococcus uberis 0140J
NC_012471	Streptococcus equi subsp equi 4047
NC_012491	Brevibacillus brevis NBRC 100599
NC_012563	Clostridium botulinum A2 str Kyoto
NC_012673	Exiguobacterium sp AT1b
NC_012781	Eubacterium rectale ATCC 33656
NC_012891	Streptococcus dysgalactiae subsp equisimilis GGS_124
NC_012914	Paenibacillus sp JDR-2
NC_012925	Streptococcus suis P1/7
NC_012984	Lactobacillus plantarum JDM1
NC_013198	Lactobacillus rhamnosus GG
NC_013216	Desulfotomaculum acetoxidans DSM 771

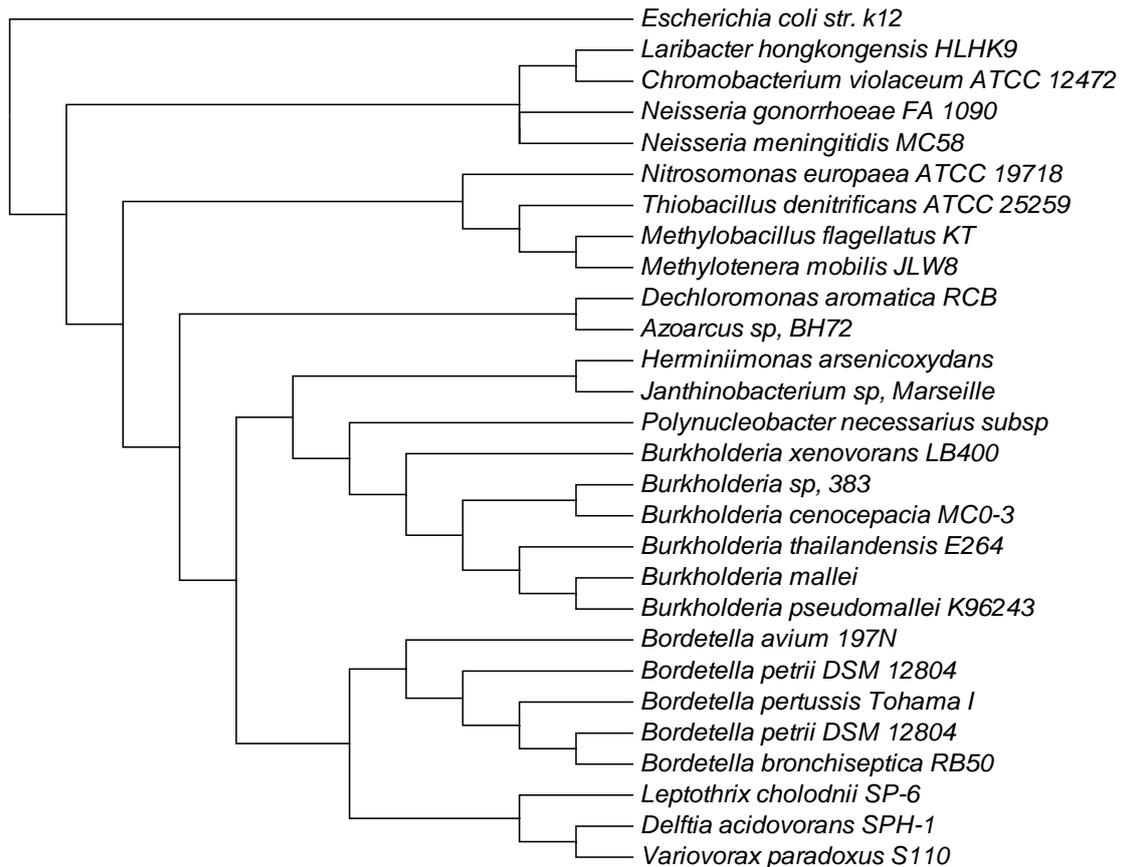


Figura S 3.1. Filogenia de la clase β -proteobacteria.

Análisis filogenético para 27 organismos de la clase β -proteobacteria (Neighbor Joining, NJ) utilizando las matrices de distancia estimadas por el modelo Jones–Thornton–Taylor (JTT), MEGA 4.0, bootstrap=1000. En árbol se realizó con el conjunto de proteínas presentes en >90% de los genomas (*core genome*). Las proteínas ortólogas de los diferentes organismos se alinearon individualmente utilizando el ClustalW y luego se concatenaron los alineamientos para la construcción de la filogenia. Los resultados concuerdan con la descripción y agrupamiento en los órdenes establecidos. Como *outgroup* se utilizó *Escherichia coli str. K-12 substr. MG1655*.

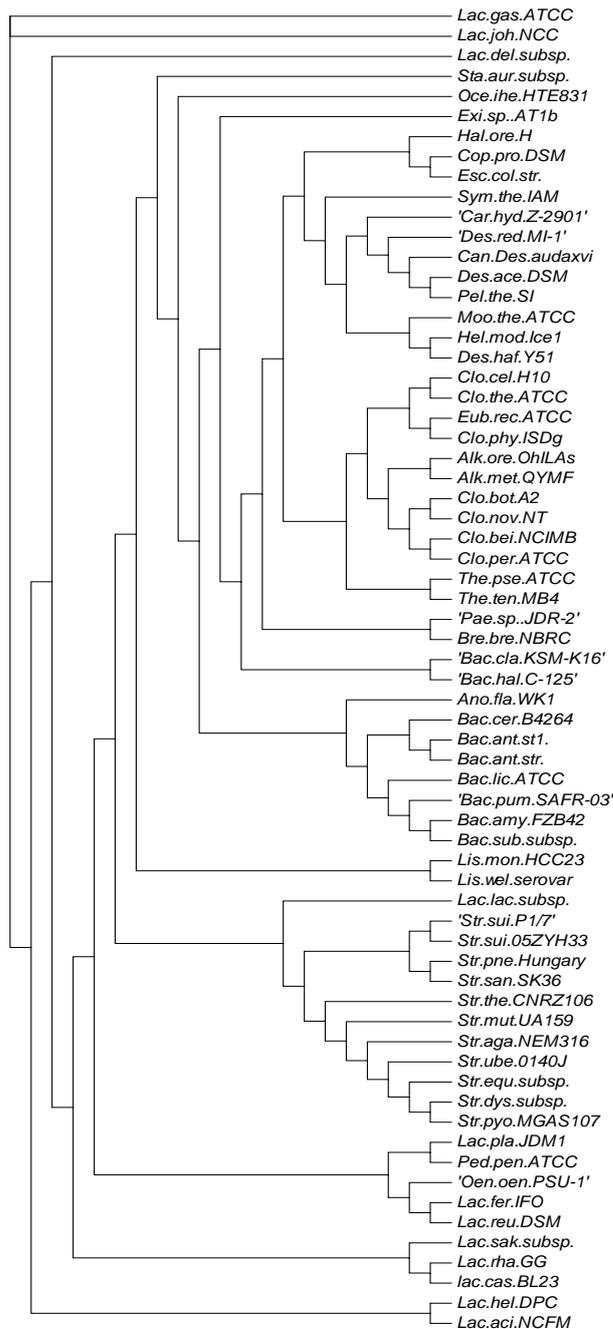


Figura S 3.5. Filogenia de Firmicutes.

Análisis filogenético para 60 organismos de Firmicutes (Neighbor Joining, NJ) utilizando las matrices de distancia estimadas por el modelo Jones–Thornton–Taylor (JTT), MEGA 4.0, bootstrap=1000. En árbol se realizó con el conjunto de proteínas presentes en >90% de los genomas (*core geome*). Las proteínas ortólogas de los diferentes organismos se alinearon individualmente utilizando el ClustalW y luego se concatenaron los alineamientos para la construcción de la filogenia. Los resultados concuerdan con la descripción y agrupamiento en los órdenes establecidos. Como *outgroup* se utilizó *Escherichia coli str. K-12 substr. MG1655*.

Tabla S 3.4. Características de los genomas de los organismos estudiados para los diferentes grupos filogenéticos.

α-proteobacteria										
Organismo	Tamaño genoma Mb	Genes	Proteínas	Genes en operones	# of operons	Media	Desvío	% genes en operon	# genes en single UT	% genes en single UT
<i>Rickettsia prowazekii</i> str. Madrid E	1,11	888	835	485	188	2,56	2,00	54,62	403	45,38
<i>Mesorhizobium loti</i> MAFF303099	7,04	6804	6743	3883	1352	2,87	1,48	57,07	2921	42,93
<i>Caulobacter crescentus</i> CB15	4,02	3819	3737	2239	813	2,75	1,45	58,63	1580	41,37
<i>Wolbachia endosymbiont of Drosophila melanogaster</i>	1,27	1309	1195	699	257	2,72	1,90	53,40	610	46,60
<i>Senorhizobium meliloti</i> 1021	6,69	6292	6218	3537	1236	2,86	1,41	56,21	2755	43,79
<i>Agrobacterium tumefaciens</i> str. C58	5,46	5220	5158	2791	1001	2,79	1,30	53,47	2429	46,53
<i>Rickettsia conorii</i> str. Malish 7	1,27	1414	1374	771	292	2,63	1,39	54,53	643	45,47
<i>Brucella melitensis</i> bv. 1 str. 16M	3,29	3264	3198	1833	647	2,83	1,56	56,16	1431	43,84
<i>Ruegeria pomeroyi</i> DSS-3	4,60	4348	4252	2818	924	3,05	1,51	64,81	1530	35,19
<i>Brucella suis</i> 1330	3,32	3447	3271	1852	653	2,84	1,50	53,73	1595	46,27
<i>Bradyrhizobium japonicum</i> USDA 110	9,11	8373	8317	4613	1598	2,89	1,51	55,09	3760	44,91
<i>Anaplasma marginale</i> str. St. Maries	1,20	1004	948	507	187	2,71	1,88	50,50	497	49,50
<i>Ehrlichia rumenantium</i> str. Welgevonden	1,52	961	888	396	137	2,89	2,69	41,21	565	58,79
<i>Rhodopseudomonas palustris</i> CGA009	5,46	4891	4813	2620	927	2,83	1,55	53,57	2271	46,43
<i>Bartonella quentana</i> str. Toulouse	1,58	1356	1142	635	235	2,70	1,35	46,83	721	53,17
<i>Bartonella henselae</i> str. Houston-1	1,93	1665	1488	878	315	2,77	1,47	52,73	787	47,27
<i>Rickettsia typhi</i> str. Wilmington	1,11	918	837	462	181	2,54	1,95	50,33	456	49,67
<i>Zymomonas mobilis</i> subsp. mobilis ZM4	2,06	1807	1736	1044	383	2,72	1,30	57,78	763	42,22
<i>Gluconobacter oxydans</i> 621H	2,70	2499	2432	1491	506	2,94	1,51	59,66	1008	40,34
<i>Ehrlichia rumenantium</i> str. Gardel	1,50	989	950	429	145	2,95	2,78	43,38	560	56,62
<i>Ehrlichia rumenantium</i> str. Welgevonden	1,51	997	958	433	150	2,89	2,65	43,43	564	56,57
<i>Wolbachia endosymbiont str. TRS of Brugia malayi</i>	1,08	940	805	452	171	2,64	2,65	48,09	488	51,91
<i>Brucella abortus</i> bv. 1 str. 9-941	3,29	3355	3084	1751	616	2,84	1,52	52,19	1604	47,81

<i>Rickettsia felis</i> URRWXCa2	1,49	1439	1400	731	292	2,50	1,58	50,80	708	49,20
<i>Ehrlichia canis</i> str Jake	1,32	984	925	431	146	2,95	2,81	43,80	553	56,20
<i>Nitrobacter wenogradskyi</i> Nb-255	3,40	3198	3122	1642	619	2,65	1,38	51,34	1556	48,66
<i>Rhodobacter sphaeroides</i> 2.4.1	4,13	3958	3857	2609	811	3,22	1,68	65,92	1349	34,08
<i>Brucella melitensis</i> biovar Abortus 2308	3,28	3418	3034	1732	617	2,81	1,38	50,67	1686	49,33
<i>Magnetospirillum magneticum</i> AMB-1	4,97	4613	4561	3111	980	3,17	2,05	67,44	1502	32,56
<i>Rhodospirillum rubrum</i> ATCC 11170	4,35	3867	3788	2205	799	2,76	1,41	57,02	1662	42,98
<i>Erythrobacter litoralis</i> HTCC2594	3,05	3059	3011	1977	680	2,91	1,87	64,63	1082	35,37
<i>Rhizobium etli</i> CFN 42	4,38	4126	4035	2879	1030	2,79	1,23	69,78	1247	30,22
<i>Rhodopseudomonas palustris</i> HaA2	5,33	4772	4683	2557	907	2,81	1,57	53,58	2215	46,42
<i>Novosphengobium aromaticivorans</i> DSM 12444	4,05	3843	3755	2606	833	3,13	1,85	67,81	1237	32,19
<i>Anaplasma phagocytophilum</i> HZ	1,47	1411	1264	547	214	2,55	1,81	38,77	864	61,23
<i>Neorickettsia sennetsu</i> str. Miyayama	0,86	973	932	563	192	2,93	2,12	57,86	410	42,14
<i>Ehrlichia chaffeensis</i> str. Arkansas	1,18	1158	1105	526	187	2,80	2,33	45,42	632	54,58
<i>Jannaschia</i> sp. CCS1	4,32	4265	4212	2783	923	3,02	1,55	65,25	1482	34,75
<i>Rhodopseudomonas palustris</i> BisB18	5,51	5016	4886	2750	952	2,89	1,59	54,82	2266	45,18
<i>Rickettsia bellii</i> RML369-C	1,52	1469	1429	845	318	2,66	1,74	57,52	624	42,48
<i>Rhodopseudomonas palustris</i> BisB5	4,89	4492	4397	2378	860	2,77	1,48	52,94	2114	47,06
<i>Nitrobacter hamburgensis</i> X14	4,41	4107	3804	1860	685	2,72	1,40	45,29	2247	54,71
<i>Ruegeria</i> sp. TM1040	4,02	3840	3758	2171	759	2,86	1,36	56,54	1669	43,46
<i>Sphengopyxis alaskensis</i> RB2256	3,35	3230	3165	1992	699	2,85	1,56	61,67	1238	38,33
<i>Roseobacter denitrificans</i> OCh 114	4,13	4007	3946	2531	862	2,93	1,49	63,16	1476	36,84
<i>Granulibacter thesedensis</i> CGDNIH1	2,71	2494	2433	1533	506	3,03	1,63	61,47	961	38,53
<i>Maricaulis maris</i> MCS10	3,37	3133	3063	1893	642	2,95	1,86	60,42	1240	39,58
<i>Hyphomonas neptunium</i> ATCC 15444	3,71	3568	3505	2059	716	2,87	1,56	57,71	1509	42,29
<i>Rhizobium leguminosarum</i> bv. viciae 3841	7,45	7018	6855	3820	1303	2,93	1,47	54,43	3198	45,57
<i>Rhodopseudomonas palustris</i> BisA53	5,51	4972	4878	2635	928	2,84	1,62	53,00	2337	47,00
<i>Paracoccus denitrificans</i> PD1222	5,24	5177	5077	3546	1091	3,25	1,81	68,50	1631	31,50
<i>Bartonella bacilliformis</i> KC583	1,45	1375	1283	688	247	2,78	1,57	50,04	687	49,96
<i>Rhodobacter sphaeroides</i> ATCC 17029	4,37	4136	4024	2735	841	3,25	1,82	66,13	1401	33,87
<i>Rhodobacter sphaeroides</i> ATCC	4,10	4030	3913	2641	810	3,26	1,97	65,53	1389	34,47

17025										
<i>Bradyrhizobium</i> sp. ORS 278	7,46	6818	6717	3574	1258	2,84	1,62	52,42	3244	47,58
<i>Acidiphilium cryptum</i> JF-5	3,39	3158	3063	2182	685	3,19	1,76	69,09	976	30,91
<i>Bradyrhizobium</i> sp. BTAi1	8,26	7550	7393	3982	1393	2,86	1,65	52,74	3568	47,26
<i>Orientia tsutsugamushi</i> str. Boryong	2,13	2216	1182	547	211	2,59	1,38	24,68	1669	75,32
<i>Brucella ovis</i> ATCC 25840	3,28	3198	2890	1579	574	2,75	1,41	49,37	1619	50,63
<i>Sphengomonas wittichii</i> RW1	5,38	4941	4850	3405	1181	2,88	1,40	68,91	1536	31,09
<i>Senorhizobium medicae</i> WSM419	6,60	6389	6064	3458	1166	2,97	1,50	54,12	2931	45,88
<i>Ochrobactrum anthropi</i> ATCC 49188	4,78	4528	4424	2572	873	2,95	1,63	56,80	1956	43,20
<i>Parvibaculum lavamentivorans</i> DS-1	3,91	3707	3636	2216	785	2,82	1,50	59,78	1491	40,22
<i>Xanthobacter autotrophicus</i> Py2	5,31	4847	4746	2980	1018	2,93	1,64	61,48	1867	38,52
<i>Rickettsia canadensis</i> str. McKiel	1,16	1126	1090	521	206	2,52	1,85	46,27	605	53,73
<i>Rickettsia akari</i> str. Hartford	1,23	1293	1258	637	253	2,52	1,25	49,27	656	50,73
<i>Rickettsia rickettsii</i> str. 'Sheila Smith'	1,26	1379	1343	714	282	2,53	1,60	51,78	665	48,22
<i>Rickettsia bellii</i> OSU 85-389	1,53	1511	1475	878	333	2,62	1,18	58,11	633	41,89
<i>Rickettsia massiliae</i> MTU5	1,36	1423	968	543	205	2,65	1,94	38,16	880	61,84
<i>Azorhizobium caulenodans</i> ORS 571	5,37	4782	4717	2960	988	2,99	1,57	61,90	1822	38,10
<i>Denoroseobacter shibae</i> DFL 12	3,79	3634	3584	2385	782	3,05	1,57	65,63	1249	34,37
<i>Brucella canis</i> ATCC 23365	3,31	3418	3251	1871	662	2,83	1,49	54,74	1547	45,26
<i>Gluconacetobacter diazotrophicus</i> PAI 5	3,94	3926	3777	2405	813	2,96	1,66	61,26	1521	38,74
<i>Bartonella tribocorum</i> CIP 105476	2,62	2148	2069	1213	410	2,95	1,92	56,47	935	43,53
<i>Brucella suis</i> ATCC 23445	3,32	3426	3241	1836	652	2,82	1,45	53,59	1590	46,41
<i>Methylobacterium extorquens</i> PA1	5,47	4956	4829	2403	873	2,75	1,44	48,49	2553	51,51
<i>Rickettsia rickettsii</i> str. Iowa	1,27	1493	1384	750	294	2,55	1,59	50,23	743	49,77
<i>Caulobacter</i> sp. K31	5,48	5133	5061	2780	1001	2,78	1,62	54,16	2353	45,84
<i>Methylobacterium radiotolerans</i> JCM 2831	6,08	5839	5686	3379	1202	2,81	1,43	57,87	2460	42,13
<i>Methylobacterium</i> sp. 4-46	7,66	7047	6609	3421	1221	2,80	1,53	48,55	3626	51,45
<i>Beijerinckia endica</i> subsp. <i>endica</i> ATCC 9039	4,17	3740	3569	1575	615	2,56	1,14	42,11	2165	57,89
β-proteobacteria										
Organismo	Tamaño genoma Mb	Genes	Proteínas	Genes en operones	# of operons	Media	Desvío	% genes en operon	# genes en single UT	% genes en single UT

<i>Bordetella bronchiseptica</i> RB50	5,34	5072	4994	3665	1080	3,39	2,08	72,26	1407	27,74
<i>Bordetella parapertussis</i> 12822	4,77	4467	4185	2918	922	3,16	1,82	65,32	1549	34,68
<i>Bordetella pertussis</i> Tohama I	4,09	3867	3436	2206	744	2,97	1,66	57,05	1661	42,95
<i>Neisseria gonorrhoeae</i> FA 1090	2,15	2069	2002	1136	422	2,69	1,15	54,91	933	45,09
<i>Neisseria meningitidis</i> MC58	2,27	2225	2063	1199	443	2,71	1,38	53,89	1026	46,11
<i>Neisseria meningitidis</i> Z2491	2,18	2065	1909	1135	410	2,77	1,57	54,96	930	45,04
<i>Nitrosomonas europaea</i> ATCC 19718	2,81	2620	2461	1639	571	2,87	1,60	62,56	981	37,44
<i>Chromobacterium violaceum</i> ATCC 12472	4,75	4528	4405	2698	877	3,08	1,68	59,58	1830	40,42
<i>Burkholderia mallei</i> ATCC 23344	5,84	5508	5024	2995	1069	2,80	1,40	54,38	2513	45,62
<i>Burkholderia pseudomallei</i> K96243	7,25	5935	5728	3457	1136	3,04	1,73	58,25	2478	41,75
<i>Dechloromonas aromatica</i> RCB	4,50	4283	4171	3109	919	3,38	2,25	72,59	1174	27,41
<i>Thiobacillus denitrificans</i> ATCC 25259	2,91	2879	2827	2115	611	3,46	2,18	73,46	764	26,54
<i>Burkholderia pseudomallei</i> 1710b	7,31	6419	6344	3621	1267	2,86	1,48	56,41	2798	43,59
<i>Burkholderia</i> sp. 383	8,68	7823	7716	4503	1522	2,96	1,70	57,56	3320	42,44
<i>Burkholderia thailandensis</i> E264	6,71	5712	5632	3570	1171	3,05	1,73	62,50	2142	37,50
<i>Methylobacillus flagellatus</i> KT	2,97	2822	2753	1970	603	3,27	1,86	69,81	852	30,19
<i>Burkholderia xenovorans</i> LB400	9,73	9043	8702	4979	1682	2,96	1,56	55,06	4064	44,94
<i>Burkholderia cenocepacia</i> AU 1054	7,28	6632	6477	4151	1406	2,95	1,63	62,59	2481	37,41
<i>Azoarcus</i> sp. BH72	4,38	4057	3989	2932	906	3,24	1,90	72,27	1125	27,73
<i>Neisseria meningitidis</i> FAM18	2,19	2046	1917	1162	427	2,72	1,43	56,79	884	43,21
<i>Burkholderia mallei</i> NCTC 10229	5,74	5623	5509	3477	1180	2,95	1,66	61,84	2146	38,16
<i>Burkholderia pseudomallei</i> 668	7,04	7194	7116	4568	1461	3,13	1,81	63,50	2626	36,50
<i>Burkholderia pseudomallei</i> 1106a	7,09	7255	7174	4497	1462	3,08	1,74	61,98	2758	38,02
<i>Burkholderia mallei</i> NCTC 10247	5,85	5925	5415	3670	1250	2,93	1,63	61,94	2255	38,06
<i>Hermeniimonas arsenicoxydans</i>	3,42	3361	3295	1982	674	2,94	1,53	58,97	1379	41,03
<i>Polynucleobacter necessarius</i> subsp. <i>asymbioticus</i> QLW-P1DMWA-1	2,16	2132	2077	1614	458	3,52	2,31	75,70	518	24,30
<i>Janthenobacterium</i> sp. Marseille	4,11	3750	3697	2379	761	3,13	1,82	63,44	1371	36,56
<i>Delftia acidovorans</i> SPH-1	6,77	6153	6040	3961	1300	3,05	1,71	64,38	2192	35,62
<i>Neisseria meningitidis</i> 053442	2,15	2122	2020	1160	434	2,67	1,35	54,67	962	45,33
<i>Bordetella petrii</i> DSM 12804	5,29	5089	5027	3511	1088	3,23	1,88	68,99	1578	31,01
<i>Burkholderia cenocepacia</i> MC0-3	7,97	7176	7008	4130	1396	2,96	1,65	57,55	3046	42,45
<i>Leptothrix cholodnii</i> SP-6	4,91	4443	4363	3314	970	3,42	2,25	74,59	1129	25,41

<i>Polynucleobacter necessarius</i> subsp. <i>necessarius</i> STIR1	1,56	1783	1508	1086	322	3,37	2,35	60,91	697	39,09
<i>Bordetella avium</i> 197N	3,73	3502	3381	2269	705	3,22	1,86	64,79	1233	35,21
<i>Laribacter hongkongensis</i> HLHK9	3,17	3333	3235	2111	686	3,08	1,73	63,34	1222	36,66
<i>Variovorax paradoxus</i> S110	6,75	6453	6279	4522	1373	3,29	2,01	70,08	1931	29,92
<i>Methylobacter mobilis</i> JLW8	2,55	2403	2338	1613	505	3,19	1,95	67,12	790	32,88
Firmicutes										
Organismo	Tamaño genoma Mb	Genes	Proteínas	Genes en operones	# de operons	Media	Desvío	%genes en operon	# genes en single UT	% genes en single UT
<i>Bacillus subtilis</i> subsp <i>subtilis</i> str 168	4,22	4422	4176	2712	848	3,20	2,30	61,33	1710	38,67
<i>Bacillus halodurans</i> C-125	4,20	4170	4065	2488	831	2,99	1,96	59,66	1682	40,34
<i>Lactococcus lactis</i> subsp <i>lactis</i> II1403	2,37	2425	2321	1433	474	3,02	1,92	59,09	992	40,91
<i>Staphylococcus aureus</i> subsp <i>aureus</i> MRSA252	2,90	2839	2650	1656	526	3,15	2,38	58,33	1183	41,67
<i>Thermoanaerobacter tengcongensis</i> MB4	2,69	2721	2588	1883	538	3,50	2,77	69,20	838	30,80
<i>Bacillus anthracis</i> str <i>Ames</i>	5,23	5636	5328	2950	1011	2,92	2,01	52,34	2686	47,66
<i>Oceanobacillus iheyensis</i> HTE831	3,63	3594	3500	2111	698	3,02	1,95	58,74	1483	41,26
<i>Streptococcus mutans</i> UA159	2,03	2042	1960	1329	431	3,08	1,83	65,08	713	34,92
<i>Streptococcus agalactiae</i> NEM316	2,21	2235	2094	1518	473	3,21	2,23	67,92	717	32,08
<i>Lactobacillus johnsonii</i> NCC 533	1,99	1918	1821	1354	400	3,39	2,55	70,59	564	29,41
<i>Bacillus anthracis</i> str <i>Sterne</i>	5,23	5417	5289	3053	1043	2,93	1,99	56,36	2364	43,64
<i>Symbiobacterium thermophilum</i> IAM 14863	3,57	3454	3337	2113	681	3,10	1,84	61,18	1341	38,82
<i>Bacillus licheniformis</i> ATCC 14580	4,22	4332	4173	2587	858	3,02	2,11	59,72	1745	40,28
<i>Streptococcus thermophilus</i> CNRZ1066	1,80	2000	1915	1316	450	2,92	1,87	65,80	684	34,20
<i>Bacillus clausii</i> KSM-K16	4,30	4204	4096	2575	821	3,14	1,76	61,25	1629	38,75
<i>Lactobacillus acidophilus</i> NCFM	1,99	1938	1864	1276	413	3,09	2,05	65,84	662	34,16
<i>Carboxydotherrmus hydrogenoformans</i> Z-2901	2,40	2707	2620	2038	534	3,82	3,14	75,29	669	24,71
<i>Lactobacillus sakei</i> subsp <i>sakei</i> 23K	1,88	1963	1879	1293	420	3,08	2,18	65,87	670	34,13
<i>Moorella thermoacetica</i> ATCC 39073	2,63	2613	2463	1684	542	3,11	2,25	64,45	929	35,55
<i>Desulfitobacterium hafniense</i> Y51	5,73	5137	5060	3010	1033	2,91	1,67	58,59	2127	41,41
<i>Streptococcus pyogenes</i> MGAS10750	1,94	2060	1979	1436	428	3,36	2,66	69,71	624	30,29

<i>Lactobacillus delbrueckii</i> subsp <i>bulgaricus</i> ATCC 11842	1,86	2184	1529	1066	339	3,14	2,46	48,81	1118	51,19
<i>Clostridium perfringens</i> ATCC 13124	3,26	3016	2876	1560	507	3,08	2,08	51,72	1456	48,28
<i>Pediococcus pentosaceus</i> ATCC 25745	1,83	1847	1755	1248	365	3,42	2,53	67,57	599	32,43
<i>Oenococcus oeni</i> PSU-1	1,78	1864	1691	1097	314	3,49	2,77	58,85	767	41,15
<i>Lactobacillus gasseri</i> ATCC 33323	1,89	1898	1755	1311	372	3,52	2,72	69,07	587	30,93
<i>Listeria welshimeri</i> serovar 6b str SLCC5334	2,81	2864	2774	1914	593	3,23	2,35	66,83	950	33,17
<i>Clostridium novyi</i> NT	2,55	2427	2315	1528	462	3,31	2,37	62,96	899	37,04
<i>Streptococcus sanguinis</i> SK36	2,39	2348	2270	1578	498	3,17	2,06	67,21	770	32,79
<i>Clostridium thermocellum</i> ATCC 27405	3,84	3363	3173	1840	632	2,91	1,78	54,71	1523	45,29
<i>Desulfotomaculum reducens</i> MI-1	3,61	3423	3276	2046	646	3,17	2,31	59,77	1377	40,23
<i>Streptococcus suis</i> 05ZYH33	2,10	2254	2186	1610	518	3,11	1,72	71,43	644	28,57
<i>Pelotomaculum thermopropionicum</i> SI	3,03	2977	2919	1822	635	2,87	1,36	61,20	1155	38,80
<i>Lactobacillus reuteri</i> DSM 20016	2,00	2027	1900	1318	400	3,30	2,79	65,02	709	34,98
<i>Clostridium beijerinckii</i> NCIMB 8052	6,00	5243	5020	2492	890	2,80	1,67	47,53	2751	52,47
<i>Alkaliphilus metalliredigens</i> QYMF	4,93	4940	4625	2925	899	3,25	2,38	59,21	2015	40,79
<i>Bacillus amyloliquefaciens</i> FZB42	3,92	3813	3693	2289	749	3,06	2,22	60,03	1524	39,97
<i>Bacillus pumilus</i> SAFR-032	3,70	3822	3679	2280	737	3,09	2,22	59,65	1542	40,35
<i>Alkaliphilus oremlandii</i> OhILAs	3,12	2991	2836	1862	603	3,09	2,30	62,25	1129	37,75
<i>Clostridium phytofermentans</i> ISDg	4,85	4023	3902	2015	741	2,72	1,40	50,09	2008	49,91
<i>Lactobacillus helveticus</i> DPC 4571	2,08	1838	1610	1006	306	3,29	2,43	54,73	832	45,27
<i>Thermoanaerobacter</i> <i>pseudethanolicus</i> ATCC 33223	2,36	2358	2239	1674	460	3,64	2,49	70,99	684	29,01
<i>Heliobacterium Modasticaldum</i> Ice 1	3,08	3137	2999	1806	621	2,91	1,83	57,57	1331	42,43
<i>Streptococcus pneumoniae</i> Hungary19A-6	2,25	2402	2155	1505	481	3,13	1,87	62,66	897	37,34
<i>Candidatus Desulforudis audaxviator</i> MP104C	2,35	2295	2157	1410	469	3,01	1,76	61,44	885	38,56
<i>Lactobacillus fermentum</i> IFO 3956	2,10	1912	1843	1218	400	3,05	1,94	63,70	694	36,30
<i>Lactobacillus casei</i> BL23	3,08	3090	3015	1913	643	2,98	1,71	61,91	1177	38,09
<i>Coprothermobacter proteolyticus</i> DSM 5265	1,42	1541	1482	1202	299	4,02	3,69	78,00	339	22,00
<i>Anoxybacillus flavithermus</i> WK1	2,85	2954	2831	2100	590	3,56	2,76	71,09	854	28,91
<i>Listeria monocytogenes</i> HCC23	2,98	3059	2974	2024	624	3,24	2,30	66,17	1035	33,83

<i>Bacillus cereus</i> B4264	5,42	5564	5398	2914	987	2,95	2,04	52,37	2650	47,63
<i>Clostridium cellulolyticum</i> H10	4,07	3569	3390	2197	700	3,14	2,21	61,56	1372	38,44
<i>Halothermothrix orenii</i> H 168	2,58	2436	2342	1597	481	3,32	2,20	65,56	839	34,44
<i>Streptococcus uberis</i> 0140J	1,85	1908	1762	1294	402	3,22	1,90	67,82	614	32,18
<i>Streptococcus equi</i> subsp <i>equi</i> 4047	2,25	2243	2001	1433	443	3,23	2,27	63,89	810	36,11
<i>Brevibacillus brevis</i> NBRC 100599	6,30	6121	5947	3633	1182	3,07	2,16	59,35	2488	40,65
<i>Clostridium botulinum</i> A2 str Kyoto	4,16	4034	3877	2234	730	3,06	2,22	55,38	1800	44,62
<i>Exiguobacterium</i> sp AT1b	3,00	3141	3020	2059	618	3,33	2,51	65,55	1082	34,45
<i>Eubacterium rectale</i> ATCC 33656	3,45	3698	3621	2545	782	3,25	1,92	68,82	1153	31,18
<i>Streptococcus dysgalactiae</i> subsp <i>equisimilis</i> GGS_124	2,11	2174	2094	1442	460	3,13	2,02	66,33	732	33,67
<i>Paenibacillus</i> sp JDR-2	7,18	6414	6213	3711	1266	2,93	1,62	57,86	2703	42,14
<i>Streptococcus suis</i> P1/7	2,01	2011	1824	1311	431	3,04	1,77	65,19	700	34,81
<i>Lactobacillus plantarum</i> JDM1	3,20	3029	2948	1741	588	2,96	1,92	57,48	1288	42,52
<i>Lactobacillus rhamnosus</i> GG	3,01	2985	2913	1789	631	2,84	1,55	59,93	1196	40,07
<i>Desulfotomaculum acetoxidans</i> DSM 771	4,55	4470	4068	2556	834	3,06	2,09	57,18	1914	42,82

Tabla S 3.5. Porcentajes de conservación por alas diferentes categorías de genes: persistentes, no-persistentes, todos los pares, alta expresión (HE), baja expresión (LE), niveles de proteínas balanceados (SIM), desbalanceados (DIF), genes esenciales (EE) y genes no esenciales (NN).

α-proteobacteria	Tamaño genoma (Mb)	>90%	<50%	Todos pares	HE	LE	SIM	DIF	EE	NN
<i>Rickettsia prowazekii</i> str, Madrid E	1,11	0,57	0,00	0,39	0,41	0,26	0,48	0,42	0,48	0,23
<i>Mesorhizobium loti</i> MAFF303099	7,04	0,58	0,33	0,34	0,37	0,32	0,42	0,35	0,44	0,32
<i>Caulobacter crescentus</i> CB15	4,02	0,58	0,29	0,36	0,40	0,30	0,42	0,36	0,40	0,39
<i>Wolbachia endosymbiont of Drosophila melanogaster</i>	1,27	0,56	0,20	0,33	0,39	0,04	0,42	0,39	0,43	0,24
<i>Sinorhizobium meliloti</i> 1021	6,69	0,57	0,33	0,37	0,40	0,29	0,45	0,36	0,44	0,37
<i>Agrobacterium tumefaciens</i> str, C58	5,46	0,46	0,33	0,34	0,36	0,28	0,43	0,31	0,39	0,34

<i>Rickettsia conorii</i> str, Malish 7	1,27	0,50	0,00	0,31	0,33	0,15	0,38	0,39	0,41	0,16
<i>Brucella melitensis</i> bv, 1 str, 16M	3,29	0,55	0,38	0,37	0,36	0,34	0,42	0,40	0,41	0,37
<i>Ruegeria pomeroyi</i> DSS-3	4,60	0,51	0,35	0,35	0,33	0,32	0,36	0,39	0,39	0,34
<i>Brucella suis</i> 1330	3,32	0,52	0,45	0,37	0,36	0,38	0,42	0,38	0,39	0,39
<i>Bradyrhizobium japonicum</i> USDA 110	9,11	0,55	0,35	0,32	0,33	0,25	0,33	0,33	0,35	0,32
<i>Anaplasma marginale</i> str, St, Maries	1,20	0,55	0,29	0,32	0,35	0,10	0,45	0,33	0,43	0,25
<i>Ehrlichia ruminantium</i> str, Welgevonden	1,52	0,55	0,40	0,37	0,43	0,00	0,54	0,45	0,51	0,28
<i>Rhodopseudomonas palustris</i> CGA009	5,46	0,54	0,29	0,33	0,35	0,22	0,38	0,34	0,40	0,31
<i>Bartonella quintana</i> str, Toulouse	1,58	0,61	0,56	0,49	0,47	0,44	0,54	0,47	0,50	0,53
<i>Bartonella henselae</i> str, Houston-1	1,93	0,61	0,52	0,49	0,47	0,45	0,55	0,46	0,47	0,55
<i>Rickettsia typhi</i> str, Wilmington	1,11	0,57	0,00	0,37	0,41	0,18	0,47	0,41	0,49	0,24
<i>Zymomonas mobilis</i> subsp, mobilis ZM4	2,06	0,33	0,47	0,30	0,29	0,27	0,35	0,21	0,26	0,32
<i>Gluconobacter oxydans</i> 621H	2,70	0,60	0,34	0,42	0,43	0,34	0,47	0,37	0,51	0,38
<i>Ehrlichia ruminantium</i> str, Gardel	1,50	0,55	0,33	0,39	0,44	0,00	0,52	0,42	0,51	0,30
<i>Ehrlichia ruminantium</i> str, Welgevonden	1,51	0,54	0,17	0,37	0,43	0,00	0,52	0,38	0,51	0,28
<i>Wolbachia endosymbiont strain TRS of Brugia malayi</i>	1,08	0,56	0,33	0,39	0,43	0,09	0,46	0,44	0,47	0,33
<i>Brucella abortus</i> bv, 1 str, 9-941	3,29	0,52	0,39	0,35	0,35	0,33	0,43	0,38	0,39	0,37
<i>Rickettsia felis</i> URRWXCa2	1,49	0,52	0,00	0,33	0,35	0,18	0,41	0,36	0,45	0,21
<i>Ehrlichia canis</i> str Jake	1,32	0,56	0,25	0,36	0,43	0,00	0,54	0,42	0,52	0,25
<i>Nitrobacter winogradskyi</i> Nb-255	3,40	0,54	0,23	0,34	0,35	0,28	0,34	0,40	0,43	0,31
<i>Rhodobacter sphaeroides</i> 2,4,1	4,13	0,52	0,51	0,43	0,44	0,41	0,44	0,41	0,38	0,46
<i>Brucella melitensis</i> biovar Abortus 2308	3,28	0,52	0,38	0,35	0,35	0,33	0,40	0,35	0,40	0,36
<i>Magnetospirillum magneticum</i> AMB-1	4,97	0,62	0,37	0,41	0,41	0,38	0,48	0,49	0,50	0,38
<i>Rhodospirillum rubrum</i> ATCC 11170	4,35	0,63	0,43	0,40	0,41	0,34	0,44	0,43	0,46	0,41
<i>Erythrobacter litoralis</i> HTCC2594	3,05	0,49	0,18	0,31	0,35	0,28	0,39	0,36	0,40	0,28
<i>Rhizobium etli</i> CFN 42	4,38	0,50	0,46	0,37	0,37	0,33	0,46	0,33	0,44	0,36
<i>Rhodopseudomonas palustris</i> HaA2	5,33	0,54	0,29	0,32	0,33	0,24	0,37	0,34	0,39	0,31

<i>Novosphingobium aromaticivorans</i> DSM 12444	4,05	0,53	0,29	0,35	0,41	0,31	0,45	0,38	0,47	0,32
<i>Anaplasma phagocytophilum</i> HZ	1,47	0,54	0,40	0,33	0,38	0,05	0,49	0,38	0,46	0,24
<i>Neorickettsia sennetsu</i> str, Miyayama	0,86	0,52	0,00	0,33	0,37	0,00	0,53	0,33	0,39	0,27
<i>Ehrlichia chaffeensis</i> str, Arkansas	1,18	0,54	0,33	0,36	0,42	0,00	0,52	0,42	0,49	0,26
<i>Jannaschia</i> sp, CCS1	4,32	0,52	0,28	0,33	0,35	0,26	0,37	0,34	0,40	0,31
<i>Rhodopseudomonas palustris</i> BisB18	5,51	0,52	0,40	0,37	0,33	0,31	0,35	0,34	0,38	0,37
<i>Rickettsia bellii</i> RML369-C	1,52	0,55	0,08	0,36	0,39	0,16	0,43	0,42	0,48	0,19
<i>Rhodopseudomonas palustris</i> BisB5	4,89	0,46	0,35	0,35	0,36	0,29	0,37	0,36	0,36	0,35
<i>Nitrobacter hamburgensis</i> X14	4,41	0,54	0,32	0,35	0,37	0,22	0,38	0,38	0,43	0,34
<i>Ruegeria</i> sp, TM1040	4,02	0,58	0,35	0,36	0,35	0,34	0,43	0,41	0,40	0,37
<i>Sphingopyxis alaskensis</i> RB2256	3,35	0,45	0,38	0,31	0,33	0,25	0,33	0,34	0,40	0,27
<i>Roseobacter denitrificans</i> OCh 114	4,13	0,50	0,30	0,36	0,39	0,31	0,35	0,39	0,43	0,33
<i>Granulibacter thesedensis</i> CGDNIH1	2,71	0,52	0,39	0,39	0,39	0,35	0,42	0,42	0,45	0,37
<i>Maricaulis maris</i> MCS10	3,37	0,68	0,28	0,43	0,49	0,33	0,52	0,51	0,53	0,41
<i>Hyphomonas neptunium</i> ATCC 15444	3,71	0,58	0,35	0,35	0,38	0,32	0,45	0,40	0,43	0,33
<i>Rhizobium leguminosarum</i> bv, viciae 3841	7,45	0,54	0,44	0,37	0,38	0,33	0,47	0,35	0,45	0,37
<i>Rhodopseudomonas palustris</i> BisA53	5,51	0,57	0,32	0,32	0,34	0,24	0,35	0,35	0,39	0,31
<i>Paracoccus denitrificans</i> PD1222	5,24	0,55	0,42	0,40	0,41	0,35	0,40	0,44	0,40	0,39
<i>Bartonella bacilliformis</i> KC583	1,45	0,64	0,65	0,50	0,47	0,48	0,56	0,44	0,48	0,56
<i>Rhodobacter sphaeroides</i> ATCC 17029	4,37	0,54	0,46	0,43	0,45	0,38	0,45	0,40	0,42	0,44
<i>Rhodobacter sphaeroides</i> ATCC 17025	4,10	0,56	0,38	0,41	0,44	0,36	0,42	0,42	0,41	0,40
<i>Bradyrhizobium</i> sp, ORS 278	7,46	0,53	0,32	0,34	0,37	0,26	0,36	0,37	0,40	0,32
<i>Acidiphilium cryptum</i> JF-5	3,39	0,64	0,55	0,49	0,52	0,35	0,54	0,43	0,51	0,50
<i>Bradyrhizobium</i> sp, BTai1	8,26	0,53	0,31	0,33	0,36	0,28	0,34	0,39	0,39	0,33
<i>Orientia tsutsugamushi</i> str, Boryong	2,13	0,63	0,50	0,52	0,52	0,39	0,53	0,48	0,50	0,58
<i>Brucella ovis</i> ATCC 25840	3,28	0,55	0,41	0,36	0,37	0,35	0,41	0,39	0,42	0,35
<i>Sphingomonas wittichii</i> RW1	5,38	0,43	0,38	0,35	0,33	0,25	0,34	0,23	0,39	0,33

<i>Sinorhizobium medicae</i> WSM419	6,60	0,52	0,35	0,36	0,38	0,30	0,44	0,32	0,42	0,37
<i>Ochrobactrum anthropi</i> ATCC 49188	4,78	0,50	0,40	0,38	0,37	0,38	0,43	0,39	0,39	0,39
<i>Parvibaculum lavamentivorans</i> DS-1	3,91	0,64	0,22	0,38	0,44	0,25	0,48	0,39	0,53	0,33
<i>Xanthobacter autotrophicus</i> Py2	5,31	0,56	0,37	0,34	0,35	0,27	0,38	0,35	0,45	0,33
<i>Rickettsia canadensis</i> str, McKiel	1,16	0,52	0,00	0,34	0,39	0,10	0,45	0,45	0,49	0,15
<i>Rickettsia akari</i> str, Hartford	1,23	0,52	0,00	0,32	0,36	0,11	0,41	0,39	0,44	0,20
<i>Rickettsia rickettsii</i> str, 'Sheila Smith'	1,26	0,52	0,20	0,36	0,37	0,16	0,41	0,40	0,45	0,20
<i>Rickettsia bellii</i> OSU 85-389	1,53	0,50	0,06	0,33	0,35	0,17	0,38	0,34	0,42	0,22
<i>Rickettsia massiliae</i> MTU5	1,36	0,50	0,00	0,34	0,37	0,17	0,43	0,42	0,42	0,23
<i>Azorhizobium caulinodans</i> ORS 571	5,37	0,62	0,42	0,39	0,39	0,30	0,41	0,37	0,50	0,35
<i>Dinoroseobacter shibae</i> DFL 12	3,79	0,55	0,29	0,37	0,37	0,38	0,36	0,43	0,41	0,34
<i>Brucella canis</i> ATCC 23365	3,31	0,54	0,45	0,39	0,37	0,40	0,43	0,39	0,41	0,40
<i>Gluconacetobacter diazotrophicus</i> PAI 5	3,94	0,64	0,30	0,38	0,46	0,33	0,52	0,47	0,47	0,37
<i>Bartonella tribocorum</i> CIP 105476	2,62	0,59	0,46	0,45	0,45	0,39	0,55	0,40	0,45	0,50
<i>Brucella suis</i> ATCC 23445	3,32	0,53	0,42	0,38	0,37	0,36	0,44	0,39	0,40	0,40
<i>Methylobacterium extorquens</i> PA1	5,47	0,50	0,30	0,29	0,30	0,24	0,33	0,31	0,37	0,25
<i>Rickettsia rickettsii</i> str, Iowa	1,27	0,52	0,20	0,35	0,37	0,13	0,42	0,40	0,45	0,18
<i>Caulobacter</i> sp, K31	5,48	0,59	0,31	0,34	0,36	0,26	0,43	0,38	0,39	0,34
<i>Methylobacterium radiotolerans</i> JCM 2831	6,08	0,46	0,32	0,30	0,30	0,23	0,30	0,28	0,33	0,30
<i>Methylobacterium</i> sp, 4-46	7,66	0,55	0,27	0,27	0,30	0,19	0,34	0,28	0,36	0,26
<i>Beijerinckia indica</i> subsp, indica ATCC 9039	4,17	0,41	0,32	0,29	0,30	0,28	0,31	0,31	0,35	0,27
β-proteobacteria	Tamaño genoma Mb	>90%	<50%	Todos pares	HE	LE	SIM	DIF	EE	NN
<i>Bordetella bronchiseptica</i> RB50	5,34	0,66	0,58	0,62	0,62	0,57	0,65	0,59	0,68	0,57
<i>Bordetella parapertussis</i> 12822	4,77	0,68	0,52	0,61	0,61	0,51	0,64	0,61	0,68	0,54

<i>Bordetella pertussis</i> Tohama I	4,09	0,66	0,50	0,62	0,62	0,54	0,67	0,60	0,67	0,55
<i>Neisseria gonorrhoeae</i> FA 1090	2,15	0,45	0,50	0,38	0,38	0,25	0,50	0,38	0,47	0,27
<i>Neisseria meningitidis</i> MC58	2,27	0,45	0,60	0,41	0,40	0,28	0,50	0,39	0,47	0,30
<i>Neisseria meningitidis</i> Z2491	2,18	0,46	0,43	0,41	0,40	0,28	0,49	0,38	0,48	0,31
<i>Nitrosomonas europaea</i> ATCC 19718	2,81	0,72	0,70	0,63	0,68	0,54	0,70	0,68	0,74	0,57
<i>Chromobacterium violaceum</i> ATCC 12472	4,75	0,68	0,77	0,64	0,64	0,60	0,70	0,62	0,74	0,59
<i>Burkholderia mallei</i> ATCC 23344	5,84	0,57	0,63	0,56	0,55	0,53	0,56	0,50	0,62	0,51
<i>Burkholderia pseudomallei</i> K96243	7,25	0,61	0,67	0,59	0,56	0,58	0,55	0,51	0,65	0,55
<i>Dechloromonas aromatica</i> RCB	4,50	0,75	0,64	0,65	0,67	0,56	0,75	0,65	0,70	0,60
<i>Thiobacillus denitrificans</i> ATCC 25259	2,91	0,72	0,82	0,68	0,64	0,64	0,69	0,69	0,69	0,64
<i>Burkholderia pseudomallei</i> 1710b	7,31	0,57	0,63	0,56	0,56	0,50	0,51	0,48	0,63	0,50
<i>Burkholderia</i> sp, 383	8,68	0,59	0,67	0,60	0,59	0,56	0,60	0,54	0,69	0,54
<i>Burkholderia thailandensis</i> E264	3,19	0,59	0,82	0,59	0,54	0,61	0,55	0,52	0,66	0,55
<i>Methylobacillus flagellatus</i> KT	2,97	0,72	0,53	0,66	0,62	0,67	0,65	0,62	0,75	0,56
<i>Burkholderia xenovorans</i> LB400	9,73	0,51	0,55	0,51	0,52	0,46	0,50	0,44	0,59	0,46
<i>Burkholderia cenocepacia</i> AU 1054	7,28	0,63	0,70	0,62	0,62	0,58	0,64	0,55	0,69	0,59
<i>Azoarcus</i> sp, BH72	4,38	0,79	0,71	0,69	0,69	0,61	0,72	0,71	0,75	0,65
<i>Neisseria meningitidis</i> FAM18	2,19	0,47	0,50	0,43	0,42	0,34	0,52	0,43	0,47	0,34
<i>Burkholderia mallei</i> NCTC 10229	5,74	0,60	0,67	0,59	0,58	0,59	0,57	0,52	0,64	0,55
<i>Burkholderia pseudomallei</i> 668	7,04	0,59	0,70	0,59	0,55	0,60	0,54	0,49	0,63	0,56
<i>Burkholderia pseudomallei</i> 1106a	7,09	0,62	0,70	0,61	0,57	0,62	0,55	0,53	0,65	0,58
<i>Burkholderia mallei</i> NCTC 10247	3,50	0,60	0,68	0,57	0,55	0,57	0,56	0,51	0,61	0,53
<i>Hermiimonas arsenicoxydans</i>	3,42	0,60	0,73	0,58	0,56	0,52	0,58	0,57	0,66	0,53
<i>Polynucleobacter necessarius</i> subsp, <i>asymbioticus</i> QLW-P1DMWA-1	2,16	0,76	0,85	0,70	0,66	0,78	0,74	0,70	0,74	0,67
<i>Janthinobacterium</i> sp, Marseille	4,11	0,63	0,75	0,64	0,62	0,65	0,63	0,62	0,67	0,63
<i>Delftia acidovorans</i> SPH-1	6,77	0,59	0,50	0,52	0,54	0,39	0,62	0,53	0,57	0,48

<i>Neisseria meningitidis</i> 053442	2,15	0,43	0,40	0,37	0,37	0,23	0,47	0,35	0,45	0,26
<i>Bordetella petrii</i> DSM 12804	5,29	0,72	0,63	0,66	0,67	0,57	0,71	0,70	0,76	0,61
<i>Burkholderia cenocepacia</i> MC0-3	7,97	0,62	0,70	0,62	0,60	0,60	0,63	0,57	0,71	0,57
<i>Leptothrix cholodnii</i> SP-6	4,91	0,67	0,66	0,61	0,61	0,57	0,69	0,56	0,65	0,57
<i>Polynucleobacter necessarius</i> subsp. <i>necessarius</i> STIR1	1,56	0,77	1,00	0,70	0,67	0,75	0,76	0,69	0,75	0,61
<i>Bordetella avium</i> 197N	3,73	0,66	0,46	0,61	0,60	0,51	0,64	0,60	0,65	0,58
<i>Laribacter hongkongensis</i> HLHK9	3,17	0,71	0,84	0,64	0,61	0,58	0,67	0,63	0,69	0,60
<i>Variovorax paradoxus</i> S110	6,75	0,63	0,65	0,58	0,61	0,45	0,61	0,51	0,65	0,55
<i>Methylobacterium mobilis</i> JLW8	2,55	0,73	0,50	0,64	0,62	0,65	0,66	0,63	0,75	0,54
Firmicutes	Tamaño genoma Mb	>90%	<50%	Todos pares	HE	LE	SIM	DIF	EE	NN
<i>Bacillus subtilis</i> subsp <i>subtilis</i> str 168	4,22	0,61	0,45	0,44	0,50	0,37	0,58	0,47	0,53	0,46
<i>Bacillus halodurans</i> C-125	4,20	0,60	0,37	0,42	0,46	0,33	0,51	0,40	0,45	0,43
<i>Lactococcus lactis</i> subsp <i>lactis</i> II1403	2,37	0,32	0,32	0,33	0,40	0,17	0,44	0,31	0,34	0,40
<i>Staphylococcus aureus</i> subsp <i>aureus</i> MRSA252	2,90	0,60	0,40	0,42	0,49	0,29	0,62	0,44	0,49	0,41
<i>Thermoanaerobacter tengcongensis</i> MB4	2,69	0,71	0,52	0,54	0,57	0,44	0,67	0,56	0,56	0,58
<i>Bacillus anthracis</i> str Ames	5,23	0,55	0,32	0,37	0,40	0,33	0,46	0,40	0,48	0,38
<i>Oceanobacillus ihyensensis</i> HTE831	3,63	0,61	0,43	0,45	0,48	0,39	0,53	0,47	0,49	0,49
<i>Streptococcus mutans</i> UA159	2,03	0,28	0,41	0,36	0,38	0,21	0,44	0,35	0,33	0,41
<i>Streptococcus agalactiae</i> NEM316	2,21	0,47	0,29	0,41	0,46	0,26	0,56	0,41	0,47	0,39
<i>Lactobacillus johnsonii</i> NCC 533	1,99	0,68	0,28	0,49	0,56	0,26	0,63	0,62	0,61	0,44
<i>Bacillus anthracis</i> str Sterne	5,23	0,56	0,37	0,39	0,42	0,34	0,47	0,42	0,48	0,41
<i>Symbiobacterium thermophilum</i> IAM 14863	3,57	0,63	0,41	0,43	0,46	0,37	0,56	0,42	0,49	0,44
<i>Bacillus licheniformis</i> ATCC 14580	4,22	0,60	0,44	0,44	0,48	0,40	0,50	0,42	0,49	0,47

<i>Streptococcus thermophilus</i> CNRZ1066	1,80	0,42	0,19	0,36	0,40	0,21	0,46	0,41	0,41	0,38
<i>Bacillus clausii</i> KSM-K16	4,30	0,60	0,39	0,42	0,47	0,32	0,52	0,49	0,45	0,44
<i>Lactobacillus acidophilus</i> NCFM	1,99	0,64	0,38	0,51	0,59	0,29	0,63	0,62	0,63	0,50
<i>Carboxydotherrnus hydrogenoformans</i> Z-2901	2,40	0,71	0,45	0,51	0,55	0,37	0,67	0,46	0,56	0,52
<i>Lactobacillus sakei</i> subsp <i>sakei</i> 23K	1,88	0,59	0,44	0,45	0,49	0,23	0,55	0,44	0,48	0,48
<i>Moorella thermoacetica</i> ATCC 39073	2,63	0,74	0,41	0,47	0,51	0,38	0,64	0,50	0,61	0,45
<i>Desulfitobacterium hafniense</i> Y51	5,73	0,63	0,36	0,40	0,46	0,30	0,54	0,42	0,49	0,37
<i>Streptococcus pyogenes</i> MGAS10750	1,94	0,39	0,22	0,35	0,40	0,27	0,44	0,37	0,43	0,34
<i>Lactobacillus delbrueckii</i> subsp <i>bulgaricus</i> ATCC 11842	1,86	0,63	0,22	0,54	0,60	0,32	0,66	0,65	0,62	0,54
<i>Clostridium perfringens</i> ATCC 13124	3,26	0,70	0,40	0,41	0,46	0,33	0,58	0,44	0,56	0,39
<i>Pediococcus pentosaceus</i> ATCC 25745	1,83	0,63	0,29	0,49	0,57	0,25	0,64	0,57	0,55	0,51
<i>Oenococcus oeni</i> PSU-1	1,78	0,69	0,36	0,51	0,56	0,29	0,69	0,52	0,58	0,51
<i>Lactobacillus gasseri</i> ATCC 33323	1,89	0,66	0,18	0,48	0,58	0,26	0,65	0,58	0,63	0,40
<i>Listeria welshimeri</i> serovar 6b str SLCC5334	2,81	0,52	0,49	0,45	0,46	0,35	0,53	0,36	0,45	0,49
<i>Clostridium novyi</i> NT	2,55	0,66	0,43	0,48	0,54	0,32	0,63	0,53	0,53	0,48
<i>Streptococcus sanguinis</i> SK36	2,39	0,43	0,30	0,37	0,42	0,28	0,57	0,28	0,44	0,40
<i>Clostridium thermocellum</i> ATCC 27405	3,84	0,60	0,35	0,42	0,44	0,35	0,51	0,41	0,45	0,40
<i>Desulfotomaculum reducens</i> MI-1	3,61	0,69	0,42	0,49	0,51	0,41	0,61	0,53	0,56	0,48
<i>Streptococcus suis</i> 05ZYH33	2,10	0,35	0,25	0,33	0,40	0,25	0,48	0,27	0,31	0,36
<i>Pelotomaculum thermopropionicum</i> SI	3,03	0,48	0,27	0,36	0,36	0,27	0,39	0,32	0,45	0,32
<i>Lactobacillus reuteri</i> DSM 20016	2,00	0,61	0,41	0,47	0,53	0,30	0,64	0,51	0,54	0,47
<i>Clostridium beijerinckii</i> NCIMB 8052	6,00	0,64	0,36	0,35	0,40	0,29	0,47	0,37	0,45	0,32
<i>Alkaliphilus metalliredigens</i> QYMF	4,93	0,69	0,34	0,39	0,46	0,28	0,54	0,44	0,55	0,36
<i>Bacillus amyloliquefaciens</i> FZB42	3,92	0,61	0,43	0,41	0,47	0,34	0,50	0,46	0,51	0,42
<i>Bacillus pumilus</i> SAFR-032	3,70	0,61	0,43	0,45	0,48	0,36	0,54	0,48	0,50	0,47
<i>Alkaliphilus oremlandii</i> OhILAs	3,12	0,75	0,43	0,48	0,56	0,35	0,62	0,53	0,61	0,43
<i>Clostridium phytofermentans</i> ISDg	4,85	0,50	0,24	0,29	0,34	0,19	0,42	0,29	0,38	0,27

<i>Lactobacillus helveticus</i> DPC 4571	2,08	0,65	0,38	0,51	0,57	0,33	0,66	0,60	0,60	0,49
<i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223	2,36	0,71	0,35	0,51	0,53	0,42	0,64	0,50	0,58	0,52
<i>Heliobacterium modesticaldum</i> Ice1	3,08	0,58	0,32	0,41	0,46	0,32	0,56	0,46	0,47	0,39
<i>Streptococcus pneumoniae</i> Hungary19A-6	2,25	0,42	0,04	0,37	0,43	0,19	0,52	0,36	0,41	0,38
<i>Candidatus Desulforudis audaxviator</i> MP104C	2,35	0,67	0,30	0,43	0,46	0,36	0,60	0,44	0,52	0,38
<i>Lactobacillus fermentum</i> IFO 3956	2,10	0,62	0,38	0,48	0,52	0,31	0,64	0,46	0,49	0,52
<i>Lactobacillus casei</i> BL23	3,08	0,52	0,45	0,42	0,47	0,28	0,55	0,38	0,43	0,45
<i>Coprothermobacter proteolyticus</i> DSM 5265	1,42	0,70	0,42	0,55	0,57	0,59	0,61	0,60	0,63	0,58
<i>Anoxybacillus flavithermus</i> WK1	2,85	0,65	0,39	0,48	0,52	0,43	0,57	0,53	0,60	0,49
<i>Listeria monocytogenes</i> HCC23	2,98	0,53	0,49	0,44	0,47	0,34	0,53	0,37	0,45	0,49
<i>Bacillus cereus</i> B4264	5,42	0,55	0,42	0,40	0,41	0,39	0,46	0,40	0,46	0,43
<i>Clostridium cellulolyticum</i> H10	4,07	0,62	0,32	0,41	0,46	0,29	0,55	0,41	0,49	0,38
<i>Halothermothrix orenii</i> H 168	2,58	0,68	0,43	0,48	0,47	0,41	0,56	0,42	0,50	0,47
<i>Streptococcus uberis</i> 0140J	1,85	0,45	0,48	0,39	0,43	0,25	0,56	0,31	0,44	0,37
<i>Streptococcus equi</i> subsp <i>equi</i> 4047	2,25	0,39	0,40	0,37	0,40	0,19	0,45	0,33	0,41	0,35
<i>Brevibacillus brevis</i> NBRC 100599	6,30	0,61	0,37	0,40	0,43	0,36	0,46	0,36	0,49	0,39
<i>Clostridium botulinum</i> A2 str Kyoto	4,16	0,63	0,30	0,39	0,44	0,27	0,52	0,41	0,48	0,39
<i>Exiguobacterium</i> sp AT1b	3,00	0,69	0,42	0,49	0,54	0,43	0,63	0,53	0,60	0,51
<i>Eubacterium rectale</i> ATCC 33656	3,45	0,50	0,24	0,35	0,38	0,29	0,47	0,33	0,43	0,34
<i>Streptococcus dysgalactiae</i> subsp <i>equisimilis</i> GGS_124	2,11	0,39	0,24	0,35	0,40	0,15	0,47	0,33	0,38	0,37
<i>Paenibacillus</i> sp JDR-2	7,18	0,60	0,36	0,40	0,45	0,31	0,49	0,38	0,48	0,41
<i>Streptococcus suis</i> P1/7	2,01	0,39	0,27	0,38	0,42	0,28	0,49	0,31	0,37	0,41
<i>Lactobacillus plantarum</i> JDM1	3,20	0,58	0,43	0,45	0,53	0,23	0,63	0,50	0,50	0,48
<i>Lactobacillus rhamnosus</i> GG	3,01	0,55	0,49	0,45	0,50	0,33	0,61	0,43	0,50	0,49
<i>Desulfotomaculum acetoxidans</i> DSM 771	4,55	0,65	0,31	0,44	0,50	0,34	0,55	0,48	0,52	0,46

Publicaciones durante el período de la tesis de Doctorado.

1. **Nuñez PA**, Soria M, Farber M (2012). The Twin-arginine Translocation Pathway in α -proteobacteria is Functionally Preserved Irrespective of Genomic and Regulatory Divergence. PLoS One.
2. Tomassone L, **Nuñez PA**, Ceballos LA, Gurtler RE, Kitron U, et al. (2010) Detection of "Candidatus Rickettsia sp. strain Argentina" and Rickettsia bellii in Amblyomma ticks (Acari: Ixodidae) from Northern Argentina. Exp Appl Acarol 52: 93-100.
3. Tomassone L, **Nuñez PA**, Gurtler RE, Ceballos LA, Orozco MM, et al. (2008) Molecular detection of Ehrlichia chaffeensis in Amblyomma parvum ticks, Argentina. Emerg Infect Dis 14: 1953-1955.
4. Silberstein C, Pistone Creydt V, Gerhardt E, **Nuñez PA**, Ibarra C (2008) Inhibition of water absorption in human proximal tubular epithelial cells in response to Shiga toxin-2. Pediatr Nephrol 23: 1981-1990.
5. Moretta R, Ruybal P, Mesplet M, Petrigh R, **Nuñez PA**, et al. (2008) Flow cytometry to evaluate Anaplasma marginale parasitemia using a fluorescent nucleic acid stain. Ann N Y Acad Sci 1149: 111-113.

Bibliografía.

1. Batut, J., S.G. Andersson, y D. O'Callaghan, *The evolution of chronic infection strategies in the alpha-proteobacteria*. Nat Rev Microbiol, 2004. **2**(12): p. 933-45.
2. McCutcheon, J.P., B.R. McDonald, y N.A. Moran, *Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont*. PLoS Genet, 2009. **5**(7): p. e1000565.
3. Dumler, J.S., A.F. Barbet, C.P. Bekker, G.A. Dasch, G.H. Palmer, et al., *Reorganization of genera in the families Rickettsiaceae and Anaplasmataceae in the order Rickettsiales: unification of some species of Ehrlichia with Anaplasma, Cowdria with Ehrlichia and Ehrlichia with Neorickettsia, descriptions of six new species combinations and designation of Ehrlichia equi and 'HGE agent' as subjective synonyms of Ehrlichia phagocytophila*. Int J Syst Evol Microbiol, 2001. **51**(Pt 6): p. 2145-65.
4. Chen, S.M., J.S. Dumler, J.S. Bakken, y D.H. Walker, *Identification of a granulocytotropic Ehrlichia species as the etiologic agent of human disease*. J Clin Microbiol, 1994. **32**(3): p. 589-95.
5. Goodman, J.L., C. Nelson, B. Vitale, J.E. Madigan, J.S. Dumler, et al., *Direct cultivation of the causative agent of human granulocytic ehrlichiosis*. N Engl J Med, 1996. **334**(4): p. 209-15.
6. Dumler, J.S., J.E. Madigan, N. Pusterla, y J.S. Bakken, *Ehrlichioses in humans: epidemiology, clinical presentation, diagnosis, and treatment*. Clin Infect Dis, 2007. **45 Suppl 1**: p. S45-51.
7. Raoult, D. y V. Roux, *Rickettsioses as paradigms of new or emerging infectious diseases*. Clin Microbiol Rev, 1997. **10**: p. 694-719.
8. Fuxelius, H.H., A. Darby, C.K. Min, N.H. Cho, y S.G. Andersson, *The genomic and metabolic diversity of Rickettsia*. Res Microbiol, 2007. **158**(10): p. 745-53.
9. Stothard, D.R., J.B. Clark, y P.A. Fuerst, *Ancestral divergence of Rickettsia bellii from the spotted fever and typhus groups of Rickettsia and antiquity of the genus Rickettsia*. Int J Syst Bacteriol, 1994. **44**(4): p. 798-804.
10. Vitorino, L., I.M. Chelo, F. Bacellar, y L. Ze-Ze, *Rickettsiae phylogeny: a multigenic approach*. Microbiology, 2007. **153**(Pt 1): p. 160-8.
11. Parola, P., C.D. Paddock, y D. Raoult, *Tick-borne rickettsioses around the world: emerging diseases challenging old concepts*. Clin Microbiol Rev, 2005. **18**(4): p. 719-56.
12. Labruna, M.B., J.W. McBride, D.H. Bouyer, L.M. Camargo, E.P. Camargo, et al., *Molecular evidence for a spotted fever group Rickettsia species in the tick Amblyomma longirostre in Brazil*. J Med Entomol, 2004. **41**(3): p. 533-7.
13. Williams, K.P., B.W. Sobral, y A.W. Dickerman, *A robust species tree for the alphaproteobacteria*. J Bacteriol, 2007. **189**(13): p. 4578-86.
14. Boussau, B., E.O. Karlberg, A.C. Frank, B.A. Legault, y S.G. Andersson, *Computational inference of scenarios for alpha-proteobacterial genome evolution*. Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9722-7.
15. Theiler, A., *Gall sickness of South Africa (anaplasmosis of cattle)*. J. Comp. Pathol. Ther, 1910a. **23**: p. 98-115.
16. Theiler, A., *Anaplasma marginale (gen. and spec. nov). The marginal points in the blood of cattle suffering from a specific disease*. Report of the Government on Veterinary Bacteriology in Transvaal. Department of Agriculture, South Africa, 1908-1909, 1910b: p. 7-64.
17. Theiler, A., *Further investigations into anaplasmosis of South African cattle*. Report of the Director of Veterinary Research. Department of Agriculture of the Union of South Africa, 1911: p. 7-46.

18. Ajayi, S., *Experimental bovine anaplasmosis: clinico-pathological and nutritional studies*. Res. Vet. Sci, 1978. **25**: p. 76-81.
19. Aubry, P. y D.W. Geale, *A review of bovine anaplasmosis*. Transbound Emerg Dis, 2011. **58**(1): p. 1-30.
20. Barbet, A.F., J. Yi, A. Lundgren, B.R. McEwen, E.F. Blouin, et al., *Antigenic variation of Anaplasma marginale: major surface protein 2 diversity during cyclic transmission between ticks and cattle*. Infect Immun, 2001. **69**(5): p. 3057-66.
21. Kieser, S.T., I.S. Eriks, y G.H. Palmer, *Cyclic rickettsemia during persistent Anaplasma marginale infection of cattle*. Infect Immun, 1990. **58**(4): p. 1117-9.
22. Aguirre D.H., G.A.B., Vinñabal A.E., Torioni de Echaide S., Guglielmone A.A., *Transmission of Anaplasma marginale with adult Boophilus microplus ticks fed as nymphs on calves with different levels of rickettsaemia*. Parasite, 1994. **1**: p. 405-407.
23. De La Fuente, J., V. Naranjo, F. Ruiz-Fons, U. Hofle, I.G. Fernandez De Mera, et al., *Potential vertebrate reservoir hosts and invertebrate vectors of Anaplasma marginale and A. phagocytophilum in central Spain*. Vector Borne Zoonotic Dis, 2005. **5**(4): p. 390-401.
24. Hawkins, J.A., J.N. Love, y R.J. Hidalgo, *Mechanical transmission of anaplasmosis by tabanids (Diptera: Tabanidae)*. Am J Vet Res, 1982. **43**(4): p. 732-4.
25. Zaugg, J.L., D. Stiller, M.E. Coan, y S.D. Lincoln, *Transmission of Anaplasma marginale Theiler by males of Dermacentor andersoni Stiles fed on an Idaho field-infected, chronic carrier cow*. Am J Vet Res, 1986. **47**(10): p. 2269-71.
26. Kocan, K.M., J. de la Fuente, A.A. Guglielmone, y R.D. Melendez, *Antigens and alternatives for control of Anaplasma marginale infection in cattle*. Clin Microbiol Rev, 2003. **16**(4): p. 698-712.
27. Johnson, E.M., S.A. Ewing, R.W. Barker, J.C. Fox, D.W. Crow, et al., *Experimental transmission of Ehrlichia canis (Rickettsiales: Ehrlichieae) by Dermacentor variabilis (Acari: Ixodidae)*. Vet Parasitol, 1998. **74**(2-4): p. 277-88.
28. Abdala A.A., M.A.J., Torioni S, *Control y erradicación de la anaplasmosis en un rodeo lechero*. Vet. Arg, 1998. **15**(342-346).
29. Guglielmone, A., A. Abdala, O. Anziani, A. Mangold, M. Volpogni, et al., *Different seasonal occurrence of anaplasmosis outbreaks in beef and dairy cattle of an area of Argentina free of Rhipicephalus microplus*. . Vet Parasitol, 1997. **19**(32-33).
30. Rossanigo, C., R. Sager, G. Ferrero, y J. Toselli, *Anaplasmosis en sistemas intensivos de áreas no endémicas*. Instituto Nacional de Tecnología Agropecuaria, Estación Experimental Regional Agropecuaria San Luis. 24° Congreso Argentino de Producción Animal, 2001. **1**(1): p. 245-246.
31. Guglielmone, A.A., *Epidemiology of babesiosis and anaplasmosis in South and Central America*. Vet Parasitol, 1995. **57**(1-3): p. 109-19.
32. Jonsson, N.N. y S.W. Reid, *Global climate change and vector borne diseases*. Vet J, 2000. **160**(2): p. 87-9.
33. Dumler, J.S. y J.S. Bakken, *Human ehrlichioses: newly recognized infections transmitted by ticks*. Annu Rev Med, 1998. **49**: p. 201-13.
34. da Costa PS, V.L., Brigatte ME, Greco DB., *More about human monocytotropic ehrlichiosis in Brazil: serological evidence of nine new cases*. J Infect Dis, 2006. **10**: p. 7-10.
35. Martinez, M.C., C.N. Gutierrez, F. Monger, J. Ruiz, A. Watts, et al., *Ehrlichia chaffeensis in child, Venezuela*. Emerg Infect Dis, 2008. **14**(3): p. 519-20.
36. Ripoll, C.M., C.E. Remondegui, G. Ordonez, R. Arazamendi, H. Fusaro, et al., *Evidence of rickettsial spotted fever and ehrlichial infections in a subtropical territory of Jujuy, Argentina*. Am J Trop Med Hyg, 1999. **61**(2): p. 350-4.

37. Machado, R.Z., J.M. Duarte, A.S. Dagnone, y M.P. Szabo, *Detection of Ehrlichia chaffeensis in Brazilian marsh deer (Blastocerus dichotomus)*. Vet Parasitol, 2006. **139**(1-3): p. 262-6.
38. Tomassone, L., P. Nunez, R.E. Gurtler, L.A. Ceballos, M.M. Orozco, et al., *Molecular detection of Ehrlichia chaffeensis in Amblyomma parvum ticks, Argentina*. Emerg Infect Dis, 2008. **14**(12): p. 1953-5.
39. Horta, M.C., M.B. Labruna, L.A. Sangioni, M.C. Vianna, S.M. Gennari, et al., *Prevalence of antibodies to spotted fever group rickettsiae in humans and domestic animals in a Brazilian spotted fever-endemic area in the state of Sao Paulo, Brazil: serologic evidence for infection by Rickettsia rickettsii and another spotted fever group Rickettsia*. Am J Trop Med Hyg, 2004. **71**(1): p. 93-7.
40. Labruna, M.B., T. Whitworth, D.H. Bouyer, J. McBride, L.M. Camargo, et al., *Rickettsia bellii and Rickettsia amblyommii in Amblyomma ticks from the State of Rondonia, Western Amazon, Brazil*. J Med Entomol, 2004. **41**(6): p. 1073-81.
41. Pinter, A., R.A. Dias, S.M. Gennari, y M.B. Labruna, *Study of the seasonal dynamics, life cycle, and host specificity of Amblyomma aureolatum (Acari: Ixodidae)*. J Med Entomol, 2004. **41**(3): p. 324-32.
42. Venzal, J.M., A. Portillo, A. Estrada-Pena, O. Castro, P.A. Cabrera, et al., *Rickettsia parkeri in Amblyomma triste from Uruguay*. Emerg Infect Dis, 2004. **10**(8): p. 1493-5.
43. Medina-Sanchez, A., D.H. Bouyer, V. Alcantara-Rodriguez, C. Mafra, J. Zavala-Castro, et al., *Detection of a typhus group Rickettsia in Amblyomma ticks in the state of Nuevo Leon, Mexico*. Ann N Y Acad Sci, 2005. **1063**: p. 327-32.
44. Estrada-Pena, A., J.M. Venzal, A.J. Mangold, M.M. Cafrune, y A.A. Guglielmo, *The Amblyomma maculatum Koch, 1844 (Acari: Ixodidae: Amblyomminae) tick group: diagnostic characters, description of the larva of A. parvitarsum Neumann, 1901, 16S rDNA sequences, distribution and hosts*. Syst Parasitol, 2005. **60**(2): p. 99-112.
45. Labruna, M.B., R.C. Pacheco, S. Nava, P.E. Brandao, L.J. Richtzenhain, et al., *Infection by Rickettsia bellii and Candidatus "Rickettsia amblyommii" in Amblyomma neumanni ticks from Argentina*. Microb Ecol, 2007. **54**(1): p. 126-33.
46. Pacheco, R.C., J. Moraes-Filho, S. Nava, P.E. Brandao, L.J. Richtzenhain, et al., *Detection of a novel spotted fever group rickettsia in Amblyomma parvum ticks (Acari: Ixodidae) from Argentina*. Exp Appl Acarol, 2007. **43**(1): p. 63-71.
47. Tomassone, L., P. Nunez, L.A. Ceballos, R.E. Gurtler, U. Kitron, et al., *Detection of "Candidatus Rickettsia sp. strain Argentina" and Rickettsia bellii in Amblyomma ticks (Acari: Ixodidae) from Northern Argentina*. Exp Appl Acarol. **52**(1): p. 93-100.
48. Oliveira, R.P., M.A. Galvao, C.L. Mafra, C.B. Chamone, S.B. Calic, et al., *Rickettsia felis in Ctenocephalides spp. fleas, Brazil*. Emerg Infect Dis, 2002. **8**(3): p. 317-9.
49. Blair, P.J., J. Jiang, G.B. Schoeler, C. Moron, E. Anaya, et al., *Characterization of spotted fever group rickettsiae in flea and tick specimens from northern Peru*. J Clin Microbiol, 2004. **42**(11): p. 4961-7.
50. Cardoso, L.D., R.N. Freitas, C.L. Mafra, C.V. Neves, F.C. Figueira, et al., *[Characterization of Rickettsia spp. circulating in a silent peri-urban focus for Brazilian spotted fever in Caratinga, Minas Gerais, Brazil]*. Cad Saude Publica, 2006. **22**(3): p. 495-501.
51. Cicuttin, G., *Primera detección de Rickettsia massiliae en la ciudad de Buenos Aires. Resultados preliminares*. Rev Argentina Zoonosis, 2004. **1**: p. 8-10.
52. Herndon, D.R., G.H. Palmer, V. Shkap, D.P. Knowles, Jr., y K.A. Brayton, *Complete genome sequence of Anaplasma marginale subsp. centrale*. J Bacteriol, 2010. **192**(1): p. 379-80.

53. Kuttler, K.L. y R.A. Todorovic, *Techniques of premunization for the control of Anaplasmosis*. Proceedings of the 6th National Anaplasmosis Conference, Las Vegas, NV, 1973: p. 106–112.
54. Abdala, A.A., E. Pipano, D.H. Aguirre, A.B. Gaido, M.A. Zurbriggen, et al., *Frozen and fresh Anaplasma centrale vaccines in the protection of cattle against Anaplasma marginale infection*. Rev Elev Med Vet Pays Trop, 1990. **43**(2): p. 155-8.
55. Brown, W.C., *Unraveling the immune regulatory mechanisms imposed by Anaplasma*. Vet J, 2008. **175**(1): p. 10-1.
56. Han, S., J. Norimine, K.A. Brayton, G.H. Palmer, G.A. Scoles, et al., *Anaplasma marginale infection with persistent high-load bacteremia induces a dysfunctional memory CD4+ T lymphocyte response but sustained high IgG titers*. Clin Vaccine Immunol, 2010. **17**(12): p. 1881-90.
57. Meeus, P.F. y A.F. Barbet, *Ingenious gene generation*. Trends Microbiol, 2001. **9**(8): p. 353-5; discussion 355-6.
58. French, D.M., W.C. Brown, y G.H. Palmer, *Emergence of Anaplasma marginale antigenic variants during persistent rickettsemia*. Infect Immun, 1999. **67**(11): p. 5834-40.
59. French, D.M., T.F. McElwain, T.C. McGuire, y G.H. Palmer, *Expression of Anaplasma marginale major surface protein 2 variants during persistent cyclic rickettsemia*. Infect Immun, 1998. **66**(3): p. 1200-7.
60. Abbott, J.R., G.H. Palmer, K.A. Kegerreis, P.F. Hetrick, C.J. Howard, et al., *Rapid and long-term disappearance of CD4+ T lymphocyte responses specific for Anaplasma marginale major surface protein-2 (MSP2) in MSP2 vaccinates following challenge with live A. marginale*. J Immunol, 2005. **174**(11): p. 6702-15.
61. Mwangi, W., W.C. Brown, G.A. Splitter, C.J. Davies, C.J. Howard, et al., *DNA vaccine construct incorporating intercellular trafficking and intracellular targeting motifs effectively primes and induces memory B- and T-cell responses in outbred animals*. Clin Vaccine Immunol, 2007. **14**(3): p. 304-11.
62. Palmer, G.H. y T.C. McGuire, *Immune serum against Anaplasma marginale initial bodies neutralizes infectivity for cattle*. J Immunol, 1984. **133**(2): p. 1010-5.
63. Visser, E.S., T.C. McGuire, G.H. Palmer, W.C. Davis, V. Shkap, et al., *The Anaplasma marginale msp5 gene encodes a 19-kilodalton protein conserved in all recognized Anaplasma species*. Infect Immun, 1992. **60**(12): p. 5139-44.
64. Noh, S.M., Y. Zhuang, J.E. Futse, W.C. Brown, K.A. Brayton, et al., *The immunization-induced antibody response to the Anaplasma marginale major surface protein 2 and its association with protective immunity*. Vaccine, 2010. **28**(21): p. 3741-7.
65. Lopez, J.E., G.H. Palmer, K.A. Brayton, M.J. Dark, S.E. Leach, et al., *Immunogenicity of Anaplasma marginale type IV secretion system proteins in a protective outer membrane vaccine*. Infect Immun, 2007. **75**(5): p. 2333-42.
66. Lopez, J.E., W.F. Siems, G.H. Palmer, K.A. Brayton, T.C. McGuire, et al., *Identification of novel antigenic proteins in a complex Anaplasma marginale outer membrane immunogen by mass spectrometry and genomic mapping*. Infect Immun, 2005. **73**(12): p. 8109-18.
67. Morse, K., J. Norimine, G.H. Palmer, E.L. Suttin, T.V. Baszler, et al., *Association and Evidence for Linked Recognition of Type IV Secretion System Proteins VirB9-1, VirB9-2, and VirB10 in Anaplasma marginale*. Infect Immun, 2012. **80**(1): p. 215-27.
68. Noh, S.M., K.A. Brayton, W.C. Brown, J. Norimine, G.R. Munske, et al., *Composition of the surface proteome of Anaplasma marginale and its role in protective immunity induced by outer membrane immunization*. Infect Immun, 2008. **76**(5): p. 2219-26.

69. Tebele, N., T.C. McGuire, y G.H. Palmer, *Induction of protective immunity by using Anaplasma marginale initial body membranes*. Infect Immun, 1991. **59**(9): p. 3199-204.
70. Palmer, G.H., D. Munodzana, N. Tebele, T. Ushe, y T.F. McElwain, *Heterologous strain challenge of cattle immunized with Anaplasma marginale outer membranes*. Vet Immunol Immunopathol, 1994. **42**(3-4): p. 265-73.
71. Brayton, K.A., L.S. Kappmeyer, D.R. Herndon, M.J. Dark, D.L. Tibbals, et al., *Complete genome sequencing of Anaplasma marginale reveals that the surface is skewed to two superfamilies of outer membrane proteins*. Proc Natl Acad Sci U S A, 2005. **102**(3): p. 844-9.
72. Suttent, E.L., J. Norimine, P.A. Beare, R.A. Heinzen, J.E. Lopez, et al., *Anaplasma marginale type IV secretion system proteins VirB2, VirB7, VirB11, and VirD4 are immunogenic components of a protective bacterial membrane vaccine*. Infect Immun, 2010. **78**(3): p. 1314-25.
73. Bagos, P.G., T.D. Liakopoulos, I.C. Spyropoulos, y S.J. Hamodrakas, *PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins*. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W400-4.
74. Ge, Y. y Y. Rikihisa, *Identification of novel surface proteins of Anaplasma phagocytophilum by affinity purification and proteomics*. J Bacteriol, 2007. **189**(21): p. 7819-28.
75. Ge, Y. y Y. Rikihisa, *Surface-exposed proteins of Ehrlichia chaffeensis*. Infect Immun, 2007. **75**(8): p. 3833-41.
76. Driessen, A.J. y N. Nouwen, *Protein translocation across the bacterial cytoplasmic membrane*. Annu Rev Biochem, 2008. **77**: p. 643-67.
77. Lee, P.A., D. Tullman-Ercek, y G. Georgiou, *The bacterial twin-arginine translocation pathway*. Annu Rev Microbiol, 2006. **60**: p. 373-95.
78. Bao, W., Y. Kumagai, H. Niu, M. Yamaguchi, K. Miura, et al., *Four VirB6 paralogs and VirB9 are expressed and interact in Ehrlichia chaffeensis-containing vacuoles*. J Bacteriol, 2009. **191**(1): p. 278-86.
79. Cheng, Z., X. Wang, y Y. Rikihisa, *Regulation of type IV secretion apparatus genes during Ehrlichia chaffeensis intracellular development by a previously unidentified protein*. J Bacteriol, 2008. **190**(6): p. 2096-105.
80. Gillespie, J.J., N.C. Ammerman, S.M. Dreher-Lesnack, M.S. Rahman, M.J. Worley, et al., *An anomalous type IV secretion system in Rickettsia is evolutionarily conserved*. PLoS One, 2009. **4**(3): p. e4833.
81. Rikihisa, Y., *Mechanisms of obligatory intracellular infection with Anaplasma phagocytophilum*. Clin Microbiol Rev, 2011. **24**(3): p. 469-89.
82. Yuan, J., J.C. Zweers, J.M. van Dijl, y R.E. Dalbey, *Protein transport across and into cell membranes in bacteria and archaea*. Cell Mol Life Sci, 2010. **67**(2): p. 179-99.
83. Bogsch, E.G., F. Sargent, N.R. Stanley, B.C. Berks, C. Robinson, et al., *An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria*. J Biol Chem, 1998. **273**(29): p. 18003-6.
84. Sargent, F., E.G. Bogsch, N.R. Stanley, M. Wexler, C. Robinson, et al., *Overlapping functions of components of a bacterial Sec-independent protein export pathway*. Embo J, 1998. **17**(13): p. 3640-50.
85. Sargent, F., N.R. Stanley, B.C. Berks, y T. Palmer, *Sec-independent protein translocation in Escherichia coli. A distinct and pivotal role for the TatB protein*. J Biol Chem, 1999. **274**(51): p. 36073-82.
86. Berks, B.C., F. Sargent, y T. Palmer, *The Tat protein export pathway*. Mol Microbiol, 2000. **35**(2): p. 260-74.
87. Xiong, Y., C.L. Santini, B. Kan, J. Xu, A. Filloux, et al., *Expression level of heterologous tat genes is crucial for in vivo reconstitution of a functional Tat translocase in Escherichia coli*. Biochimie, 2007. **89**(5): p. 676-85.

88. Jack, R.L., F. Sargent, B.C. Berks, G. Sawers, y T. Palmer, *Constitutive expression of Escherichia coli tat genes indicates an important role for the twin-arginine translocase during aerobic and anaerobic growth*. J Bacteriol, 2001. **183**(5): p. 1801-4.
89. Chanal, A., C. Santini, y L. Wu, *Potential receptor function of three homologous components, TatA, TatB and TatE, of the twin-arginine signal sequence-dependent metalloenzyme translocation pathway in Escherichia coli*. Mol Microbiol, 1998. **30**(3): p. 674-6.
90. Dilks, K., M.I. Gimenez, y M. Pohlschroder, *Genetic and biochemical analysis of the twin-arginine translocation pathway in halophilic archaea*. J Bacteriol, 2005. **187**(23): p. 8104-13.
91. Jongbloed, J.D., U. Grieger, H. Antelmann, M. Hecker, R. Nijland, et al., *Two minimal Tat translocases in Bacillus*. Mol Microbiol, 2004. **54**(5): p. 1319-25.
92. Pop, O., U. Martin, C. Abel, y J.P. Muller, *The twin-arginine signal peptide of PhoD and the TatAd/Cd proteins of Bacillus subtilis form an autonomous Tat translocation system*. J Biol Chem, 2002. **277**(5): p. 3268-73.
93. Wu, L.F., B. Ize, A. Chanal, Y. Quentin, y G. Fichant, *Bacterial twin-arginine signal peptide-dependent protein translocation pathway: evolution and mechanism*. J Mol Microbiol Biotechnol, 2000. **2**(2): p. 179-89.
94. Blaudeck, N., P. Kreutzenbeck, M. Muller, G.A. Sprenger, y R. Freudl, *Isolation and characterization of bifunctional Escherichia coli TatA mutant proteins that allow efficient tat-dependent protein translocation in the absence of TatB*. J Biol Chem, 2005. **280**(5): p. 3426-32.
95. Punginelli, C., B. Maldonado, S. Grahl, R. Jack, M. Alami, et al., *Cysteine scanning mutagenesis and topological mapping of the Escherichia coli twin-arginine translocase TatC Component*. J Bacteriol, 2007. **189**(15): p. 5482-94.
96. Hinsley, A.P., N.R. Stanley, T. Palmer, y B.C. Berks, *A naturally occurring bacterial Tat signal peptide lacking one of the 'invariant' arginine residues of the consensus targeting motif*. FEBS Lett, 2001. **497**(1): p. 45-9.
97. Ignatova, Z., C. Hornle, A. Nurk, y V. Kasche, *Unusual signal peptide directs penicillin amidase from Escherichia coli to the Tat translocation machinery*. Biochem Biophys Res Commun, 2002. **291**(1): p. 146-9.
98. Pradel, N., J. Delmas, L.F. Wu, C.L. Santini, y R. Bonnet, *Sec- and Tat-dependent translocation of beta-lactamases across the Escherichia coli inner membrane*. Antimicrob Agents Chemother, 2009. **53**(1): p. 242-8.
99. Alami, M., I. Luke, S. Deitermann, G. Eisner, H.G. Koch, et al., *Differential interactions between a twin-arginine signal peptide and its translocase in Escherichia coli*. Mol Cell, 2003. **12**(4): p. 937-46.
100. Meloni, S., L. Rey, S. Sidler, J. Imperial, T. Ruiz-Argueso, et al., *The twin-arginine translocation (Tat) system is essential for Rhizobium-legume symbiosis*. Mol Microbiol, 2003. **48**(5): p. 1195-207.
101. Bronstein, P.A., M. Marrichi, S. Cartinhour, D.J. Schneider, y M.P. DeLisa, *Identification of a twin-arginine translocation system in Pseudomonas syringae pv. tomato DC3000 and its contribution to pathogenicity and fitness*. J Bacteriol, 2005. **187**(24): p. 8450-61.
102. Caldelari, I., S. Mann, C. Crooks, y T. Palmer, *The Tat pathway of the plant pathogen Pseudomonas syringae is required for optimal virulence*. Mol Plant Microbe Interact, 2006. **19**(2): p. 200-12.
103. De Buck, E., E. Lammertyn, y J. Anne, *The importance of the twin-arginine translocation pathway for bacterial virulence*. Trends Microbiol, 2008. **16**(9): p. 442-53.
104. De Buck, E., L. Maes, E. Meyen, L. Van Mellaert, N. Geukens, et al., *Legionella pneumophila Philadelphia-1 tatB and tatC affect intracellular replication and biofilm formation*. Biochem Biophys Res Commun, 2005. **331**(4): p. 1413-20.

105. Tringe, S.G., C. von Mering, A. Kobayashi, A.A. Salamov, K. Chen, et al., *Comparative metagenomics of microbial communities*. Science, 2005. **308**(5721): p. 554-7.
106. Joshi, M.V., S.G. Mann, H. Antelmann, D.A. Widdick, J.K. Fyans, et al., *The twin arginine protein transport pathway exports multiple virulence proteins in the plant pathogen Streptomyces scabies*. Mol Microbiol, 2010. **77**(1): p. 252-71.
107. Lavander, M., S.K. Ericsson, J.E. Broms, y A. Forsberg, *The twin arginine translocation system is essential for virulence of Yersinia pseudotuberculosis*. Infect Immun, 2006. **74**(3): p. 1768-76.
108. McDonough, J.A., K.E. Hacker, A.R. Flores, M.S. Pavelka, Jr., y M. Braunstein, *The twin-arginine translocation pathway of Mycobacterium smegmatis is functional and required for the export of mycobacterial beta-lactamases*. J Bacteriol, 2005. **187**(22): p. 7667-79.
109. Ochsner, U.A., A. Snyder, A.I. Vasil, y M.L. Vasil, *Effects of the twin-arginine translocase on secretion of virulence factors, stress response, and pathogenesis*. Proc Natl Acad Sci U S A, 2002. **99**(12): p. 8312-7.
110. Rossier, O. y N.P. Cianciotto, *The Legionella pneumophila tatB gene facilitates secretion of phospholipase C, growth under iron-limiting conditions, and intracellular infection*. Infect Immun, 2005. **73**(4): p. 2020-32.
111. Saint-Joanis, B., C. Demangel, M. Jackson, P. Brodin, L. Marsollier, et al., *Inactivation of Rv2525c, a substrate of the twin arginine translocation (Tat) system of Mycobacterium tuberculosis, increases beta-lactam susceptibility and virulence*. J Bacteriol, 2006. **188**(18): p. 6669-79.
112. Zhang, L., Z. Zhu, H. Jing, J. Zhang, Y. Xiong, et al., *Pleiotropic effects of the twin-arginine translocation system on biofilm formation, colonization, and virulence in Vibrio cholerae*. BMC Microbiol, 2009. **9**: p. 114.
113. Ding, Z. y P.J. Christie, *Agrobacterium tumefaciens twin-arginine-dependent translocation is important for virulence, flagellation, and chemotaxis but not type IV secretion*. J Bacteriol, 2003. **185**(3): p. 760-71.
114. Lindenstrauss, U. y T. Bruser, *Conservation and variation between Rhodobacter capsulatus and Escherichia coli Tat systems*. J Bacteriol, 2006. **188**(22): p. 7807-14.
115. Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, et al., *Clustal W and Clustal X version 2.0*. Bioinformatics, 2007. **23**(21): p. 2947-8.
116. Hicks, M.G., E. de Leeuw, I. Porcelli, G. Buchanan, B.C. Berks, et al., *The Escherichia coli twin-arginine translocase: conserved residues of TatA and TatB family components involved in protein transport*. FEBS Lett, 2003. **539**(1-3): p. 61-7.
117. Buchanan, G., E. de Leeuw, N.R. Stanley, M. Wexler, B.C. Berks, et al., *Functional complexity of the twin-arginine translocase TatC component revealed by site-directed mutagenesis*. Mol Microbiol, 2002. **43**(6): p. 1457-70.
118. Hicks, M.G., D. Guymer, G. Buchanan, D.A. Widdick, I. Caldelari, et al., *Formation of functional Tat translocases from heterologous components*. BMC Microbiol, 2006. **6**: p. 64.
119. Ize, B., N.R. Stanley, G. Buchanan, y T. Palmer, *Role of the Escherichia coli Tat pathway in outer membrane integrity*. Mol Microbiol, 2003. **48**(5): p. 1183-93.
120. Stanley, N.R., K. Findlay, B.C. Berks, y T. Palmer, *Escherichia coli strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope*. J Bacteriol, 2001. **183**(1): p. 139-44.
121. Bolhuis, A., *Protein transport in the halophilic archaeon Halobacterium sp. NRC-1: a major role for the twin-arginine translocation pathway?* Microbiology, 2002. **148**(Pt 11): p. 3335-46.

122. Rose, R.W., T. Bruser, J.C. Kissinger, y M. Pohlschroder, *Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway*. Mol Microbiol, 2002. **45**(4): p. 943-50.
123. Jenner, D.C., E. Dassa, A.M. Whatmore, y H.S. Atkins, *ATP-Binding Cassette Systems of Brucella*. Comp Funct Genomics, 2009: p. 354649.
124. Holland, I.B. y M.A. Blight, *ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans*. J Mol Biol, 1999. **293**(2): p. 381-99.
125. Heikkila, M.P., U. Honisch, P. Wunsch, y W.G. Zumft, *Role of the Tat transport system in nitrous oxide reductase translocation and cytochrome cd1 biosynthesis in Pseudomonas stutzeri*. J Bacteriol, 2001. **183**(5): p. 1663-71.
126. Seleem, M.N., S.M. Boyle, y N. Sriranganathan, *Brucella: a pathogen without classic virulence genes*. Vet Microbiol, 2008. **129**(1-2): p. 1-14.
127. Krehenbrink, M. y J.A. Downie, *Identification of protein secretion systems and novel secreted proteins in Rhizobium leguminosarum bv. viciae*. BMC Genomics, 2008. **9**: p. 55.
128. Pickering, B.S. y I.J. Oresnik, *The twin arginine transport system appears to be essential for viability in Sinorhizobium meliloti*. J Bacteriol, 2010. **192**(19): p. 5173-80.
129. Jiang, X. y M.A. Fares, *Functional Diversification of the Twin-Arginine Translocation Pathway Mediates the Emergence of Novel Ecological Adaptations*. Mol Biol Evol.
130. Massung, R.F., S.L. Hiratzka, K.A. Brayton, G.H. Palmer, y K.N. Lee, *Succinate dehydrogenase gene arrangement and expression in Anaplasma phagocytophilum*. Gene, 2008. **414**(1-2): p. 41-8.
131. Andersson, S.G., D.R. Stothard, P. Fuerst, y C.G. Kurland, *Molecular phylogeny and rearrangement of rRNA genes in Rickettsia species*. Mol Biol Evol, 1999. **16**(7): p. 987-95.
132. Rurangirwa, F.R., K.A. Brayton, T.C. McGuire, D.P. Knowles, y G.H. Palmer, *Conservation of the unique rickettsial rRNA gene arrangement in Anaplasma*. Int J Syst Evol Microbiol, 2002. **52**(Pt 4): p. 1405-9.
133. Sargent, F., B.C. Berks, y T. Palmer, *Pathfinders and trailblazers: a prokaryotic targeting system for transport of folded proteins*. FEMS Microbiol Lett, 2006. **254**(2): p. 198-207.
134. Shruthi, H., M.M. Babu, y K. Sankaran, *TAT-pathway-dependent lipoproteins as a niche-based adaptation in prokaryotes*. J Mol Evol, 2010. **70**(4): p. 359-70.
135. De Buck, E., L. Vranckx, E. Meyen, L. Maes, L. Vandersmissen, et al., *The twin-arginine translocation pathway is necessary for correct membrane insertion of the Rieske Fe/S protein in Legionella pneumophila*. FEBS Lett, 2007. **581**(2): p. 259-64.
136. Begg, K.J. y W.D. Donachie, *Cell shape and division in Escherichia coli: experiments with shape and division mutants*. J Bacteriol, 1985. **163**(2): p. 615-22.
137. Van den Ent, F., C.M. Johnson, L. Persons, P. de Boer, y J. Lowe, *Bacterial actin MreB assembles in complex with cell shape protein RodZ*. Embo J, 2010. **29**(6): p. 1081-90.
138. Figge, R.M., A.V. Divakaruni, y J.W. Gober, *MreB, the cell shape-determining bacterial actin homologue, co-ordinates cell wall morphogenesis in Caulobacter crescentus*. Mol Microbiol, 2004. **51**(5): p. 1321-32.
139. Stich, R.W., G.A. Olah, K.A. Brayton, W.C. Brown, M. Fechheimer, et al., *Identification of a novel Anaplasma marginale appendage-associated protein that localizes with actin filaments during intraerythrocytic infection*. Infect Immun, 2004. **72**(12): p. 7257-64.
140. Jacob, F. y J. Monod, *Genetic regulatory mechanisms in the synthesis of proteins*. J Mol Biol, 1961. **3**: p. 318-56.

141. Stahl, F.W. y N.E. Murray, *The evolution of gene clusters and genetic circularity in microorganisms*. Genetics, 1966. **53**(3): p. 569-76.
142. Rocha, E.P., *The organization of the bacterial genome*. Annu Rev Genet, 2008. **42**: p. 211-33.
143. Price, M.N., K.H. Huang, A.P. Arkin, y E.J. Alm, *Operon formation is driven by co-regulation and not by horizontal gene transfer*. Genome Res, 2005. **15**(6): p. 809-19.
144. Thanaraj, T.A. y P. Argos, *Ribosome-mediated translational pause and protein domain organization*. Protein Sci, 1996. **5**(8): p. 1594-612.
145. Danchin, A., P. Guerdoux-Jamet, I. Moszer, y P. Nitschke, *Mapping the bacterial cell architecture into the chromosome*. Philos Trans R Soc Lond B Biol Sci, 2000. **355**(1394): p. 179-90.
146. Dandekar, T., B. Snel, M. Huynen, y P. Bork, *Conservation of gene order: a fingerprint of proteins that physically interact*. Trends Biochem Sci, 1998. **23**(9): p. 324-8.
147. Lawrence, J.G. y J.R. Roth, *Selfish operons: horizontal transfer may drive the evolution of gene clusters*. Genetics, 1996. **143**(4): p. 1843-60.
148. Fang, G., E.P. Rocha, y A. Danchin, *Persistence drives gene clustering in bacterial genomes*. BMC Genomics, 2008. **9**: p. 4.
149. Lynch, M., *Streamlining and simplification of microbial genome architecture*. Annu Rev Microbiol, 2006. **60**: p. 327-49.
150. McCutcheon, J.P. y N.A. Moran, *Extreme genome reduction in symbiotic bacteria*. Nat Rev Microbiol, 2011. **10**(1): p. 13-26.
151. Wernegreen, J.J., *For better or worse: genomic consequences of intracellular mutualism and parasitism*. Curr Opin Genet Dev, 2005. **15**(6): p. 572-83.
152. McCutcheon, J.P. y N.A. Moran, *Extreme genome reduction in symbiotic bacteria*. Nat Rev Microbiol. **10**(1): p. 13-26.
153. Cherry, J.L., *Genome size and operon content*. J Theor Biol, 2003. **221**(3): p. 401-10.
154. Guell, M., V. van Noort, E. Yus, W.H. Chen, J. Leigh-Bell, et al., *Transcriptome complexity in a genome-reduced bacterium*. Science, 2009. **326**(5957): p. 1268-71.
155. Yus, E., T. Maier, K. Michalodimitrakis, V. van Noort, T. Yamada, et al., *Impact of genome reduction on bacterial metabolism and its regulation*. Science, 2009. **326**(5957): p. 1263-8.
156. Konstantinidis, K.T. y J.M. Tiedje, *Trends between gene content and genome size in prokaryotic species with larger genomes*. Proc Natl Acad Sci U S A, 2004. **101**(9): p. 3160-5.
157. Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki, y H. Ishikawa, *Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.* APS. Nature, 2000. **407**(6800): p. 81-6.
158. Papp, B., C. Pal, y L.D. Hurst, *Dosage sensitivity and the evolution of gene families in yeast*. Nature, 2003. **424**(6945): p. 194-7.
159. Ohta, T., *Slightly deleterious mutant substitutions in evolution*. Nature, 1973. **246**(5428): p. 96-8.
160. Felsenstein, J., *Phylogenies and the comparative method*. American Naturalist, 1985. **125**: p. 1-15.
161. Conesa, A., S. Gotz, J.M. Garcia-Gomez, J. Terol, M. Talon, et al., *Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research*. Bioinformatics, 2005. **21**(18): p. 3674-6.
162. Bendtsen, J.D., H. Nielsen, G. von Heijne, y S. Brunak, *Improved prediction of signal peptides: SignalP 3.0*. J Mol Biol, 2004. **340**(4): p. 783-95.
163. Torioni de Echaide, S., D.P. Knowles, T.C. McGuire, G.H. Palmer, C.E. Suarez, et al., *Detection of cattle naturally infected with Anaplasma marginale in a region of endemicity by nested PCR and a competitive enzyme-linked*

- immunosorbent assay using recombinant major surface protein 5*. J Clin Microbiol, 1998. **36**(3): p. 777-82.
164. Higuchi, R., *Rapid, efficient DNA extraction for PCR from cells or blood*. 1989. **Amplifications**(2): p. 1-3.
165. Pfaffl, M.W., G.W. Horgan, y L. Dempfle, *Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR*. Nucleic Acids Res, 2002. **30**(9): p. e36.
166. Talavera, G. y J. Castresana, *Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments*. Syst Biol, 2007. **56**(4): p. 564-77.
167. Kumar, S., M. Nei, J. Dudley, y K. Tamura, *MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences*. Brief Bioinform, 2008. **9**(4): p. 299-306.
168. Stothard, P. y D.S. Wishart, *Circular genome visualization and exploration using CGView*. Bioinformatics, 2005. **21**(4): p. 537-9.
169. Ruybal, P., R. Moretta, A. Perez, R. Petri, P. Zimmer, et al., *Genetic diversity of Anaplasma marginale in Argentina*. Vet Parasitol, 2009. **162**(1-2): p. 176-80.
170. Wieser, M. y H.J. Busse, *Rapid identification of Staphylococcus epidermidis*. Int J Syst Evol Microbiol, 2000. **50 Pt 3**: p. 1087-93.
171. Jack, R.L., G. Buchanan, A. Dubini, K. Hatzixanthis, T. Palmer, et al., *Coordinating assembly and export of complex bacterial proteins*. Embo J, 2004. **23**(20): p. 3962-72.
172. Stanley, N.R., T. Palmer, y B.C. Berks, *The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in Escherichia coli*. J Biol Chem, 2000. **275**(16): p. 11591-6.
173. Silvestro, A., J. Pommier, y G. Giordano, *The inducible trimethylamine-N-oxide reductase of Escherichia coli K12: biochemical and immunological studies*. Biochim Biophys Acta, 1988. **954**(1): p. 1-13.
174. Bendtsen, J.D., H. Nielsen, D. Widdick, T. Palmer, y S. Brunak, *Prediction of twin-arginine signal peptides*. BMC Bioinformatics, 2005. **6**: p. 167.
175. Dilks, K., R.W. Rose, E. Hartmann, y M. Pohlschroder, *Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey*. J Bacteriol, 2003. **185**(4): p. 1478-83.
176. Bagos, P.G., E.P. Nikolaou, T.D. Liakopoulos, y K.D. Tsirigos, *Combined prediction of Tat and Sec signal peptides with hidden Markov models*. Bioinformatics, 2010. **26**(22): p. 2811-7.
177. Mao, F., P. Dam, J. Chou, V. Olman, y Y. Xu, *DOOR: a database for prokaryotic operons*. Nucleic Acids Res, 2009. **37**(Database issue): p. D459-63.
178. Yin, Y., H. Zhang, V. Olman, y Y. Xu, *Genomic arrangement of bacterial operons is constrained by biological pathways encoded in the genome*. Proc Natl Acad Sci U S A, 2010. **107**(14): p. 6310-5.
179. Allen, T.E., M.J. Herrgard, M. Liu, Y. Qiu, J.D. Glasner, et al., *Genome-scale analysis of the uses of the Escherichia coli genome: model-driven analysis of heterogeneous data sets*. J Bacteriol, 2003. **185**(21): p. 6392-9.
180. Covert, M.W., E.M. Knight, J.L. Reed, M.J. Herrgard, y B.O. Palsson, *Integrating high-throughput and computational data elucidates bacterial networks*. Nature, 2004. **429**(6987): p. 92-6.
181. Corbin, R.W., O. Paliy, F. Yang, J. Shabanowitz, M. Platt, et al., *Toward a protein profile of Escherichia coli: comparison to its transcription profile*. Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9232-7.
182. Lu, P., C. Vogel, R. Wang, X. Yao, y E.M. Marcotte, *Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation*. Nat Biotechnol, 2007. **25**(1): p. 117-24.
183. Lopez-Campistrous, A., P. Semchuk, L. Burke, T. Palmer-Stone, S.J. Brokx, et al., *Localization, annotation, and comparison of the Escherichia coli K-12*

- proteome under two states of growth*. Mol Cell Proteomics, 2005. **4**(8): p. 1205-9.
184. Masuda, T., N. Saito, M. Tomita, y Y. Ishihama, *Unbiased quantitation of Escherichia coli membrane proteome using phase transfer surfactants*. Mol Cell Proteomics, 2009. **8**(12): p. 2770-7.
185. Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection*. Mol Syst Biol, 2006. **2**: p. 2006 0008.