



TÍTULO

ANÁLISIS BIOINFORMÁTICO DE GENES APLICADOS EN REPLICACIÓN DE GENOMAS DE BACTERIAS ENDOSIMBIONTES

AUTOR

María del Rocío González Soltero

Director
Curso
ISBN

©
©

Esta edición electrónica ha sido realizada en 2010

Enrique Viguera Mínguez

II Máster en Bioinformática

978-84-693-6140-5

María del Rocío González Soltero

Para esta edición, la Universidad Internacional de Andalucía



Reconocimiento-No comercial-Sin obras derivadas 2.5 España.

Usted es libre de:

- Copiar, distribuir y comunicar públicamente la obra.

Bajo las condiciones siguientes:

- **Reconocimiento.** Debe reconocer los créditos de la obra de la manera especificada por el autor o el licenciadador (pero no de una manera que sugiera que tiene su apoyo o apoyan el uso que hace de su obra).
 - **No comercial.** No puede utilizar esta obra para fines comerciales.
 - **Sin obras derivadas.** No se puede alterar, transformar o generar una obra derivada a partir de esta obra.
-
- *Al reutilizar o distribuir la obra, tiene que dejar bien claro los términos de la licencia de esta obra.*
 - *Alguna de estas condiciones puede no aplicarse si se obtiene el permiso del titular de los derechos de autor.*
 - *Nada en esta licencia menoscaba o restringe los derechos morales del autor.*



II MAESTRÍA EN BIOINFORMÁTICA

UNIVERSIDAD INTERNACIONAL DE ANDALUCÍA

SEDE TECNOLÓGICA DE MÁLAGA

AUTOR: MARÍA DEL ROCÍO GONZÁLEZ SOLTERO

“Análisis bioinformático de genes implicados en replicación en genomas de bacterias endosimbiontes”

DIRECTOR/TUTOR DE TESIS:
DR. ENRIQUE VIGUERA MÍNGUEZ



**MEMORIA DE TESIS PRESENTADA PARA OPTAR AL TÍTULO DE MÁSTER EN
BIOINFORMÁTICA**

MÁLAGA, 18 DE DICIEMBRE DE 2009

ÍNDICE

ÍNDICE	5
Glosario	7
Introducción	9
1. Hacia la búsqueda del genoma mínimo: concepto de tamaño de genoma bacteriano	9
2. Evolución reductiva del genoma y concepto de genoma mínimo	10
2.1. La vida, desde una concepción reduccionista: concepto de Biología Sintética .	10
2.2. Proceso de reducción genómica en Bacterias Endosimbiontes.....	11
3. Genómica comparativa	13
3.1. Métodos para establecer correspondencia de genes entre genomas.....	14
3.2. Sistemas automáticos para el análisis comparativo	16
4. Caso de estudio	17
5. Reducción genómica del componente 3R en bacterias endosimbiontes obligados	19
6. Replicación del DNA en bacterias	20
6.1. Inicio de la replicación	20
6.2. Elongación de la replicación	20
6.3. Terminación	24
7. Evolución de los sistemas de replicación	25
8. Sistemas de reparación y recombinación en procariontes	25
9. Redes metabólicas mínimas	26
Materiales	29
1. Grupo de genomas endosimbiontes utilizados	29
2. Herramientas bioinformáticas utilizadas para la determinación de proteínas ortólogas en genomas de bacterias endosimbiontes de insectos	30
2.1. Herramientas para el análisis comparativo de los genomas mínimos	31
Perspectiva actual y objetivos	39
1. El número de componentes del sistema 3R depende del tamaño del genoma del endosimbionte	43
2. Conjunto mínimo de proteínas implicadas en procesos de replicación del DNA	45
2.1. Maquinaria de inicio de replicación	45
2.2. Elongación de la replicación: la Holoenzima PolIII	50
2.3. Terminación de la replicación	51
3. Topología del DNA	51

3.1. Topología del DNA en genomas mínimos.....	52
4. Síntesis y modulación de nucleótidos.....	53
5. Proteínas de recombinación y reparación del DNA.....	58
5.1. Reparación por escisión de bases (BER).....	59
5.2. Reversión directa del daño.....	60
5.3. Reparación de bases mal apareadas (MMR).....	60
5.4. Escisión de nucleótidos (NER).....	61
5.5. Proteínas de unión a cadena sencilla.....	62
5.6. Factores de reparación acoplados a la transcripción.....	63
5.7. Proteínas de tolerancia al daño: TLS DNA polimerasas DinB, DinG y UmuCD..	63
5.8. Reparación por recombinación RecA.....	63
5.9. Reparación independiente de RecA-recombinación.....	66
6. <i>Strand bias</i> y componente 3R en bacterias endosimbiontes.....	66
7. Conservación de dominios estructurales y de su arquitectura en proteínas implicadas en transferencia de información.....	68
7.1. Ausencia de proteínas de Recombinación y Conservación de helicasas y endo- y exo- nucleasas y glicosidasas.....	69
Discusión.....	71
1. Conservación de la maquinaria de replicación en bacterias endosimbiontes	73
2. La proteína NrdA como fósil molecular para la filogenia de poblaciones de endosimbiontes.....	74
3. Conservación de la maquinaria de recombinación y reparación en bacterias endosimbiontes.....	75
4. Adaptación hacia otros modos de replicación: replicación dependiente de transcripción y círculo rodante.....	76
5. Evolución de la minimización de genomas hasta su conversión en orgánulo: justificación de la Teoría Endosimbiótica.....	76
Conclusiones.....	79
Bibliografía.....	81
ANEXOS.....	89
Agradecimientos.....	91

Glosario

BACTERIOMA (BACTERIOCITO): orgánulo o célula especializada que presentan algunos insectos y que se utilizan para albergar bacterias simbiotes. Normalmente se encuentran localizados en el citoplasma.

BIOLOGÍA SINTÉTICA: campo de investigación que combina ciencia e ingeniería. Su objetivo es el diseño y construcción de sistemas biológicos.

BIOLOGÍA DE SISTEMAS: término usado para describir a una nueva rama de la biología que estudia los sistemas biológicos desde una disciplina multidisciplinar y trata de indagar las complejas redes de interacciones que tienen lugar en los mismos. Esta nueva disciplina, a diferencia de otras, tiene una perspectiva holística en lugar de la reduccionista que es la más habitual.

CONCEPTOS BASADOS EN LA TEORÍA NEUTRAL DE LA EVOLUCIÓN MOLECULAR

(Propuesta por Kimura (1968) y Kingy Jukes(1969)):

Según esta teoría, la tasa de evolución observada a nivel molecular es muy rápida para ser exclusivamente debida a selección natural. En el nivel molecular, la mayoría de los cambios evolutivos se debe a la deriva genética de genes mutantes selectivamente equivalentes.

*TASA DE EVOLUCIÓN MOLECULAR: la tasa de evolución molecular puede calcularse comparando el número de diferencias aminoacídicas o nucleotídicas entre dos especies que derivan hace "X" tiempo de un antecesor común.

* HIPÓTESIS DEL RELOJ MOLECULAR: deduce el tiempo pasado a partir del número de diferencias entre dos secuencias de ADN. La técnica del reloj molecular es una herramienta importante en la sistemática molecular y estima que la divergencia molecular entre dos especies es proporcional al tiempo de separación de las mismas. Su tasa de evolución es relativamente constante a lo largo de períodos largos y toma valores semejantes en distintas especies. El conocimiento de la tasa aproximada de evolución molecular en ciertos grupos de linajes facilita el establecimiento de fechas de eventos filogenéticos no documentados en el registro fósil, como la divergencia de taxones vivos y la formación del árbol filogenético.

ENDOSIMBIOTES PRIMARIOS (P-endosimbiontes): microorganismos simbiotes que son heredados estrictamente por transmisión vertical y, en la mayoría de los casos, residen en células especializadas denominadas bacteriocitos. La asociación obligada entre los áfidos y su endosimbionte *Buchnera aphidicola* que es transmitido intracelularmente ofrece un modelo para analizar la reducción genómica e identificar los mecanismos moleculares potencialmente implicados en dicho proceso. La reducción genómica en bacterias endosimbiontes es un proceso continuo derivado de su adaptación a la vida intracelular. Los tamaños de los genomas de las diferentes cepas de *B. aphidicola* aisladas de diferentes subfamilias de áfidos oscilan en tamaño entre 450 Kb a 641 Kb, siendo el endosimbionte del áfido *Cinara cedri* (*B. aphidicola Cc*) la que presenta un genoma más reducido. Todas estas estirpes de *Buchnera* muestran una arquitectura genómica conservada con su último ancestro simbiote, por lo que la pérdida de genes en linajes extensos debe estar relacionada con la especificidad de huésped. La pregunta que queda por resolver es si esta reducción alcanzará un límite o continuará hasta la extinción de la bacteria y su reemplazamiento por un nuevo simbiote.

ENDOSIMBIOTES SECUNDARIOS (S-Endosimbiontes): ocasionalmente, los hospedadores pueden tolerar la existencia de un hospedador secundario o facultativo (S-simbionte) que puede coexistir con el P-endosimbionte. Aunque son transmitidos normalmente por transmisión vertical, podría existir una esporádica transmisión horizontal entre hospedadores individuales. Un ejemplo es el caso de S-simbionte *Sodalis glossinidius* que coexisten en el lumen del estómago de las moscas tse-tse con el P-endosimbionte *Wigglesworthia glossinidia*. En los endosimbiontes secundarios la reducción genómica es menos evidente, presentan mayor número de pseudogenes y genomas de mayor tamaño.

GENOMA MÍNIMO: estimación del tamaño de genoma compuesto por el menor número de genes suficientes para constituir un organismo celular de vida libre.

TASA DE MUTACIÓN: es la probabilidad de que se de una mutación en un organismo o gen en cada generación.

TOPRIM: dominio proteico de aproximadamente 100 aminoácidos y que presenta dos motivos conservados, uno de los cuales presenta Glutamato y dos Aspartatos conservados. El Glutamato actúa como base para la polimerización de nucleótidos por la Primasa. Es común en la Familia de ATPasas tipo AAA+.

POLIPLOIDÍA: fenómeno por el cual se originan células, tejidos u organismos con tres o más de juegos completos de cromosomas o genomas.

Introducción

1. Hacia la búsqueda del genoma mínimo: concepto de tamaño de genoma bacteriano

El genoma bacteriano está compuesto por un replicón generalmente circular (endogenoma) y por moléculas extracromosómicas de tipo plasmídico. El tamaño de los genomas se representa por el valor C que indica la cantidad total de DNA en un genoma haploide. Se denomina C , por constante o característico, para indicar el hecho de que el tamaño es prácticamente constante dentro de una especie. Sin embargo, este valor varía ampliamente, desde menos de 10^6 pares de bases (pb) para algunos genomas de arqueas, bacterias y protistas, hasta más de 10^{11} pb para algunos protistas, plantas y animales. Esto se conoce como *paradoja del valor C* dado que se espera que la cantidad de DNA se correlacione positivamente con la **complejidad genética** del organismo.

Son muchos los mecanismos mutacionales que pueden producir cambios en el tamaño del genoma. Algunos de ellos ocurren a gran escala (duplicación de todo el genoma), mientras que otros ocurren a una escala muy pequeña (pérdida o ganancia de unos pocos nucleótidos). Los elementos genéticos móviles, o elementos transponibles, son otros de los causantes de grandes variaciones en el tamaño del genoma. Se considera que las inserciones o deleciones espontáneas (llamadas *indels*) de unos pocos nucleótidos son una de las causas más importantes de la evolución del tamaño del genoma a largo plazo. La fijación de estas mutaciones es muy improbable si el *indel* afecta a un gen, pero es más probable si afecta a pseudogenes (genes no funcionales, inactivados recientemente) o a otras secuencias de DNA sin función. La desintegración de los genes, o desaparición de un gen no esencial del genoma, suele transcurrir en un primer paso con su inactivación mediante una mutación puntual (formación de un pseudogen). Posteriormente, el DNA que forma los pseudogenes va siendo eliminado hasta perderse cualquier vestigio de su presencia en el genoma (Pérez-Brocal *et al.*, 2006).

Uno de los procesos de mayor importancia en el aumento del tamaño del genoma en organismos unicelulares, especialmente procariotas, es la transferencia genética horizontal (Ochman *et al.*, 2000). Este proceso consiste en la introducción en el genoma de una especie de un fragmento de DNA de otra especie, que contiene uno

o varios genes. El DNA se introduce en la célula por diferentes mecanismos y, para poder integrarse en el genoma de la célula ha de producirse un evento de recombinación. Si los genes introducidos confieren alguna ventaja al organismo, éstos suelen mantenerse, pero si no es el caso, pueden adquirir nuevas mutaciones lo que conduce a su inactivación. Un ejemplo de la importancia de estos procesos se puede observar en la bacteria *Escherichia coli*, algunas de cuyas cepas han adquirido un carácter patogénico por la transferencia de genes de virulencia.

Mientras que la ganancia de genes permite a las bacterias colonizar nuevos nichos, la pérdida de genes les ha permitido refinar esa adaptación a los nuevos nichos. Éste es el caso de las bacterias que establecen una relación simbiótica con un hospedador (Moran y Wernegreen, 2000). Entre los casos de endosimbiosis mejor estudiados se encuentran las gamma-proteobacterias simbiotes de insectos (Wernegreen, 2002), de las que disponemos de numerosos genomas secuenciados hasta la fecha. Dentro de este phylum se incluyen los organismos modelo de la genética y patogénesis bacteriana, *E. coli* y *Salmonella typhimurium*. En este grupo encontramos también algunas de las bacterias con los tamaños de genomas secuenciados más pequeños que oscilan entre los 450 Kb de *Buchnera aphidicola* (Pérez-Brocá et al., 2006) y los 7000 Kb de *Pseudomonas fluorescens* (Paulsen et al., 2005).

2. Evolución reductiva del genoma y concepto de genoma mínimo

2.1. La vida, desde una concepción reduccionista: concepto de Biología Sintética

Los adelantos de la Genómica han permitido afrontar la secuenciación de genomas de bacterias no cultivables, patógenas o endosimbiontes.

Una característica principal de la gran mayoría de bacterias endosimbiontes es que presentan tamaños genómicos reducidos con respecto a otras bacterias de vida libre con las que están emparentadas filogenéticamente. El análisis del tamaño y composición de dichos genomas ha abierto una serie de estudios encaminados a determinar las bases moleculares del proceso de reducción genómica y, consecuentemente, de los límites teóricos en cuanto a tamaño y contenido génico de estos genomas mínimos. A su vez, este concepto se ha desarrollado a la par de las técnicas de síntesis de cadenas de nucleótidos *in vitro*, dando paso a la

posibilidad de sintetizar genomas mínimos artificiales en el laboratorio. Estos avances han supuesto la aparición de un nuevo concepto que se conoce como *Biología Sintética*. Entre las perspectivas de la Biología Sintética están el diseño y fabricación de componentes y sistemas biológicos que no existen hoy día en la naturaleza, así como el rediseño de los sistemas biológicos ya existentes. Una consecuencia del progreso en el campo de la Biología Sintética es la concepción de las células como partes ensambladas que configuran un organismo con un fenotipo deseado. Es el reduccionismo molecular en su máxima expresión.

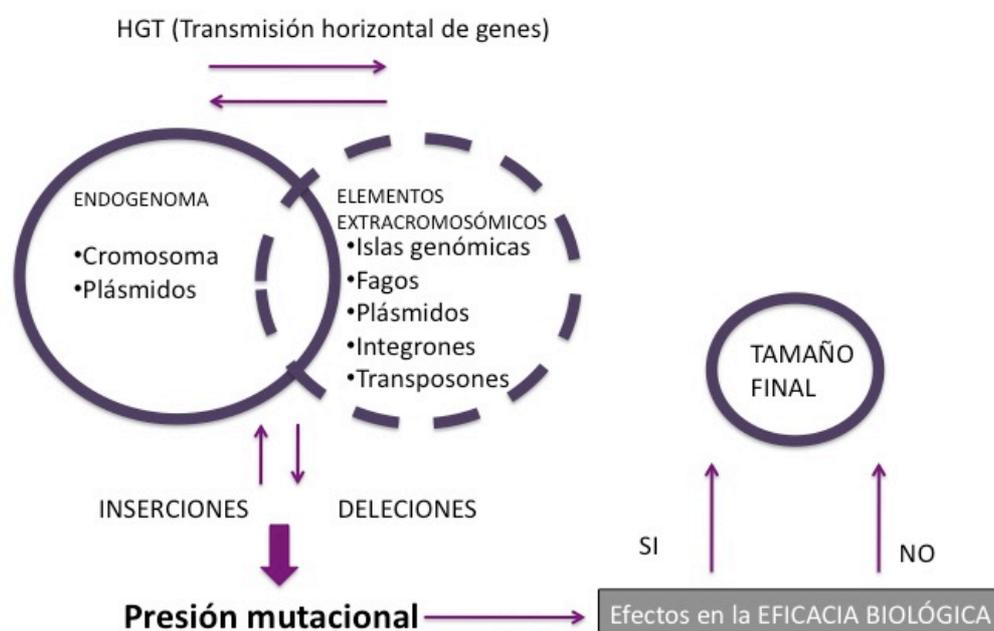
Para lograr el objetivo de síntesis de vida artificial es necesario un requisito previo: identificar la configuración mínima de genes necesarios para sustentar la vida que permitiera a dicha célula artificial replicarse de manera autónoma. Son varias las aproximaciones experimentales seguidas hasta la fecha para obtener este «genoma mínimo». Por un lado, la identificación de genes esenciales por mutagénesis con transposones, por otro lado el análisis bioinformático de genes compartidos entre taxones diferentes. Una tercera alternativa consiste en el estudio de sistemas biológicos que, de forma natural, han experimentado una reducción de su material genético.

En este trabajo hemos analizado el componente replicativo de genomas de bacterias endosimbiontes con el objetivo de determinar el componente mínimo esencial que requeriría una bacteria artificial para replicar su material genético. A su vez, este estudio podría aportar indicios que permitan plantear una hipótesis acerca del mecanismo molecular por el que se produce un sesgo hacia alto contenido en A+T en los genomas de bacterias endosimbiontes, así como el proceso por el que se producirían deleciones en el DNA. En base a los conocimientos experimentales de los que disponemos hoy día, cabe pensar que ambos procesos estarían relacionados con defectos en las maquinarias de replicación reparación y recombinación, de ahí que el estudio se haya hecho también extensible a estos sistemas.

2.2. Proceso de reducción genómica en Bacterias Endosimbiontes

El estudio comparado de las secuencias genómicas de algunas bacterias endosimbiontes de insectos ha puesto de manifiesto los drásticos cambios que han experimentado estos organismos durante su proceso de adaptación a la vida

intracelular. Entre otros, cabe citar la espectacular reducción genómica, el cambio en la regulación de algunos de sus operones, la proliferación de plásmidos de biosíntesis de aminoácidos, el incremento en el contenido de A+T, la aceleración de las tasas evolutivas y la pérdida del sesgo en el uso de codones. El conjunto de procesos que ha podido sufrir un organismo hasta llegar a su tamaño final se resume en la figura 1.



Fuerzas que modelan el tamaño de los genomas.

Fig. 1. Reducción genómica en endosimbiontes. Procesos de reducción de tamaño que ha sufrido el componente génico en algunas bacterias, así como algunos de los procesos que han tenido lugar sobre el genoma y que han dado lugar al tamaño de genoma actual.

En la adaptación a la vida intracelular determinados genes esenciales para la vida libre pasarían a ser prescindibles, bien porque el producto génico es fabricado por el hospedador, bien porque ya no son necesarios en el nuevo nicho ecológico, por lo que no existiría una presión selectiva sobre ellos.

Se ha descrito la existencia de procesos de mutación que ocurren durante la colonización de nuevos hospedadores por parte de la bacteria endosimbionte. Estos

procesos pueden resumirse en dos etapas. La primera etapa, o *etapa inicial*, vendría caracterizada por una elevada inestabilidad genómica que promueve un gran número de cambios genómicos y como consecuencia de ello, acumulación de mutaciones en genes no esenciales. La incorporación de nuevo material génico por procesos de transferencia génica horizontal se ve impedida en el nuevo contexto celular, lo que unido al pequeño tamaño poblacional y cuellos de botella aumenta la fijación de mutaciones ligeramente deletéreas (**Trinquete de Müller**). La *etapa final* está caracterizada por la estabilidad genómica y la reducción del genoma. Esta estabilidad es resultado de la pérdida de genes de recombinación y de secuencias repetidas durante las etapas tempranas y de las reducidas oportunidades de intercambio genético, debido al *secuestro* en un ambiente intracelular.

Otros procesos que caracterizan a los genomas que han sufrido pérdida de material genético son, junto con una elevada tasa de sustitución, el sesgo en A+T, la ausencia de sesgo adaptativo en el uso de codones y la pérdida de elementos repetidos. Todos ellos constituyen los síntomas de la degradación genómica a la que está sometida esta bacteria, cuya principal consecuencia es la drástica reducción del tamaño genómico.

Dado que la gran mayoría de bacterias endosimbiontes comparten un ancestro común con bacterias de vida libre de las cuales disponemos de su secuencia genómica, el **análisis comparativo de genomas** nos permite detectar los genes conservados y los diferentes estadios de degradación genómica: delección de genes, transformación en pseudogenes o conservación del gen ancestral.

3. Genómica Comparativa

La **Genómica Comparada** es el estudio de las relaciones entre los genomas de diferentes especies. Se basa en el análisis de la secuencia genómica, desde aspectos comparativos relacionados con la composición como de los factores que promueven la evolución de los genomas.

A día de hoy se ha secuenciado un elevado número de genomas de bacterias simbiotes. La disponibilidad de estas secuencias permite realizar estudios con el objeto de comprender el proceso evolutivo por el que se ha producido una reducción en el tamaño de los genomas. Así, diversos estudios basados en la comparación de genomas de endosimbiontes con otros de bacterias de vida libre filogenéticamente

relacionados, han revelado los cambios en el contenido y tamaño genómico, comenzado a revelar los mecanismos que han tenido lugar y sus consecuencias funcionales.

El primer paso en genómica comparativa consiste en hacer corresponder los segmentos cromosómicos del nuevo genoma con elementos funcionales en las especies comparadas. Esto implica determinar los segmentos de **DNA ortólogos** - que descienden de un mismo ancestro común en las especies comparadas- o **parálogos**, surgidos a partir de eventos de duplicación anteriores a la divergencia de las especies comparadas. Existen numerosos sistemas y métodos computacionales que permiten realizar esta tarea de caracterización de los genomas así como el modo en que debe organizarse la información para facilitar la interpretación y la comparación de los resultados. En realidad, estos sistemas utilizan un conjunto de herramientas bioinformáticas que conlleva la anotación automatizada de los genomas y sus genes. Sin embargo, determinar qué genes comparten y cuáles no los distintos genomas o la reconstrucción y la comparación de las rutas metabólicas, aún está pendiente de ser automatizada.

3.1. Métodos para establecer correspondencia de genes entre genomas

- *Best Bidirectional Hits (BBH)* (Fitch 1970; Fitch 1995). Este método identifica los pares de genes que establecen una mejor correspondencia recíproca en dos genomas dados y los marca como ortólogos. En el caso de eventos recientes de duplicación génica sólo uno de los genes duplicados será marcado como ortólogo, sin señalar la presencia de homólogos adicionales. Este método no garantiza, por tanto, que las correspondencias representen relaciones de ortología y se pueden encontrar algunas asignaciones incorrectas.
- *Clusters of Orthologous Genes (COG)* (Tatusov et al. 1997; Tatusov et al. 2001). Este método establece correspondencia entre grupos de genes entre varios genomas. Los grupos de genes ortólogos se corresponden normalmente con familias génicas que se expandieron antes de la divergencia de las especies comparadas. Este método, sin embargo, es incapaz de distinguir eventos de duplicación recientes de aquellos más antiguos.

- *KEGG Orthology (KO) System* (Goto *et al.*, 1997). El sistema de Ortología KEGG (KO) consiste en una clasificación de genes ortólogos, incluidas las relaciones de parálogos en grupos de genes ortólogos. Este sistema constituye la base de los mapas KEGG PATHWAY y de la Ontología KEGG BRITE. De esta forma los genes identificados son mapeados automáticamente en KEGG pathways y asignados a las jerarquías BRITE.
- *Perfiles filogenéticos* (Pellegrini *et al.*, 1999). Este método se basa en la identificación de genes que tienen el mismo patrón de ausencia/presencia en varios genomas.
- *Conservación del contexto o vecindario génico* (Sintenia). Se basa en la observación de la existencia de un cierto porcentaje de genes cuyo vecindario está conservado entre diferentes organismos. La situación más evidente de conservación de grupos de genes es la de los operones conservados de procariontes. Por tanto, se asume que dos genes que estén ocupando la misma o una posición parecida en un grupo de genes conservados entre distintos genomas, pueden estar desarrollando la misma función (aunque no sean BBHs), con una probabilidad más alta que la que se esperaría por azar (Overbeek *et al.*, 1999; Huynen *et al.*, 1998). El caso más acusado de conservación del contexto se da cuando también se conserva el orden de los genes. A esta situación se la denomina SINTENIA y se observa especialmente en operones que codifican varios genes de una misma vía metabólica. Las desventajas de este método son que, en general, sólo se puede aplicar a genomas procariontes, y que, aún en procariontes, el orden de los genes es una característica que se pierde muy rápidamente.
- *Fusiones génicas*. Dos proteínas, o dominios de proteínas, codificadas en genes diferentes, pueden tener una función relacionada si en algunas especies están codificadas por un único gen, que presumiblemente ha sido originado en un evento de fusión génica. Esta situación también se observa en proteínas que forman parte de rutas metabólicas.

- *Mirror trees*. Este método, así como el método de las *mutaciones* correlacionadas, se basa en la idea de que proteínas que interactúan funcionalmente evolucionan al mismo tiempo. Por tanto, los árboles filogenéticos de las familias correspondientes debieran de parecerse más que los de familias de proteínas no relacionadas. Por último, asumimos que si las dos proteínas están relacionadas funcionalmente, pueden formar parte del mismo proceso.
- *Mutaciones correlacionadas*. Este método parte, como el anterior, de alineamientos múltiples de familias de proteínas, con el objetivo de identificar mutaciones compensatorias. La identificación de dichas mutaciones implica interacción funcional y pertenencia al mismo proceso celular.

3.2. Sistemas automáticos para el análisis comparativo

La existencia de sistemas automáticos es el *talón de Aquiles* de este tipo de estudios. La mayor parte de ellos poseen copias locales de herramientas bioinformáticas y bases de datos, y agrupan el acceso a todas ellas, de modo que el usuario simplemente tiene que facilitarles la secuencia o secuencias que quiere analizar. El sistema se encarga de ejecutar todos los programas y dar el formato adecuado a los datos para la presentación al usuario.

Sin embargo, los requerimientos de este tipo de sistemas son grandes: gran capacidad de almacenamiento y gran poder de cálculo. Por ello, no han sido muchos los sistemas de este tipo que han sido desarrollados para el uso público. Las páginas web de algunos sistemas de análisis de genomas se presentan a continuación.

- *Integr8 (EBI)*. Este portal proporciona un acceso integrado a la información de genomas secuenciados y sus correspondientes proteomas. Estos datos incluyen secuencias de DNA (de las bases de datos EMBL *Nucleotide Sequence Database*, *Genome Reviews* y Ensembl), secuencias de proteínas (UniProt, Knowledgebase y IPI) y análisis genómico y proteómico de genomas (InterPro; CluSTr y GOA), así como información sobre ortología, paralogía y sintenia. En este caso, los ortólogos son identificados como el mejor BBHs, usando la base de datos CluSTr para los datos de similitud de proteínas (PORC (*Putative Orthologous Clusters*)). CluSTr usa

el algoritmo Smith-Waterman para la medida de la similitud de proteínas. La puntuación obtenida del algoritmo es convertida a una medida estadística de similitud. La similitud entre proteínas es, por tanto, referida a nivel de ortólogo.

- COGs. www.ncbi.nih.gov/COG . La base de datos COGs agrupa a las proteínas provenientes de microorganismos secuenciados en su totalidad en *clusters* de proteínas ortólogas, las cuales han surgido un proceso de especiación, conservando la misma función en todos los miembros del grupo (Tatusov et al., 1997). Considera una categoría COG cualquier *cluster* formado por al menos tres proteínas que pertenecen a linajes diferentes y que presentan mayor grado de similitud unas con otras que con cualquier otra proteína perteneciendo a los mismos genomas.
- The comprehensive Microbial Resource (CMR) comparative Tools. <http://cmr.jcvi.org/cgi-bin/CMR/CmrHomePage.cgi>. Dispone de 509 genomas bacterianos secuenciados (6/12/09). Esta herramienta incluye comparaciones entre contenidos GC, secuenciaciones entre regiones genómicas, comparaciones de homología multigenoma, así como permitir mostrar datos de homología entre genoma.
- *Prolinks Database 2.0*. Esta base de datos es una colección de métodos de inferencia usados para predecir relaciones funcionales entre proteínas. Estos métodos incluyen el método del Perfil Filogenético, que determina la presencia y ausencia de proteínas a partir de alineamientos múltiples.

4. Caso de estudio

Como se ha mencionado anteriormente, el caso de estudio que vamos a considerar en este trabajo, es el componente mínimo replicativo en bacterias endosimbiontes que han sufrido un proceso de reducción genómica.

El origen de la asociación bacteria-insecto tiene, en la mayoría de los casos, un componente bioquímico. La asociación ha permitido la especialización de los insectos en determinados nichos ecológicos y comportamientos alimenticios, al proporcionar nutrientes que son deficientes en sus dietas debido a la amplia variabilidad de rutas metabólicas presentes en los microorganismos procariontes. La

mayoría de estas bacterias han sufrido un proceso de reducción genómica, conservando un grupo de genes que determina el inicio de la asociación parasítica o mutualística.

En el caso de las bacterias con genomas reducidos, la mayoría consisten en patógenos intracelulares o bien simbioses que mantienen una relación obligada con un hospedador eucariota en la que el hospedador les suplementa nutrientes. La mayoría de los endosimbiontes estudiados hasta el momento corresponden al grupo de gamma-proteobacterias que viven en asociación obligada con insectos y proporcionan a su hospedador aminoácidos esenciales, vitaminas y cofactores que están ausentes en la dieta del hospedador. La mayoría de estas especies son endosimbiontes heterotróficos.

El proceso de reducción genómica, viene derivado de su adaptación a la vida intracelular. El debate actual se centra en si esta reducción es resultado de un proceso de *deletion bias* o de selección natural (Mira *et al.*, 2001). El concepto de *deletion bias* se explica como un aumento en el número de deleciones genómicas en lugar de inserciones que ocurre cuando la selección no es lo suficientemente fuerte como para mantener gran cantidad de genes sin función en un determinado ambiente. Estas deleciones suelen implicar la pérdida o inactivación de grandes grupos de genes. El *deletional bias* observado en pseudogenes bacterianos trata de explicar porqué los genomas de bacterias endosimbiontes son compactos y ricos en genes y porqué las regiones no funcionales no se acumulan en estos genomas. Sin embargo, la comprensión de los procesos que han sido responsables de una reducción extrema de los genomas de simbioses es difícil de evaluar usando los métodos convencionales que alinean y comparan secuencias homólogas. La extensa reducción genómica podría proceder de un proceso lento y continuo de erosión de genes individuales o de grupos de genes. Para determinar la escala de los eventos de deleción, en magnitud y frecuencia, Nilsson *et al.*, (2005) monitorizaron los cambios en una estirpe silvestre y en un mutante de reparación (*mutS*) de *Salmonella typhimurium*, encontrando que las deleciones a gran escala podrían haber jugado un papel en los primeros estadios de la reducción genómica.

5. Reducción genómica del componente 3R en bacterias endosimbiontes obligados

La estricta transmisión vertical de los endosimbiontes obligados hace que sólo determinadas bacterias sean transmitidas, lo que genera un cuello de botella continuo que favorece la acción de la deriva genética. Este hecho nos permite plantear la hipótesis de que la combinación de estos factores permitiría la acumulación de mutaciones en genes necesarios, pero no esenciales como determinados genes implicados en reparación del DNA y recombinación. El sesgo hacia un mayor contenido en A+T es probablemente también un efecto de la desaparición de genes implicados en reparación que conduce a una presión mutacional de G+C a A+T. Además, el grado de ambos efectos, reducción del tamaño genómico y el incremento en el contenido en A+T de los endosimbiontes se correlaciona con el tiempo en que se produjo la asociación. Los S-endosimbiontes presentan mayor tamaño y menor porcentaje A+T, al tratarse de asociaciones más recientes en el tiempo, al contrario que los P-endosimbiontes, con tamaños muy reducidos y un marcado sesgo hacia A+T.

Como ya se ha comentado, el proceso de reducción genómica ha tenido lugar en dos etapas. Al comienzo de la relación simbiótica se produciría un proceso masivo de la reducción en el número de genes, probablemente mediante largas deleciones de genes continuos. Durante la segunda etapa, la reducción pudo haber ocurrido por una pérdida gradual de genes distribuidos a lo largo de todo el genoma (Moran, 2003). En lo referente a las regiones intergénicas, hay una reducción leve en cuatro especies de *B. aphidicola* cuando se comparan con su pariente *E. coli*. Sin embargo esta reducción es evidente en *Carsonella ruddii*, en la cual la pérdida es tan extrema que ha conducido al solapamiento de genes (Nakabachi *et al.*, 2006).

Aunque el establecimiento de la simbiosis conlleva una pérdida de genes, la bacteria simbiótica debe retener determinadas funciones que le permitan sobrevivir dentro del hospedador. Un ejemplo es la pérdida del gen *dnaA* en algunos endosimbiontes que podría indicar un control directo de la replicación del DNA en la población bacteriana por parte del hospedador.

Entre los genes conservados por todos los endosimbiontes secuenciados se encuentran genes implicados en funciones esenciales como la replicación del DNA, transcripción y traducción y que constituyen casi un tercio del tamaño genoma del endosimbionte. El sistema de chaperonas se mantiene también en todos los

genomas para asegurar un ensamblaje y localización adecuado de los componentes proteicos.

6. Replicación del DNA en bacterias

La replicación cromosómica tiene un papel esencial en la célula y posibilita otras funciones esenciales en la célula como la expresión génica y la división celular.

La maquinaria de replicación ha sido extensamente estudiada en la bacteria *E. coli*, organismo modelo para las bacterias Gram-, y *Bacillus subtilis*, organismo modelo para las bacterias Gram+. En este trabajo nos ceñiremos al aparato replicativo en las bacterias Gram-, y, en esencia, al de la bacteria *E. coli*, que nos servirá como organismo modelo de referencia para realizar estudios de genómica comparativa.

6.1. Inicio de la replicación

En *E. coli*, la replicación del DNA se inicia en el sitio específico *oriC* donde se une la forma activa de la proteína DnaA (ATP-DnaA) y permite la apertura local de la doble hélice de DNA. La helicasa DnaB puede acceder entonces al inicio de replicación, posibilitando la apertura del DNA, generando una región de cadena sencilla y permitiendo la entrada de la primasa, DnaG, que genera el cebador necesario para la entrada de la abrazadera de DNA del complejo replicativo β 2.

Una vez ensamblado el complejo de replicación en *oriC*, la progresión de la horquilla de replicación requiere la acción coordinada de cuatro actividades básicas que conforman el replisoma: proteína de unión a cadena sencilla (SSB), DNA Polimerasa, DNA helicasa y DNA primasa.

6.2. Elongación de la replicación

El replisoma se define como el complejo de proteínas necesarias para activar una horquilla de replicación y posibilitar que se lleve a cabo la síntesis de DNA de las cadenas continua y discontinua (Sandler, 2000). La polimerasa que lleva a cabo la síntesis principal del cromosoma de *E. coli* es la DNA polimerasa III, que forma parte de un complejo multiprotéico denominado holoenzima Pol III. La unión de la holoenzima Pol III al complejo cebador determina el final de la etapa de

iniciación y el comienzo de la elongación (Seufert y Messer, 1987). La holoenzima Pol III está compuesta por diez polipéptidos: el núcleo enzimático (subunidades α , ϵ y θ), la abrazadera deslizante β y el complejo γ (dnaX (γ - τ), cargador de la enzima en el molde. El complejo γ está formado por cinco subunidades que se unen con una estequiometría definida $\gamma, \delta, \delta', \chi, \psi$. Para la síntesis de las cadenas continua y discontinua se necesitan, al menos, dos polimerasas que se encuentran en la forma de dos núcleos enzimáticos que se mantienen unidos por la subunidad τ (McHenry, 1982). La estructura de la holoenzima Pol III en la horquilla de replicación se muestra en la Figura 2.

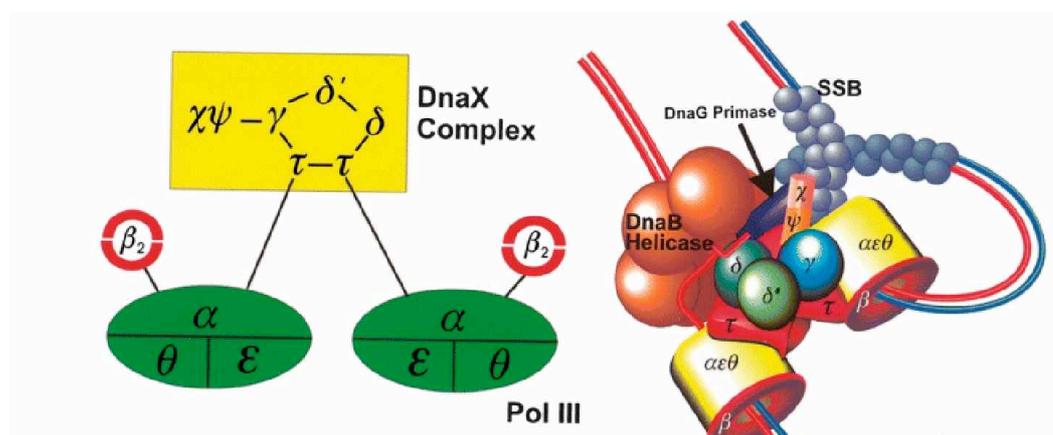


Figura 2. Esquema de la maquinaria de síntesis de DNA en la horquilla de replicación de la bacteria *E. coli* (tomado de Mc Henry, 2002).

Existe una clara conservación de la arquitectura general del complejo abrazadera a lo largo de todas las especies de Eubacterias, aunque la extensión del uso de la abrazadera por polimerasas implicadas en replicación, reparación o recombinación de la doble hélice difiere entre especies. La abrazadera β es necesaria para que el núcleo de la polimerasa aumente su procesividad. El complejo γ proporciona la energía necesaria para unir la abrazadera β al DNA mediante la hidrólisis de ATP. Una vez que la subunidad β es cargada, el complejo γ se separa del DNA dejando la abrazadera ensamblada, unida al núcleo de la Pol III y al DNA (Hingorani y O'Donnell, 1998).

Debido a la naturaleza antiparalela de la molécula de DNA, la elongación de las nuevas cadenas crea una asimetría en el punto de replicación por el hecho de que la DNA polimerasa sólo polimeriza nucleótidos en el sentido 5'→3'. De esta forma una de las cadenas es sintetizada de forma continua mientras que la complementaria se sintetiza de manera discontinua en fragmentos de unas 2 Kb, los llamados fragmentos de Okazaki. Las DNA polimerasas no pueden iniciar la síntesis de DNA *de novo* sin un cebador. La DNA primasa (DnaG) sintetiza pequeñas secuencias de RNA que son utilizadas para iniciar cada fragmento de Okazaki. Dado que un nuevo fragmento de Okazaki ha de iniciarse cada varios segundos, para asegurar que a esta velocidad la primasa tenga pleno acceso a las cadenas abiertas, se asocia al punto de replicación uniéndose a la helicasa DnaB (Tougu *et al.*, 1994). La proteína SSB es necesaria para mantener de forma de DNA de hebra simple las cadenas recién abiertas por la helicasa previo al paso de la horquilla de replicación (Glover y McHenry, 1998). Tras la síntesis de cada fragmento de Okazaki se eliminan los cebadores por la acción combinada de la RNasa HI y de la actividad exonucleasa-5'-3' de la DNA polimerasa I (Lehman y Uyemura, 1976). Esta polimerasa sintetiza DNA para rellenar el hueco generado. Finalmente, la DNA ligasa une los fragmentos contiguos de DNA.

Se ha propuesto un modelo de la elongación de la replicación en *E. coli* que permite la integración de todos los elementos implicados en esta etapa. Las horquillas de replicación estarían ancladas a la membrana y el DNA que se replica es desplazado a través de la maquinaria de replicación estacionaria o *factoría de replicación* (Lemon y Grossman, 1998; Koppes *et al.*, 1999; Sawitzke y Austin, 2001; den Blaauwen *et al.*, 2006). La factoría de replicación estaría formada por dos replisomas y permanecerían unidas por dos hexámeros de DnaB (den Blaauwen, 2006).

Aunque la holoenzima Pol III lleva a cabo la síntesis mayoritaria del DNA durante la replicación cromosómica, en *E. coli* se ha descrito la actividad de otras 4 polimerasas: Pol I (*polA*), Pol II (*polB*), Pol IV (*dinB*) y Pol V (*umuDC*).

La DNA polimerasa I (Pol I) es la polimerasa más abundante en *E. coli*, con unas 400 moléculas por célula. Esta enzima está constituida por dos dominios: (i) uno

combina la actividad polimerasa y la actividad exonucleasa 3', denominado fragmento Klenow y (ii) otro que incluye la actividad ExoII o exonucleasa 5'. Ambos dominios son funcionalmente independientes, aunque *in vivo* actúan de forma coordinada. Pol I presenta una procesividad muy limitada de 15 a 20 nucleótidos, que se correlaciona *in vivo* con la síntesis de pequeños fragmentos de DNA como son los que unen los fragmentos de Okazaki (revisado en Patel *et al.*, 2001).

Pol II forma parte de los sistemas de reparación del DNA. Estirpes mutantes en el gen *polB* no muestran defecto ni en el crecimiento ni en la replicación. La síntesis de Pol II es inducida en fase estacionaria, donde tanto el crecimiento como la replicación son procesos lentos y donde se incrementan los daños sobre el DNA, acumulándose pequeños huecos que pueden bloquear a Pol III. En estas circunstancias, Pol II soluciona el problema al reiniciar la síntesis de DNA en los pequeños huecos. Esta polimerasa tiene una tasa de error baja, pero su velocidad de síntesis de DNA es más lenta que la de la replicación normal. Cuando células de *E. coli* son expuestas a niveles elevados de radiación o a la presencia de ciertos mutágenos, se induce la respuesta SOS. Entre los genes inducidos se encuentran *dinB*, *umuC* y *umuD*, cuyos productos conforman la DNA polimerasa IV (Pol IV) y la DNA polimerasa V (Pol V). Ambas polimerasas reparan el DNA y su actividad es propensa a error al no presentar actividad exonucleasa 3' correctora de pruebas. Además, requiere las subunidades β y la subunidad γ de Pol III para aumentar su procesividad y alcanzar una actividad óptima.

Por último, la circularidad del cromosoma de *E. coli* impide que la molécula de DNA tenga libertad de giro a medida que la DnaB helicasa va abriendo la doble hélice. Esto supone un incremento del superenrollamiento positivo delante de la horquilla según avanza la replicación. En este paso es necesaria la acción de las topoisomerasas.

En *E. coli*, los niveles de superenrollamiento son mantenidos dentro de límites precisos durante la fase de crecimiento exponencial por la acción de cuatro topoisomerasas: girasa, topoisomerasa I (Topo I), topoisomerasa III (Topo III) y topoisomerasa IV (Topo IV). La girasa y Topo IV son topoisomerasas tipo II dependientes de ATP que introducen superenrollamiento negativo y eliminan superenrollamiento positivo respectivamente (Champoux, 2001). La Topo I (*topA*) es

una topoisomerasa tipo IA que elimina superenrollamiento negativo. Estas enzimas actúan conjuntamente para eliminar los cambios topológicos resultantes de la actuación de enzimas que se mueven sobre el DNA como la RNA polimerasa, manteniendo el superenrollamiento cromosómico en un nivel estacionario. Las topoisomerasas tipo I hidrolizan un enlace fosfodiéster en una cadena, hacen pasar la otra cadena a través del corte y vuelven a unir los extremos de la primera; como resultado, el índice de enlace de la molécula de DNA aumenta o disminuye en una unidad. Las topoisomerasas de tipo II hidrolizan sendos enlaces fosfodiéster en ambas cadenas y hacen pasar la otra cadena a través del corte, resellando éste de nuevo; de este modo, el índice de enlace varía en dos unidades de una sola vez.

El nivel de superenrollamiento regula la expresión de los genes de la girasa (*gyrA* y *gyrB*) y Topo I. La relajación del superenrollamiento induce la expresión de *gyrA* y *gyrB* y reprime la expresión de *topA* (Neumann y Quinones, 1997; Weinstein-Fischer y Altuvia, 2007).

6.3. Terminación

Cuando el complejo de replicación que se formó en *oriC* llega a la región diametralmente opuesta, la región del término, la replicación se detiene. Ésta es la etapa de terminación de la replicación. Si un punto de replicación alcanza la zona del término antes que el otro es frenado por los complejos Tus-*Ter*. Estos complejos se forman por la unión de la proteína Tus (*tus*) a las secuencias terminadoras *Ter*. Estas secuencias se han localizado tanto en el cromosoma como en algunos plásmidos. Estos complejos actúan sobre el replisoma bloqueando su avance de forma polar. Existen evidencias que Tus posee la actividad actividad contraria a la helicasa y que interacciona físicamente con DnaB (Neylon *et al.*, 2005).

Una vez que la replicación llega a su fin se produce la segregación de los cromosomas recién replicados, fenómeno que depende frecuentemente de procesos de recombinación. Los cromosomas hermanos se encuentran en forma de círculos concatenados y Topo IV es la enzima que lleva a cabo la decatenación de los mismos (Li *et al.*, 2002). En bacterias con cromosomas circulares, el entrecruzamiento por recombinación homóloga genera cromosomas diméricos que

no pueden ser segregados a las células hijas a menos que se conviertan en monómeros previamente a la división celular. Esto se consigue por el sistema de recombinación de sitio específico Xer altamente conservado en la evolución. La resolución de dímeros también requiere FtsK, una proteína localizada en el septo de división celular, que utiliza la energía generada por la hidrólisis de ATP para activar la reacción de resolución de dímeros (Sherratt *et al.*, 2004). La ausencia de proteínas de recombinación resulta en una segregación inadecuada y en fenómenos de inestabilidad cromosómica.

7. Evolución de los sistemas de replicación

La replicación del DNA es un mecanismo central para la vida celular. El estudio de la historia evolutiva de este proceso es especialmente relevante para la comprensión de los primeros eventos en la evolución de la vida celular. Las características comunes de todos los sistemas de replicación conocidos son: (i) la replicación es semi-conservativa; (ii) la replicación siempre se inicia a partir de orígenes definidos con la participación de un sistema de reconocimiento de orígenes; (iii) los cebadores de RNA son necesarios para iniciar la replicación; (vi) las nucleasas, polimerasas y ligasas reemplazan el cebador de RNA con DNA y sellan el hueco remanente (Baker and Bell, 1998; Kornberg and Baker, 1991).

Sin embargo, aunque se trata de procesos homólogos, la mayoría de las proteínas de replicación no son homólogas entre los dominios *Bacteria* y *Archaea/Eukarya*. Existen ortólogos dentro del mismo dominio, pero no existen ortólogos entre dominios sugiriendo un desplazamiento masivo que sucedió poco después de la separación de los dominios *Bacteria* y *Archaea/Eukarya*. Si la raíz del árbol universal se localiza en la rama *Bacteria*, como generalmente ha sido aceptado (Olsen and Woese, 1997), el desplazamiento por genes no ortólogos podría haber ocurrido en la rama bacteriana *Archaea/Eukarya*.

8. Sistemas de reparación y recombinación en procariontes

Los procesos de recombinación homóloga son esenciales para el mantenimiento de la integridad cromosómica y para la variabilidad genética. Sin estos mecanismos, la acumulación excesiva de lesiones conduciría a la

degradación total del genoma y la pérdida de información genética vital. La mayoría de los estudios han descrito la existencia de rutas de reparación muy reducidas y donde cada proceso se ha visto afectado de forma diferente por la reducción genómica.

Usando técnicas de genética comparativa, Rocha y Michel estudiaron la presencia de proteínas implicadas en recombinación homóloga en genomas bacterianos. En el grupo de bacterias que estudiaron, algunos genes como RecA y RecR están siempre presentes, sugiriendo que la recombinación homóloga está presente en la mayor parte de los genomas bacterianos. Sin embargo, el paso central de invasión e intercambio de hebra catalizado por RecA o un homólogo está ausente en la mayoría de genomas mínimos estudiados como *Buchera* o *Blochmania* (Rocha *et al.*, 2005).

Algunos grupos de proteínas pre-sinápticas como RecFOR son más frecuentes que las proteínas RecBCD/AddAB, dándose una co-ocurrencia entre estos sistemas y los sistemas de reparación de desapareamientos (*mismatch repair*) y SbcB, pero no la recombinación de extremos no homólogos (NHEJ) que parece ser raro en bacterias. Sin embargo, muchos genomas han perdido estos componentes pre-sinápticos. Aunque RecF, RecR, y RecO no forman *foci* que permitan deducir la existencia de una posible interacción, RecF y RecR a menudo si localizan en el cromosoma junto a proteínas de proteínas de replicación. Muchos genomas tienen operones cercanos al origen de replicación que contienen cuatro genes: *dnaA* (proteína de inicio de replicación), *dnaN* (subunidad β -clamp de la polimerasa III), *recF* y *gyrB* (DNA girasa).

9. Redes metabólicas mínimas

Comprender los procesos evolutivos que conducen a la reducción genómica requiere determinar la forma en la que tuvo lugar la pérdida de genes y consecuentemente proponer un posible mecanismo molecular.

La diversidad del contenido génico en bacterias refleja tanto la variación en las fuerzas selectivas como la pérdida selectiva de algunas rutas metabólicas. Las redes biológicas son de naturaleza robusta frente a la eliminación aleatoria de determinados nodos, es decir, no todos los genes tienen la misma probabilidad de

ser eliminados en un proceso de reducción genómica. El conocimiento de las redes genómicas mínimas es un paso previo a la biología sintética, de ahí la importancia de conocer los sistemas biológicos previo a la síntesis de una nueva función biológica. Recientemente se han publicado tres artículos científicos en la prestigiosa revista *Science* que proporcionan la primera idea de lo que sería una célula mínima, estudiando el microorganismo *Mycoplasma pneumoniae* revelando que el sistema es más complejo de lo que se pensaba. Al estudiar tanto el metaboloma como el proteoma, se ha encontrado que muchas moléculas son multifuncionales, pudiendo algunas enzimas catalizar reacciones múltiples, es decir, forman parte de más de un complejo proteico y que el transcriptoma de estos microorganismos puede tener una complicación elevada y funcionar de forma similar al Eucariota (Güell *et al.*, 2009; Kühner *et al.*, 2009; Yus *et al.*, 2009).

Materiales

El hecho de que los genomas de bacterias endosimbiontes hayan evolucionado por reducción génica ha sido utilizado por diferentes autores para identificar genes esenciales y realizar aproximaciones hacia la identificación del menor número mínimo de genes necesarios para el mantenimiento de la vida en un nicho determinado (revisado en Gil et al., 2004).

1. Grupo de genomas endosimbiontes utilizados

Como grupo de análisis hemos seleccionado un grupo de bacterias endosimbiontes caracterizadas por poseer un genoma de tamaño reducido. A día de hoy, a pesar de que hay numerosos proyectos de secuenciación de genomas microbianos, el número de genomas de endosimbiontes pertenecientes a la categoría Gamma-Proteobacteria completados asciende a 12 (cuando se completó este estudio se habían secuenciado dos genomas más de *B. aphidicola* y no se han considerado en esta memoria). El estudio de las especies de la clase Gamma-Proteobacteria resulta interesante al encontrarse en el mismo linaje que *E. coli* o *Salmonella*, organismos intensamente estudiados desde el punto de vista genético.

En este estudio hemos incluido además a algunos simbioses correspondientes a la categoría de Alpha-Proteobacteria (tres especies de *Wolbachia*), Epsilon-Proteobacterias (*Sulfurovorum*), y al genoma del primer simbionte facultativo de vida libre secuenciado: *Elusimicrobium minutum*, para comprobar si el proceso de conservación o erosión genómica es característico de un grupo bacteriano o es extensible al estilo de vida endosimbiontes. Hemos incluido también en el estudio los genomas de los endosimbiontes *Carsonella*, *Sulcia* y *Hodgkinia*, considerados en el límite entre célula y orgánulo, intentando responder a la pregunta de si el destino último de estos genomas es convertirse en orgánulos celulares especializados. En el caso de estos últimos, existen controversias en cuanto a que se traten de organismos vivos o por el contrario ya han entrado en la categorías de orgánulos celulares.

En la Tabla 1 se muestran los datos de la especie, el código de acceso del NCBI, el tamaño del genoma, la referencia bibliográfica y el número de proteínas que poseen los genomas seleccionados. Sólo se han incluido en este estudio los genes del cromosoma bacteriano y no los plasmídicos dado que no se han identificado genes del componente de replicación entre estos últimos.

Genoma	Referencia	GenBank	Length (Mbp)	Contenido G+C	Proteínas
<i>Buchnera aphidicola str. Cc</i>	NC_008513	CP000263	0.42	20.00%	357
<i>Buchnera aphidicola str. APS</i>	NC_002528	BA000003	0.64	26.00%	564
<i>Buchnera aphidicola str. Bp</i>	NC_004545	AE016826	0.61598	25.00%	504
<i>Buchnera aphidicola str. Sg</i>	NC_004061	AE013218	0.64145	25.00%	546
<i>Candidatus Blochmannia floridanus</i>	NC_005061	BX248583	4	25.00%	546
<i>Candidatus Blochmannia pennsylvanicus str. BPEN</i>	NC_007292	CP000016	0.71	27.00%	583
<i>Candidatus Carsonella ruddii PV</i>	NC_008512	AP009180	0.16	16.00%	182
<i>Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis</i>	NC_004344	BA000021	0.69772	22.00%	611
<i>Sodalis glossinidius str. 'morsitans' morsitans</i>	NC_007712	AP008232	4	22.00%	611
<i>Baumannia cicadellinicola str. Hc</i>	NC_007984	CP000238	4.17115	54.00%	2432
<i>Candidatus Hodgkinia cicadicola Dsem</i>	NC_012960	CP001226	0.69	33.00%	595
<i>Candidatus Sulcia muelleri GWSS</i>	NC_010118	CP000770	0.14	58.00%	169
<i>Candidatus Ruthia magnifica</i>	NC_008610	CP000488	0.25	22.00%	227
<i>Calyptogenia okutanii</i>	NC_009465	AP009247	1.16078	34.00%	976
<i>Elusimicrobium minutum Pei19</i>	NC_010644	CP001055	1.02215	31.00%	937
<i>Sulfurovum sp. NBC37-1</i>	NC_009663	AP009179	1.64356	39.00%	1529
			2.56228	43.00%	2438

Tabla 1. Características genómicas principales de los genomas seleccionados.

Según su modo de vida dichos organismos podrían clasificarse como:

- Simbiontes obligados (Bacterioma-Asociados):
- Simbiontes facultativos-manipuladores de la reproducción (*Wolbachia sp.*)².

Herramientas bioinformáticas utilizadas para la determinación de proteínas ortólogas en genomas de bacterias endosimbiontes de insectos

Para la realización de este estudio hemos realizado una aproximación de genómica comparativa. Para ello utilizamos distintas herramientas bioinformática que nos han permitido identificar el conjunto de genes compartidos entre los distintos genomas, que serían necesarios para realizar funciones básicas de transferencia de información en genomas mínimos de endosimbiontes de bacterias. La identificación de ortólogos es una herramienta poderosa para la anotación de genomas, estudios evolutivos de genes/proteínas, genómica comparativa e identificación de secuencias y asignación a taxones.

2.1. Herramientas para el análisis comparativo de los genomas mínimos

Para la búsqueda de los genes y proteínas ortólogos hemos utilizado las bases de datos que se indican a continuación. Con estos datos se han construido las tablas que figuran en esta memoria.

-**COG** ("Clusters of Orthologous Groups of Proteins") (<http://www.ncbi.nlm.nih.gov/COG/>). Base de datos del NCBI que contiene unas 2800 familias de proteínas conservadas o grupos de proteínas que han evolucionado a partir de un mismo ancestro –ortólogos–.

-**MBDG** ("Microbial Genome Database for Comparative Analysis") (<http://mbgd.genome.ad.jp>). MBDG es una base de datos para el análisis comparativo de genomas de microorganismos completamente secuenciados. Permite, entre otros, la identificación de ortólogos, parálogos, análisis de motivos proteicos y comparación del orden génico.

-**KEGG** (<http://www.genome.ad.jp/kegg2.html>). La Enciclopedia Kyoto de Genes y Genomas permite comparar genomas en cuanto a su capacidad de codificar diferentes rutas metabólicas. También permite identificar grupos –clusters– de genes conservados entre dos especies.

-**Integr8** (<http://www.ebi.ac.uk/integr8/>). El portal web Integr8 proporciona un acceso rápido a información almacenada de genomas y proteomas. Los datos disponibles incluyen secuencias de DNA (de las bases de datos del EMBL *Nucleotide Sequence Database*, *Genome Reviews* y *Ensembl*); secuencias de

proteínas (bases de datos UniProt e IPI); estadísticas de análisis de Genomas y Proteomas (obtenidas usando InterPro, CluSTr, y GOA); así como información sobre ortología, paralogía y sintenia.

Para la **asignación de dominios a secuencias de proteínas** se utilizaron las siguientes bases de datos:

- **ExPASy** (for "Expert Protein Analysis System"). <http://us.expasy.org>. Servidor de Proteómica para el análisis de secuencias y estructuras.

- **Pfam** (for "Protein Families Database of Alignments and HMMs") (<http://www.sanger.ac.uk/Software/Pfam>)- Base de datos de familias de proteínas disponibles a partir de alineamientos múltiples de secuencias y modelos ocultos de Markov (HMMs).

También se han tomado en consideración los datos aportados por diferentes estudios que han sido publicados durante el tiempo de realización del presente trabajo:

- **BBHs** tomados del trabajo *Choosing BLAST options for better detection of orthologs as reciprocal best hits* (Gabriel Moreno-Hagelsieb y Kristen Latimer, 2008).

-Así como los datos recogidos en Gil *et al.*, 2004.

- Para asignar funciones a los genes seleccionados en *E. coli* se utilizó el Entry Point disponible en el servidor de la Universidad de Berkeley: http://coli.berkeley.edu/cgi-bin/ecoli/coli_entry.pl

Las herramientas utilizadas para la identificación de **dominios conservados** en genomas de endosimbiontes fueron:

- Alineamiento de secuencias con **ClustalW** (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>).

Se usaron los parámetros por defecto presentes en el formulario online. Para establecer el formato de los alineamientos múltiples se utilizó el programa Jalview,

realizándose tomas de pantalla para salvar los datos. Con los resultados del alineamiento se construyeron los árboles filogenéticos en los casos en los que fue necesario.

```
Default parameters:  
When "def" values are used, we let ClustalW (1.82)  
use its own default values:  
DNA Gap Open Penalty = 15.0  
DNA Gap Extension Penalty = 6.66  
DNA Matrix = Identity  
Protein Gap Open Penalty = 10.0  
Protein Gap Extension Penalty = 0.2  
Protein matrix = Gonnet  
Protein/DNA ENDGAP = -1  
Protein/DNA GAPDIST = 4
```

Figura 3. Parámetros por defecto para la realización de alineamientos múltiples de secuencia.

2.2. Herramientas utilizadas para la validación de ortólogos:

Para la asignación de genes ortólogos se han tenido en cuenta algunos criterios adicionales:

- Perfiles Filogenéticos: genes con igual patrón de presencia/ausencia en varios genomas.
- Conservación del contexto o vecindario génico (Sintenia)
- Alineamiento múltiple de familias de proteínas para identificar mutaciones compensatorias. Conservación de la arquitectura de dominios y el plegamiento de la estructura en el espacio tridimensional (fold).

Métodos

Existe un intenso debate respecto en torno a la asignación de genes ortólogos. Los ortólogos son aquellos genes que provienen de un mismo gen (son homólogos) y cuya divergencia se debe a un proceso de especiación. En palabras más sencillas, son ortólogos aquellos genes (o proteínas) que tienen la misma identidad en distintas especies.

En este trabajo se plantea la búsqueda de genes ortólogos en genomas de bacterias endosimbiontes. Posteriormente estos genomas se compararán con el de la bacteria de vida libre, *E. coli*, puesto que está emparentada filogenéticamente con la mayor parte de organismos endosimbiontes de insectos conocidos, las Gamma-Proteobacterias y por ser el modelo bacteriano del que más conocimiento genético tenemos. En el caso de las estirpes endosimbiontes los problemas principales que se han planteado a lo largo de este trabajo han sido, entre otras, la presencia de anotaciones incompletas para algunos genomas, la diferencia de información según la base de datos analizada y la diversidad de algoritmos existentes para la determinación de ortólogos según el tipo de aproximación utilizada para su estudio dependiendo de la base de datos analizada, aparezcan o no anotadas como ortólogos.

Por todo lo expuesto, en este trabajo hemos tratado de realizar una metodología homogénea, sin llegar a ser sistemática dada la complejidad de las anotaciones, considerando toda o la mayor parte de la información genómica disponible hasta la fecha. Este sistema ha dado lugar a un esquema de trabajo que se recoge en la figura 4.

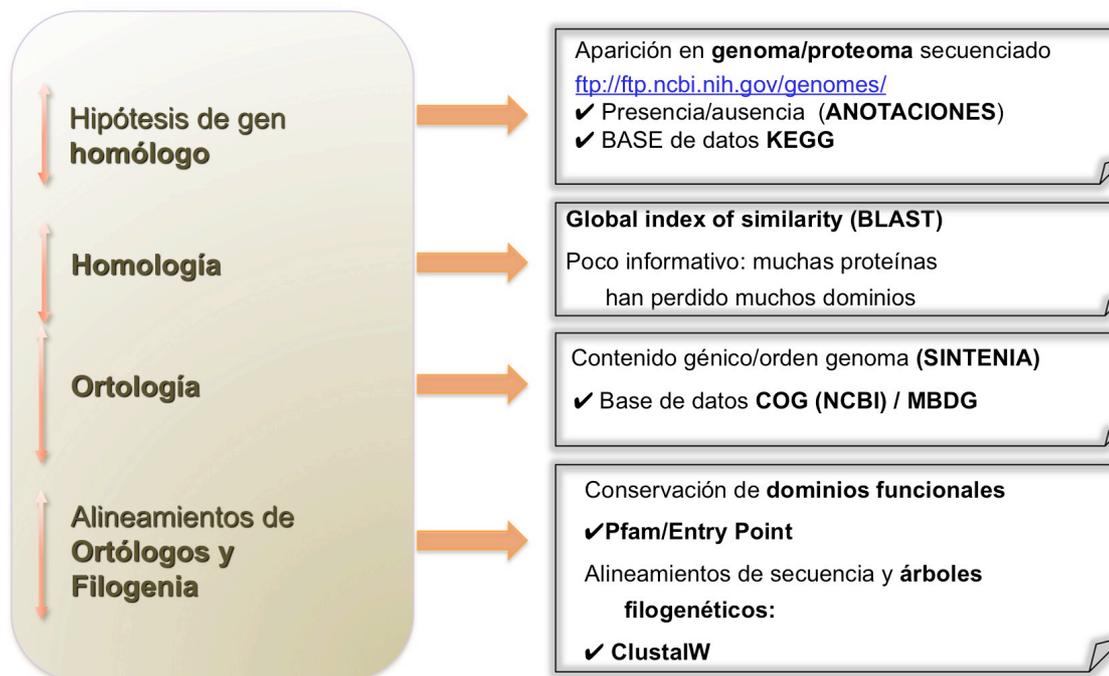


Figura 4. Esquema de trabajo planteado para la asignación de genes ortólogos.

El primer paso se basa en el análisis manual de las anotaciones de los genomas seleccionados. Durante el desarrollo del trabajo, dado que se iban secuenciando nuevos genomas de endosimbiontes se iban anotando, se decidió ir incorporándolos al esquema general del trabajo. Los genomas se obtuvieron a partir del repositorio <ftp://ftp.ncbi.nih.gov>. Como referencia para la anotación de posibles ortólogos se utilizaron las rutas metabólicas recogidas en KEGG, así como las descritas en el libro Friedberg *et al.*, 2006.

Una vez seleccionado el grupo inicial de “posibles” ortólogos se hizo un análisis de homología utilizando el algoritmo BLAST como en la consulta en las bases de datos sobre cuáles de los genes de la clase replicación, reparación y recombinación habían sido identificados como verdaderos ortólogos. Hay que señalar que para muchos de los genomas recogidos no existe esta información por lo que este último análisis no se pudo realizar en muchos casos. Las bases de datos utilizadas fueron COG (NCBI) y sobre todo la base de datos de microorganismos MGDB.

Se ha descrito que muchos de los genes que codifican para proteínas implicadas en replicación forman parte de operones definidos como es el caso de *dnaA*, que se encuentra en el mismo operón de *dnaN* y *recF* (Macian *et al.*, 1994). Se examinó la sintenia en algunos casos especiales, como en *dnaA*, *dnaN* ó *dnaX*, para ver si

conservaban el vecindario génico y, por tanto, reflejar una función y evolución común. Para el análisis de la sintenia se utilizó la base de datos MGDB.

Algunas proteínas, como el ortólogo de *nrdA* de *C. ruddii* fueron añadidos a la lista al comprobar la presencia del dominio funcional de la proteína. Para el análisis de presencia/ausencia de determinados dominios se utilizó el servidor Pfam.

Por último, y en el caso de la proteína NrdA que aparecía representada en la mayor parte de los grupos analizados y con el propósito de estudiar si la misma podría tratarse de un fósil molecular, se realizaron árboles filogenéticos para determinar su filogenia.

Perspectiva actual y objetivos

El concepto de **Genoma Mínimo** ha estado siempre unido intrínsecamente a la de esencialidad de los genes en las bacterias, tomando como base que los genes esenciales para el metabolismo celular serían los mejor conservados.

La primera aproximación bioinformática al estudio del set mínimo de genes fue realizada por Mushegian y Koonin en 1996 buscando los genes conservados entre los tres genomas completados en esa época (*H. influenzae*, *M. Genitalium* y *E. coli*). El resultado inicial fue un set de 256 genes (Mushegian y Koonin, 1996).

Posteriormente, se determinó que en *Bacillus subtilis* (Gram+) con un genoma de 4.2 Mb tendría 271 genes esenciales y que 150 de estos genes son compartidos con *E. coli* (Gram-). Sin embargo, 83 genes esenciales en *E. coli* no están presentes en *Bacillus subtilis* por lo que es difícil asignar un genoma mínimo, incluso dentro de las Eubacterias (Kobayashi, 2003).

Esta complejidad nos ha llevado a concentrar nuestro estudio en el grupo de bacterias Gram-, centrándonos en bacterias endosimbiontes. Otros estudios siguiendo un enfoque similar al seguido en este trabajo, pero considerando un menor número de genomas, dispusieron que 5 endosimbiontes compartían 277 proteínas ortólogas. Cuando éstos se comparan con el genoma de *M. genitalium*, se reducen a 180 genes constitutivos y a 156 cuando se compara con los parásitos *Clamidia tracomatis* y *Rickettsia prowazekii* (Sakharkar et al., 2004).

Gil et al., 2004 proponen un genoma mínimo compuesto por 206 genes codificantes para proteínas necesarias para el mantenimiento del genoma y la reproducción, metabolismo de nutrientes esenciales, en un sistema en ausencia de estrés ambiental. Proponen también cómo estarían distribuidas estas funciones en una hipotética célula mínima.

Belda et al., 2005, mostraron cómo 30 genomas de gamma-proteobacterias secuenciados en esa fecha comparten 244 proteínas ortólogas (Belda et al., 2005). Posteriormente Merhej et al., 2009 registraron la pérdida de los mismos 100 genes ortólogos en todas las bacterias endosimbiontes obligados, indicando una evolución selectiva de determinados grupos de genes.

El componente 3R está integrado por los genes implicados en procesos de replicación, recombinación y reparación y se encuentran interconectados funcionalmente y constituyen la misma categoría COG: la categoría L. *Escherichia coli* presenta 237 genes en la categoría COG L (Replicación, Recombinación y Reparación). Muchos de los componentes de este conjunto de genes resultan esenciales para la vida del microorganismo. En el caso de endosimbiontes y parásitos con genomas reducidos la definición de gen esencial cambia, ya que el microorganismo adquiere algunas funciones del hospedador al que le compensa con la síntesis de algún aminoácido esencial para el metabolismo del mismo. De esta forma, el estudio de la composición del sistema 3R en estas bacterias resulta de gran importancia para el estudio de rutas metabólicas mínimas, especialmente en situaciones de endosimbiosis donde la situación de estrés metabólico se reduce de forma muy importante.

La pregunta principal formulada al comienzo de este trabajo se corresponde precisamente con este planteamiento, el análisis del componente mínimo 3R en bacterias endosimbiontes. A partir de estos datos se pretende proponer modelos que expliquen cómo estos microorganismos han compensado la pérdida de determinados genes y presentar una replicación genómica y transmisión horizontal del genoma, que a día de hoy está casi en equilibrio y al límite de la reducción genómica entre organismo vivo-orgánulo celular.

Éste tema ha sido objeto de otros trabajos y revisiones frecuentes en los últimos años (revisado en Moran et al., 2009). De hecho, mientras se escribía esta memoria, se ha publicado un trabajo con objetivos similares a los planteados aquí (Sharples, 2009).

Llegados a este punto, con el objeto de iniciar esta investigación se plantearon dos cuestiones:

- ¿Cuál es el conjunto mínimo de genes requeridos para la replicación/conservación del genoma en organismos endosimbiontes que han sufrido reducción genómica?
- Y de este conjunto, ¿cuántos de estos genes pertenecen al componente básico 3R (Replicación, Recombinación y Reparación)?

Resultados

1. El número de componentes del sistema 3R depende del tamaño del genoma del endosimbionte

En primer lugar se llevó a cabo un recuento de los genes presentes en estos genomas y que se encuadrarían dentro del sistema 3R (Figura 5). Estos datos se representaron frente al tamaño del genoma de las bacterias elegidas para este estudio y que se recogen en el apartado de Métodos.

En la figura 5A, se observa que los genomas con menor tamaño presentan menor número de genes 3R y viceversa. Los genomas de los organismos *Carsonella*, *Sulcia* y *Hodgkinia*, que se encuentran en la frontera organismo-orgánulo presentan un número muy limitado de genes, en torno a 10, teniendo muy limitadas la replicación del genoma y casi inexistentes las capacidades de recombinación y reparación del DNA. En el caso de la Epsilon-Proteobacteria *Sulfurovum*, a pesar de tener un genoma de tamaño muy superior al de otros endosimbiontes, el número de genes se sitúa en torno a 30, diferenciándose claramente dos grupos. Esto implica que probablemente la conservación/pérdida de este grupo de genes depende del estadio de reducción genómica en el que se encuentre el microorganismo.

En las figuras 5B, 5C y 5D podemos ver la representación del número de genes componentes del sistema 3R con respecto al tamaño del genoma. En los grupos de replicación y recombinación, la distribución parece similar a la del total del componente 3R pero muy diferente, sin embargo, en el grupo correspondiente a genes de reparación. Este grupo parece haber sufrido el mayor proceso de reducción genómica dentro del componente 3R, con 10 genes como máximo en algunos organismos (figura 5D).

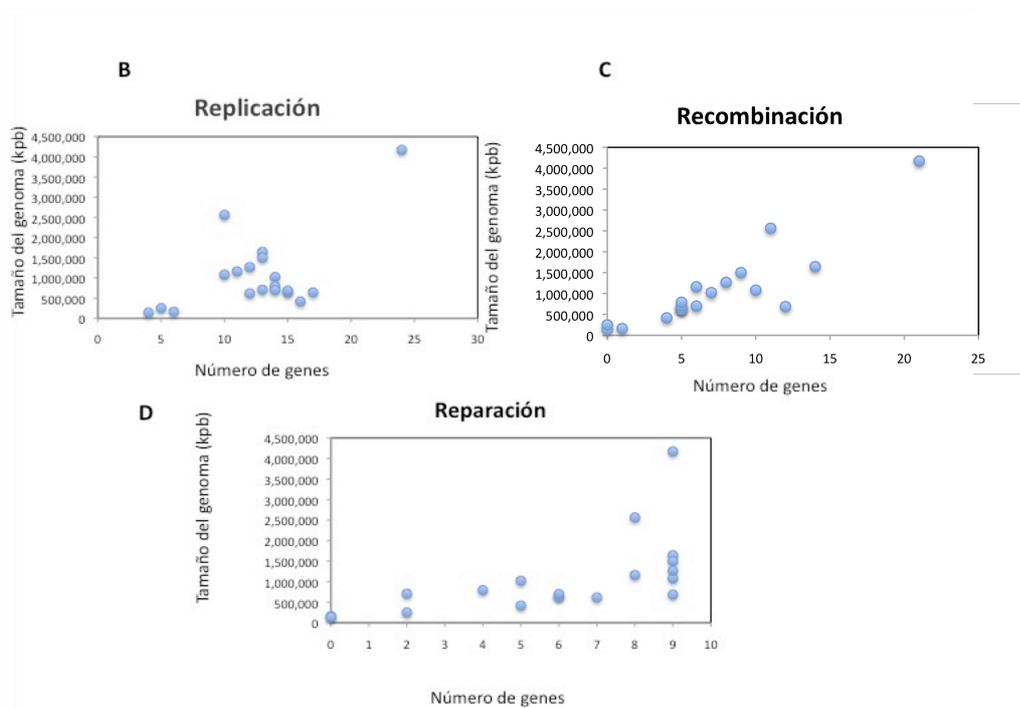
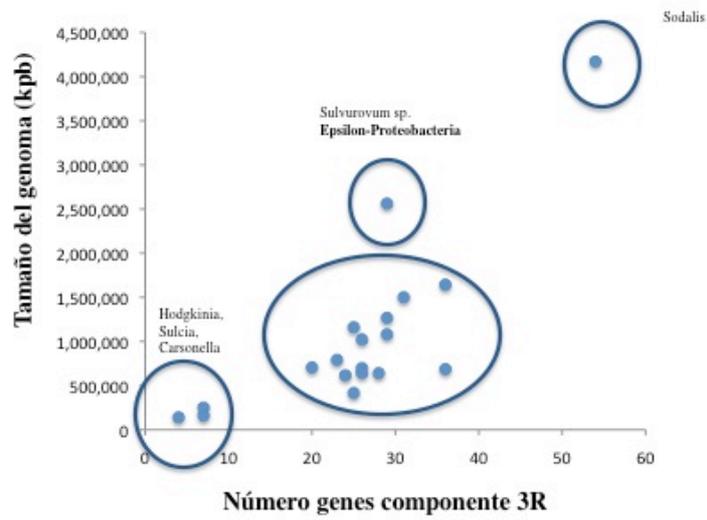


Figura 5. Número de genes del componente 3R frente al tamaño del genoma. La distribución se conserva en el caso de la replicación y recombinación, pero no en el caso de la reparación que parece haber sufrido un proceso mayor de pérdida de número de genes diferente.

2. Conjunto mínimo de proteínas implicadas en procesos de replicación del DNA

2.1. Maquinaria de inicio de replicación

En Eubacteria, la región del origen de replicación (*oriC*), es relativamente pequeña, de 100 a 1000pb. En γ -Proteobacteria, la región del origen de replicación se localiza frecuentemente dentro en el cluster de genes *mpA-rmpH-dnaA-dnaN-recF-gyrB*, normalmente próximo a *dnaA* (Messer, 2002).

Esta proteína muestra una representación bastante ubicua dentro del dominio Eubacteria. Sin embargo, se ha visto que los géneros *Blochmannia*, *Baumannia* y *Wigglesworthia* no poseen ortólogo de la proteína *dnaA* (Tabla 3). Para explicar cómo estas bacterias podrían solventar la pérdida de la proteína DnaA podremos recurrir al comportamiento que presenta *E. coli* en ausencia de esta proteína. Se ha propuesto que en *E. coli* se podría aliviar la ausencia de DnaA mediante un mecanismo de **Replicación dependiente de Recombinación**, descrito como mecanismo alternativo en determinadas situaciones (Kogoma, 1997). Este mecanismo presenta una dependencia de la proteína RecA que sí está presente en algunos de estos genomas. El inicio de la replicación en algunos géneros de endosimbiontes por un mecanismo de inicio de replicación RecA-dependiente podría estar apoyado por la presencia de proteína de recombinación RecA en *Wigglesworthia* y *Baumannia*.

Es de destacar la presencia de la helicasa DnaB la presencia de esta proteína en la mayor parte de los genomas analizados (sólo ausente en *Hogkinia* y *Sulcia*).

En el caso del género *Buchnera*, donde el gen *dnaA* aparece en todas las especies secuenciadas hasta la fecha, el inicio de la replicación podría ser explicado, al igual que en el caso de *E. coli*, por un mecanismo de inicio DnaA-dependiente que permitiría la entrada de DnaB a la horquilla de replicación. Para comprobar esta hipótesis se realizó un alineamiento múltiple con las secuencias de la proteína DnaA de *E. coli* frente a la de *B. aphidicola*, para ver el grado de conservación del dominio de interacción con DnaB (Figura 6).

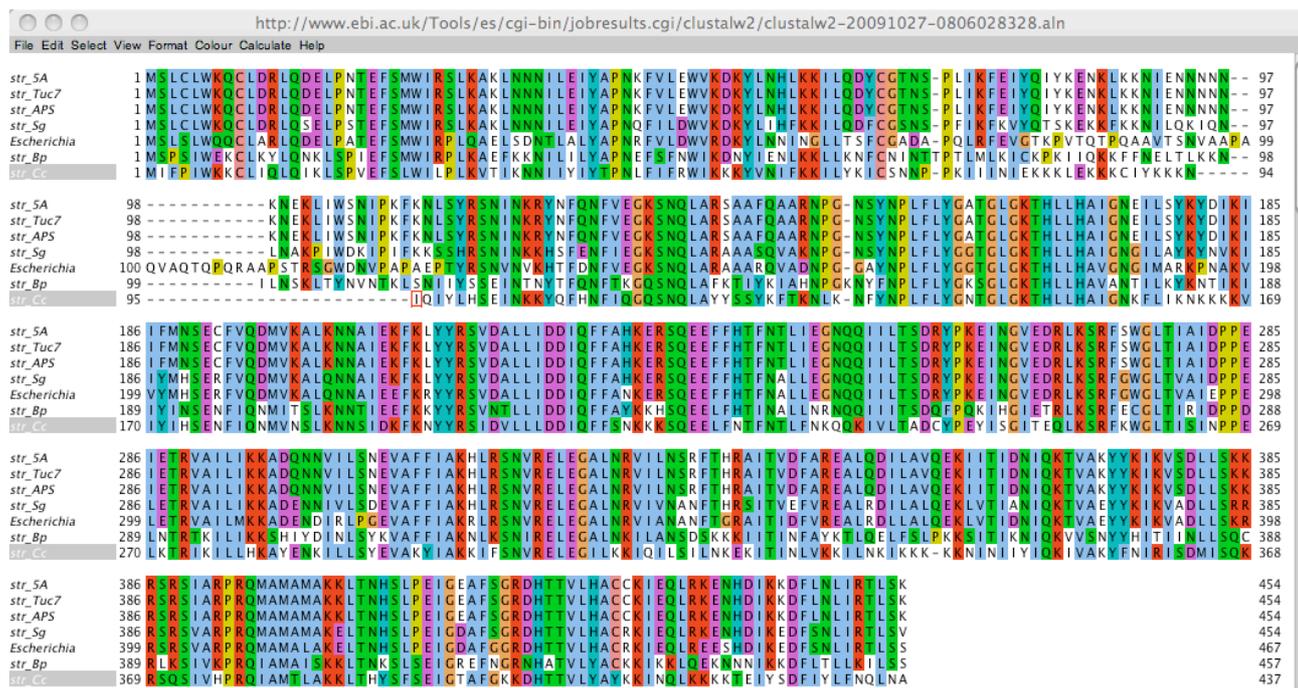


Figura 6. Alineamiento múltiple de secuencia de la proteína DnaA en los siete genomas secuenciados hasta la fecha dentro del género *Buchnera*.

Como se observa en la Fig. 6, la proteína DnaA está altamente conservada excepto en la región de los aminoácidos 80-110 aproximadamente. Esta zona se corresponde con el dominio funcional 2 de DnaA que comprende los aminoácidos 87–134, y constituye un bucle flexible muy variable en secuencia y poco conservado evolutivamente (Messer *et al.*, 1999). La región de interacción con DnaB está perfectamente conservada, por lo que el mecanismo de inicio de la replicación en *Buchnera* debe ser muy similar al que tiene lugar en *E. coli* (apartado 6 de Introducción).

En *E. coli*, además de la interacción con DnaA, es necesaria la presencia de un cargador de la helicasa, la proteína DnaC. Sin embargo, la proteína DnaC sólo está presente en *Buchnera* y *Sodalis*, por lo que parece ser un factor accesorio en la maquinaria de replicación. Se ha propuesto que DnaC es un parálogo cercano de DnaA, y de forma sorprendente conserva la capacidad de formar un ensamblado helicoidal de forma similar a ATP-DnaA (Mott *et al.*, 2008).

La función principal de DnaC es permitir cargar la helicasa, por lo que su ausencia implicaría la existencia de un mecanismo de carga algo diferente al de *E. coli*, como debe suceder en *Wolbachia*, *Ruthia* y *Vesicomysocius*, que presentan

DnaA, pero no DnaC, donde la interacción directa de DnaB con DnaA podría ser suficiente para la entrada a la horquilla de replicación. En el caso de DnaC de *B. aphidicola*, el grado de conservación de la secuencia es también muy alto, lo cual sugiere un inicio de replicación similar a *E. coli* donde DnaB realizaría la misma función de cargador de la helicasa (Figura 7).

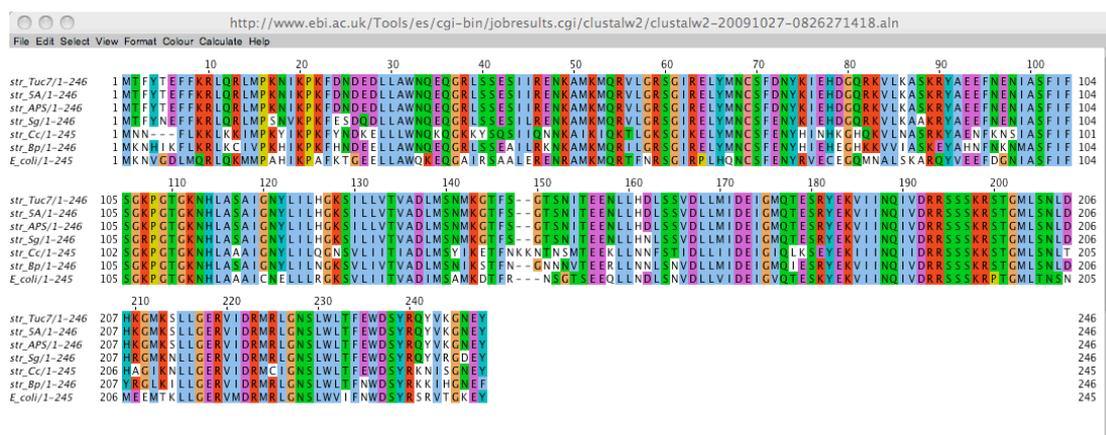


Figura 7. Alineamiento múltiple de secuencias de la proteína DnaC en *E. coli* y en los siete genomas secuenciados del género *Buchnera*.

El cuadro resumen que explicaría las maquinarias de inicio de replicación en estos sistemas se muestra en la Tabla 2.

Bacteria	Sistema inicio replicación
<i>Ruthia magnifica</i>	DnaA/DnaB
<i>Vesicomycosocius okutanii</i> HA	DnaA/DnaB
<i>Elusimicrobium minutum</i>	DnaA/DnaB
<i>Sulfurovum</i> sp. NBC37-1	DnaA/DnaB
<i>Wolbachia pipientis</i> Culex	
<i>quinquefasciatus</i> Pel wPip	DnaA/DnaB
<i>Wolbachia pipientis</i> brujia malayi	DnaA/DnaB
<i>Wolbachia pipientis</i> Drosophila	
<i>melanogaster</i>	DnaA/DnaB
<i>Buchnera aphidicola</i> Ap	DnaA/DnaB-C
<i>Buchnera aphidicola</i> Bp	DnaA/DnaB-C
<i>Buchnera aphidicola</i> Cc	DnaA/DnaB-C
<i>Buchnera aphidicola</i> Sg	DnaA/DnaB-C
<i>Sodalis glossinidius</i>	DnaA/DnaB-C
<i>Blochmannia floridanus</i>	ni DnaA ni RecA RecBCD- dependiente?
<i>Blochmannia pennsylvanicus</i>	ni DnaA ni RecA RecBCD- dependiente?
<i>Candidatus Hodgkinia cicadicola</i>	
<i>Dsem</i>	ni DnaA ni RecA
<i>Sulcia muelleri</i>	ni DnaA ni RecA
<i>Baumannia cicadellinicola</i>	RecA-dependiente
<i>Wigglesworthia glossindia</i>	RecA-dependiente
<i>Candidatus Carsonella ruddii</i> PV	RecA-dependiente

Tabla 2. Resumen de los mecanismos de inicio de replicación propuestos para las diferentes especies de bacterias en función de los genes y dominios proteicos conservados.

La **síntesis de un cebador** para el inicio de la replicación parece ser que es mediada por DnaG en *Buchnera* al igual que en *E. coli*. Se ha localizado un gen ortólogo de *dnaG* en todos los genomas secuenciados excepto en *Sulcia* y *Hodgkinia*. Los dominios Toprim presentes en DnaG se han conservados evolutivamente (hemos identificado una enzima con actividad Toprim en el genoma de *Carsonella*. Tabla 3). Estos dominios están presentes en la familia RecR/M (topoisomerasea-primase) que es el dominio catalítico en las topoisomerasas tipo IA y II (tipo DnaG). Se ha propuesto que el ancestro común de todas las formas de vida podría poseer una enzima tipo Toprim con funciones nucleotidil transferasa y corte de cadenas polinucleótídicas (Gil et al., 2004). DnaG está prácticamente conservada en todos los genomas. En *C. ruddii*, la proteína homóloga a DnaG que

presenta un dominio TOPRIM (CRP_010) está bastante degenerada, al igual que sucede con otras proteínas en este simbionte (Tamames et al., 2007).

El ensamblaje de horquillas de replicación de forma **independiente a DnaA** se ha descrito en procesos de restauración de horquillas de replicación paradas en *E. coli* (Sandler, 2005). En estos procesos, son necesarias proteínas accesorias para el reensamblaje de un nuevo replisoma. Este proceso lo llevan a cabo los factores de ensamblaje del primosoma PriA/PriB/DnaT o PriC, junto con DnaC, en el caso de *E. coli*. Este complejo queda reducido a PriA en la mayoría de las especies, si bien, no está presente ni en *Blochmania* ni *Wigglesworthia*, ni tampoco en algunas estirpes de *Buchnera*. De hecho, en *Blochmania* no existe ni DnaA ni PriA. DnaT, que codifica la proteína que sería responsable de la carga de la helicasa replicativa junto a DnaC sólo está presente en aquellas estirpes de *Buchnera* que conservan PriA, y además en *Baumannia*. Podemos decir que esta ruta ha sufrido un proceso de evolución diferente según los linajes. Tanto DnaT como DnaC parecen ser adquirida del antecesor común pero ambas proteínas presentan similitud con otras de bacteriófagos. Sin embargo, los genomas de endosimbiontes parecen haber perdido muchas de las proteínas de origen vírico, manteniendo la maquinaria estricta para llevar a cabo los procesos de transferencia de la información, de ahí que estas proteínas estén ausentes en la mayoría de genomas.

La proteína PriB está ausente en muchos genomas. Los análisis de secuencia detallados han mostrado que la organización de PriB evoluciona a partir de la proteína de unión a cadena sencilla SSB. Se ha propuesto que ambas proteínas podrían ser resultantes de un mecanismo de duplicación génica en el ancestro común (Ponomarev et al., 2003). SSB tiene una representación constante en todos los genomas y esto podría explicarse por un papel predominante en estas vías y otras relacionadas con mecanismos de recombinación y reparación de DNA, donde la protección del ssDNA es necesaria dada la ubicuidad de un gran número de nucleasas celulares que han sido conservadas a lo largo de la evolución.

Los **mecanismos de replicación a partir de R-loops** han sido descritos en presencia de mutaciones del gen *rnhA*. RNasa HI es una proteína bastante conservada en todos los genomas, por lo que este tipo de intermediarios deben formarse asiduamente. Debido a esto, no debe descartarse el papel que pudiera

jugar la RNA Polimerasa en el inicio de la replicación incluso de mayor importancia al desarrollado en *E. coli* (Skarstad, 2009).

Respecto al **control de la replicación**, el inicio de la replicación es un proceso altamente regulado en procariotas. En *E. coli*, la replicación necesariamente ocurre una vez por ciclo celular y esto es garantizado por múltiples mecanismos, como el secuestro del origen de replicación mediado por SeqA, la represión autógena o la titulación de la proteína DnaA. Proteínas como SeqA o HdaA no están presentes en estos genomas indicando una pérdida casi total del control que la replicación ejerce sobre el ciclo celular en casi todos estos genomas (Revisado en Messer, 2002). Sin embargo, la presencia de orígenes de replicación, similares al de *E. coli*, en bacterias endosimbiontes no parece del todo clara.

En el caso de *Buchnera* que parece tener una maquinaria de replicación bastante conservada, el origen de replicación presenta una sola caja de interacción con DnaA, por lo que su función podría verse limitada como secuencia iniciadora. La presencia de un genoma rico en A+T podría aliviar la ausencia de secuencias de reconocimiento específicas.

Otro de los indicadores de la pérdida de control del ciclo celular sobre la replicación de estos genomas es la presencia de **poliploidía** en algunos de estos genomas. *B. aphidicola* presenta una media de 120 copias de su cromosoma, aunque depende del desarrollo del hospedador y de la demanda de moléculas esenciales proporcionadas por el microorganismo (Komaki y Ishikawa, 2000).

2.2. Elongación de la replicación: la Holoenzima Pol III

Con la excepción de unas cuantas subunidades accesorias (χ , ψ , θ), y excluyendo a *Carsonella*, *Ruthia* y *Hogdkinia*, la mayor parte de los genomas analizados poseen una Holoenzima DNA polimerasa III completa. El esquema de la polimerasa replicativa procesiva de *E. coli* se resume en la figura 2 de Introducción.

Resulta llamativa la estructura del cargador de la polimerasa (γ - τ). La estructura de los cargadores de polimerasas son homólogos en los tres dominios **Bacteria/Archea/Eucaria**. En *E. coli*, el cargador consta de las subunidades τ , γ , δ , δ' que son homólogas a las proteínas RFC de Archeas o Eucariotas. En *E. coli*, a

partir del gen *dnaX* se sintetizan las unidades τ y γ . La proteína τ (71 kDa) es la proteína completa, mientras que γ es una versión truncada de 47 kDa debido a un cambio en la transcripción que hace que se sobrepase el codón de stop. El C-terminal de τ tiene una extensión extra de 24 kDa lo que permite conectar a DnaB con las polimerasas (Gao y McHenry, 2001). Se ha propuesto que esta extensión C-terminal es una estrategia de estas proteínas para forzar a los componentes de la horquilla replicativa a interactuar entre ellos. Sin embargo, éste extremo C-terminal no está presente en todos los grupos. En el caso de *Buchnera*, el gen *dnaX* tiene una extensión menor y codifica una sola proteína de menor tamaño, disminuyendo probablemente la eficiencia del sistema de ensamblaje de horquillas de replicación. De hecho, se ha visto en *E. coli* que la región C-terminal no es estrictamente necesaria para la carga de abrazadera, pero si es esencial para la viabilidad celular (Blinkova et al., 1993), debido probablemente a su capacidad para organizar el replisoma.

2.3. Terminación de la replicación

Al igual que para la región del inicio (*oriC*), diferentes estudios han tratado de caracterizar mediante diferentes herramientas computacionales evidencias de sitios de terminación (*terC*) en genomas procariontes (Sernova et al., 2008). En el pasado, el papel primordial de la terminación fue asignado a los sitios *Ter*, sin embargo, recientemente se ha propuesto que la terminación debe ocurrir aproximadamente una kilobase antes de dichos sitios *Ter*, en la proximidad de los sitios *dif* (Hendrickson H et al., 2007). La proteína Tus, responsable de unión a *Ter* no ha sido identificada en ninguno de los genomas de endosimbiontes estudiados en este trabajo.

3. Topología del DNA

El estado de la superhélice es esencial para el mantenimiento de muchas funciones esenciales que tienen lugar en el cromosoma. Un aspecto fundamental es el ensamblaje y el grado de compactación del DNA cromosómico, resultado del superenrollamiento negativo impuesto sobre el DNA (Holmes y Cozzarelli, 2000). El grado de superenrollamiento es crítico para el inicio de determinados procesos como la replicación o la transcripción (Pruss y Drlica, 1989). La replicación, por sí

misma, causa también relajación y aumento en el grado de superenrollamiento negativo. Para la conservación del nivel de superenrollamiento adecuado es necesario la presencia de Topoisomerasas de DNA.

3.1. Topología del DNA en genomas mínimos

Los genes que codifican para la girasa, *gyrA* y *gyrB*, están presentes en todos los organismos analizados en este estudio. Los genes *parC* y *parE* (Topo IV) no se encuentran presentes en ninguno de los endosimbiontes analizados, mientras que *topA* (Topo I) sólo está presente en dos estirpes de *B. aphidicola* (Tabla 4). La girasa es la encargada de restaurar el estado de superenrollamiento óptimo del DNA tras el paso, por ejemplo, de la horquilla de replicación. De acuerdo al modelo "twin-supercoiled-domain", se genera superenrollamiento positivo (SE+) en la zona delantera de la RNA polimerasa durante la transcripción y negativo (SE-) por detrás. La ausencia de Topo I en los genomas de endosimbiontes sería equiparable a los mutantes de *E. coli topA-*. El exceso de SE+ sería eliminado por la girasa, pero el SE-, al estar ausente la Topo I, sería acumulado y podría servir como fuerza de apertura de la hélice, necesaria para la entrada de la maquinaria de replicación (Benham, 2004).

El exceso de superenrollamiento negativo también podría deberse a que fuera necesario mantener un mayor grado de compactación de los genomas, probablemente al presentar muchos de ellos poliploidía y, por lo tanto, un alto número de cromosomas. Estos datos unidos al sesgo A+T que presentan estos genomas podría ser suficiente para explicar inicios de replicación sin la necesidad de una zona determinada en el cromosoma y explicar también una mayor facilidad para restaurar horquillas de replicación paradas. Se han descrito vías de restauración de la replicación que implican inicios de replicación en presencia de superenrollamiento negativo en *E. coli* (Grompone et al., 2004).

Es posible, por tanto, que las proteínas mencionadas anteriormente, GyrA/B, presentes en todas las bacterias, lleven a cabo todas las funciones topoisomerasas requeridas para la replicación del DNA y, también en la segregación cromosómica.

Otras proteínas de mantenimiento de la estructura del cromosoma y determinadas características como la curvatura del DNA, por ejemplo IHF y FIS, se han conservado en algunos de los genomas y no en otros. Estos genes no son esenciales para el crecimiento de mutantes en *E. coli* por lo que existe cierta relación entre la esencialidad en *E. coli* y el hecho de que se conserve o no en genomas reducidos.

4. Síntesis y modulación de nucleótidos

Aunque no consideradas estrictamente como pertenecientes al sistema 3R, en este estudio se han incluido también proteínas que interactúan con las proteínas de la maquinaria de replicación, de transcripción y con proteínas responsables de la síntesis de nucleótidos, así como otras proteínas reguladoras de estos procesos (Tabla 5).

Diversos estudios han mostrado que proteínas implicadas en la síntesis de nucleótidos como la RNR (NrdA) en *E. coli*, juegan un papel fundamental en el mantenimiento de la horquilla de replicación a lo largo del genoma, ya que un mutante *nrdA101* es capaz de generar horquillas de replicación paradas que necesitan funciones de recombinación para continuar su camino a lo largo del cromosoma (Guarino *et al.*, 2007).

La enzima NrdA lleva a cabo la reducción de nucleótidos y este proceso fue esencial evolutivamente ya que constituyó un paso clave en la transición de RNA a DNA. La conservación del dominio *ATP-cone* es común en muchas bacterias. En el caso de *E. coli* se ha dado una duplicación de esta región (Torrens, 2008).

La reducción de nucleótidos fue esencial para la transición de RNA a DNA, al proporcionar los precursores de los deoxirribonucleótidos. Los dNTPs constituyen unas moléculas limitantes para la replicación del DNA. Algunos metabolitos y enzimas que necesitan estos microorganismos son proporcionados por el hospedador, entre ellos los nucleótidos que son limitantes, de ahí el grado de conservación de estas enzimas. Los niveles en la célula son aproximadamente $\approx 1\%$ de la cantidad necesaria para una sola ronda de replicación. Además, la reducción

en la concentración de un solo dNTP podría causar un descenso de la velocidad de replicación *in vivo*.

El hecho de que esta enzima aparezca casi perfectamente conservada en todos los genomas analizados, incluso en el genoma de *Carsonella ruddii*, hace que esta proteína sea considerada un fósil molecular (Figura 8). Considerando este hecho, se ha utilizado esta proteína como base para la construcción del árbol filogenético de las diferentes especies utilizadas dentro de gamma-proteobacterias (Figura 8).

CLUSTAL 2.0.12 multiple sequence alignment

BfFLO	MNQILFVTKRNGHQEPINLDKIHVRVNWAAKGLEKISVSQVELRSHIQFYNGIKTSDIQE	60
BPEN	MNQTLVTKRNGHHEP INLDKIHVRVNWAAKGLEQVSVSQVELRSHIQFYNGIKTSDIQE	60
Escherichia	MNQNLVTKRNGDSTERINLDKIHVRVLDWAAEGLHNVISQVELRSHIQFYDGIKTSDIHE	60
BU_APS	MKKNLFVTKRDRGRKEKINLDKIHKVLNWAASEGLDDVSVSQVELCSRIQFYNNITTTINIE	60
BU_Sg	MKNSLFVTKRNGKKEKINLDKIHKVLNWAAGLENI SVSQVELRSRIQFYNNISTVNIHE	60
BU_Bp	MNQMLFVTKRNGKIELINLDKIHVRVNWAAKGLENISISQVALKSRIQFYNNIKTTIIE	60
	*: *:*:*:*:* *	
BfFLO	TIKAAADLISQEEPDYQYLAARLVFHLRKKAYGQFEPKLYNHVHLVKGKIDKLV	120
BPEN	TIKAAADLISEEEDYQYLAARLVFHLRKKAYGQFEPKLYNHVLRLIKIGKYDQHLL	120
Escherichia	TIKAAADLISRDADYQYLAARLAI FHLRKKAYGQFEPALYDHVVKMVMEMGYDNHLL	120
BU_APS	TIKAAADLISQDTPDYQYMAARLAI FHLRKKAYGQFEPKLYDHVKKMVDLEKYDENLL	120
BU_Sg	TIKAAADLISENTPDYQYMAARLAI FHLRKKAYGQFEPNLYYHVKKMVLGKYDEKLL	120
BU_Bp	TIKAAADLISENSPDYQYMAARLTI FHLRKKAFGQFNPNLYDHVKNMIKLEKYDQHLL	120
	* *	
BfFLO	NDYNHKEFEEMDTFIDHWRDMNFSYAAIKQLEGGKYLQVNRVTGKIYESAQFLYILIAACL	180
BPEN	DDYNAEEFNQMDTFIDHWRDMNFSYAAVQLEGGKYLQVNRVTGEIYESAQFLYILTAACL	180
Escherichia	EDYFEEEFKQMDTFIDHWRDMNFSYAAVQLEGGKYLQVNRVTGEIYESAQFLYILVAACL	180
BU_APS	KNYSYKEFLQMNSTFIDHWRDMNFSYAAVQLEGGKYLQVNRVNGKIYESAQFLYILISACL	180
BU_Sg	QNYSFEEYVEMNSFVDHRRDMNFSYAAVQLEAKYLLQVNRVTGQIYESAQFLYILISACL	180
BU_Bp	NDYPKQDFLIMNSFIKHRDMNFSYAAVQLEGGKYLQVNRVTGKIYESAQFLYILISACL	180
	.:* :*: *:*:*:* *	
BfFLO	FAKYPEHIRMKYIKNFYNAISTFKISLATPIMAGVTPTRQFSFSSCVLIECDNLDLSINAT	240
BPEN	FSKYPKTRMHIYIQFYDAISTFKISLATPIMSGVTPTRQFSFSSCVLIECDSDLSINAT	240
Escherichia	FSNYPRETRLQYVKRFYDAVSTFKISLPTPIMSGVTPTRQFSFSSCVLIECGDSDLSINAT	240
BU_APS	FSQYKPNVRRMNYIHRFYNAISTFKISLPTPIMSGVTPTRQFSFSSCVLIECGDSDLSINAT	240
BU_Sg	FSKYKTRMNYIQRFYNAISTFKISLPTPIMSGVTPTRQFSFSSCVLIECDNLDLSINAT	240
BU_Bp	FSKYPNETRMKYIKNFYDAISTFKISLPTPIMAGLTPTRQFSFSSCVLIESEDNLDLSINAT	240
	: * . * : * : *	
BfFLO	SSAIVKYVAQRAGIGINAGRIRAVGSPIRNGEAFHTGCIPFYKHFQTAVKSCSQGGVRRG	300
BPEN	SSAIVKYVAQRAGIGINAGRIRALGSPIRNGEAFHTGCIPFYKHFQTAVKSCSQGGVRRG	300
Escherichia	SSAIVKYVSRAGIGINAGRIRALGSPIRNGEAFHTGCIPFYKHFQTAVKSCSQGGVRRG	300
BU_APS	TSSIVKYVSRAGIGVAGQIRALGSPIRNGEAFHTGCIPFYKHFQSAVKSCSQGGVRRG	300
BU_Sg	TSAIVKYVSRAGIGINAGQIRALGSPIRNGEAFHTGCIPFYKHFQSAVKSCSQGGVRRG	300
BU_Bp	TSAIVKYVSRAGIGINAGRIRALGSPIRNGDTHLHTGCIPFYKHFQSAVKSCSQGGVRRG	300
	:*:* *	
BfFLO	AATLFYPIWHLEIENLLVKNRNGIDSNRIRHLDYGVQINKLMYQRLITGKYITLFSFSD	360
BPEN	AATLFYPIWHLEIENLLVKNRNGIDSNRIRHLDYGVQINKLMYQRLIEGKYITLFSFSD	360
Escherichia	AATLFYPMWHLEVESLLVKNRNGVEGNRVRHMDYGVQINKLMYTRLLKGEDITLFSFSD	360
BU_APS	AATLFYPIWHLEVESLLVKNRNGIENRVRHIDYAIQINKLMYQRLISGDIITLFSFSD	360
BU_Sg	AATVFYPIWHLEIESLLVKNRNGIENRVRHLDYAVQINKLMYQRLISGDIITLFSFSD	360
BU_Bp	AATLFYPIWHFEIESLLVKNRNGIENRVRHMDYSVQINKLMYQRLILGEHITLFSFSD	360
	* *	
BfFLO	VPDLYENFFSDQNKFEYLHYKYEQNHNIRHRKIKSIELFSLMMQERASTGRIYIQNDHC	420
BPEN	VPGLYEAFFNQNEFEHLYNIYENDHNIRHRKIKSIELFSLMMQERASTGRIYIQHTDHC	420
Escherichia	VPGLYDAFFADQEEFERLYTKYEKDDSIKQVQKAVELFSLMMQERASTGRIYIQNVDC	420
BU_APS	VPDLYKAFSDQKFKKIYLYQYKKNIRKTIKALDLFSLMMQERTSTGRIYIQNVDC	420
BU_Sg	VPKLYEYFFSDQKFKKIYLYQYKKNIRKTIKALDLFCLIMRERASTGRIYIQHVDHC	420
BU_Bp	VPNLYDSFFNNQEKFERLYKYENDKTIKQKVKASLYFSLMMQERTSTGRIYIQNVDC	420
	** *	
BfFLO	NHSPFNSALAPIRQSNLCLEITLPTKPLNNINDINGEIALCTLSALNLGNINNLQDLSE	480
BPEN	NTHSAFDSIAAPVRQSNLCLEITLPTKPLNNINDHNGEIALCTLSAFNLGNINLDDLSE	480
Escherichia	NTHSPFPAIAPVRQSNLCLEIALPTKPLNDVNDENGEIALCTLSAFNLGAINNLDELEE	480
BU_APS	NLHSAFNPESPIRQSNLCLEITLPTKSLNDVHDTGDEIALCTLSAFNLGKIKSLDDFKE	480
BU_Sg	NHSAFNSKLATIRQSNLCLEITLPTKPLNNIDDKNGEIALCTLSALNLGLINDLHDLKE	480

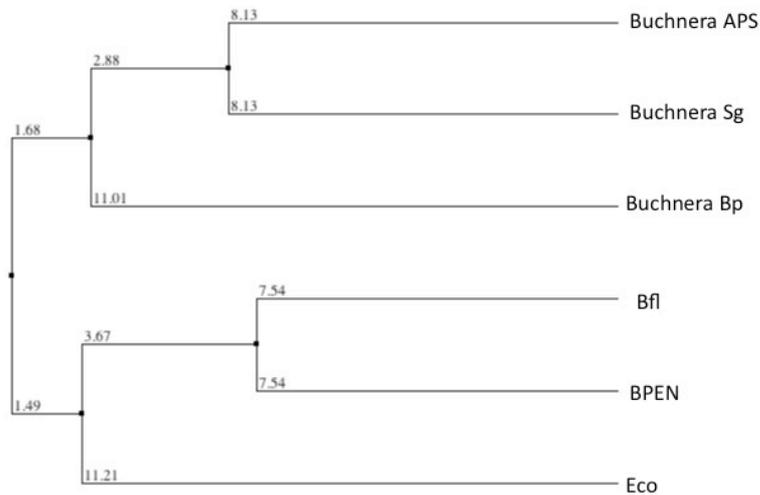


Figura 8. Árbol filogenético construido a partir del alineamiento de NrdA (Figura 7). BIFLO (*Blochmannia floridanus*), BPEN (*Blochmannia pensylvanicus*), BU (*Buchnera aphidicola*).

En la figura 8 se representa el árbol que agrupa por un lado al género *Buchnera* y por otro a las *Blochmannia*, que según este árbol estaría más cercana evolutivamente a *E. coli*. Resulta interesante por tanto el uso de esta enzima como marcador filogenético en Bacterias. El uso de esta enzima como fósil molecular se ha descrito ya en anteriores trabajos (Torrents, 2002).

La pérdida de la Ndk (difosfato nucleosido kinasa) en *Escherichia coli* da lugar a un aumento de la mutación espontánea y desbalancea el *pool* de nucleótidos. Curiosamente, esta enzima aparece conservada en todos los grupos, excepto en las cuatro especies de *Buchneras* y los tres citados genomas de menor tamaño, recalcando la importancia del mantenimiento de un *pool* de nucleótidos óptimo en estos sistemas.

El gen *gidA*, descrito inicialmente como implicado en la división celular, parece estar implicado en la **hipermodificación de tRNAs** y ha sido renombrado como *mnmG*. Este gen aparece representado en todos los genomas analizados y constituiría parte del set mínimo de genes. Sin embargo, *gidA* no es un gen esencial para *E. coli*, de hecho, los mutantes de este gen en *E. coli*, aunque presentan algún problema replicativo, son perfectamente viables (Molina *et al.*, 1999).

5. Proteínas de recombinación y reparación del DNA

Los genomas de una población asexual acumulan mutaciones deletéreas de forma irreversible, como consecuencia del denominado “Trinquete de Müller”. En el caso de las bacterias endosimbiontes que muestran poblaciones pequeñas y sin recombinación hace que se acumulan mutaciones deletéreas en un corto espacio de tiempo. Esas mutaciones se reflejan en el sesgo que presentan estos genomas en su composición nucleotídica.

La replicación del DNA es una de las fuentes principales de inestabilidad genómica. Sin embargo, en estas bacterias, la presencia de un genoma drásticamente reducido, disminuye la proporción de lesiones que pueden darse como resultado de errores del paso de las horquillas de replicación, bloqueo de la RNA polimerasa o disociación de la maquinaria replicativa. Deben existir, por tanto, un conjunto de funciones destinadas a mantener el sistema en equilibrio. La ausencia de estrés ambiental, conlleva, por otro lado, que parte de los genes necesarios para la vida libre sean eliminados en el momento de adaptación a la vida intracelular.

Las bacterias endosimbiontes obligados de insectos se caracteriza por una pérdida casi total de los sistemas de recombinación y reparación (Shigenobu, S *et al.*, 2000; Tamas, I, 2002). Esta pérdida de genes podía verse justificada por el reducido tamaño del genoma y su facilidad para replicarlo. Sin embargo, la facilidad de replicación no tendría que justificar la ausencia de intercambio genético, porque RecA aparece conservada en genomas de muy reducido tamaño como *C. ruddii* (Table 6), sino por eventos específicos de cada linaje. Esto, unido a que algunos genomas presentan determinados componentes del aparato de recombinación induce a pensar que ciertos errores deben ser corregidos por la maquinaria de replicación que presentan estas bacterias. Así, la presencia de un genoma reducido no implica que no se dé intercambio genético ya que son muchos de estos genomas los que conservan RecA.

En la tabla 6 se presentan los genes presentes en los genomas analizados, recogidos en categorías que son un compendio de las registradas en la base de datos KEGG y las recogidas en el Anexo de Fiedberg *et al.*, 2006.

Además de los genes recogidos en las categorías anteriores, se han añadido otros genes si bien, sea por estudios genéticos en *E. coli* o por estudios bioquímicos, se haya determinado un papel de alguna de estas proteínas en reparación. Acorde con esto, los mayores representantes en estos sistemas son: endo – y exo- nucleasas, DNA helicasas, ATPasas (no helicasas) implicadas en migración y carga de complejos multiproteicos en el DNA, DNA ligasas, DNA polimerasas y nucleotidiltransferasas. Entre los dominios conservados abundan sobre todos los implicados en interacción proteína-proteína y proteína-DNA que están recogidos en algunos apartados de esta memoria y cuya explicación en este punto puede resultar redundante.

5.1. Reparación por escisión de bases (BER)

La reparación del DNA por escisión de bases es el modo más frecuente de reparación del DNA en la naturaleza. Esta función es llevada a cabo fundamentalmente por DNA glicosilasas que catalizan la hidrólisis de enlaces N-glicosídico, uniendo bases alteradas químicamente o inapropiadamente unida al esqueleto de la molécula de DNA.

DNA Glicosilasas

Estas enzimas reconocen sólo algunas bases incorporadas inapropiadamente (como la incorporación de Uracilo durante la síntesis de DNA semiconservativa o un daño químico). La representación en los diferentes grupos de la presencia/ausencia de genes se resume en la Tabla 6. A grandes rasgos estos datos pueden resumirse:

- *mutY* (*adenin- DNA glicosilasa*): elimina adeninas apareadas con 8-oxo-G (resultantes de un daño químico sobre la guanina. Está conservada en *Buchnera* y *Blochmannia*, pero no en el resto de grupos analizados.
- *ung* (*uracil-DNA glicosilasa*): la función principal de esta proteína es eliminar uracilos resultantes de la deaminación de la citosina o por incorporación de dUTP durante la síntesis de DNA. Entre otras hipótesis que se han propuesto para explicar la asimetría entre hebras en genomas mínimos está la deaminación de citosinas. Esta proteína está conservada en todos los genomas analizados excepto

Carsonella, *Sulcia* y *Ruthia*. Algunas de estas enzimas eliminan también bases metiladas, este puede ser la razón de que estos organismos no conserven proteínas como AlkA, que presentan esta función en *E.coli*.

- *mutM* (formamidopirimidina-DNA glicosilasa). Elimina 8-oxo-G y formamidopiridinas. Trabajan asociadas a enzimas AP liasas. Poco conservada, sólo presente en *Wolbachia* y *Vesicomysocius*.

- *nth* (endonucleasa III). Esta enzima es una DNA glicosilasa con un amplio espectro de sustratos (derivados de pirimidinas) específicos. Está conservada en todos los grupos excepto *Buchnera Cc*, *Carsonella* y *Sulcia*.

- *AP endonucleasas:*

- recJ* (exonucleasa específica de cadena sencilla). La exonucleasa RecJ se piensa que actúa en el procesamiento de extremos de DNA en la vía RecF. No conservado en *Buchnera* ni *Blochmannia*, probablemente porque éstas tienen *mutY*.

5.2. Reversión directa del daño

Este sistema ha sido diseñado por las células para responder al daño directo generado por exposición a luz UV y es llevado a cabo por la enzima fotoliasa.

- *Phr* (deoxyribodipirimidina-fotoliasa): las fotoliasas de DNA reparan dímeros de DNA inducidos tras exposición a luz UV, como los dímeros de ciclobutano pirimidina. Se ha identificado este gen en 3 genomas de *Buchnera* (no en *Cc*) y *Wigglesworthia*. La escasa representación de este sistema podría deberse a que el daño UV al que están sometidas estas bacteria en un ambiente intracelular es prácticamente inexistente.

5.3. Reparación de bases mal apareadas (MMR)

Este mecanismo de reparación reconoce y re-empareja bases mal apareadas resultado de inserciones erróneas, deleciones o incorporaciones

incorrectas durante los procesos de replicación y recombinación del DNA. Ejemplos de bases mal apareadas incluyen G/T o A/C.

Estos mecanismos son específicos de hebra, capaces de reconocer entre la hebra que se acaba de sintetizar y donde estarán los errores, gracias a la presencia, en bacterias Gram negativas de fenómenos de hemimetilación del DNA.

- *mutS*: los dímeros MutS2 torsionan la hélice de DNA aproximadamente 20 pares de bases. Este gen está bastante conservado, aunque no aparece ni en *Blochmannia* ni *Vesicomysocius*.
- *mutL*: MutL forma un complejo con MutS y MutH incrementando la actividad de MutS sobre el DNA. No está presente en *Blochmannia*, *Ruthia* ni *Vesicomysocius*.
- *mutH*: MutH tiene actividad exonucleasa y genera roturas de hasta 600 pb entre la zona de unión del complejo y la zona de desapareamiento, exponiendo grandes regiones de DNA a la escisión por nucleasas. A esto se debe probablemente el hecho de que esté poco conservada en genomas mínimos (*Buchnera* Bp).
- *uvrD (mutU)*: UvrD además de esta ruta, participa también en la reparación por escisión de nucleótidos (NER) y juega un papel fundamental en el mantenimiento de la estabilidad genómica. Aparece en *Blochmannia*, *Wigglesworthia* y *Wolbachia*.

5.4. Escisión de nucleótidos (NER)

El mayor producto resultante del daño sobre el DNA causado por radiación UV es la presencia de dímeros de pirimidina ciclobutano (CPD) y PP. En *E. coli* estos productos son eliminados por escisión de fragmentos oligonucleotídicos en lugar de bases mal apareadas como hacen los sistemas NER.

Este sistema está mejor conservado en alpha-proteobacterias y gamma-proteobacterias endosimbiontes de moluscos (*Ruthia* y *Vesicomysocius*). Se

aprecia una pérdida del sistema completo en *Buchnera* y *Blochmannia*, aunque estaría mejor conservada en el resto. El sistema está compuesto por un complejo proteico UvrABC, que curiosamente son también miembros del sistema de transporte ABC. La presencia de un dominio de unión a ATP para transporte de proteínas puede llevar a incluir como ortólogas proteínas con similitud de dominios, pero con función diferente. El sistema de transporte ABC está bien conservado en genomas mínimos, por lo que hay que tratar de distinguirlo en la asignación de proteínas ortólogas.

- *uvrA* (ABC excinucleasa subunidad A): este gen se coordina junto a otros tras la inducción de la respuesta SOS (apartado 8 de Introducción). Sin embargo, la respuesta SOS tiene escasa representación en genomas reducidos ya que la presencia del activador LexA queda reducida a un par de genomas. La proteína UvrA es una ATPasa independiente de DNA, aunque une DNA. presente en *Wolbachia*, *Ruthia*, *Vesicomysocius*.
- *uvrB* (subunidad excinucleasa UvrB): Está presente en *Wolbachia* y *Ruthia*.
- *uvrC* (subunidad endonucleasa UvrC). Está presente en *Wolbachia*, *Ruthia*, *Baumannia* y *Sodalis*.
- *uvrD* (descrita en el apartado 5.3).
- *ligA*. Encargada de unir extremos de doble cadena. Está conservada en todos los genomas analizados y con función en replicación y recombinación.

5.5. Proteínas de unión a cadena sencilla

La función de estas proteínas es estabilizar la apertura de la doble hélice en todos los procesos como el inicio de la replicación y la transcripción, así como todos los fenómenos de estabilización de cadenas sencillas en presencia de daño sobre el DNA. Debido a esta función relevante, el gen *ssb* se haya localizado en todos los genomas estudiados.

5.6. Factores de reparación acoplados a la transcripción

La proteína Mfd está ampliamente conservada en bacterias al acoplar la reparación del DNA con la transcripción. Esta proteína reconoce los sitios donde la RNA polimerasa se ha parado por daño en el DNA, tras la alteración del complejo transcripcional o la liberación del transcrito y enzima. Esta proteína recluta la maquinaria de escisión de DNA, causando un movimiento de la RNA polimerasa, usando la energía generada del gasto de ATP. Su actividad es homóloga a la que presenta RecG en la migración de hebras en los Intermediarios de Holliday (HJs). Esta proteína está conservada en *Buchnera*, *Blochmannia* y *Wigglesworthia*, organismos donde se ha perdido RecG. Podemos decir que probablemente Mfd y RecG sean parálogos en el antecesor común. Estas proteínas presentan también similitud de secuencia y estructural con *uvrB*. Como se ha dicho en varias ocasiones a lo largo de esta memoria, estos organismos han reducido la presencia de parálogos, excepto en el caso de *Sodalis*, de ahí que no sea de extrañar que la eficiencia del sistema haya llevado a conservar un solo gen para una sola función.

5.7. Proteínas de tolerancia al daño: TLS DNA polimerasas DinB, DinG y UmuCD

Este sistema se ha perdido totalmente en genomas de endosimbiontes. El sesgo mutacional al que se vieron sometidos estos genomas en los primeros estadios de reducción genómica hizo que probablemente estas proteínas jugaran un papel fundamental para luego ser eliminados y mantener así la homeostasis genómica. Esto explicaría porqué se han perdido todos estos genes a pesar de ser sistemas bastante estables en ausencia de estrés.

5.8. Reparación por recombinación RecA

Durante la realización de este trabajo, se publicó un estudio en el cual se ha analizado utilizando los mismos genomas que en nuestro estudio, el componente de reparación por recombinación (Rocha y Michel., 2005; Sharples, 2009). La función de los genes descritos en este apartado aparece recogido en el apartado 7 de la Introducción de esta memoria.

La proteína LexA es el regulador y activador principal de la respuesta SOS en bacterias. En los genomas analizados sólo se ha localizado en: *Ruthia*, *Vesicomysocius*, *Baumannia* y *Sodalis*. Su escasa presencia en estos genomas, a pesar de la importancia de este sistema de reparación en microorganismos de vida libre vendría dado por la ausencia de estrés al que se ven sometidas estas bacterias en el interior celular.

Es de resaltar el caso de *Vesicomysocius* que tiene LexA, pero no RecA. Sin embargo tiene algunos miembros del sistema UvrABC, por tanto su presencia debe ser un resquicio del antecesor común.

RecA, proteína clave en el fenómeno de reparación por recombinación, está poco conservada en Gamma-proteobacterias. Se trata de un gen no esencial en el crecimiento de *E. coli*, aunque éste sí se ve afectado. En los organismos analizados, está ausente en *Buchnera*, *Blochmania* y *Baumannia*. En el caso de *Buchnera* donde está presente una maquinaria de replicación prácticamente intacta, RecA está ausente por lo que los fenómenos de RDR, si es que existen, deberían llevarse a cabo por otra proteína. Resulta llamativa presencia de RecA en *Carsonella* (41% similitud respecto a RecA en *E. coli*) y sólo pierde parte del N- y C-terminal (Figura 9).

El complejo RecBCD *está conservado en los genomas de gamma-proteobacterias*. Este complejo podría jugar un papel en el inicio de la replicación cuando se ha perdido DnaA. Esto explicaría porqué están presentes en todos los endosimbiontes incluidos algunos que perdieron RecA. SbcB está presente cuando aparece RecBCD, probablemente para limitar la recombinación ilegítima eliminando los extremos 3' (Tabla 6).

El complejo *RecFOR* aparece conservado en *Wolbachia* spp. (alpha-proteobacterias). Éste es también el sistema presente en Rickettsiales (datos no mostrados). Respecto a las gamma-proteobacterias, RecFOR se conserva en *Sodalis*. Curiosamente *Wolbachia* también conserva a RecB y todos los genomas que conservan el componente RecFOR también presentan RecA, siendo esta última esencial para completar su función. Las proteínas del sistema RecFOR promueven

el ensamblaje de los filamentos de RecA en estructuras formadas por dsDNA-ssDNA (HJ) y compensan el efecto inhibitorio de SSB que está unida a cadena sencilla (Tabla 6).

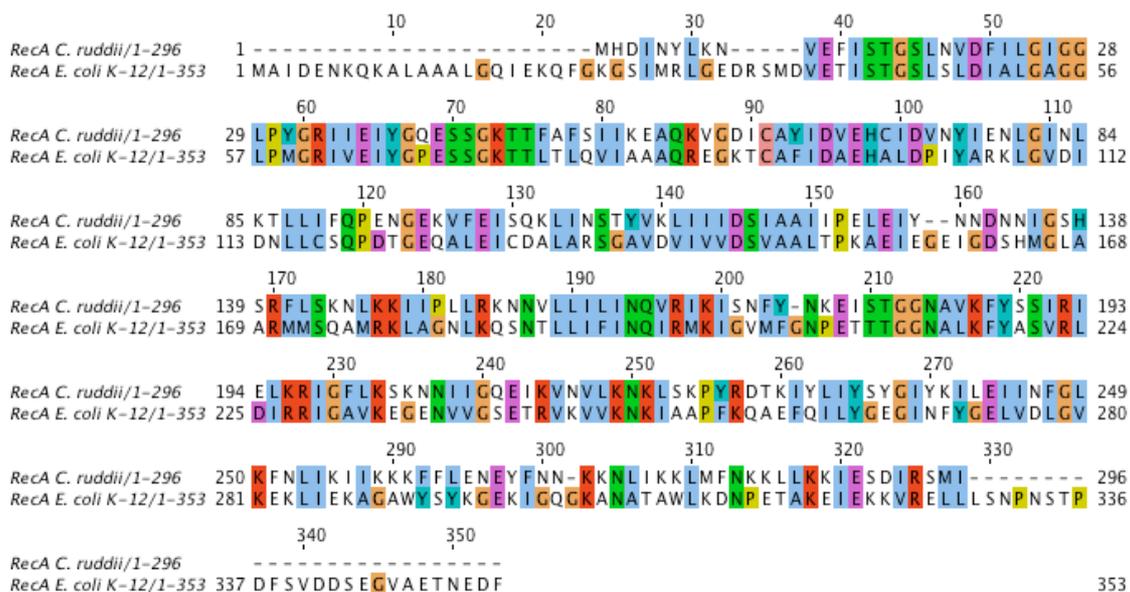


Figura 9. Alineamiento de secuencia entre las proteínas RecA (*E. coli*) y RecA (*C. ruddii*).

La proteína RdgC es un regulador negativo de la función de RecA que sólo está presente en *Sodalis*.

En *Escherichia coli*, la regresión de horquillas de replicación tras la generación de una parada en la horquilla de replicación que da lugar a una rotura de doble hebra, se lleva a cabo por el complemento RuvABC ó RecG. Sin embargo, se ha descrito reversión directa de estas paradas de la replicación. Considerando la ausencia de estos sistemas en la mayor parte de los genomas donde sí aparece RecBCD, se asume que es una reversión directa, espontánea del daño, la que tiene lugar en estos sistemas.

En *Ruthia* y *Vesicomysocius*, se ha perdido el sistema RecBCD, a diferencia de sus compañeras endosimbiontes de insectos. En estos géneros se ha mantenido el componente RecFOR, como en alpha-proteobacterias (donde no se ha encontrado

RecBCD en los genomas analizados). *Vesicomysocius* conserva además el componente RuvABC.

5.9. Reparación independiente de RecA-recombinación

La recombinación independiente de RecA es un fenómeno eficiente, pero limitado por la presencia de exonucleasas de cadena sencilla (ssExos) en la célula (Dutra et al., 2007).

Muchos de los genomas de endosimbiontes han perdido RecA. Las consecuencias de la pérdida de RecA en la evolución de los genomas puede hacerse comparando *Baumannia* y *Wigglesworthia* que la conservan, frente a *Buchnera*, organismo que ha perdido esta proteína.

Sin embargo, la presencia de nucleasas en la célula es común en estos genomas, probablemente para evitar fenómenos de recombinación ilegítima. Se ha propuesto que el sesgo de delección es la fuerza mayor que modela estos genomas de ahí que las **exonucleasas** jueguen un papel muy importante en la estabilidad de estos genomas. La ssExo más importante es RecJ (exonucleasa 5'→3') y SbcB/ExoI (3'→5'). Entre las que degradan las extremidades 5' está también la ExoVII y para las 3' ExoVII y ExoX. RecJ está ausente en *Buchnera* y *Blochmannia*, donde está también ausente RecA. Su ausencia en estos sistemas vendría dada por una limitación del intercambio de hebra cuando RecA está ausente (Tabla 6).

En muchas de estas estirpes se observa la pérdida de la actividad exonucleasa 3'-5' de la Pol I. La ausencia de actividad correctora de pruebas podría suponer un aumento de la tasa de mutación en estos genomas, sin embargo su ausencia supone también un aumento en la velocidad de replicación de una horquilla, ya de por sí bastante sesgada en cuanto a sus componentes, por lo que el sistema parece estar diseñado de forma que tienda al equilibrio entre fidelidad y eficiencia.

6. *Strand bias* y componente 3R en bacterias endosimbiontes

Los tendencia a la composición nucleotídica de los genomas microbianos es resultado de las fuerzas mutacionales y selectivas asociadas con procesos celulares fundamentales como la replicación, transcripción y traducción. Así, diversos

estudios han mostrado como el uso preferente de un conjunto de codones en bacterias de vida libre para asegurar una traducción eficaz de genes altamente expresados (Sharp *et al.*, 2005). El uso de codones preferente en otras especies resultado de la mutación ha hecho que los genomas deriven a concentraciones extremas de AT ó GC, comúnmente observado en bacterias intracelulares (Andersson y Sharp 1996a; Andersson y Sharp 1996b; Sharp *et al.* 2005). Muchos genomas intracelulares tienen fuerte sesgo en las concentraciones de GC (Rocha, 2004). Estos genomas intracelulares tienen también mayores diferencias en la composición de bases entre una cadena y otra de la doble hélice. Estas diferencias en la composición de nucleótidos podrían ayudar a desvelar los procesos que han llevado a la situación actual en la que se encuentran esos genes actualmente en esas bacterias.

Una de las causas propuestas para ese sesgo ha sido la ausencia de genes de reinicio de la replicación y de la maquinaria de reparación de daños directos. Por ejemplo, en *E. coli* el primer paso para la actuación de la maquinaria MMR es la actuación de la DNA metilasa que no se conserva en estos genomas. La ausencia de mutY sería también una de las causas del sesgo hacia A+T (Apartado 5.1. de Resultados). La ausencia de esta enzima, caso común en los genomas endosimbiontes conlleva la escisión de nucleótidos al azar, que en *E. coli* produce a altas frecuencias de sustitución de nucleótidos (Lobner-Olesen *et al.*, 2005).

MutY está presente en genomas como *Buchnera* spp., que presentan una riqueza en A+T bastante considerable, por lo que deben ser también otras fuerzas la que dirijan este sesgo en la composición de bases.

Otras hipótesis que se han propuesto para explicar la asimetría entre hebras están: la deaminación de citosinas, la procesividad de la DNA polimerasa, reordenamientos genómicos, longitud de fragmentos de Okazaki, preferencia de hebra, uso de codones o, relacionados con las mutaciones relacionadas con la transcripción. En el caso de la deaminación de la citosina, que causa mutaciones de C->T, aparece más veces en la hebra molde (Tillier *et al.*, 2000). El codón UGA ha evolucionado en algunos genomas reducidos de codón de stop a reconocer al triptófano (esto también sucede en *Mycoplasmas* y en la mitocondria), de ahí que la aparición de uracilos debe haber sido bastante frecuente a lo largo de la evolución

de estos genomas. Sin embargo, la proteína Ung está bastante conservada en estos genomas (aunque no sabemos qué grado de actividad debe presentar), de ahí que debe ser un equilibrio entre todas estas fuerzas lo que ha llevado a estas composiciones en A+T.

Para estudiar la posible relación entre pérdida de componentes 3R y el contenido GC, se representaron ambos parámetros para todos los genomas analizados (Figura 10).

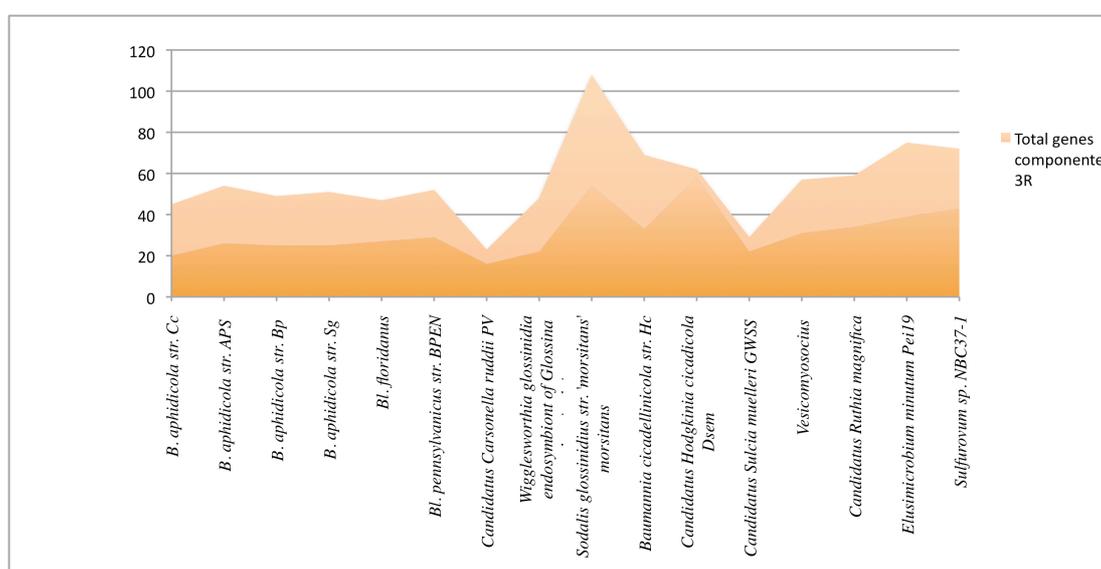


Figura 10. Número de genes 3R en distintos genomas en función del contenido en G+C.

Como se observa en la figura 10 existe una correspondencia entre el número de genes del componente 3R y el porcentaje G+C, por lo que ambos fenómenos deben estar estrechamente interconectados. Una excepción la constituye el endosimbionte *Hodgkinia*. En el caso de *Hodgkinia* se ha descrito la presencia de un uso alternativo del codón UGA, no actuando como codón de STOP, paliando el sesgo en G+C (McCutcheon et al., 2009).

7. Conservación de dominios estructurales y de su arquitectura en proteínas implicadas en transferencia de información

El repertorio de arquitecturas proteicas en los genomas está evolutivamente conservado y constituye una evidencia importante de la historia genómica.

Consecuentemente, los repertorios de estructuras en los proteomas pueden ser considerados como una colección de fósiles moleculares de la historia genómica (Wang et al., 2007). La Genómica Evolutiva trata de aunar la estructura molecular, los genomas y las redes biológicas.

Como se ha venido mencionando, para asignar función a las proteínas conservadas en genomas mínimos, es necesario determinar cuáles de estas proteínas identificadas presentan alguna pérdida de dominio frente el antecesor común. A lo largo del texto, se han dado datos de pérdidas de determinados dominios en proteínas como DnaC (Apartado 1.2. de Resultados). En el caso de *C. ruddii*, la pérdida de dominios proteicos es muy evidente. Un ejemplo lo constituye la presencia de una proteína homóloga de DnaG que conserva un dominio TOPRIM (topoisomerasa-primasa), pero en la que el resto de la proteína se ha perdido (Tamames et al., 2007).

Los genomas de endosimbiontes de insectos han perdido muchos genes, sin embargo, aquellos que conservan lo hacen organizados de una forma sorprendentemente similar. Entre los diez estructuras proteicas (recogidas en la clasificación SCOP) mejor conservadas en organismos con un genoma mínimo, dos están presentes en todos los genomas analizados en este estudio: DNA primase core (e.13), DNA clamp (d.131) Kim, 2006).

7.1. Ausencia de proteínas de recombinación y conservación de helicasas y endo- y exo- nucleasas y glicosidasas

Como ya se ha mencionado, existe una asociación entre la pérdida de proteínas de recombinación y la ausencia/presencia de terminadas endo- y exonucleasas que reducirían los eventos de recombinación ilegítima.

La proteína UvrC constituye la subunidad endonucleasa del complejo UvrABC (ATPasa tipo ABC) (Lin et al., 1992 a y b). Esta familia de endonucleasas contiene una secuencia RX3[YH], dos tirosinas conservadas separadas por 10 residuos y un glutamato. Estos residuos polares poseen capacidad catalítica como se ha demostrado por mutagénesis dirigida (Derbyshire *et al.*, 1997). UvrC está presente

en *Wolbachia* que presenta un sistema de escisión de nucleótidos bien conservado (Tabla 6).

La endonucleasa V (*E. coli nfi*) se localizó sólo en *Elusimicrobium*. Los alineamientos múltiples de representantes de la familia EndoV con la secuencia de UvrC muestran la conservación de 2 aspartatos y una lisina, implicados en la catálisis, así como varios elementos potencialmente estructurales. Esto podría suponer que ambas proteínas son parálogos y por eso se han conservado en unos sistemas y no en otros.

Una de las funciones mejor conservadas es la correspondientes a las Helicasas de DNA. Miembros de esta familia lo constituyen las proteínas: DnaB, PriA, rep, UvrD, RecQ, RecG, RuvAB. DnaB es la principal helicasa replicativa y está representada en todos los genomas estudiados hasta la fecha, excepto en los genomas mínimos *Hodgkinia* y *Sulcia*. Helicasas como Rep, DinG, UvrD son esenciales para la resolución de bloqueos entre horquillas de replicación y transcripción (Boubakri *et al.*, 2010). Rep está conservada en *Buchnera*, que como se ha mencionado anteriormente posee una maquinaria de replicación prácticamente intacta.

La conservación de algunos de los sistemas que relacionan íntimamente la replicación con la transcripción en estos genomas mínimos nos lleva a pensar que ambos sistemas están interconectados para llevar a cabo sus funciones de manera más eficiente.

Discusión

“ Sé perfectamente que apenas se discute en este libro un solo punto acerca del cual no puedan aducirse hechos que con frecuencia llevan, al parecer, a conclusiones diametralmente opuestas a aquellas a las que yo he llegado”.

Charles Darwin

Un gran número de insectos y bivalvos marinos presentan bacterias endosimbiontes obligadas, que viven en el interior de células especializadas llamadas **Bacteriocitos** y que se transmiten verticalmente de generación en generación mediante su inoculación a huevos o embriones. La función de estas células especializadas es generalmente la de proporcionar nutrientes al hospedador, que normalmente tienen una dieta muy pobre o muy limitada en el que viven. Esta asociación, es obligada y mutualista al mismo tiempo, ya que la bacteria no puede crecer y reproducirse fuera del huésped.

Estas poblaciones se caracterizan por haber sufrido un proceso de reducción genómica extrema lo que lleva a ser considerados casi genomas mínimos lo que aporta una información muy valiosa de cara a determinar el contenido génico mínimo.

La evolución bacteriana se refleja directamente en la composición y estructura de los genomas. Procesos como la duplicación, ganancia o pérdida de genes pueden ser inferidos de las relaciones evolutivas entre grupos de bacterias. Existen diferentes estudios que confirman que la evolución de la simbiosis, unida al proceso de reducción genómica ha seguido un proceso de pérdida genética influenciado principalmente por procesos de selección, en los cuales los genes mejor conservados son aquellos con mayor importancia funcional y aquellos que suponen una selección positiva dentro del cuello de botella que han seguido las poblaciones de endosimbiontes (Klasson y Andersson, 2006).

En bacterias Gram negativas, en el que se incluyen muchos patógenos, la mayoría

de los genes seleccionados durante miles de años de evolución se corresponden con genes esenciales. La importancia de la determinación de genes esenciales en bacterias viene dada por su posible uso como dianas de drogas antimicrobianas. De ahí la necesidad de conocer cuáles son esos genes y de realizar estudios que determinen el conjunto mínimo de genes que pueden llevar a cabo una determinada función en la célula.

Dentro de los genes esenciales para la supervivencia de la célula muchos se encuadran dentro del **componente 3R (Replicación, Recombinación y Reparación)**. Precisamente por esta razón es interesante estudiar estas maquinarias en genomas que han sufrido un proceso de disminución drástica de su tamaño y contenido génico y que presentan un componente 3R básico. El objetivo principal de este trabajo ha sido caracterizar el componente 3R mínimo en bacterias Gram negativas endosimbiontes que han sufrido una disminución drástica del contenido de su genoma y de ahí poder inferir los mecanismos de supervivencia de la bacteria dentro de un ambiente libre de estrés con un conjunto reducido de genes.

El hecho de que estas bacterias sobrevivan en un ambiente intracelular y que la gran mayoría no pueden ser cultivadas fuera de este ambiente, implica la imposibilidad de realizar experimentos genéticos con las mismas que puedan validar las hipótesis planteadas. Además, la expresión de genes de endosimbiontes como *Buchnera spp.* en otros organismos se ve imposibilitada dado que, debido al sesgo mutacional, estas proteínas se encuentran al borde del colapso molecular (Bastolla et al., 2004). Esto implica que la mayor parte de las aproximaciones que se realizan desde una aproximación **Bioinformática**.

Los problemas comunes que suelen presentar en este tipo de estudios son, en primer lugar, el escaso número de genomas secuenciados disponibles actualmente, y, por otro lado, la imposibilidad de caracterizar estadios de degeneración genómica intermedios a lo largo de la escala evolutiva ya que estos organismos no se encuentran disponibles hoy en día para su análisis, ni tampoco en el registro fósil. Esto hace difícil identificar los ortólogos de especies muy lejanas evolutivamente. De hecho, algunas funciones han sido desplazadas por genes no ortólogos (análogos). Esto es común en los sistemas de reparación, donde la ausencia de una

vía hace que el organismo intente solucionar el problema presentado a partir de una vía paralela (desplazamiento no ortólogo=NOD). El escaso número de genomas hace también que los datos no pueden tomarse como definitivos y pueden cambiar con cada nuevo genoma secuenciado. En estos genomas reducidos escasean los pseudogenes que son los que permitirían seguir la reducción genómica real. Para intentar captar estos eventos intermedios de evolución de los genomas, se ha incluido en el estudio a *Sodalis glosinidia*, a medio camino entre la simbiosis y la vida libre y cuyo genoma presenta un alto número de pseudogenes. También hemos incluido el genoma del primer endosimbionte cultivado, *Elusimicrobium minutum*, perteneciente al grupo de Epsilon-proteobacterias.

La presencia de pseudogenes es explicada por la “Teoría Dominó” de la muerte genética que relata un proceso paulatino de pérdida de función génica, inactivación del gen y, por último, pérdida de la secuencia génica (Dagan et al., 2006). En el caso de los P-endosimbiontes no existen prácticamente pseudogenes, mientras que existe un número elevado en los S-endosimbiontes como *Sodalis*.

1. Conservación de la maquinaria de replicación en bacterias endosimbiontes

En algunas bacterias como *Borrelia burgdorferi*, aunque se ha identificado un origen de replicación, no se han observado cajas de unión a DnaA. Otros como *Helicobacter pylori* poseen secuencias diferentes a las de *E. coli*. En los genomas de endosimbiontes considerados en este estudio no se han identificado ninguno de estos sistemas de control del ciclo celular. Algunos genomas, como las diferentes especies de *Buchnera*, presentan DnaA, pero la secuencia del origen está prácticamente reducida a una caja de unión de esta proteína. Todos los genomas carecen además de proteínas como Hda, Dam y SeqA que garantizan que sólo tenga lugar un inicio de replicación por ciclo celular. Sólo algunas proteínas de unión a DNA como HU ó IHF presentan una distribución más uniforme en estos genomas mínimos, y su función parece ser más de mantenimiento de la topología del DNA que de regulación de la replicación.

Esto podría ser explicado por el sesgo en A+T que presentan los genomas de bacterias simbiotes. La apertura de la doble hélice está termodinámicamente

favorecida frente a un genoma rico en G+C. Por lo tanto, no se requeriría la presencia de proteínas específicas de unión a DNA para la apertura de la doble hélice en el origen de replicación, aunque en algunos casos sí se conserva su secuencia. Procesos de apertura de la doble hélice en presencia de estrés térmico en *E. coli* conducen a la apertura de la doble hélice con dependencia parcial de DnaA (González-Soltero *et al.*, 2006).

Diversos estudios estructurales y de filogenómica demuestran que los componentes claves de la maquinaria de replicación del DNA (DNA Polimerasas y Primasas) evolucionaron dos veces, para dar el aparato bacteriano y el de arqueas/eucariotas (Leipe *et al.*, 1999; Forterre, 2002; Eilée y Millykallio, 2004).

En el caso de *Buchnera spp.* se conserva una maquinaria de replicación similar a la de *E. coli*. Dada la conservación de DnaA, DnaC-DnaB y de una secuencia de origen de replicación (esta última muy degenerada), el inicio tendría lugar de forma similar a la que sucede en *E. coli*. En el caso del linaje *Buchnera*, el proceso replicativo parece tener, por lo tanto, mayor importancia para su ciclo celular y esto ha hecho que conserve una maquinaria de replicación con respecto a otros genomas que han sufrido un proceso de degeneración parecido.

En el caso de *Blochmannia*, *Wigglesworthia* y *Baumannia*, la ausencia de DnaA podría ser compensada por la presencia de RecA. Mecanismos de RDR similar a los descritos en *E. coli* en situaciones de horquillas de replicación bloqueadas podrían explicar el inicio de replicación en algunos de estos genomas.

La presencia de RecBCD en la mayor parte de los genomas de Gamma-Proteobacterias parece subrayar la importancia de este sistema, que, por otro lado, permitiría la consecución de replicación vía RDR (Kogoma, 1997). En el caso de los endosimbiontes pertenecientes al género *Wolbachia* el sistema presente es RecFOR, además todos ellos presentan también RecA.

2. La proteína NrdA como fósil molecular para la filogenia de poblaciones de endosimbiontes

Las Ribonucleotido reductasas (RNRs) pertenecen a una familia de

complejos enzimáticos que juegan un papel esencial en los organismos vivos, catalizando la conversión de los cuatro nucleótidos en deoxinucleótidos esenciales para la replicación y reparación del DNA. En *E. coli* las RNRs existen como proteínas homodiméricas NrdA (2) y NrdB (2) dispuestas como un heterotetrámero.

El grado de conservación de la secuencia primaria de estas proteínas en los genomas de endosimbiontes indica que esta proteína constituye un fósil molecular y que, por tanto, podría utilizarse para estimar el tiempo de divergencia de estas especies. La conservación a nivel de estructura terciaria es también evidente (Torrents et al., 2002). Por tanto, estudios filogenéticos utilizando como base esta familia de proteínas podrían definir el tiempo de divergencia de los linajes de endosimbiontes respecto al antecesor común. Un ejemplo del empleo de RNRs como reloj molecular se muestra en Torrents et al., 2002, donde se estudiaron los tiempos de divergencia usando algoritmos de Máxima Probabilidad.

3. Conservación de la maquinaria de recombinación y reparación en bacterias endosimbiontes

La composición diferencial de genes de reparación entre distintas bacterias endosimbiontes es un sistema modelo ideal para estudiar el efecto a largo plazo de la pérdida de genes de reparación. De hecho, la presencia o ausencia observada de genes parece ser específica de linaje y es una respuesta a la adaptación al medio intracelular.

Además, la ausencia de muchos componentes de la maquinaria de replicación proporciona a estas bacterias una aparente incapacidad para llevar a cabo procesos mediados por recombinación, confiriéndoles estabilidad genómica al prevenir cualquier tipo de reordenación genómica o eventos de transferencia lateral. La ausencia de RecA es también la causante de estabilidad en la arquitectura genómica. La pérdida de secuencias repetidas de más de 30pb disminuye la posibilidad de reorganización, al limitar el número de reordenamientos posibles mediante recombinación intra-cromosómica.

4. Adaptación hacia otros modos de replicación: replicación dependiente de transcripción y círculo rodante

Es predecible que complementos mínimos para la replicación, transcripción y traducción resulten en baja velocidad de elongación y altas frecuencias de error. Debe existir, por tanto, un delicado balance entre velocidad, precisión y procesividad que conlleva a largo plazo la pérdida de genes correspondientes a los sistemas de información.

Considerando que la mayoría de estos organismos presentan una maquinaria de replicación incompleta de ahí que probablemente hayan adoptado mecanismos de replicación más sencillos.

La presencia de RNasa HI, junto a helicasas como RecG, lleva a pensar que algunos de estos organismos hayan adoptado un mecanismo de replicación dependiente de transcripción a partir de estructuras híbridas de RNA-DNA denominadas *R-Loops*. Dichos mecanismos de replicación a partir de *R-Loops* han sido descritos en *E. coli* en situaciones de estrés (Kogoma, 1997).

La ausencia de proteínas de terminación de replicación habría repercutido en la aparición de sistemas de replicación similar al círculo rodante en plásmidos. Estos sistemas podrían iniciar la replicación a partir de un cebador de RNA (DnaG está conservada en todos los genomas). Este tipo de replicación explicaría la presencia de poliploidía en determinados genomas de endosimbiontes. La poliploidía estaría además facilitada por la ausencia de proteínas de control sobre el ciclo celular.

5. Evolución de la minimización de genomas hasta su conversión en orgánulo: justificación de la Teoría Endosimbiótica

La Teoría Endosimbiótica fue propuesta por la Dra. Lynn Margulis en los años 60s, pero no fue hasta 1981 cuando fue oficialmente recogida en su libro "Simbiosis en la Evolución Celular".

La vida de estos endosimbiontes en el interior de células especializadas y el hecho de que su transmisión sea también vía materna, transmitida en el huevo, como en el caso de *Buchnera*, o a través de la leche, como en el caso de *Wigglesworthia*, hace que recuerde al tipo de herencia mitocondrial, también vía materna, por lo que este

tipo de adaptaciones podrían ser una confirmación más de la teoría endosimbiótica de Lynn Margulis. En las mitocondrias la maquinaria de replicación se ve reducida a una helicasa (Twinkle) que no tiene dominio Primasa, una Polimerasa y SSB. La maquinaria de transcripción juega un papel importante en el inicio de replicación del genoma mitocondrial (tipo círculo rodante). Tal y como se ha descrito a lo largo de esta memoria estos organismos parecen también adquirido un tipo de replicación similar a los plásmidos como es el círculo rodante de las mitocondrias.

Los genomas mínimos parecen haber sufrido una disminución de tamaño que tiende al colapso funcional. Algunas funciones perdidas en estos endosimbiontes primarios son reemplazadas por endosimbiontes secundarios. Según la teoría endosimbiótica estos microorganismos tendrían una evolución similar a la sufrida por las alpha-proteobacterias que dieron lugar a la mitocondria y darían lugar a orgánulos especializados en la célula. Algo parecido es lo que ya ha ocurrido en endosimbiontes como *Carsonella*, *Sulcia* o *Hodgkinia*.

Conclusiones

1. El número de componentes del sistema 3R depende del tamaño del genoma del endosimbionte.
2. El sistema de replicación que utilizan estos endosimbiontes está muy degenerado respecto al de bacterias de vida libre. La mayoría de ellas no presentan región para el inicio de replicación y en los casos donde existe, como *B. aphidicola*, está muy degenerado y su función se ve posiblemente simplificada. Además, el proceso de reducción genómica que ha sufrido este sistema parece ser específico de linaje, presentando algunos géneros un aparato de replicación más complejo que otros, e incluso existen cambios de composición dentro del mismo linaje como sucede en el caso de *Buchnera spp.*
3. En los casos donde no se ha descrito presencia de origen de replicación, el genoma ha superado este problema probablemente adquiriendo otros tipos de replicación como son la RDR o replicación tipo círculo rodante, como es el caso de muchos plásmidos de *E. coli*.
4. El mantenimiento de un *pool* de nucleótidos óptimo parece ser esencial en estos sistemas de ahí que se conserven en prácticamente todos los genomas proteínas como RNR ó Ndk, encargadas de mantener este *pool*. Las proteínas RNR están muy conservadas en estos genomas pudiendo utilizarse en estudios filogenéticos como **Reloj Molecular**.
5. La pérdida de proteínas de recombinación y reparación es muy evidente y los procesos que han llevado a la conservación de genes parecen ser específicos de linaje.
6. Existe una relación entre el tamaño del genoma y el número de genes conservados, excepto para el componente de reparación.
7. La composición nucleotídica hacia bajos porcentajes de G+C parece sufrir una evolución similar a la degeneración mostrada por el tamaño del genoma.

8. El limitado número de genomas secuenciados impide sacar conclusiones sobre los procesos que conllevan la pérdida de determinados genes y no otros, como consecuencia a la adaptación a determinado nicho ecológico. Las nuevas técnicas de ultrasecuenciación permitirán obtener la secuencia de un mayor número de genomas y, por tanto, en mayor número de datos para sacar conclusiones más evidentes.

Bibliografía

- Adékambi T, Shinnick TM, Raoult D, Drancourt M. Complete *rpoB* gene sequencing as a suitable supplement to DNA-DNA hybridization for bacterial species and genus delineation. *Int. J. Syst. Evol. Microbiol.* 2008;58(Pt 8):1807-1814.
- Akman Gündüz E, Douglas AE. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proc. Biol. Sci.* 2009. 7;276(1658):987-991.
- Basler G, Nikoloski Z, Ebenhöf O, Handorf T. Biosynthetic potentials from species-specific metabolic networks. *Genome Inform.* 2008 ;20135-148.
- Bastolla U, Moya A, Viguera E, Roel, Ham CV. Genomic Determinants of Protein Folding Thermodynamics in Prokaryotic Organisms. *Journal of Molecular Biology.* 2004;343(5):1451-1466.
- Belda E, Moya A, Silva FJ. Genome rearrangement distances and gene order phylogeny in gamma-Proteobacteria. *Mol. Biol. Evol.* 2005; 22(6):1456-1467.
- Benham CJ, Bi C. The analysis of stress-induced duplex destabilization in long genomic DNA sequences. *J. Comput. Biol.* 2004;11(4):519-543.
- Bhattacharya D, Archibald JM, Weber APM, Reyes-Prieto A. How do endosymbionts become organelles? Understanding early events in plastid evolution. *Bioessays.* 2007;29(12):1239-1246.
- Blinkova A, Hervas C, Stukenberg PT, Onrust R, O'Donnell ME, Walker JR. The *Escherichia coli* DNA polymerase III holoenzyme contains both products of the *dnaX* gene, tau and gamma, but only tau is essential. *J. Bacteriol.* 1993;175(18):6018-6027.
- Boubaki H, Langlois de Septenville A, Viguera E, Michel, B. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *EMBO J.* 2010; 29, 145-157.
- Calcutt MJ. Gene organization in the *dnaA-gyrA* region of the *Streptomyces coelicolor* chromosome. *Gene.* 1994. 30;151(1-2):23-28.
- Champoux JJ. DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* 2001 ;70369-413.
- Dagan T, Blekhman R, Graur D. The “Domino Theory “of Gene Death: Gradual and Mass Gene Extinction Events in Three Lineages of Obligate Symbiotic Bacterial Pathogens. 2006 ;23(2):310-6.
- Dagan T, Blekhman R, Graur D. The Domino Theory of Gene Death: Gradual and Mass Gene Extinction Events in Three Lineages of Obligate Symbiotic Bacterial Pathogens. *Molecular Biology and Evolution.* 2006. 20;23310-316.
- Dale C, Wang B, Moran N, Ochman H. Loss of DNA recombinational repair

- enzymes in the initial stages of genome degeneration. *Mol. Biol. Evol.* 2003; 20(8):1188-1194.
- Degnan PH, Leonardo TE, Cass BN, Hurwitz B, Stern D, Gibbs RA, Richards S, Moran NA. Dynamics of genome evolution in facultative symbionts of aphids *Environ. Microbiol.* 2009 .
 - den Blaauwen T, Aarsman MEG, Wheeler LJ, Nanninga N. Pre-replication assembly of *E. coli* replisome components. *Mol. Microbiol.* 2006; 62(3):695-708.
 - Enault F, Suhre K, Abergel C, Poirot O, Claverie J. Annotation of bacterial genomes using improved phylogenomic profiles. *Bioinformatics.* 2003 ;19 Suppl 1i105-107.
 - Escarmís C, Dávila M, Charpentier N, Bracho A, Moya A, Domingo E. Genetic Lesions Associated with Muller's Ratchet in an RNA Virus. *Journal of Molecular Biology.* 1996;264(2):255-267.
 - Friedberg et al. .DNA repair and mutagenesis, 2d ed. Ed. by Errol C. Friedberg et al. *Am. Soc. for Microbiology.* 2006
 - Fitch WM. Uses for evolutionary trees. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 1995;349(1327):93-102.
 - Fitch WM, Yasunobu KT. Phylogenies from amino acid sequences aligned with gaps: the problem of gap weighting. *J. Mol. Evol.* 1975;5(1):1-24.
 - Fitches E, Wiles D, Douglas AE, Hinchliffe G, Audsley N, Gatehouse JA. The insecticidal activity of recombinant garlic lectins towards aphids. *Insect Biochem. Mol. Biol.* 2008;38(10):905-915.
 - Flåtten I, Morigen, Skarstad K. DnaA protein interacts with RNA polymerase and partially protects it from the effect of rifampicin. *Mol. Microbiol.* 2009; 71(4):1018-1030.
 - Fuglsang A. Impact of bias discrepancy and amino acid usage on estimates of the effective number of codons used in a gene, and a test for selection on codon usage. *Gene.* 2008;410(1):82-88.
 - Gil R, Silva FJ, Peretó J, Moya A. Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* 2004;68(3):518-537.
 - Glover BP, McHenry CS. The chi psi subunits of DNA polymerase III holoenzyme bind to single-stranded DNA-binding protein (SSB) and facilitate replication of an SSB-coated template. *J. Biol. Chem.* 1998;273(36):23476-23484.
 - González-Soltero R, Botello E, Jiménez-Sánchez A. Initiation of heat-induced replication requires DnaA and the L-13-mer of oriC. *J. Bacteriol.* 2006; 188(23):8294-8298.
 - González-Soltero R, Jiménez-Sánchez A, Botello E. Functional requirements

- for heat induced genome amplification in *Escherichia coli*. *Process Biochemistry*. 2008;43(10):1162-1170.
- Gosalbes MJ, Lamelas A, Moya A, Latorre A. The striking case of tryptophan provision in the cedar aphid *Cinara cedri*. *J. Bacteriol.* 2008;190(17):6026-6029.
 - Goto S, Bono H, Ogata H, Fujibuchi W, Nishioka T, Sato K, Kanehisa M. Organizing and computing metabolic pathway data in terms of binary relations. *Pac Symp Biocomput.* 1997;175-186.
 - Gómez-Valero L, Latorre A, Silva FJ. The evolutionary fate of nonfunctional DNA in the bacterial endosymbiont *Buchnera aphidicola*. *Mol. Biol. Evol.* 2004;21(11):2172-2181.
 - Gómez-Valero L, Latorre A, Gil R, Gadau J, Feldhaar H, Silva FJ. Patterns and rates of nucleotide substitution, insertion and deletion in the endosymbiont of ants *Blochmannia floridanus*. *Mol. Ecol.* 2008
 - Grompone G, Ehrlich D, Michel B. Cells defective for replication restart undergo replication fork reversal. *EMBO Rep.* 2004;5(6):607-612.
 - Guarino E, Jiménez-Sánchez A, Guzmán EC. Defective ribonucleoside diphosphate reductase impairs replication fork progression in *Escherichia coli*. *J. Bacteriol.* 2007;189(9):3496-3501.
 - Güell M, van Noort V, Yus E, Chen W, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kühner S, Rode M, Suyama M, Schmidt S, Gavin A, Bork P, Serrano L. Transcriptome complexity in a genome-reduced bacterium. *Science.* 2009 Nov;326(5957):1268-1271.
 - Hendrickson H, Lawrence JG. Mutational bias suggests that replication termination occurs near the dif site, not at Ter sites. *Mol. Microbiol.* 2007; 64(1):42-56.
 - Hingorani MM, O'Donnell M. Toroidal proteins: running rings around DNA. *Curr. Biol.* 1998;8(3):R83-86.
 - Holmes VF, Cozzarelli NR. Closing the ring: links between SMC proteins and chromosome partitioning, condensation, and supercoiling. *Proc. Natl. Acad. Sci. U.S.A.* 2000;97(4):1322-1324.
 - Huang C, Lee C, Wu H, Kuo M, Lai C. Interactions of chaperonin with a weakly active anthranilate synthase from the aphid endosymbiont *Buchnera aphidicola*. *Microb. Ecol.* 2008;56(4):696-703.
 - Huynen M, Doerks T, Eisenhaber F, Orengo C, Sunyaev S, Yuan Y, Bork P. Homology-based fold predictions for *Mycoplasma genitalium* proteins. *J. Mol. Biol.* 1998;280(3):323-326.
 - Jouselin E, Desdevises Y, Coeur d'acier A. Fine-scale cospeciation between *Brachycaudus* and *Buchnera aphidicola*: bacterial genome helps define species and evolutionary relationships in aphids. *Proc. Biol. Sci.* 2009;

276(1654):187-196.

- Klasson L, Andersson SGE. Strong asymmetric mutation bias in endosymbiont genomes coincide with loss of genes for replication restart pathways. *Mol. Biol. Evol.* 2006;23(5):1031-1039.
- Kogoma T. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol. Mol. Biol. Rev.* 1997; 61(2):212-238.
- Komaki K., Ishikawa H.[1]. Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. *Insect Biochemistry and Molecular Biology.* 2000;30:253-258.
- Koppes LJ, Woldringh CL, Nanninga N. *Escherichia coli* contains a DNA replication compartment in the cell center. *Biochimie.* 1999;81(8-9):803-810.
- Kühner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, Castaño-Diez D, Chen W, Devos D, Güell M, Norambuena T, Racke I, Rybin V, Schmidt A, Yus E, Aebersold R, Herrmann R, Böttcher B, Frangakis AS, Russell RB, Serrano L, Bork P, Gavin A. Proteome organization in a genome-reduced bacterium. *Science.* 2009;326(5957):1235-1240.
- Lamelas A, Pérez-Brocal V, Gómez-Valero L, Gosalbes MJ, Moya A, Latorre A. Evolution of the secondary symbiont "*Candidatus serratia symbiotica*" in aphid species of the subfamily Lachninae. *Appl. Environ. Microbiol.* 2008; 74(13):4236-4240.
- Lehman IR, Uyemura DG. DNA polymerase I: essential replication enzyme. *Science.* 1976;193(4257):963-969.
- Lemon KP, Grossman AD. Localization of bacterial DNA polymerase: evidence for a factory model of replication. *Science.* 1998; 282(5393):1516-1519.
- Li Y, Sergueev K, Austin S. The segregation of the *Escherichia coli* origin and terminus of replication. *Mol. Microbiol.* 2002;46(4):985-996.
- Margulis L. *Symbiosis in cell evolution.* New York: W.H. Freeman, (1981). 452 p.
- Martínez-Torres D, Buades C, Latorre A, Moya A. Molecular Systematics of Aphids and Their Primary Endosymbionts. *Molecular Phylogenetics and Evolution.* 2001;20(3):437-449.
- McCutcheon JP, McDonald BR, Moran NA. Origin of an Alternative Genetic Code in the Extremely Small and GC-Rich Genome of a Bacterial Symbiont. *PLoS Genet.* 2009;5(7):e1000565.
- McHenry CS. Purification and characterization of DNA polymerase III'. Identification of tau as a subunit of the DNA polymerase III holoenzyme. *J. Biol. Chem.* 1982;257(5):2657-2663.

- Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol. Direct.* 2009;4:13.
- Messer W. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol. Rev.* 2002; 26(4):355-374.
- Messer W, Blaesing F, Majka J, Nardmann J, Schaper S, Schmidt A, Seitz H, Speck C, Tüngler D, Wegrzyn G, Weigel C, Welzeck M, Zakrzewska-Czerwinska J. Functional domains of DnaA proteins. *Biochimie.* 1999;81(8-9):819-825.
- Mira A, Ochman H, Moran NA. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* 2001;17(10):589-596.
- Mira A, Ochman H, Moran NA. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* 2001;17(10):589-596.
- Molina F, Jiménez-Sánchez A, Zyskind JW, Guzmán EC. Chromosomal insertions localized around oriC affect the cell cycle in *Escherichia coli*. *Biochimie.* 1999;81(8-9):811-818.
- Moran NA. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr. Opin. Microbiol.* 2003;6(5):512-518.
- Moran NA. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr. Opin. Microbiol.* 2003;6(5):512-518.
- Moran, Wernegreen. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol. Evol. (Amst.)*. 2000;15(8):321-326.
- Moran NA, McCutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* 2008 ;42:165-190.
- Moran NA, McLaughlin HJ, Sorek R. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science.* 2009;323(5912):379-382.
- Moreno-Hagelsieb G, Latimer K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics.* 2008;24(3):319-324.
- Mott ML, Erzberger JP, Coons MM, Berger JM. Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell.* 2008;135(4):623-634.
- Mushegian AR, Koonin EV. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U.S.A.* 1996;93(19):10268-10273.
- Najar-Rodríguez AJ, McGraw EA, Mensah RK, Pittman GW, Walter GH. The microbial flora of *Aphis gossypii*: patterns across host plants and geographical space. *J. Invertebr. Pathol.* 2009;100(2):123-126.
- Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori

- M. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science*. 2006;314(5797):267.
- Neumann S, Quiñones A. Discoordinate gene expression of *gyrA* and *gyrB* in response to DNA gyrase inhibition in *Escherichia coli*. *J. Basic Microbiol.* 1997 ;37(1):53-69.
 - Neylon C, Kralicek AV, Hill TM, Dixon NE. Replication termination in *Escherichia coli*: structure and antihelicase activity of the Tus-Ter complex. *Microbiol. Mol. Biol. Rev.* 2005;69(3):501-526.
 - Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JCD, Andersson DI. Bacterial genome size reduction by experimental evolution. *Proc. Natl. Acad. Sci. U.S.A.* 2005;102(34):12112-12116.
 - Ochman H, Jones IB. Evolutionary dynamics of full genome content in *Escherichia coli*. *EMBO J.* 2000;19(24):6637-6643.
 - Olsen GJ, Woese CR. Archaeal genomics: an overview. *Cell.* 1997;89(7):991-994.
 - Ortiz-Rivas B, Moya A, Martínez-Torres D. Molecular systematics of aphids (Homoptera: Aphididae): new insights from the long-wavelength opsin gene. *Molecular Phylogenetics and Evolution.* 2004;30(1):24-37.
 - Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. Use of contiguity on the chromosome to predict functional coupling. *In Silico Biol. (Gedruckt).* 1999 ;1(2):93-108.
 - Patel PH, Suzuki M, Adman E, Shinkai A, Loeb LA. Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection. *J. Mol. Biol.* 2001;308(5):823-837.
 - Paulsen IT, Press CM, Ravel J, Kobayashi DY, Myers GSA, Mavrodi DV, DeBoy RT, Seshadri R, Ren Q, Madupu R, Dodson RJ, Durkin AS, Brinkac LM, Daugherty SC, Sullivan SA, Rosovitz MJ, Gwinn ML, Zhou L, Schneider DJ, Cartinhour SW, Nelson WC, Weidman J, Watkins K, Tran K, Khouri H, Pierson EA, Pierson LS, Thomashow LS, Loper JE. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat. Biotechnol.* 2005;23(7):873-878.
 - Pérez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A. A small microbial genome: the end of a long symbiotic relationship? *Science.* 2006;314(5797):312-313.
 - Pérez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A. A small microbial genome: the end of a long symbiotic relationship? *Science.* 2006;314(5797):312-313.
 - Ponomarev VA, Makarova KS, Aravind L, Koonin EV. Gene duplication with displacement and rearrangement: origin of the bacterial replication protein PriB from the single-stranded DNA-binding protein Ssb. *J. Mol. Microbiol. Biotechnol.* 2003 ;5(4):225-229.

- Pruss GJ, Drlica K. DNA supercoiling and prokaryotic transcription. *Cell*. 1989; 56(4):521-523.
- Rocha EPC, Cornet E, Michel B. Comparative and Evolutionary Analysis of the Bacterial Homologous Recombination Systems . *PLoS Genet*. 2005;1(2):e15.
- Sakharkar KR, Dhar PK, Chow VTK. Genome reduction in prokaryotic obligatory intracellular parasites of humans: a comparative analysis. *Int. J. Syst. Evol. Microbiol*. 2004;54(Pt 6):1937-1941.
- Sandler SJ. Multiple genetic pathways for restarting DNA replication forks in *Escherichia coli* K-12. *Genetics*. 2000;155(2):487-497.
- Sandler SJ. Requirements for replication restart proteins during constitutive stable DNA replication in *Escherichia coli* K-12. *Genetics*. 2005;169(4):1799-1806.
- Sawitzke J, Austin S. An analysis of the factory model for chromosome replication and segregation in bacteria. *Mol. Microbiol*. 2001;40(4):786-794.
- Sernova NV, Gelfand MS. Identification of replication origins in prokaryotic genomes. *Brief. Bioinformatics*. 2008;9(5):376-391.
- Seufert W, Messer W. Start sites for bidirectional in vitro DNA replication inside the replication origin, *oriC*, of *Escherichia coli*. *EMBO J*. 1987;6(8):2469-2472.
- Sherratt DJ, Søballe B, Barre F, Filipe S, Lau I, Massey T, Yates J. Recombination and chromosome segregation. *Philos. Trans. R. Soc. Lond., B, Biol. Sci*. 2004;359(1441):61-69.
- Silva FJ, Latorre A, Moya A. Why are the genomes of endosymbiotic bacteria so stable? *Trends in Genetics*. 2003;19(4):176-180.
- Snyder LAS, Loman NJ, Fütterer K, Pallen MJ. Bacterial flagellar diversity and evolution: seek simplicity and distrust it? *Trends Microbiol*. 2009;17(1):1-5.
- Sorfová P, Skeríková A, Hypsa V. An effect of 16S rRNA intercistronic variability on coevolutionary analysis in symbiotic bacteria: molecular phylogeny of *Arsenophonus triatominarum*. *Syst. Appl. Microbiol*. 2008; 31(2):88-100.
- Tamames J, Gil R, Latorre A, Peretó J, Silva FJ, Moya A. The frontier between cell and organelle: genome analysis of *Candidatus Carsonella ruddii*. *BMC Evol. Biol*. 2007; 7181.
- Tamas I, Wernegreen JJ, Nystedt B, Kauppinen SN, Darby AC, Gomez-Valero L, Lundin D, Poole AM, Andersson SGE. Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. *Proc. Natl. Acad. Sci. U.S.A*. 2008;105(39):14934-14939.
- Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families.

Science. 1997;278(5338):631-637.

- Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 2001;29(1):22-28.
- Toft C, Fares MA. The evolution of the flagellar assembly pathway in endosymbiotic bacterial genomes. *Mol. Biol. Evol.* 2008;25(9):2069-2076.
- Toft C, Fares MA. Selection for translational robustness in *Buchnera aphidicola*, endosymbiotic bacteria of aphids. *Mol. Biol. Evol.* 2009;26(4):743-751.
- Torrents E, Aloy P, Gibert I, Rodríguez-Trelles F. Ribonucleotide reductases: divergent evolution of an ancient enzyme. *J. Mol. Evol.* 2002;55(2):138-152.
- Torres-Puente M, Cuevas JM, Jiménez-Hernández N, Bracho MA, García-Robles I, Wrobel B, Fern, Carnicer O, Olmo JD, Ortega E, Moya A, Fern, González-C O, elas. Using evolutionary tools to refine the new hypervariable region 3 within the envelope 2 protein of hepatitis C virus. *Infection, Genetics and Evolution.* 2008;8(1):74-82.
- Tougu K, Peng H, Marians KJ. Identification of a domain of *Escherichia coli* primase required for functional interaction with the DnaB helicase at the replication fork. *J. Biol. Chem.* 1994;269(6):4675-4682.
- van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernández JM, Jiménez L, Postigo M, Silva FJ, Tamames J, Viguera E, Latorre A, Valencia A, Morán F, Moya A. Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl. Acad. Sci. U.S.A.* 2003;100(2):581-586.
- van Hoek MJA, Hogeweg P. The role of mutational dynamics in genome shrinkage. *Mol. Biol. Evol.* 2007;24(11):2485-2494.
- Weinstein-Fischer D, Altuvia S. Differential regulation of *Escherichia coli* topoisomerase I by Fis. *Mol. Microbiol.* 2007;63(4):1131-1144.
- Wernegreen JJ. Genome evolution in bacterial endosymbionts of insects. *Nat. Rev. Genet.* 2002;3(11):850-861.
- Wilkinson TL, Fukatsu T, Ishikawa H. Transmission of symbiotic bacteria *Buchnera* to parthenogenetic embryos in the aphid *Acyrtosiphon pisum* (Hemiptera: Aphidoidea). *Arthropod Struct Dev.* 2003;32(2-3):241-245.
- Wixon J. Featured Organism: Reductive Evolution in Bacteria: *Buchnera* sp., *Rickettsia prowazekii* and *Mycobacterium leprae*. *Comp. Funct. Genomics.* 2001;2(1):44-48.
- Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, Chen W, Wodke JAH, Güell M, Martínez S, Bourgeois R, Kühner S, Raineri E, Letunic I, Kalinina OV, Rode M, Herrmann R, Gutiérrez-Gallego R, Russell RB, Gavin A, Bork P, Serrano L. Impact of genome reduction on bacterial metabolism. *Science* 27 November 2009: Vol. 326. no. 5957, pp. 1263 - 1268.

ANEXOS

CLAVES DE INTERPRETACIÓN DE LAS TABLAS

En las Tablas se presenta el gen con el nombre que recibe en *Escherichia coli*, la función del gen y la presencia/ausencia en los organismos considerados en el estudio. La casilla en blanco significa que no se ha encontrado gen homólogo respecto al considerado en *E. coli*, con el número asignado en la secuenciación del genoma.

1.ABREVIATURAS MICRORGANISMOS:

- B_APS: *Buchnera aphidicola* str. APS
- B_Bp: *Buchnera aphidicola* str. BP (Baigonzia Pistaceae)
- B_Cc: *Buchnera aphidicola* str. Cc (*Cinara cedri*)
- B_Sg: *Buchnera aphidicola* str. APS
- Bfl: *Blochmannia floridanus*
- BPEN: *Blochmannia pennsylvanicus*
- Rma: *Ruthia magnifica*
- Vok: *Vesicomysocius okutanii*
- Wgl: *Wigglesworthia glossinidae*
- Bci: *Baumannia cicadelifolia*
- Sgl: *Sodalis glossinidae*
- W_Bm: *Wolbachia endosymbiont Brujia malayi*
- W_C: *Wolbachia endosymbiont Culex*
- W_Dm: *Wolbachia endosymbiont Drosophila melanogaster*
- E_min: *Elusimicrobium minutum*

- Sulfurov: *Sulfurovorum*
- C_rudii: *Carsonella ruddii*
- Sulcia: *Sulcia muelleri*
- Hodgkinia: *Hodgkinia cicadelifolia*

Tablas 3, 4, 5 y 6 (contactar en caso de interés en los datos: eviguera@uma.es; rociogs2006@gmail.com)

Agradecimientos

Y por último terminar esta memoria agradeciendo a todos los que me han ayudado a la realización de esta memoria. Este trabajo no hubiera sido posible sin la ayuda del Dr. Enrique Viguera que ha estado ahí sin rechistar cuando enviaba correos en Navidades, Semana Santa y vacaciones de verano, que era cuando me era posible seguir con este trabajo, de ahí que se haya alargado por más de tres años...Gracias Enrique por confiar que se terminaría algún día sin mostrar desconfianza en algún momento.

Por otro lado, quiero señalar toda la ayuda prestada por el personal de la UNIA en Málaga, sobre todo a Rocío González Aguilar, por todos los mails, las solicitudes de prórroga, etc. Nunca hubo una negativa ni una mala palabra de su parte. Con gente así da gusto hablar.

También me gustaría agradecer a mis compañeros de la Unidad de Biocomputación del CNB que me han prestado ayuda durante la realización de este trabajo, a Javi Díez por enseñarme algunos servidores de comparación múltiple de secuencias y a todos en general. También a mis anteriores compis de laboratorio de Badajoz que siempre me han animado para que termine esto.

Volviendo la vista atrás, ya unos cuantos años, quería agradecer a mis compañeros y profesores del Máster, a los que me ayudaron a tener los ejercicios hechos para el fin de semana, por todo lo que aprendí, por demostrar que puede haber compañerismo sin vernos las caras ...

Y sí, no me olvido de vosotros, que sois mi apoyo diario y los que sabéis perfectamente el esfuerzo que ha conllevado esto. A Miguel por el día a día, por su ayuda y su apoyo...y mis padres y a mi hermana, por todo su cariño y su apoyo incondicional, a ellos sólo puedo decir GRACIAS.