

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**Facultad de Ciencias**

**Análisis Evolutivo de la Proteína UvrA  
de Reparación de DNA por Escisión de Nucleótidos  
en Genomas Completamente Secuenciados**

**T E S I S**  
**que para obtener el título de**  
**B I Ó L O G O**  
**p r e s e n t a:**

**German Bonilla Rosso**

**Director de Tesis**  
**Dr. Antonio Eusebio Lazcano-Araujo Reyes**

**2004**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

*A Vicky y Germán por la materia prima  
A Vicky y Guillermo por el contenido*

*A don Mario Rosso, por el rizo de locura en la frente italiana.  
te hubiera encantado el siglo, viejo loco*

*A Dios, que de existir , nada tendría sentido.*

Gracias...

Paola, porque a pesar del tanto blanco y tanto morado, todavía tenemos el clima y los vegetales. Porque mataste mis arlequines y me paraste cuerdo y entero; porque me has enseñado gran parte de lo que es vivir de a de veras y por reparar mis espejos rotos. Y más que nada, por el plasma intangible ese que Dios nos envidia porque no se le ocurrió a él primero.

Eduwigis, por la terquedad, la carne de gallina y el carácter que nos sube el color al rostro. Gracias por todo eso que hoy prefieres no haberlo inculcado, gracias por esa vida tuya demasiado entregada pagando en tiempos y sueños para que yo pudiera escribir éste trabajo que a medias es tuyo.

Guillermo, por la paciencia que no aprendí, la libertad espiritual que sí aprendí y el único apoyo incondicional que he conocido. Pocos escogen a sus hijos y yo sí te escogía de nuevo.

Germán, por el idealismo inútil y utópico que me mantiene vivo; de no haber muerto, quién sabe quién sería.

Queta por las migas de domingo y por toda la sapiencia existencial que todavía no le aprendo.

## Agradecimientos Académicos...

Al master de masters Gerardo Rivas Lechuga, por hacer ciencia sin sacrificar integridad

A mi maestro Jedi Zenón Cano Santana, el rigor científico de un chapulín no cabe en corazones de actuario

A Su Alteza Serenísima Toño Lazcano, a los autodidactas aislados nos sale mejor eso que hace la ciencia

A Álvaro Chaos, que me enseñó que al final no somos nada....

A Luis Delaye, porque The Devil Is In The Details

A Ana y Edgardo, por el tinto en cuernavaca

Al Shaolin y al Cuervo por su ayuda con la programación

A Nadia por la ayuda con las imágenes

A mi mamá y a Massyel por la ayuda con las citas

A Probetel por el apoyo económico durante el desarrollo de éste trabajo.

Nothing is true,  
everything is possible  
----*Hassan I Sabbah*

Nous ne sommes pas encore nés  
Nous ne sommes pas encore au monde,  
Il n'y a pas encore de monde  
Les choses ne sont pas encore faites.  
La raison d'être n'est pas encore trouvée  
--- *Antonin Artaud*

...la Biblioteca es total y sus anaqueles registran todas las posibles combinaciones de los veintitantos símbolos ortográficos, o sea todo lo que es dable expresar.  
----*Jorge Luis Borges*

nec natura potest iusto discernere iniquum,  
dividit ut bona diversis, nec vincet ratio hoc.  
---*Horacio*



## CONTENIDO

<b>RESUMEN</b>	<b>0</b>
<b>I. INTRODUCCIÓN</b>	<b>1</b>
<b>II. METODOLOGÍA</b>	<b>10</b>
<b>III. RESULTADOS</b>	<b>16</b>
<b>IV. DISCUSIÓN</b>	<b>40</b>
<b>1. Plásmidos y Organelos Celulares</b>	<b>40</b>
<b>2. Genomas Bacterianos Carentes de UvrA</b>	<b>41</b>
<b>3. Evolución Temprana de UvrA</b>	<b>42</b>
<b>4. Reconstrucción Filogenética de la Familia UvrA</b>	<b>44</b>
<b>4.1 UvrA <i>sensu stricto</i></b>	<b>45</b>
<b>4.2 DrrC</b>	<b>48</b>
<b>4.3 UvrA2</b>	<b>51</b>
<b>4.4 Uvr<math>\beta</math></b>	<b>54</b>
<b>4.5 UvrChlam</b>	<b>59</b>
<b>4.6 UvrPlmy</b>	<b>60</b>
<b>4.7 Proteínas huérfanas</b>	<b>64</b>
<b>5. UvrA en Archaea</b>	<b>66</b>
<b>6. Estructura Terciaria</b>	<b>68</b>
<b>7. Modelo Funcional</b>	<b>71</b>
<b>8. Relaciones Evolutivas con otras Familias</b>	<b>73</b>
<b>9. Historia Evolutiva de la Familia UvrA</b>	<b>80</b>
<b>V. CONCLUSIONES</b>	<b>84</b>
<b>VI. APÉNDICE</b>	<b>86</b>
<b>VII. LITERATURA CITADA</b>	<b>95</b>

## RESUMEN

La reparación del DNA es uno de los procesos fundamentales de los seres vivos, y una gran cantidad de lesiones son reparadas mediante la ruta de escisión de nucleótidos (REN). En bacterias, la identificación de la lesión es realizada por UvrA, un miembro de la superfamilia de transportadores ABC sin dominio transmembranal. Utilizando modelos ocultos de Markov, buscamos todos los homólogos de las proteínas UvrA caracterizadas experimentalmente en todos los genomas con secuencia completa. La proteína se encuentra ausente en plásmidos, genomas de eucariontes, organelos, y endosimbiontes obligados. La encontramos en cinco genomas arqueanos presumiblemente por transferencia horizontal. El análisis de todos los homólogos encontrados indica la existencia de al menos cuatro subfamilias naturales: (a) **UvrA** *sensu stricto*, que contiene todos los ortólogos a UvrA, (b) **DrrC**, involucrada en la resistencia a daunorribucina, (c) **UvrA2**, homólogas a DRA0188 de *D. radiodurans*, y (d) **UvrAβ**, exclusiva de las β-proteobacterias. La filogenia de UvrA *sensu stricto* conserva los principales linajes bacterianos y apoya la existencia del grupo Chlamydiae-Planctomycetales. El análisis de secuencias apoya la existencia de tres unidades funcionales como fue reportado por Doolittle *et al.* (1986), y nos permite formular un nuevo modelo evolutivo. La extrapolación de la estructura terciaria a partir de la comparación con los transportadores ABC cristalizados y la predicción de la estructura secundaria sugiere la interacción de las unidades X-Y de un mismo monómero, y la interacción de las unidades Z de las dos unidades del homodímero, cada interacción mediada por un ion de Zinc coordinado por un dedo de Zinc. Finalmente, la presencia y organización (sintenia) de los motivos descritos, no apoyan la existencia de una superfamilia de proteínas de unión a nucleótidos similares a UvrA que incluyen a la familia de proteínas eucariontes y arqueanas Rad50 y la familia SbcC de proteínas bacterianas propuestas por Gorbalenya y Koonin (1990). Éstas tres familias de proteínas (UvrA, SbcC y Rad50) simplemente pertenecen a la superfamilia de transportadores ABC.

## I. INTRODUCCIÓN

Uno de los procesos fundamentales en todos los seres vivos es la reparación de daños en el DNA, generados tanto por su inestabilidad natural, como por la acción de metabolitos celulares, reactivos químicos y diversos tipos de radiación. Hay ciertos daños que, de no ser reparados, pueden causar mutaciones puntuales y éstas a su vez pueden ser pasadas por alto e incluso resultan importantes como fuente de variación. Sin embargo, existen otro tipo de daños que resultan letales al modificar la estructura del DNA y bloquear su replicación o transcripción. Un gran número de lesiones de ésta naturaleza y sin relación entre ellas es reparado en los tres dominios celulares mediante la llamada vía de Reparación por Escisión de Nucleótidos (NER por sus siglas en inglés). Ésta vía se caracteriza por la eliminación del oligonucleótido dañado por medio de un corte a ambos lados de la lesión y después su retiro en una función de excinucleasa y no endonucleasa. En las Bacterias, la NER depende de seis proteínas (UvrA, UvrB, UvrC, UvrD, DNAPolII, DNA ligasa). El proceso es ATP-dependiente y sigue los siguientes pasos (al parecer análogos al NER eucarionte) (Goosen y Moolenaar, 2001) (Fig. 1):

1. Los dímeros de UvrA (posiblemente formando un complejo con UvrB) exploran constantemente el DNA en busca de lesiones, evaluando varios parámetros helicoidales como flexibilidad, desenrollamiento o la apilación de las bases.
2. Al encontrar una lesión, UvrB se une en forma estable al sitio dañado y el dímero de UvrA es liberado con la hidrólisis de una molécula de ATP (Thiagalingam y Grossman, 1991)

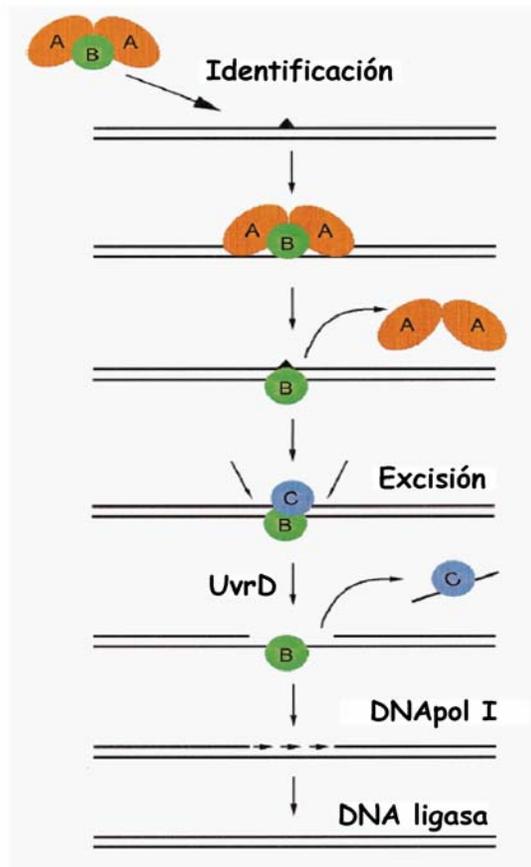


Figura 1. Esquema del mecanismo NER bacteriano. Las letras corresponden a UvrA, UvrB y UvrC. (Tomado de Goosen y Moolenaar, 2001)

3. La UvrC se une al complejo de preincisión UvrB-DNA, y realiza primero un corte en el 4<sup>to</sup>/5<sup>to</sup> enlace fosfodiéster en dirección 3' de la lesión, y después en el 8<sup>vo</sup> enlace fosfodiéster en dirección 5' de la lesión (Verhoeven *et al.*, 2000).
4. La helicasa UvrD retira el oligonucleótido dañado con UvrC.
5. El espacio es repolimerizado por la DNApol I
6. Los fragmentos son unidos por la DNA ligasa.

La fase crucial durante la NER sin duda es el reconocimiento de la lesión por la proteína UvrA, pues debe presentar una especificidad tal que le permita identificar un amplio espectro de lesiones como cambios en una sola base, distorsiones voluminosas, modificaciones no covalentes, enlaces intra e intercadenas (Goosen y Moolenaar, 2001), sin tener una afinidad hipersensible que genere escisiones aleatorias en regiones no dañadas del DNA, y sin una disminución a su sensibilidad lo que la llevaría a pasar por alto algunas lesiones.

**UvrA:** La proteína UvrA es una proteína de unión a DNA con actividad ATPasa que en presencia de ATP tiene mayor afinidad por el DNA dañado que por DNA sin daño (Seeberg y Steinum, 1982). La secuencia de UvrA en *Escherichia coli* tiene 2820pb y codifica para una proteína de 940 aminoácidos (Husain *et al.*, 1986). Al igual que otras

proteínas de reparación de DNA, se encuentran 20-25 moléculas de UvrA por célula (Sancar *et al.*, 1981; Husain *et al.*, 1986), cantidad que puede incrementarse hasta 10 veces durante la respuesta SOS (Kenyon y Walker, 1981; Markham *et al.*, 1984). El gen es regulado por el represor *lexA* (Sancar *et al.*, 1982). Por lo tanto, el bajo número de moléculas se debe a la represión de la transcripción de *uvrA* por *lexA* (Husain *et al.*, 1986).

UvrA es un miembro de la superfamilia de módulos de unión a nucleótidos de transportadores ABC (Higgins *et al.*, 1986; Saurin *et al.*, 1999) que tiene actividad de ATPasa (Oh *et al.*, 1989) y presenta dos sitios de unión de ATP que corresponden a los motivos de unión a nucleótidos tipo WalkerA (WA, consenso GxxGxGKST) y WalkerB (WB, consenso DE/D) (Walker *et al.*, 1982; Gorbalenya y Koonin, 1990). Asimismo, presenta la firma de la familia ABC, (consenso LSGGQ), determinada originalmente como péptido de unión (Ames *et al.*, 1990; Higgins *et al.*, 1988).

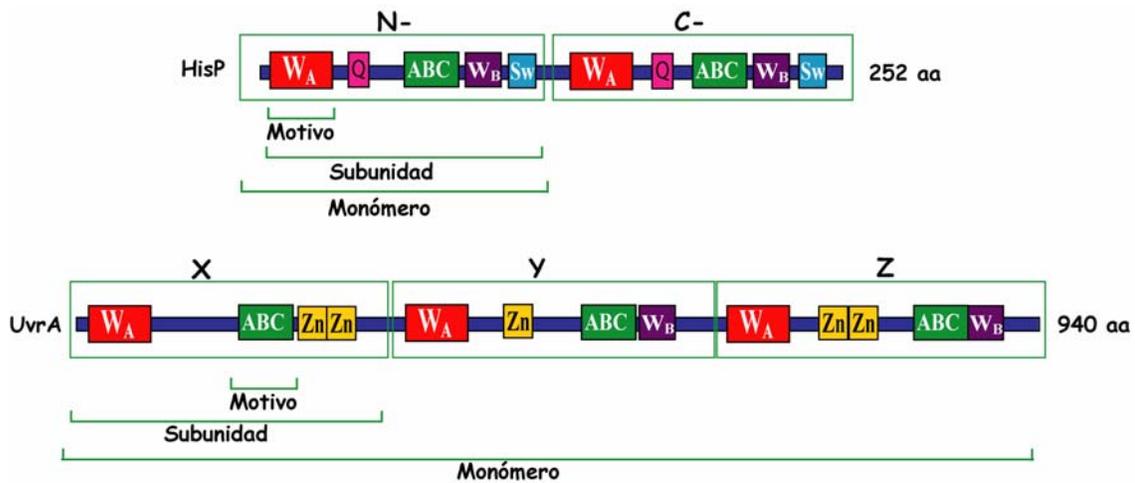


Figura 2. Estructura general de la proteína UvrA, ilustrando su naturaleza tripartita y la organización de los motivos que posee. En la parte superior se esquematiza la estructura general del transportador de histidina "HisP". En ambos casos se define gráficamente la terminología utilizada en el texto.

La proteína posee tres regiones homólogas entre sí que corresponden a tres subunidades, cada una de las cuales posee un WA, un WB y la firma ABC (Doolittle *et al.*, 1986) (Fig. 2). Lo que la diferencia de los módulos de transporte es la presencia de dos motivos relacionados con la unión a DNA y unión a metales conocidos como dedos de zinc (consenso CxxC-(x)n-CxxC) (Doolittle *et al.*, 1986; Miller *et al.*, 1985; Berg, 1986) y una “mitad” (CxxC) (Doolittle *et al.*, 1986). La región del ABC-WB de la subunidad X (ABC<sub>X</sub>-WB<sub>X</sub>) y la región del WA de la subunidad Y (WA<sub>Y</sub>) se encuentran sumamente degeneradas (Doolittle *et al.*, 1986), lo que llevó a proponer que uno de los sitios con actividad ATPasa en UvrA se forma a partir de los segmentos WA<sub>X</sub> con ABC<sub>Y</sub>/WB<sub>Y</sub> y el otro a partir de WA<sub>Z</sub> con ABC<sub>Z</sub>/WB<sub>Z</sub> (Myles *et al.*, 1991), y que UvrA se comporta funcionalmente como un homodímero en el segmento amino-terminal (subunidades X-Y), mientras que el segmento carboxilo terminal (subunidad Z) no se dimeriza o no puede hacerlo independientemente y pierde su actividad de ATPasa (Myles y Sancar, 1991).

Los motivos funcionales WA<sub>X</sub> y WA<sub>Z</sub> pueden hidrolizar el ATP independientemente, pero no están involucrados en la unión del nucleótido (Thiagalingam y Grossman, 1991). Cada subunidad juega un papel diferente en el proceso de escisión: la subunidad Z tiene una mayor afinidad por la unión de ATP que las subunidades X-Y y está involucrado en la disociación de la unión del complejo nucleoprotéico de los sitios sin lesión (Thiagalingam y Grossman, 1991). En cambio, el segmento N-terminal está involucrado en la dimerización de la proteína inducida por la unión de ATP, pero no en la supervivencia a radiación UV (Myles *et al.*, 1991; Navaratnam *et al.*, 1989) ni en la unión a DNA (Wang *et al.*, 1994). Recientemente se ha propuesto su participación en un nuevo proceso alternativo de replicación en ausencia de DNAPolI (Moolenaar *et al.*, 2000a)

**Dedos de Zinc:** Los “dedos de zinc” (ZnF) son motivos que contienen pares de Cys e His coordinando iones de zinc se denominan y son probablemente el motivo más común y versátil (Matthews y Sunde, 2002), pero todos están involucrados en la interacción de proteínas con otras biomoléculas, incluyendo DNA, RNA, proteínas y lípidos (Matthews y Sunde, 2002). Originalmente el término fue aplicado a las repeticiones de ~30 residuos descubiertas en el factor de transcripción TFIIIA de *Xenopus laevis* (Miller, 1985), actualmente el término es utilizado para identificar cualquier dominio compacto estabilizado por un ion de zinc (Krishna *et al.*, 2003), generalmente caracterizado en su estructura primaria por repeticiones de Cys e His principalmente, y en ocasiones de los ácidos Asp y Glu (Vallee *et al.*, 1991). Los dedos de zinc tienen un número relativamente pequeño de residuos conservados fuera de las cisteínas e histidinas (Michael *et al.*, 1992), pues toda la estabilidad estructural es provista por la coordinación del zinc y el núcleo de residuos conservados que rodean el sitio de unión a zinc. De aquí se ha derivado la necesidad de construir una clasificación natural y funcional de los dedos de zinc, pero la tarea se ha complicado porque la estructura primaria no es suficiente para detectarlos (Berg, 1986), y es necesario evidencia experimental y/o cristalográfica (Krishna *et al.*, 2003), e incluso contando con la estructura tridimensional de la proteína parecen formar dominios estructurales muy diferentes (Grishin, 2001).

En UvrA de *E. coli*, cada uno de los átomos de zinc está coordinado por cuatro Cys en una distancia de  $2.32 \pm 0.2 \text{ \AA}$  (Navaratnam *et al.*, 1989). El dedo de Zinc del segmento C-terminal (Zn3) se encuentra entre los 740 – 766 residuos de aminoácidos, y está involucrado en la resistencia a radiación UV posiblemente presentando un sitio de unión a DNA, pues existen ocho residuos básicos a su alrededor (R736, R739, K750, H754, K767, K769, R770, R772) que podrían formar un surco de unión al esqueleto de fosfatos de la

cadena de DNA (Wang *et al.*, 1994). El dedo de Zinc del segmento N-terminal (Zn1) se encuentra entre los residuos de aminoácidos 253-280 en y no es determinante para la resistencia a UV (Navaratnam *et al.*, 1989), pero es esencial para el proceso alterno de replicación, de manera que se ha propuesto que ha evolucionado específicamente para ésta función (*sic* Moolenaar *et al.*, 2000a).

**Homología con transportadores ABC:** Los transportadores ABC cuentan con un módulo transmembranal (TM) y un módulo de unión a ATP ((ATP-Binding Cassette). Los miembros de la superfamilia ABC son en su mayoría éstos módulos ABC y se encuentran conservados entre transportadores y otras proteínas con funciones celulares como la división celular (FtsE), la nodulación en *Rhizobium* (NodI), factores de elongación en *Saccharomyces* (EF3) y reparación de DNA en bacterias (UvrA) (Higgins *et al.*, 1986).

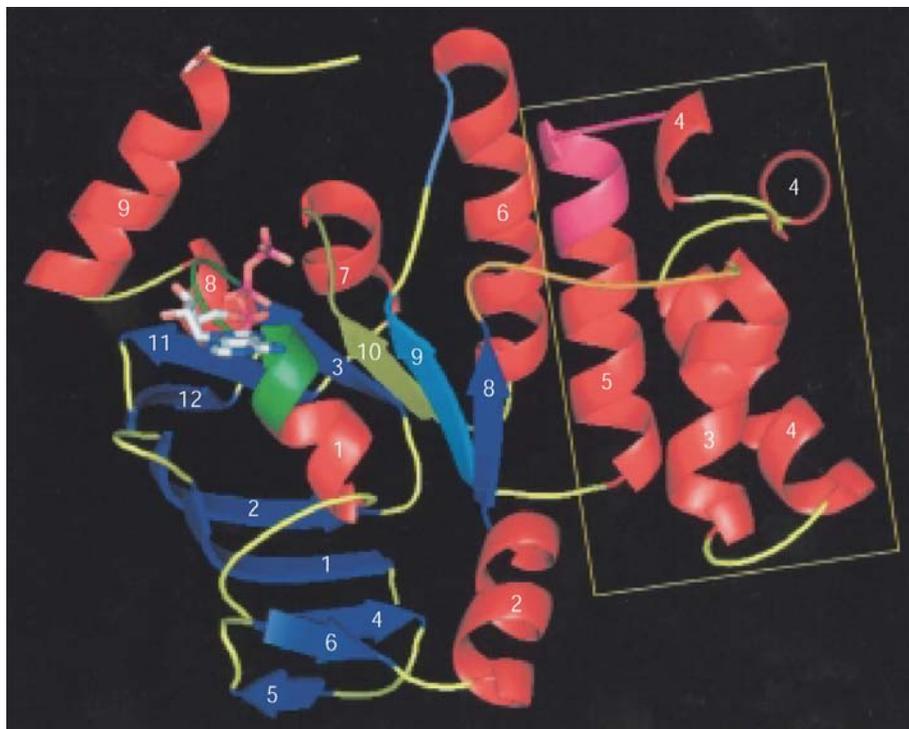


Figura 3. Estructura terciaria del módulo ABC del transportador de histidina HisP. Los motivos WalkerA, WalkerB y la firma ABC están resaltados. En el recuadro se indica la región helicoidal. (Tomado de Schmitt y Tampé, 2002).

Los miembros de ésta superfamilia acoplan la energía liberada por la hidrólisis del ATP al transporte de solutos en contra de un gradiente de concentración. Funcionan generalmente a manera de un complejo homo/hetero-dimérico y transportan una sorprendente diversidad de substratos (de iones y aminoácidos a polipéptidos y ácidos grasos ramificados) (Higgins, 1992).

Existen sólo dos cristales de transportadores ABC resueltos en ausencia de sustrato y nucleótidos a la fecha: el MsbA de *E.coli* cristalizado por Chang y Roth (2001), y el importador de vitamina B12 BtuCD de *E.coli* cristalizado por Locher *et al.*(2002). La organización de los módulos ABC se ilustra en las Figuras 2 y 3. Cada módulo presenta los sitios de unión de nucleótidos WA y WB (Walker *et al.*, 1982; Higgins *et al.*, 1986); la firma ABC (Ames *et al.*, 1990; Higgins *et al.*, 1988), que parece estar involucrado en el acoplamiento de la hidrólisis del ATP al transporte (Ames y Lecar, 1992); los residuos que conforman el motivo entre el Walker A y la firma ABC, y que interactúan directamente con el dominio transmembranal conocido como Q-loop ( FQx<sub>7</sub>); y, finalmente, el pequeño motivo localizado hacia el extremo C-terminal del WB conocido como Switch II (SW), caracterizado por una Tre y una His prácticamente invariables precedida de cinco de los siguientes aminoácidos: Val, Leu, Ile o Met ( [V/L/M/I]<sub>5</sub>TH ), que podría actuar como el sensor de fosfatos- $\gamma$  del ATP (Schmitt y Tampé, 2002) o estar involucrado en la propagación de los cambios conformacionales desatados por su hidrólisis (Yoshida y Amano, 1995).

Asimismo, entre el Q-loop y la firma ABC existe una región que a pesar de tener muy pocos residuos conservados, en la estructura secundaria tiene  $\alpha$ -hélices bien conservadas en cuanto a número y longitud; y por lo tanto se le conoce como el dominio helicoidal

(Schneider y Hunke, 1998). El dominio helicoidal interactúa con tres residuos bien conservados del dominio transmembranal conocido como el loop EAA (Mourez *et al.*, 1997) y parece conferir especificidad al proveer un sitio de interacción único entre el dominio transmembranal y el módulo ABC particular del sistema (Schneider y Hunke, 1998).

Las proteínas UvrA de reparación de DNA han sido consideradas como miembros de la superfamilia de los transportadores ABC por su homología en los motivos WA, WB y la firma ABC (Higgins *et al.*, 1986; 1988). Sin embargo, Gorbalenya y Koonin (1990) proponen una nueva superfamilia de proteínas de unión a nucleótidos trifosfatados (NTP) relacionadas con UvrA. Ésta superfamilia presenta cuatro segmentos conservados: WA, la firmaABC, WB y el SW. En esta superfamilia se incluyen (a) dos dominios de UvrA; (b) proteínas de replicación de fagos D13 y gp46; (c) las proteínas de recombinación de *E. coli* RecF y RecN; (d) la proteína de reparación de errores en la correspondencia de bases MutS en *E. coli* y HexA en *Streptococcus pneumoniae*; (e) la ATPasa DNA-dependiente del fago  $\lambda$  EA59; y (f) la proteína de reparación y replicación en *Saccharomyces cerevisiae* Rad50.

Las condiciones de la tierra primitiva sugieren que la reparación de daños al DNA por radiación UV pudo surgir tempranamente en la evolución de los sistemas vivos y por lo tanto presentarse como un buen marcador molecular para reconstruir filogenias profundas. A pesar de que UvrA no ha sido reportado en eucariontes, si la proteína UvrA surgió antes de la divergencia de los tres dominios celulares, puede estar presente en los genomas eucariontes y/o arqueanos, inclusive con una función diferente. Asimismo, UvrA presenta tres subunidades homólogas y por lo tanto también es un buen candidato para la reconstrucción de filogenias universales enraizadas con genes parálogos. En el presente trabajo se pretende reconstruir mediante métodos bioinformáticos la historia evolutiva de

---

las proteínas de reparación por escisión de nucleótidos UvrA de los genomas que están completamente secuenciados, analizando su distribución taxonómica y la organización de sus dominios, así como su relación evolutiva con los transportadores ABC y otras ATPasas involucradas en unión a DNA.

## II. METODOLOGÍA

La búsqueda en los genomas completos y la ejecución del programa se llevaron a cabo en la supercomputadora AlphaServer SC 45 Bakliz9 ([bakliz.supercomputo.unam.mx](http://bakliz.supercomputo.unam.mx)), con 32 procesadores Alpha EV68 a 1.25 GHz, en el Departamento de Supercómputo de la Dirección General de Servicios de Cómputo Académico (DGSCA). El resto de los análisis se realizó en un sistema con dos procesadores IntelPentiumIII<sup>®</sup> en plataforma DebianV ([microbio.fciencias.unam.mx](http://microbio.fciencias.unam.mx)) y los programas de plataforma Windows se realizaron en un sistema de procesador AMDAthlon<sup>®</sup> ([progenote.fciencias.unam.mx](http://progenote.fciencias.unam.mx)), ambos en el Laboratorio de Microbiología de la Facultad de Ciencias.

**Construcción de Alineaciones y Perfiles:** Se obtuvieron las secuencias de las proteínas UvrA que han sido caracterizadas experimentalmente y reportadas en la base de datos de módulos ABC ABCIISE v.3.0. (Bouige *et al.*, 2002; Dassa y Bouige, 2001), con fecha de corte del 5 de junio de 2003. Con éstas secuencias se construyó una alineación múltiple utilizando ClustalX v.1.83 (Thompson *et al.*, 1997).

La alineación de las proteínas completas fue dividida con la ayuda de BioEdit v.5.9.0 (Hall, 1999) en tres alineaciones más pequeñas correspondientes a cada una de las subunidades de UvrA. Las subunidades para el análisis filogenético se definieron en la región mejor conservada de cada una, que corresponde a 30 residuos de aminoácidos antes de la primera Gly del WA ([-30]GxsGaGKST), como lo determinaron Saurin *et al.* (1999), y 15 residuos de aminoácidos después del último Asp del WB (DEptSalD[+15]), determinado a partir del análisis de miembros de todas las familias de transportadores ABC (nuestro laboratorio, sin publicar). Las alineaciones así obtenidas fueron reeditadas en

BioEdit v.5.9.0 (Hall, 1999), y con ellas se construyó y calibró un perfil de cada subunidad utilizando un algoritmo de Modelos Ocultos de Markov con la familia de programas HMMER (Eddy, 1998). El modelo se configuró para detectar dominios múltiples por secuencia, donde cada dominio puede ser una alineación local (fragmentaria, ver Apéndice), es decir, una alineación local respecto al modelo y local respecto a la secuencia. El modelo fue configurado de ésta forma para detectar posibles homólogos de las subunidades de UvrA que se encontrasen formando parte de otras proteínas con dominios diferentes.

**Búsqueda en Genomas Completos:** Se buscaron homólogos de cada una de las tres unidades en cada uno de los 152 genomas completamente secuenciados disponibles hasta septiembre del 2003 en la Kyoto Encyclopedia of Genes and Genomes (Kanehisa, 1997; Kanehisa y Goto, 2000) y en la base de datos de cloroplastos, mitocondrias y plásmidos que se presentan en la Tabla 1.

De cada secuencia así localizada se tomó el dominio de interés, y se incorporó al perfil original para reiterar la búsqueda hasta obtener todos los miembros contenidos en cada genoma. A partir del análisis y calibración de perfiles de 43 familias de módulos de unión a nucleótidos de transportadores ABC (nuestro laboratorio, sin publicar), se determinó un método empírico para incorporar automáticamente al perfil las secuencias homólogas a la subfamilia inicial excluyendo las secuencias también homólogas pero pertenecientes a otras subfamilias. En el método implementado, la incorporación de secuencias al perfil se detiene cuando el cociente entre el valor-E de una secuencia y la siguiente es mayor a  $1 \times 10^{-15}$ .

Tabla 1. Lista de los genomas completos analizados en el presente trabajo. Los genomas se encuentran agrupados de acuerdo al grupo taxonómico al que pertenecen. Los genomas señalados con (\*) son aquellos que fueron analizados durante la búsqueda final. T/D = *Thermus/Deinococcus*; A/T = Aquificales/Thermotogales; PI = Planctomycetales; F = Fusobacteria; Chl = Chlorobia; Bact = Bacteroidetes; Na = Nanoarchaeota.

BACTERIA		BACTERIA		ARCHAEA	
<i>Agrobacterium tumefaciens</i>	atc	<i>Bacillus anthracis</i>	ban	<i>Archaeoglobus fulgidus</i>	aae
<i>Agrobacterium tumefaciens_c</i>	atd	<i>Bacillus cereus</i>	bce	<i>Methanococcus jannaschii</i>	mja
<i>Agrobacterium tumefaciens_d</i>	atu	<i>Bacillus halodurans</i>	bha	* <i>Methanococcus marisaludis</i>	mmp
<i>Bradyrhizobium japonicum</i>	bja	<i>Bacillus subtilis</i>	bsu	<i>Methanosarcina acetivorans</i>	mac
<i>Brucella melitensis</i>	bme	* <i>Bacillus cereus</i> ATCC10987	bca	<i>Methanosarcina mazei</i>	mma
<i>Brucella suis</i>	bms	<i>Listeria innocua</i>	lin	<i>Halobacterium sp.</i>	hal
<i>Caulobacter crescentus</i>	ccr	<i>Listeria monocytogenes</i>	lmo	<i>Methanobacterium thermoautotrophicum</i>	mth
<i>Mesorhizobium loti</i>	mlo	<i>Staphylococcus aureus_mu50</i>	sav	<i>Methanopyrus kandleri</i>	mka
<i>Rickettsia conorii</i>	rco	<i>Staphylococcus aureus_mw2</i>	sam	<i>Pyrococcus abyssii</i>	pab
* <i>Wolbachia wMel</i>	wol	<i>Staphylococcus aureus_n315</i>	sau	<i>Pyrococcus furiosus</i>	pfu
* <i>Rhodopseudomonas palustris</i>	rpa	<i>Staphylococcus epidermidis</i>	sep	<i>Pyrococcus horikoshii</i>	pho
<i>Rickettsia prowazekii</i>	rpr	<i>Oceanobacillus ihayensis</i>	oih	<i>Thermoplasma acidophilum</i>	tac
<i>Sinorhizobium meliloti</i>	sme	<i>Lactobacillus plantarum</i>	lpl	<i>Thermoplasma volcanium</i>	tvo
<i>Bordetella bronchisepticum</i>	bbr	<i>Lactococcus lactis</i>	lla	<i>Sulfolobus solfataricus</i>	sso
<i>Bordetella parapertussis</i>	bpa	* <i>Lactococcus johnsonii</i>	ljo	<i>Sulfolobus tokodaii</i>	sto
<i>Bordetella pertussis</i>	bpe	<i>Streptococcus agalactiae</i>	sag	<i>Aeropyrum pernix</i>	ape
<i>Ralstonia solanacearum</i>	rso	<i>Streptococcus agalactiae_nem316</i>	san	<i>Pyrobaculum aerophilum</i>	pae
<i>Neisseria meningitidis</i>	nme	<i>Streptococcus pneumoniae</i>	spn	* <i>Nanoarchaeum equitans</i>	neq
<i>Neisseria meningitidis_A</i>	nma	<i>Streptococcus pneumoniae_r6</i>	spr		
<i>Nitrosomonas europaea</i>	neu	<i>Streptococcus pyogenes</i>	spy		
* <i>Chromobacterium violaceum</i>	cvi	<i>Streptococcus pyogenes_m18</i>	spm		
<i>Buchnera aphidicola APS</i>	buc	<i>Streptococcus pyogenes_m3</i>	spg		
<i>Buchnera aphidicola BP</i>	bab	<i>Streptococcus pyogenes_ssi1</i>	sps		
<i>Buchnera aphidicola SG</i>	bas	<i>Streptococcus xmutans</i>	smu		
* <i>Blochmania floricola</i>	bfl	<i>Enterococcus faecalis</i>	efa		
<i>Coxiella burnetii</i>	cbu	<i>Clostridium acetobutylicum</i>	cac		
<i>Escherichia coli</i>	ecc	<i>Clostridium perfringens</i>	cpe		
<i>Escherichia coli_cft073</i>	ece	<i>Clostridium tetani</i>	ctc		
<i>Escherichia coli_j</i>	ecj	<i>Thermoanaerobacter tengcongensis</i>	tte		
<i>Escherichia coli_o157</i>	eco	<i>Mycoplasma gallisepticum</i>	mga		
<i>Escherichia coli_o157j</i>	ecs	<i>Mycoplasma genitalium</i>	mge		
<i>Haemophilus ducreyi</i>	hdu	<i>Mycoplasma penetrans</i>	mpe		
<i>Haemophilus influenzae</i>	hin	<i>Mycoplasma pneumoniae</i>	mpn		
<i>Pasteurella multocida</i>	pmu	* <i>Mycoplasma mycoides</i>	myy		
<i>Phytoplasma sp.</i>	poy	<i>Mycoplasma pulmonis</i>	mpu		
<i>Pseudomonas aeruginosa</i>	pae	<i>Ureaplasma urealyticum</i>	uur		
<i>Pseudomonas putida</i>	ppu	<i>Rhodopirellula baltica</i>	rba		
* <i>Photorhabdus luminiscens</i>	plu	<i>Chlamydia muridarum</i>	cmu		
<i>Pseudomonas syringae</i>	pst	<i>Chlamydia trachomatis</i>	ctr		
<i>Salmonella typhi</i>	sty	<i>Chlamydia pneumoniae</i>	cca		
<i>Salmonella typhi_ty2</i>	ett	<i>Chlamydia pneumoniae</i>	cpn		
<i>Salmonella typhimurium</i>	stm	<i>Chlamydia pneumoniae_ar39</i>	cpa		
<i>Shewanella oneidensis</i>	son	<i>Chlamydia pneumoniae_j138</i>	cpj		
<i>Shigella flexneri</i>	efl	<i>Chlamydia pneumoniae_tw183</i>	cpt		
<i>Shigella flexneri_2457t</i>	sfx	<i>Mycobacterium bovis</i>	mbo		
<i>Vibrio cholerae</i>	vch	<i>Mycobacterium tuberculosis</i>	mtu		
<i>Vibrio parahaemolyticus</i>	vpa	<i>Mycobacterium avium</i>	mpa		
<i>Vibrio vulnificus</i>	vvu	<i>Mycobacterium tuberculosis C.</i>	mtc		
* <i>Vibrio vulnificus_YJ016</i>	vvv	<i>Mycobacterium leprae</i>	mle		
<i>Wigglesworthia brevipalpis</i>	wbr	<i>Corynebacterium efficiens</i>	cef		
<i>Xanthomonas axonopodis</i>	xac	<i>Corynebacterium glutamicum</i>	cgl		
<i>Xanthomonas campestris</i>	xcc	* <i>Corynebacterium diptheriae</i>	cdi		
<i>Xylella fastidiosa</i>	xfa	<i>Streptomyces coelicolor</i>	scs		
<i>Xylella fastidiosa_T</i>	xft	<i>Streptomyces avermitilis</i>	sma		
<i>Yersinia pestis</i>	ype	<i>Tropheryma whipplei</i>	twh		
<i>Yersinia pestis_kim</i>	ypk	<i>Tropheryma whipplei_S</i>	tws		
* <i>Yersinia pestis_mediev</i>	ypm	<i>Bifidobacterium longum</i>	blo		
<i>Geobacter sulfurreducens</i>	gsu	<i>Anabaena sp.</i>	ana		
* <i>Edelovibrio bacteriovorus</i>	bbs	<i>Synechocystis sp.</i>	syn		
<i>Helicobacter hepaticus</i>	hhe	<i>Thermosynechococcus elongatus</i>	tel		
<i>Helicobacter pylori</i>	hpy	<i>Prochlorococcus marinus</i>	pma		
* <i>Wolinella succinogenes</i>	wsu	* <i>Prochlorococcus marinus MED4</i>	pmm		
<i>Helicobacter pylori_j99</i>	hpj	* <i>Prochlorococcus marinus MIT9313</i>	pmt		
<i>Campylobacter jejuni</i>	cje	* <i>Gloeobacter violaceus</i>	gvi		
<i>Deinococcus radiodurans</i>	dra	* <i>Synechococcus sp. PCC6301</i>	syc		
<i>Aquifex aeolicus</i>	aae	* <i>Synechococcus sp. WH8102</i>	syw		
<i>Thermotoga maritima</i>	tma	<i>Fusobacterium nucleatum</i>	fnu		
<i>Borrelia burgdorferi</i>	bbu	<i>Chlorobium tepidum</i>	cte		
<i>Treponema pallidum</i>	tpa	<i>Bacteroides thetaiotaomicron</i>	bth		
<i>Leptospira interrogans</i>	lil	* <i>Porphyromonas gingivalis</i>	pgi		
* <i>Leptospira interrogans C</i>	lic				
* <i>Treponema denticola</i>	tde				

Se elaboró un programa en el Practical Extraction and Report Language (Perl) para la automatización de la búsqueda reiterativa en los genomas completos y el corte y reincorporación de los dominios (Apéndice). La frecuencia de aminoácidos representados en todas las proteínas obtenidas en la alineación múltiple fue analizada con el programa Molecular Evolutionary Genetics Analysis v.2.1 (MEGA2, Kumar *et al.*, 2001).

**Análisis de las Nuevas Unidades:** Los dominios resultantes de la búsqueda fueron revisados manualmente. Cada una de las proteínas que aparecieron en cualquiera de las búsquedas fue analizada manualmente. Se construyó una alineación múltiple utilizando T-COFFEE (Notredame *et al.*, 2001), se editó, ajustó y cortó con BioEdit v.5.9.0 (Hall, 1999).

**Reconstrucción Filogenética de la Familia:** Se realizó un análisis filogenético de las proteínas completas con el programa MEGA2 (Kumar *et al.*, 2001) por Neighbour-Joining (Saitou y Nei, 1987), utilizando la corrección de Poisson y un Bootstrap de 10000 repeticiones (Felsenstein, 1985).

**Asignación a Subfamilias:** Las nuevas proteínas se asignaron directamente a una subfamilia cuando aparecían anidadas en clados con miembros conocidos y valores de bootstrap por encima de 75. En los casos en que esto no ocurrió, se realizó una inspección visual detallada.

**Análisis de Duplicaciones:** En los casos que lo requirieron, se utilizó el software en línea DotLet (<http://www.isrec.isb-sib.ch/java/dotlet/Dotlet.html>, Junier y Pagni, 2000) con una matriz de Gonnet, y la base de datos de dominios ProDom (Servant *et al.*, 2002) para identificar los posibles eventos de duplicación.

**Historia Evolutiva:** Se tomaron los miembros más representativos de cada clado de proteínas UvrA y se realizó una alineación múltiple sensible a alineaciones locales (T-COFFEE (Notredame *et al.*, 2001). La alineación se cortó sólo 30 aminoácidos antes del WA sin eliminar ninguna parte de la secuencia, porque éstas regiones también presentaron residuos conservados. Las alineaciones se realinearon independientemente utilizando ClustalX v.1.83 (Thompson *et al.*, 1997) y después se incorporaron a una sola alineación utilizando la opción de alineación por perfiles del T-COFFEE (Notredame *et al.*, 2001). La alineación final fue inspeccionada y editada en BioEdit v.5.9.0 (Hall, 1999). Los árboles se enraizaron utilizando miembros de las subfamilias de módulos ABC de resistencia a antibióticos ART y del inhibidor L de la RNAsa RLI, pues pertenecen al mismo grupo de módulos ABC involucrados en funciones celulares diferentes al transporte (Saurin *et al.*, 1999).

**Relaciones Evolutivas con otras Proteínas:** Se realizó una alineación múltiple (ClustalX v.1.83, Thompson *et al.* 1997) entre las cuatro proteínas UvrA que conservan mejor todos los motivos y los representantes de las familias de proteínas involucradas en mantenimiento de cromosomas y telómeros, recombinación homóloga y asociación de cromátidas hermanas Rad50 y SbcC identificadas en la base de datos SwissProt (Boeckman *et al.*, 2003) para analizar las relaciones evolutivas entre los tres grupos.

**Estructura Secundaria:** Se predijo la estructura secundaria de las tres unidades de la UvrA eco\_b4058 independientemente, utilizando para ello el programa en línea JPred dentro de JNet (<http://www.compbio.dundee.ac.uk/~www-jpred/>, Cuff y Barton, 2000). Se realizó una alineación múltiple sensible a alineaciones locales (T-COFFEE (Notredame *et al.*, 2001) de siete transportadores ABC cristalizados (GlcV de *S. solfataricus*, HisP de *S. typhi*, MsbA de *V. cholerae*, LivG de *M. jannaschii*, HlyB y MalK de *E. coli* y Tap1 de *H.*

*sapiens*,) con cada una de las unidades de eco\_b4058 para determinar elementos de la estructura secundaria por medio de similitudes locales. Los resultados se analizaron y las figuras se prepararon con el programa PyMOL Molecular Graphics System v.93 (DeLano, 1998-2003).

**Búsqueda por Subfamilias:** Se construyeron alineaciones múltiples de cada unidad de cada subfamilia, y se buscaron de nuevo en los genomas completamente secuenciados disponibles en el KEGG hasta Marzo del 2004 con 24 nuevos genomas (Tabla 1). Los análisis filogenéticos de cada subfamilia que presentamos incluyen todas las proteínas encontradas en las dos etapas. Las nuevas proteínas fueron incluidas sólo en los análisis filogenéticos por subfamilia.

### III. RESULTADOS

La base de datos de módulos ABC ABCIISE v.3.0. reporta 61 proteínas homólogas a UvrA dentro de la subfamilia UVA de módulos ABC (Bouige *et al.*, 2002; Dassa y Bouige, 2001) de las cuales sólo seis están caracterizadas experimentalmente. De éstas, tres están involucradas en la reparación de lesiones de DNA (DR1771 en *D. radiodurans*: Agostini *et al.*, 1996; b4058 en *E. coli* K12:Husain *et al.*, 1986; STY4450 en *S. typhimurium*: Alberti,M., Li,Y.F., Sancar,A. y Hearst,J.E. 1992,. sin publicar), y las demás determinan resistencia a mitramicina en *Streptomyces argillaceus* (gi 1163120:Fernández *et al.*, 1996); a daunorrubicina en *S. peucetius* (DrrC:Lomovskaya *et al.*, 1996) y a nogalamicina en *S. nogalater* (SnorO:Ylihonko *et al.*, 1996; Palmu *et al.*, sin publicar). Para la construcción de los perfiles, se seleccionaron únicamente las tres subunidades de las tres proteínas caracterizadas experimentalmente como reparadoras de DNA, y como los perfiles suelen ser sensibles al número de secuencias incluidas en la alineación múltiple, se construyó un perfil alternativo con las cinco secuencias más similares a UvrA de *E.coli*, identificados visualmente a partir de una alineación múltiple aleatoria (gi154417 *S.typhimurium*; gi9654814 *V. cholerae*, gi1573215 *H. influenza*, gi12722387 *P.multocida*, gi6899057 *U. urealyticum*) corriendo en forma paralela al perfil de proteínas caracterizadas experimentalmente. Los dos perfiles arrojaron los mismos resultados al buscar homólogos en los genomas completos.

**Búsqueda de homólogos en los genomas completos.** Se identificaron 179 proteínas homólogas a UvrA en 143 genomas procariontes (Tabla 2), de las cuales 120 no están reportadas en ABCISSE. Es posible que no hayan sido detectadas con anterioridad debido al grado de sensibilidad de los métodos utilizados durante la búsqueda.

Tabla 2. Proteínas homólogas a UvrA obtenidas a partir de la búsqueda en los genomas completos. Las tres primeras letras corresponden a la clave KEGG del organismo al que pertenecen. Las claves resaltadas pertenecen a proteínas de longitud inusual (>1030aa), las claves en cursiva indican proteínas que fueron eliminadas del análisis.

Homólogos a UvrA (108)		
aae_aq_686	hpj_jhp0644	sfl_SF4146
ana_alr3716	hpy_HP0705	sfx_S3583
atc_AGR_C_2790	lic_LIC11717_	sme_SMc01235
atd_Atul513	lil_LA2212	smu_SMU.1851
atu_Atul513	lla_L0256	son_SO4030
ban_BA5395	mbo_Mb1664	spg_SpyM3_1577
bbu_BB0837	mga_MGA_0778	spm_spyM18_1890
bca_BC5271_	mge_MG421	spn_SP0186
bce_BC5167	mle_ML1392	spr_spr0171
bja_blr4702	mlo_mlr0750	sps_SPs0290
bme_BMEI0878	mmy_MSC_0943_	spy_SPy1825
bms_BR1104	mpa_MAP1341	stm_STM4254
bsu_BG12697	mpe_MYPE9100	stt_t4160
cbu_CBU0274	mpn_C12_orf948	sty_STY4450
<b>cca_CCA00676</b>	mpu_MYPU_7100	syc_syc0692_d
ccr_CC2590	mtc_Mt1675	syn_slr1844
cje_Cj0342c	mtu_Rv1638	syw_SYNW2522
<b>cmu_TC0610</b>	neu_NE2455	tde_TDE1010
<b>cpa_0678</b>	nma_NMA1159	tel_tll0184
cpe_CPE0348	nme_NMB0962	tma_TM0480
<b>cpj_uvrA</b>	pae_PA4234	tpa_TP0514
<b>cpn_CPn0096</b>	plu_plu4350	tte_TTE1970
<b>cpt_CpB0096</b>	pma_Pro1881	twh_TW296
ctc_CTC02503	pmm_PMM1712	tws_TW476
<b>ctr_CT333</b>	pmt_PMT2270	uur_UU102
cvi_CVI1893	pmu_PM1951	vch_VC0394
ecc_c5048	poy_PAM450	vpa_VP2712
ece_Z5657	pst_PSPTO0654	vvu_VV11427
ecj_JW4019	rco_RC1294	vvy_VV2956_
eco_b4058	rpa_RPA2816	wol_WD0916_
ecs_ECs5040	rpr_RP835	wsu_WS1628_
fnu_FN1103	sag_SAG1709	xfa_XF2426
gvi_glr4236	sam_MW0721	xft_PD1444
hdu_HD1286	san_gbs1754	ype_YPO0324
hhe_HH0132	sau_SA0714	ypk_y0580
hin_HI0249	sav_SAV0759	ypm_YP0479

Duplicaciones (66)		
<i>bba</i>	<i>Bd0159</i>	<i>Bd2442</i>
bbr	BB0075	<b>BB4032</b>
bha	BH0714	BH3594
bja	blr4702	blr8051
blo	BL0702	BL1083
bpa	BPP0075	<b>BPP3597</b>
bpe	BP3660	<b>BP3240</b>
bth	BT0578	BT1739
cac	CAC0503	CAC1464
cdi	DIP1159	DIP2031
cef	CE1140	CE1507
cgl	NCgl11040	NCgl11322
cte	CT0527	CT1689
dra	DR1771	DRA0188
efa	EF0763	EF1032
gsu	GSU3325	<b>GSU2758</b>
lin	lin2156	lin2631
ljo	LJ0864	LJ1150
lmo	lmo2050	lmo2488
oih	OB0281	OB2487
pgi	PG1036	PG2210
ppu	PP0483	PP3087
rso	RS03388	<b>RS02326</b>
sep	SE0542	SE0707
sma	SAV2097	SAV6286
xac	XAC1186	XAC1247
xcc	XCC1088	XCC1148
rba	RB10876	RB5606
		<b>RB7959</b>
lpl	lp_0773	lp_1364
		lp_2915
sco	SCO0918	SCO1958
	SCO6517	SCO6719

Archaea (5)
hal_VNG2636G
mac_MA3325
mma_MM3291
mmp_MMp0729
mth_MTH443

La mayoría fueron anotadas automáticamente como similares a UvrA durante la secuenciación del genoma completo. En 28 genomas se encontraron dos proteínas homólogas a UvrA; en los genomas de *R. baltica* y *L. plantarum* se detectaron tres y en el genoma de *S. coelicolor* se detectaron cuatro proteínas homólogas a UvrA.

Las secuencias obtenidas tienen un largo de entre 912 y 2034 residuos de aminoácidos, con una mediana de 940 residuos.

No se encontró ninguna proteína homóloga a UvrA en genomas eucariontes y se detectaron únicamente en cinco genomas arqueanos (Euryarchaeota) que corresponden a las metanógenas, *M. acetivorans*, *M. mazei*, *M. maripaludis*, *M. thermoautotrophicum* y al

halófilo *Halobacterium* sp. Se calculó el contenido G+C de éstas cinco proteínas arqueanas y se comparó con el contenido G+C promedio de todo el genoma, con la intención de encontrar un contraste que indicara más contundentemente un evento HT, pero no se encontraron diferencias.

Tampoco se hallaron homólogos en plásmidos, genomas de organelos ni genomas virales. De los 142 genomas bacterianos completamente secuenciados, sólo en cinco genomas enterobacterianos no se identificaron homólogos a UvrA: en el simbiote de la mosca tsé-tsé *Wigglesworthia brevipalpis*, el simbiote de la hormiga carpintera *Blochmania floridanus* y en las tres cepas del simbiote de áfidos *Buchnera aphidicola* (APS, Sg y Bp).

En los genomas de doce organismos se detectó la presencia de proteínas anormalmente largas, de entre 1790 y 2034 aminoácidos. Tras detectar la presencia de nuevas repeticiones de los motivos WA, WB, ABC y ZnF, las proteínas fueron analizadas con el programa en línea DotLet, que sugiere la presencia de duplicaciones completas (Fig 4). Al compararlas contra la base de datos ProDom (Servant *et al.*, 2002) se identificaron hasta seis repeticiones de los motivos WA, WB y la firma ABC; así como de dedos de zinc en un orden relativamente conservado. Los detalles de la composición de éstas proteínas se analizan más adelante, pero es necesario señalar que se determinó que éstas proteínas presentan repetido dos veces todo el gen *uvrA*, ya sea como producto de una duplicación completa del gen o como resultado de la fusión de dos copias homólogas de *uvrA*. Éstas proteínas fueron divididas en sus dos mitades durante la reconstrucción filogenética de las proteínas completas con el fin de identificar el origen evolutivo de cada dominio.

Se identificó un motivo más que está conservado en al menos las subunidades Y y Z de todas las proteínas. Éste motivo posee quince residuos de aminoácidos hacia el extremo

carboxilo de la histidina conservada en el SW. El motivo, de diecinueve residuos de longitud, posee la secuencia consenso IDLGPEGGDHGGEVVGVT y se le denominará **UVG** por ser un motivo rico en glicinas exclusivo de UvrA.

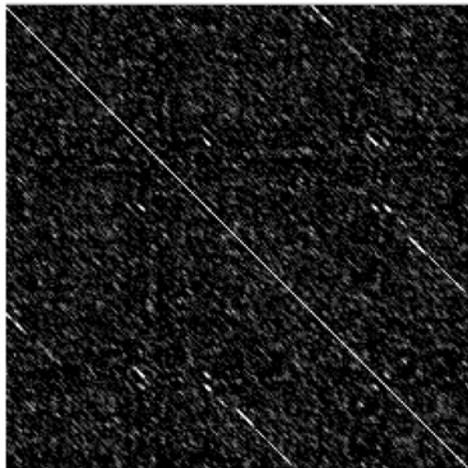
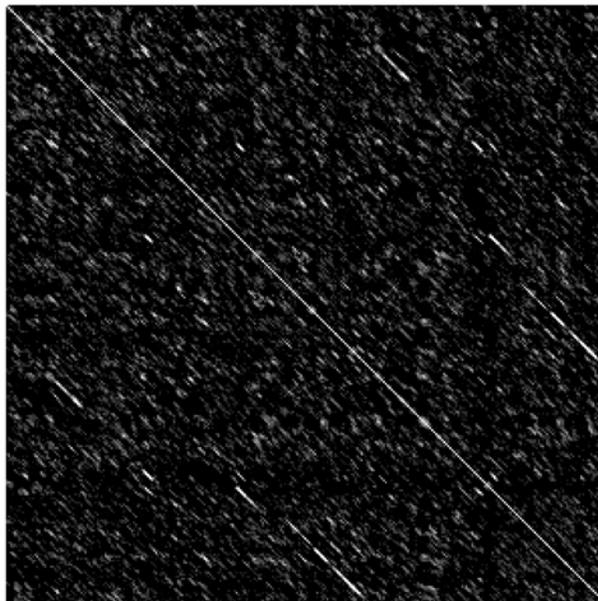
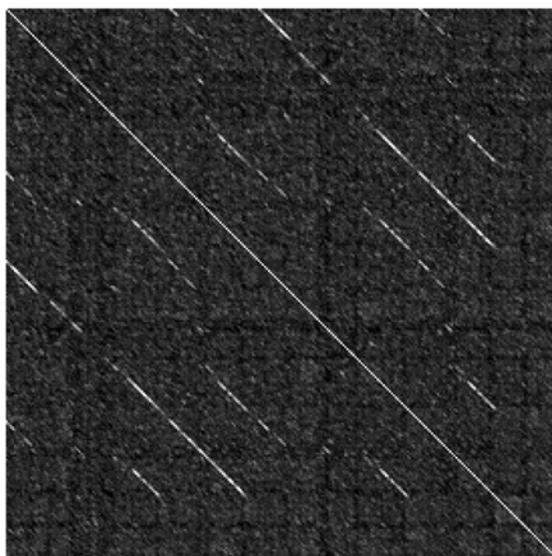
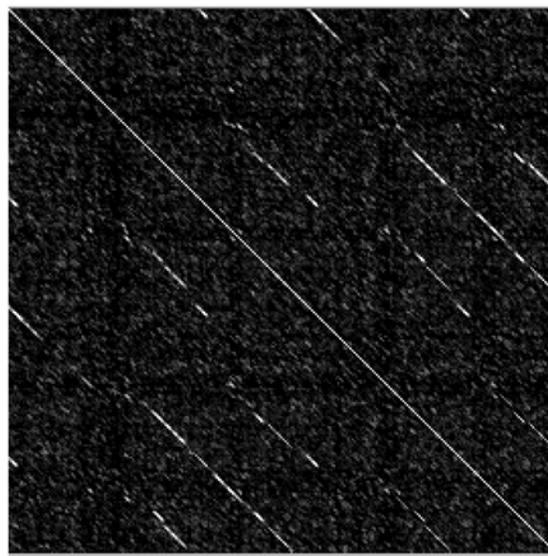
(a) *E. coli* b4058(b) *D. radiodurans* DRA0188(c) *R. baltica* RB7959(d) *B. paraptusis* BPP3597

Figura 4. Gráficas DotPlot utilizadas para la identificación de duplicaciones internas en las proteínas homólogas a UvrA inusualmente largas obtenidas de la búsqueda en genomas completos. Las superficies fueron construidas con una matriz de Gonnet. Las superficies muestran: (a) una proteína UvrA con las tres unidades características a manera de ejemplo, (b) una proteína DrrC con tres unidades, (c) una proteína UvrPlmy con duplicación interna y (d) proteína UvrA $\beta$  también con una duplicación interna. Las superficies (a) y (b) se muestran a escala normal para contrastar la longitud de las dos subfamilias, mientras que las superficies (c) y (d) se muestran a escalas diferentes.

**Reconstrucción Filogenética de la Familia UvrA.** De los 179 homólogos obtenidos de la familia UvrA *sensu lato*, se eliminaron las proteínas Bd0159, Bd2442 de *B. bacteriovorus* y CV1893 de *C. violaceum* porque sólo es posible alinearlas utilizando un gran número de espacios. La edición de la alineación múltiple de las 176 proteínas completas restantes generó un árbol filogenético en donde se puede identificar la existencia de al menos cuatro subfamilias de proteínas (Fig 5):

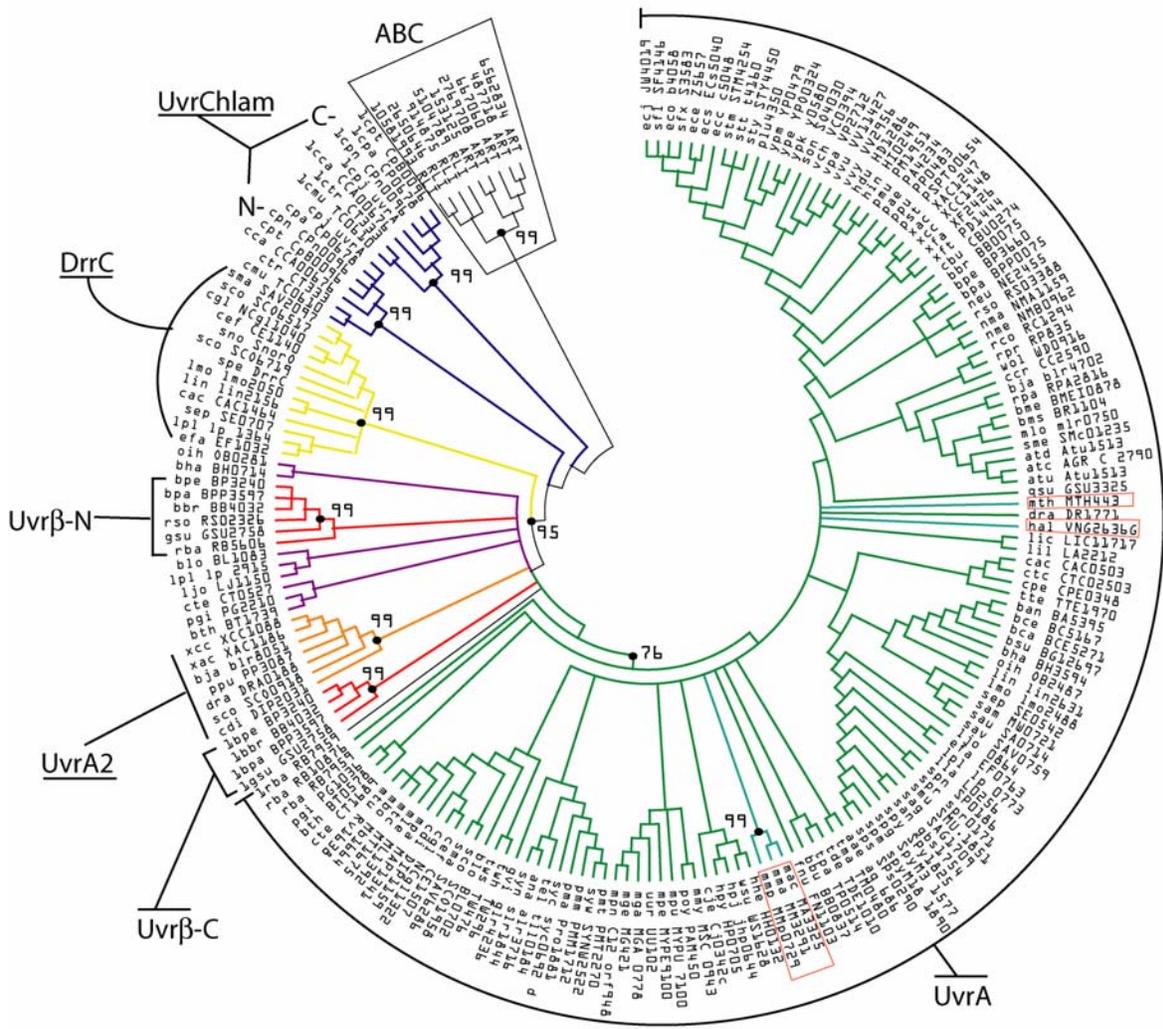


Figura 5. Árbol filogenético de las 176 proteínas homólogas a UvrA obtenidas a partir de la búsqueda por perfiles en los genomas completos. La reconstrucción se realizó por medio de NJ con 10,000 réplicas de bootstrap. Cada taxon contiene la identificación de la proteína y la clave KEGG de tres letras del organismo al que pertenece. El clado UvrA contiene todas las proteínas involucradas en NER caracterizadas experimentalmente. El clado DrrC incluye a *DrrC* y *SnoR*, involucrados en la resistencia a antibióticos. Los miembros de UvrChlam y Uvrβ se presentan divididos en las dos mitades equivalentes a tres subunidades cada una para permitir su comparación con el resto de las proteínas. Las proteínas de genomas arqueanos están señaladas en un recuadro.

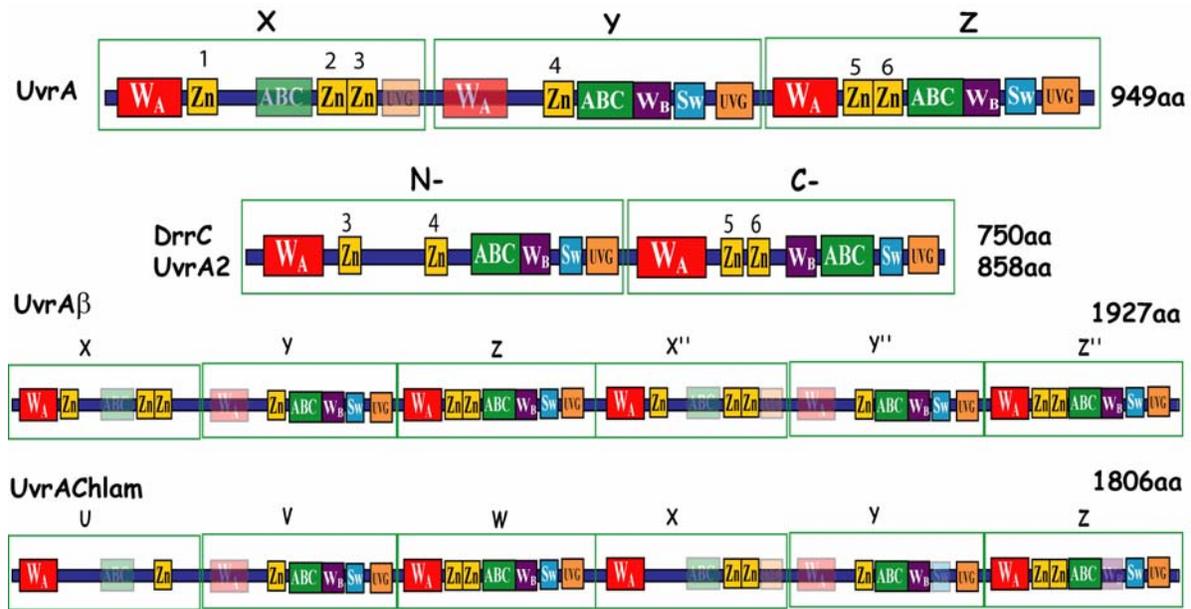


Figura 6. Representación esquemática de la organización propia de cada subfamilia de proteínas homólogas a UvrA descritas en el texto. Las regiones poco conservadas se muestran semitransparentes. La longitud promedio de cada subfamilia se muestra a la izquierda. Las subunidades son identificadas con letras diferentes para evidenciar su aparición por diferentes eventos evolutivos. WA = WalkerA, SW = región Switch II, Zn = par de cisteínas de un dedo de zinc (1-6), UVG = motivo rico en glicinas, ABC = firma ABC, WB = WalkerB.

**1.UvrA *sensu stricto*:** La primer subfamilia agrupa la mayor parte de las proteínas identificadas (133 homólogos), con un representante de cada uno de los genomas bacterianos salvo el grupo de las *Chlamydia*. En éste conglomerado se encuentran las tres proteínas caracterizadas experimentalmente como reparadoras de DNA (DR1771, b4058 y STY4450), lo que aunado al grado de conservación entre los miembros de ésta subfamilia sugiere que corresponde al grupo de ortólogos de UvrA. También incluye las cinco proteínas de genomas arqueanos (MTH443, VNG2636G, MA3325, MM3291 y MMp0729). Al interior del conglomerado se mantienen los principales grupos taxonómicos bacterianos y cabe mencionar que su topología reproduce las politomías reconocidas para las ramas correspondientes a éstos grupos cuando se utiliza el rRNA como marcador (Woese, 1987) o proteínas estructurales (Brown *et al.*, 2001). Los valores de bootstrap del conglomerado son altos (85), si se excluye del grupo a RB10876. En caso contrario, se obtiene un valor bajo (53). Al interior los valores son altos siempre que agrupan homólogos

de los principales grupos bacterianos (>70). Todos los miembros de éste conglomerado o subfamilia tienen una organización similar a la reportada por Doolittle *et al.* (1986), con tres subunidades de unión a nucleótidos organizadas como se muestra en la Fig. 6.

En los tres árboles obtenidos por métodos distintos (Fig. 7) se mantienen cohesionados los principales linajes bacterianos: las proteobacterias  $\alpha$ ,  $\beta$  y  $\gamma$  forman un grupo monofilético, al igual que el grupo Clostridia-Lactobacillales-Bacillales y Bacteroides-*C. tepidum*; Cyanobacteria, Actinobacteria (aunque por MP las dos cepas de *T. whipplei* se apartan), Mollicutes,  $\epsilon$ -proteobacteria y el clado hipertermófilo de *A. aeolicus* - *T. maritima*. En ninguno de los árboles se agrupan las dos  $\delta$ -proteobacterias: *B. bacteriovorus* y *G. sulfurreducens*, muy probablemente debido a que los hábitos endoparásitos del primero hayan erosionado la señal filogenética del marcador. En los tres árboles el *phylum* Spirochaeta aparece como polifilético en dos grupos: uno perteneciente a las dos especies de *L. interrogans*, y el otro con *B. burgdorferi*, *T. pallidum* y *T. denticola*. Cabe señalar que éstas últimas han perdido una gran cantidad de genes (Fraser *et al.*, 1997; Fraser *et al.*, 1998; Seshadri *et al.*, 2004) En los árboles por NJ y ME, el segundo grupo aparece junto con *F. nucleatum* y cerca de Mollicutes. En el árbol MP aparece con Actinobacteria y las arqueas. El grupo de *Leptospira* aparece con los Firmicutes en NJ y ME, y cerca de ellos en MP pero junto con Cyanobacteria. Las  $\epsilon$ -proteobacteria aparecen siempre cerca de la base del árbol, pero los datos no nos permite identificar su cercanía con ningún grupo en particular. En ME y NJ, las Cyanobacteria se agrupan con Actinobacteria, pero en MP se agrupan con los Firmicutes. Los Mollicutes siempre se encuentran cercanos a la base del árbol, aunque dependiendo del método utilizado varía la profundidad de divergencia.

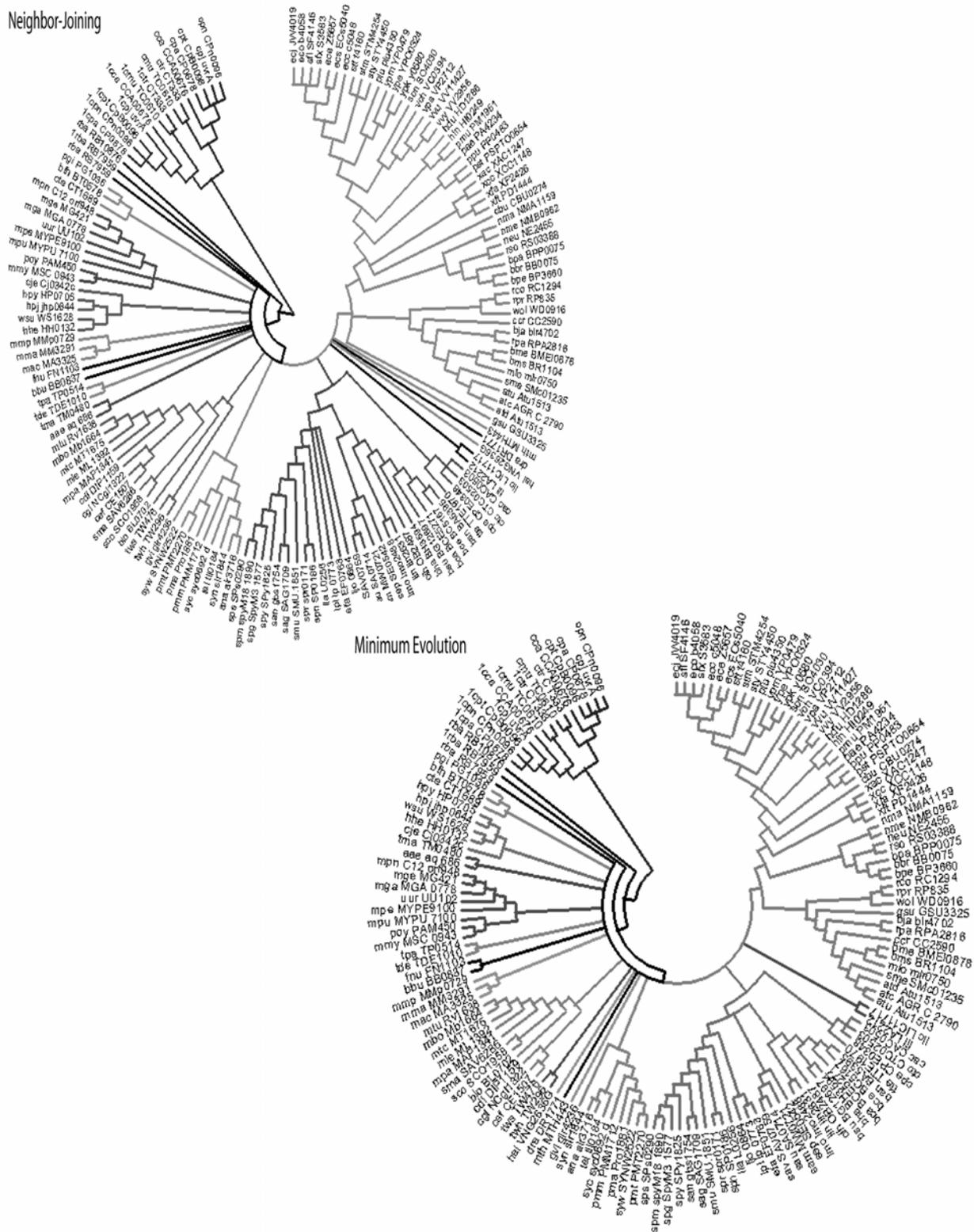


Figura 7. Árboles filogenéticos de la subfamilia UvrA *sensu stricto* obtenidos por tres métodos diferentes. Los árboles obtenidos por NJ y ME son el consenso de 10,000 réplicas de bootstrap. Cada taxón incluye la clave de tres letras del genoma y el número de acceso de la proteína. Aunque cada método provee diferentes topologías, los principales linajes bacterianos se mantienen. *Reconstrucciones preparadas con MEGA2.*



residuos. En el sitio donde esperaríamos hallar el motivo WB y el SwitchII, se encuentran los motivos de unión a zinc Zn2 y Zn3, conservados en todas las proteínas. No hay evidencia contundente de conservación del motivo UVG.

**Unidad Y.** El WA de la subunidad Y (WA<sub>Y</sub>) está bien conservado en las proteobacterias, particularmente en las bacterias entéricas, mientras que en el resto de los organismos apenas se conserva la primera glicina. En este dominio se conservan bastante bien los motivos WA (WB<sub>Y</sub>), SwitchII (SW<sub>Y</sub>) y la firma ABC (ABC<sub>Y</sub>). Entre éstos y el WA<sub>Y</sub> se encuentra el Zn4, que aunque es identificable en todas las proteínas, suele substituir una Cys por una His. Hacia el final de la subunidad se encuentra bien conservado el motivo rico en glicinas UVG<sub>Y</sub>.

**Unidad Z.** La subunidad Z es la más conservada en UvrA y por lo tanto parecería ser funcionalmente la más importante. Presenta un WA<sub>Z</sub> seguido de dos motivos de unión a zinc, Zn5 y Zn6 bien conservados; la firma ABC<sub>Z</sub>, el motivo de unión a nucleótidos WB<sub>Z</sub> y el motivo SwitchII (SW<sub>Z</sub>). Mantiene bien conservado el motivo UVG<sub>Z</sub>.

**2.DrrC:** En la segunda subfamilia se agrupan nueve proteínas junto con DrrC (resistencia a daunorrubicina en *S. peucetius*) y SnorO (resistencia a nogalamicina en *S. nogalater*). Las proteínas que conforman éste conglomerado provienen de Actinobacteria y Firmicutes, con una longitud promedio de 750 aminoácidos (entre 670 y 796 residuos), y solo presentan dos copias de los motivos WA-WB, la firma ABC; y cuatro pares de cisteínas. En las DrrC sólo se

reconocen dos subunidades de unión a nucleótidos: (a) la que corresponde al extremo amino terminal se conforma por el WA de X ( $WA_N$ ), la firma ABC de Y ( $ABC_N$ ) y el WB de Y ( $WB_N$ ); y (b) la que corresponde a la subunidad Z ( $WA_C$ ,  $WB_C$ ,  $ABC_C$ ). Conserva en la subunidad N- los motivos de unión a zinc Zn3 y Zn4, y en la subunidad C- los motivos Zn5 y Zn6. En ambas subunidades conserva el motivo SwitchII ( $SW_N$  y  $SW_C$ ) y los motivos ricos en glicinas  $UVG_N$  y  $UVG_C$ .

**3.UvrA2.** En este conglomerado se agrupan las seis proteínas que se agrupan en el mismo clado que la proteína DRA0188 de *D. radiodurans*, la cual fue originalmente descrita como una segunda copia de UvrA producto de la duplicación total del genoma en éste organismo. Una segunda hipótesis sugiere que ésta segunda copia está involucrada en un mecanismo particular de *D. radiodurans* en el que se exportan los oligonucleótidos dañados fuera de la célula para evitar su reincorporación al genoma (Boling y Setlow, 1966). El tamaño de las proteínas fluctúa entre los 853 y los 867 residuos de aminoácidos. Sólo presentan dos copias de los motivos WA-WA y de la firma ABC, y cuatro pares de cisteínas. Sólo se reconocen dos subunidades de unión a nucleótidos: (a) la que corresponde al extremo amino terminal se conforma por el WA de X ( $WA_N$ ), (b) la firma ABC de Y ( $ABC_N$ ) y el WB de Y ( $WB_N$ ); y (c) la correspondiente a la subunidad Z ( $WA_C$ ,  $WB_C$ ,  $ABC_C$ ). En la subunidad N- se conservan los motivos de unión a zinc Zn3 y Zn4, y en la subunidad C- los motivos Zn5 y Zn6. En ambas subunidades se conserva el motivo SwitchII ( $SW_N$  y  $SW_C$ ).

La proteína DRA0188 posee una inserción de 70 residuos de aminoácidos en el extremo N-terminal. Ésta proteína fue relacionada inicialmente con los miembros de DrrC (Jonathan Eisen, <http://www.tigr.org/~jeisen/UvrA/UvrA.html>), pero al realizar un análisis filogenético conjunto de DrrC y UvrA2 ambos conglomerados se mantienen separados (Fig. 8), con un valor de bootstrap de 90 para el clado de UvrA2 y de 100 para el clado DrrC (con 1000 réplicas). Ambas subunidades conservan los motivos ricos en glicinas UVG<sub>N</sub>- y UVG<sub>C</sub>-.

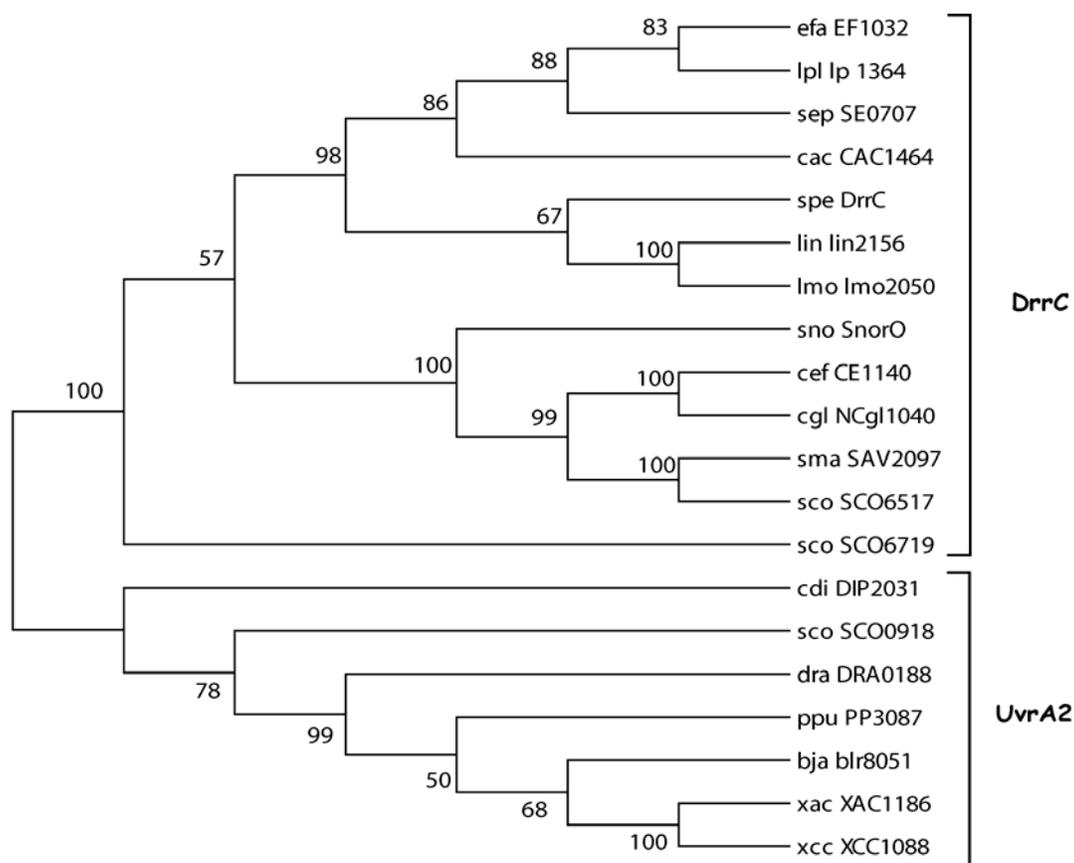


Figura 8. Árbol filogenético de las subfamilias DrrC y UvrA2, obtenido mediante NJ con 1000 réplicas de bootstrap. La topología evidencia la existencia de dos clados independientes que corresponden a cada una de las subfamilias, con un valor de bootstrap en la divergencia de 100.

**4.UvrChlam:** Siete de las once proteínas duplicadas corresponden a los genomas secuenciados del grupo de las *Chlamydia*. Las *Chlamydia* presentan sólo esta copia y no tienen homólogos en el resto del genoma. Los dos clados de siete proteínas cada uno (correspondientes a cada mitad de la proteína), aparecen contiguos en la base del árbol (Fig. 6), pero no se encuentran anidados.

**Unidad U:** La primera subunidad presenta un  $WA_U$  bien conservado en los siete genomas, pero carece de los motivos Zn1 y Zn2. El motivo Zn3 está presente junto con una región rica en Cys NCCSFAGNCSS, tan solo 15 residuos hacia el extremo carboxilo del Zn3. El ABC no se reconoce. La región comprendida entre el  $WA_U$  y el Zn3 es la más divergente de todas las proteínas identificadas en todas las subfamilias, con tres Cys conservadas en las siete proteínas. No conserva el motivo UVG.

**Unidad V:** La región correspondiente al  $WA_V$  sólo puede inferirse posicionalmente. El Zn4 se conserva bastante bien, al igual que la firma  $ABC_V$  y el  $WB_V$ . La región correspondiente al  $SW_V$  presenta un ácido glutámico antes de la histidina, conservando bien la histidina y los cinco aminoácidos alifáticos precedentes. Se mantiene conservado el motivo  $UVG_V$ .

**Unidad W:** El  $WA_W$  de la tercera subunidad está bien conservado, al igual que los motivos Zn5 y Zn6. Entre éstos motivos llama la atención una región que podría funcionar como dedo de zinc, con la secuencia consenso  $LHFEWGCIGRLIHI(X_{36})AHFSF(X_6)$ . La firma  $ABC_W$  y el  $WB_W$  están bien conservados. El  $SW_W$  se conserva de la misma forma que en  $SW_V$ . Después del  $SW_W$ , se encuentra bien conservado el motivo  $UVG_W$ .

**Unidad X:** El  $WA_X$  se caracteriza por tener una prolina entre las dos primeras glicinas, y se conserva bien en las siete proteínas. La subunidad carece del Zn1 y sólo conserva la primera leucina de la firma  $ABC_X$ . Al igual que en las demás X no existe un WB y la región del SW la ocupa el Zn3. Aunque no se identifica con claridad el motivo  $UVG_X$ , es posible identificar una región corta rica en Gly justo al terminar el Zn3 (CLTCGGRGE) que podría ser un relictos del motivo  $UVG_X$ .

**Unidad Y:** El  $WA_Y$  conserva la Gly y la Ser solamente. Presenta el Zn4 conservado, y la firma de la familia adquiere la forma **LSDGE**. El  $WB_Y$  sustituye el ácido aspártico por un ácido glutámico **FEEPL**. En la región del  $SW_Y$  se conservan tres aminoácidos alifáticos, pero se pierde la His. Hacia el extremo carboxilo de la región correspondiente al SW se encuentra un motivo que podría actuar como unión a zinc, con la secuencia **VCPSVDLHA(X<sub>6</sub>)VCP(X<sub>10</sub>)HT(X<sub>13</sub>)IHHI**. La región  $UVG_Y$  se mantiene conservada.

**Unidad Z:** Posee bien conservado el  $WA_Z$ , y a pesar de conservar los motivos Zn5 y Zn6, la región intermedia es muy divergente respecto a las demás proteínas identificadas. Presenta una  $ABC_Z$  modificada **LSVSE**, y un  $WB_Z$  poco conservado en **LIDELFSS**. Es la única subunidad que no es rica en cisteínas y que conserva bien el  $SW_Z$ . Al final de la proteína se conserva bastante bien el motivo rico en glicinas  $UVG_Z$ .

**5.Uvr $\beta$ :** Cuatro de las proteínas duplicadas restantes pertenecen al grupo de las  $\beta$ -

Proteobacterias, particularmente a los genomas de *B. pertussis*, *B. paraptentussis*, *B. bronchisepticum* y *R. solanacearum*. El dominio N-terminal de éstas proteínas se agrupa en el mismo clado que *R. baltica* (RB5606), el único genoma representante de Planctomycetes.

**Unidad X:** Preservan el WA<sub>X</sub> bien conservado, así como el motivo Zn1. La firma ABC no se reconoce y sólo se infiere posicionalmente. Los motivos Zn2 y Zn3 ocupan el lugar del WB y SW. No hay evidencia del motivo UVG.

**Unidad Y:** No se detecta el WA, salvo por la primera glicina conservada, pero sí se mantiene el Zn4, el ABC<sub>Y</sub>, el WB<sub>Y</sub> y el SW<sub>Y</sub>. El motivo rico en glicinas UVG<sub>Y</sub> se conserva bien.

**Unidad Z:** El WA<sub>Z</sub> está bien conservado, junto con los Zn5 y Zn6. También se conservan los motivos WB<sub>Z</sub>, SW<sub>Z</sub> y UVG<sub>Z</sub> hacia el final de la subunidad. La firma ABC<sub>Z</sub> conservada tiene la particularidad de presentar dos prolinas LGQPVPTLSGGEA.

**Unidad X'':** Mantiene bien conservado el WA<sub>X''</sub> y el sitio de unión a zinc Zn1. La firma ABC sólo puede suponerse posicionalmente. Conserva los motivos Zn2 y Zn3, y toda la región del Zn3 es rica en glicinas, por lo que podría ser un relicto de un UVG<sub>X''</sub> (WCGGCFGTGLQLQG). Hacia el final de la subunidad, ha perdido una región de 65 residuos de aminoácidos.

**Unidad Y'':** Posee un WA<sub>Y''</sub> rico en ácidos glutámicos y sólo conserva la primera Gly TGEET, muy cercanos al motivo Zn4. Conserva muy bien la firma

ABC<sub>Y</sub>'', el WB<sub>Y</sub>'', el SW<sub>Y</sub>'', y el motivo rico en glicinas UVG<sub>Y</sub>''. La subunidad ha sufrido una pérdida de 58 residuos de aminoácidos entre el WA<sub>Y</sub>'', y el Zn4.

**Unidad Z''**: Presenta un WA<sub>Z</sub>'', bien conservado, junto con los motivos Zn5 y Zn6. En la región comprendida entre el WA<sub>Z</sub>'', y el Zn5, se presenta un par de cisteínas bien conservadas separadas por 26 residuos de aminoácidos, que podrían conformar un dedo de zinc atípico. Matienen conservados una firma ABC<sub>Z</sub>'', con dos prolinas inusuales GQPSPTLSGGEA, el WB<sub>Z</sub>'', y el SW<sub>Z</sub>''. El UVG<sub>Z</sub>'', es de los que cuentan con mayor número de Gly de todas las proteínas encontradas LDLGPEGGSGGGQ.

- 6.UvrPlmy**: Este conglomerado polifilético contiene una proteína de tres subunidades, RB10876, y las tres subunidades del extremo carboxilo terminal de la última proteína duplicada con seis subunidades, RB7959. Ambas corresponden al único miembro de *Planctomyces*, *R. baltica*. Las tres subunidades del extremo amino terminal del RB7959 se encuentra anidada con los demás miembros UvrA.
- RB7959** en la cuarta subunidad conserva un WA<sub>R</sub> atípico con una cisteína FCGPSGSGK. Presenta el Zn1 y la firma ABC<sub>R</sub> conservando sólo la leucina y la glicina final LGVG, y los dos motivos Zn2 y Zn3. En la quinta subunidad (WA<sub>B</sub>) se conservan las dos glicinas GLG y el Zn4, junto con el ABC<sub>B</sub>, WB<sub>B</sub>, SW<sub>B</sub> y UVG<sub>B</sub>. En la sexta subunidad conserva el WA<sub>A</sub>, los Zn5 y Zn6, y muy bien conservados el ABC<sub>A</sub>, WB<sub>A</sub>, SW<sub>A</sub> y UVG<sub>A</sub>.
- RB10876** Mantiene conservado su WA<sub>X</sub> y el Zn1. Tiene un ABC<sub>X</sub> modificado LAGGMC con una cisteína atípica, seguido de los motivos Zn2 y Zn3. Extrañamente, la proteína ha perdido una región de 140 residuos de aminoácidos

correspondientes al segmento final de la subunidad X, el  $WA_Y$  y el segmento comprendido entre el  $WA_Y$  y el Zn4, motivo que presenta bien conservado. Mantiene bien conservado el  $ABC_Y$ ,  $WB_Y$ ,  $SW_Y$  y  $UVG_Y$ . En la tercera subunidad conserva el  $WA_Z$  y los motivos Zn5 y Zn6, junto con una Cys que comparte con las UvrBor antes del Zn5 **RSCPATYAG**. La firma  $ABC_Z$  se mantiene conservada, sin residuos de prolina que la relacionen con UvrBor. También conserva los motivos  $WB_Z$ ,  $SW_Z$  y  $UVG_Z$ .

**7.UvrBGS:** Las copias parálogas a UvrA de Bacteroides, *B. thetaiotaomicron* (BT1739), y de la única bacteria verde sulfurosa de la muestra, *C. tepidum* (CT0527), se agrupan en un solo clado. Posee en la primera subunidad el  $WA_X$  bien conservado, pero no presentan Zn1. El  $ABC_X$  está muy poco conservado, mas en *C. tepidum* es posible identificar cuatro residuos conservados en la subfamilia: **AESGGQ**. De los motivos de unión a zinc, sólo mantienen el Zn3. En la segunda subunidad sólo se conserva la primera glicina del  $WA_Y$  y los motivos Zn4,  $ABC_Y$ ,  $WB_Y$ ,  $SW_Y$  y  $UVG_Y$ . En la tercera subunidad conserva muy bien el  $WA_Z$ , Zn5 y Zn6,  $ABC_Z$ ,  $WB_Z$ ,  $SW_Z$  y  $UVG_Z$ .

**8.UvrBac:** Éste pequeño conglomerado contiene las copias parálogas a UvrA de *Oiheyensis* (OB0281) y *B. halodurans* (BH0714) (Firmicutes-Bacillales). En la primera subunidad presentan un  $WA_X$  bien conservado, junto con los motivos Zn1, Zn2 y Zn3. En la segunda subunidad sólo se conserva la primera glicina del  $WA_Y$ , pero el Zn4 se mantiene conservado, al igual que el  $ABC_Y$ ,  $WB_Y$ ,

SW<sub>Y</sub> y UVG<sub>Y</sub>. En la tercera subunidad, se mantiene el WA<sub>Z</sub>, los motivos Zn5 y Zn6 y los motivos ABC<sub>Z</sub>, WB<sub>Z</sub>, SW<sub>Z</sub> y UVG<sub>Z</sub>.

**9.UvrLac:** Éste clado está conformado por las copias parálogas a UvrA de dos organismos filogenéticamente distantes: *L. plantarum* (Firmicutes) y *B. longum* (Actinobacteria). En la primera subunidad conservan el WA<sub>X</sub>, pero el Zn1 sólo lo conserva *B. longum*, pues en *L. plantarum* sólo se conserva una Cys CPNGH, aunque la Hys puede funcionar como sustituto. Los motivos Zn2 y Zn3 se conservan bien, pero entre el Zn1 y el Zn2 hay una pérdida de 111 residuos de aminoácidos que incluyen a la firma ABC. En la segunda subunidad sólo se conserva la primera glicina del WA<sub>Y</sub>, y el Zn4 entero. La firma ABC<sub>Y</sub> es atípica LSTGE, pero se conservan los motivos WB<sub>Y</sub>, SW<sub>Y</sub> y UVG<sub>Y</sub>. En la tercera subunidad se conserva el WA<sub>Z</sub> y los motivos Zn5 y Zn6. Entre éstos y la firma conservada ABC<sub>Z</sub> existe una inserción de 28 aminoácidos en *B. longum*. También se mantienen conservados los motivos WB<sub>Z</sub>, SW<sub>Z</sub> y UVG<sub>Z</sub>.

**10.UvrArchaea:** Ésta es una subfamilia polifilética creada con el fin de analizar todas las proteínas de genomas arqueanos por separado. Las metanógenas *M. acetivorans*, *M. maripaludis* y *M. mazei* forman un solo clado, separadas de *M. thermoautotrophicum* y *Halobacterium* sp., las cuales aparecen sin agruparse en la base politómica del árbol de UvrA. Cabe señalar que en todas las topologías aparecen intercaladas por la UvrA de *D. radiodurans*, DR1771, e incluso en ciertas reconstrucciones ésta proteína formaba un clado con *Halobacterium* sp. (no mostradas). El Zn1 conserva las dos Cys en *Halobacterium* sp. y en *M. thermoautotrophicum*, pero la segunda se encuentra substituida por una His en

las dos especies de *Methanosarcina*. Sólo *M. thermoautotrophicum* conserva la primera Leu del ABC<sub>X</sub>, todas las demás conservan sólo la última Gly. Las cuatro conservan el Zn<sub>2</sub> y Zn<sub>3</sub>. En la segunda subunidad conservan dos aminoácidos YGTD, y los motivos Zn<sub>4</sub>, ABC<sub>Y</sub>, WB<sub>Y</sub>, SW<sub>Y</sub> y UVG<sub>Y</sub>. La tercera subunidad se encuentra totalmente conservada (WA<sub>Z</sub>, Zn<sub>5</sub>, Zn<sub>6</sub>, ABC<sub>Z</sub>, WB<sub>Z</sub>, SW<sub>Z</sub>, UVG<sub>Z</sub>).

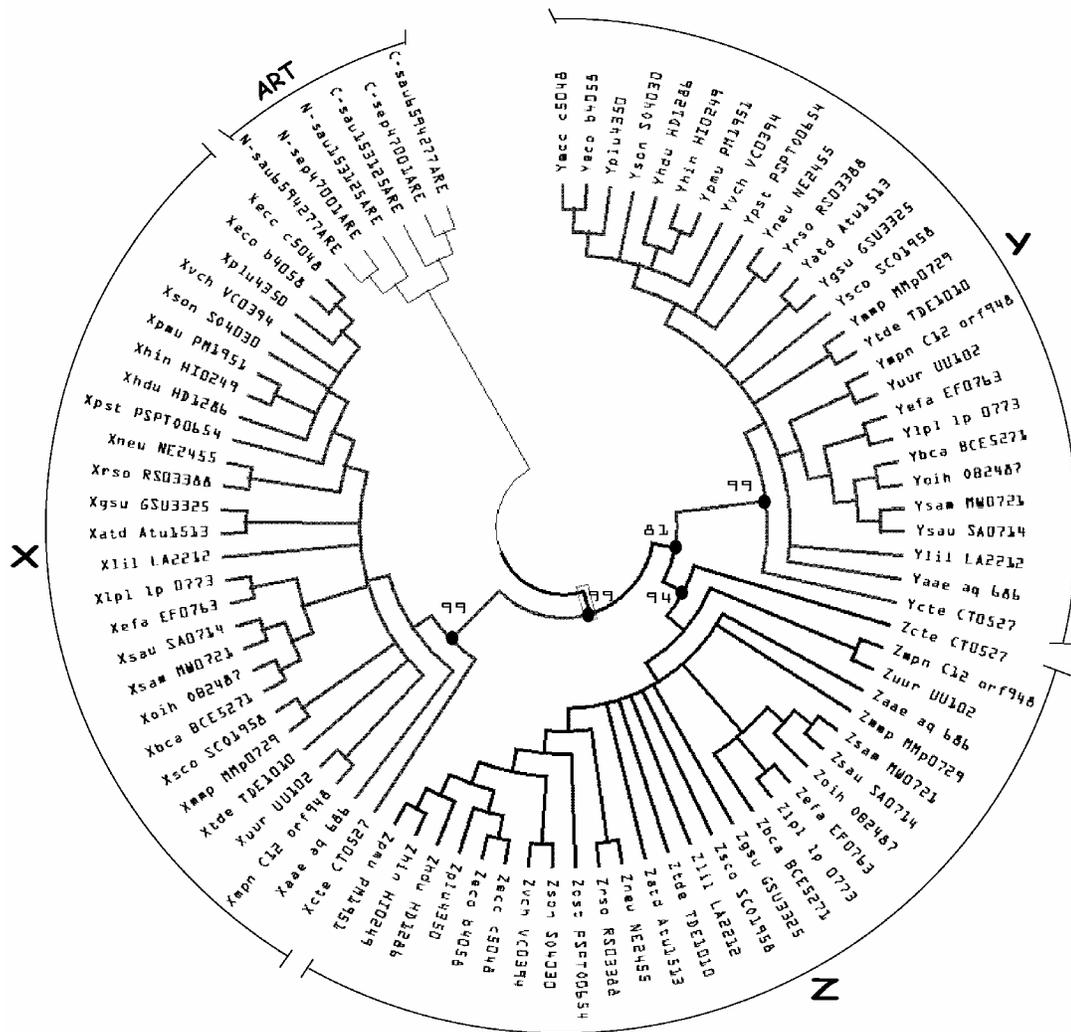


Figura 9. Reconstrucción filogenética de las tres unidades de la proteína UvrA realizada por NJ y un bootstrap de 1000 réplicas. La topología señala la divergencia entre las unidades Y y Z como el evento de duplicación más reciente. Como grupo externo se utilizaron las unidades N-terminal y C-terminal de los módulos de unión a nucleótidos ABC de la subfamilia ART.

Tabla 3. Proteínas incluídas en el análisis filogenético por subunidades individuales.

GRUPO	GENOMA	PROTEÍNA	GRUPO	GENOMA	PROTEÍNA
$\alpha$ -proteobacteria	atd	Atu1513	Mollicutes	mpn	C12orf948
$\delta$ -proteobacteria	gsu	GSU3325		uur	UU102
$\gamma$ - proteobacteria	plu	plu4350	Spirochaeta	tde	TDE1010
	son	SO4030		lil	LA2212
	vch	VC0394	Chlorobia	ete	CT0527
	pst	PSPTO0654	Aquificales	aac	aq 686
	pmu	PM1951	Archaea	mmp	MMp0729
	ecc	C5048	Bacillales	bca	BCE5271
	eco	b4058		oih	OB2487
	hdu	HD1286		sau	SA0714
hin	HI0249	sam		MW0721	
$\beta$ -proteobacteria	rso	RSO03388	Lactobacillales	lpl	lp0773
	neu	NE2455		efa	EF0763
Actinobacteria	sco	SCO1958			

**Historia Evolutiva de la Familia UvrA.** Se seleccionaron del clado de las UvrA únicamente las 27 proteínas que presentan la mayoría de los motivos conservados (Tabla 3). En la topología del árbol filogenético de las subunidades individuales de las proteínas analizadas (Fig. 9) se observan claramente definidos los tres clados correspondientes a cada una de las subunidades. A pesar de representar una divergencia muy antigua, los nodos de cada subunidad poseen valores de bootstrap por arriba de 95 , y el nodo que agrupa las subunidades Z con las Y tiene un valor de 81. Éste es, pues, el evento de duplicación más reciente, y la duplicación de la subunidad X como la divergencia más temprana.

**Relaciones Evolutivas con otras Proteínas.** La alineación múltiple se realizó con las proteínas que conservan mejor todos los motivos analizados, descritas en la Tabla 4.

La mayoría de las proteínas Rad50 se encuentran en organismos eucariontes, aunque también poseen representantes arqueanos. Tienen una longitud promedio de 1320 residuos de aminoácidos. El  $WA_{Rad50}$  se encuentra localizado en los primeros 40 residuos, y la firma  $ABC_{Rad50}$  y el  $WB_{Rad50}$  en los últimos 100 residuos de aminoácidos. El resto de la proteína tiene dos largas series de  $\alpha$ -hélices (*coiled-coils*).

Tabla 4. Proteínas utilizadas en la comparación de la familia UvrAsensu stricto con las proteínas de recombinación y mantenimiento de cromosomas Rad50 y SbcC.

UvrA		Rad50 / SbcC	
<i>E. coli</i>	eco_b4058	<i>S. cerevisiae</i>	sp P12753 RA50_YEAST
<i>S. typhi</i>	sty_STY4450	<i>S. pombe</i>	sp Q9UTJ8 RA50_SCHPO
<i>V. cholerae</i>	vch_VC0394	<i>S. solfataricus</i>	sp Q97WH0 RA50_SULSO
<i>U. urealyticum</i>	uur_UU102	<i>M. jannaschii</i>	sp Q58718 RA50_METJA
		<i>P. furiosus</i>	sp P58301 RA50_PYRFU
		<i>E. coli</i>	sp P13458 SBCC_ECOLI
		<i>D. radiodurans</i>	sp Q9RT44 SBCC_DEIRA
		<i>L. lactis</i>	sp Q9CFZ0 SBCC_LACLA

Justo a la mitad de las proteínas se localiza un motivo con un par de Cys CxxC conservado conocido como el motivo gancho o *hook construct* (Hk) (Hopfner *et al*, 2002). Éstas proteínas también presentan muy bien conservados los cuatro aminoácidos alifáticos: Tre, His, y los dos residuos ácidos del motivo SW<sub>Rad50</sub>. Las SbcC caracterizadas son proteínas bacterianas de 1280 residuos de longitud, con la presencia de un WA en los primeros 50 residuos. En los últimos 100 residuos se localizan la firma ABC<sub>SbcC</sub> y el WB<sub>SbcC</sub>. A la mitad de la proteína se encuentra una estructura homóloga al *hook construct* de Rad50, que incluye una His conservada CPLC(X<sub>4</sub>)H. También mantienen conservada la firma ABC<sub>SbcC</sub>, el WA<sub>SbcC</sub> y el SW<sub>SbcC</sub>. Al alinear las proteínas Rad50/SbcC con las tres subunidades de UvrA, se eliminaron los segmentos entre los residuos 100-650, y entre los residuos 750-1210, pues estas regiones corresponden a la región de  $\alpha$ -hélices y no fue posible alinearlas con el resto de las subunidades de las UvrA. En la alineación múltiple se pudo observar que el grupo Rad50-SbcC comparte con las UvrA el WA, ABC, el WB, el SwitchII y aparentemente el motivo Hk, que puede alinearse con los motivos de unión a zinc Zn1, Zn4 ó Zn5. No presentan regiones homólogas al motivo rico en glicinas UVG.

### Estructura Secundaria

.Las alineaciones múltiples con los módulos ABC cristalizados revelaron que conservan un buen número de los elementos de estructura secundaria con UvrA. Una alineación que agrupa las tres subunidades y algunas de las proteínas cristalizadas se muestra en la Fig. 10. El análisis de los sitios conservados concordó con la estructura predicha por JPred en la subunidad Z. Sin embargo, las predicciones para las subunidades X e Y resultaron contradictorias con los cristales conocidos y no lograron aproximar predicciones en las regiones de los motivos Zn. Por esta razón, los análisis de las subunidades X e Y en los que nos hemos basado son los de similitudes locales. Usando JPred hemos predicho la estructura terciaria correspondiente a los motivos de unión a zinc Zn5 y Zn6. En el caso de Zn5, todos los análisis concordaron en señalar hojas- $\beta$  plegadas, pero en el Zn6 sólo dos módulos, *jfr<sub>eq</sub>* y *jnet*, identificaron la presencia de  $\alpha$ -hélices. No podemos descartar la homología en los sitios con baja similitud local a nivel de estructura secundaria, pero simplemente no existe evidencia concluyente al respecto.

**Unidad Z:** Ésta subunidad es la que se conserva mejor y en la que tenemos indicios sobre la estructura de los motivos de unión a zinc, Zn5 y Zn6 particularmente. El predictor sugiere la presencia de hojas- $\beta$  plegadas alrededor de Zn5, y de  $\alpha$ -hélices alrededor de Zn6. La bolsa de unión a ATP se conserva bien, con el juego de hojas- $\beta$  encerrando la  $\alpha$ -hélice que contiene el WA. Están ausentes los elementos B5, B6 y H2, que en la secuencia corresponden a los motivos Zn5 Zn6. En las proteínas ABC, éstos elementos se encuentran orientados hacia abajo y la izquierda del sitio de unión de ATP.

También carece de la hélice H4'', aunque es posible que el método utilizado no posea la resolución necesaria para identificar éstos elementos pequeños. Por último, no presenta la región final de elementos que se encuentran dispuestos hacia atrás del sitio de unión de nucleótidos.

**Unidad Y.** Presenta una buena conservación del núcleo central de unión a ATP. Pierde las hélices B4, B5 y B6, pero conserva la hélice H2, orientando la pérdida hacia abajo y la derecha del sitio de unión a nucleótidos. Pierde también las hélices H3 y H4'', ambas en lo que se ha identificado como la región de señalización. Al igual que las otras dos subunidades, pierde la región final de la proteína que corresponde a elementos orientados hacia atrás del sitio de unión de nucleótidos.

**Unidad X.** Aunque es la subunidad más degenerada, conserva la topología general del “sandwich”  $\beta$ - $\alpha$ - $\beta$  y el dominio helicoidal. Conserva la hoja B4 pero no las B5 y B6, ni la hélice H2, y en ésta región se encuentra el motivo Zn1, orientado hacia abajo y a la izquierda del sitio de unión a nucleótidos. Tampoco presenta el giro de H4', en donde no hay motivos identificables necesarios para diferenciar entre éste elemento y el H4''. Es notable que la pérdida de la región final comienza con la ausencia de la hoja B10 que corresponde al núcleo central de unión a nucleótidos. Es en ésta región donde se encuentran los motivos Zn2Zn3, orientados hacia atrás del núcleo de unión a nucleótidos.

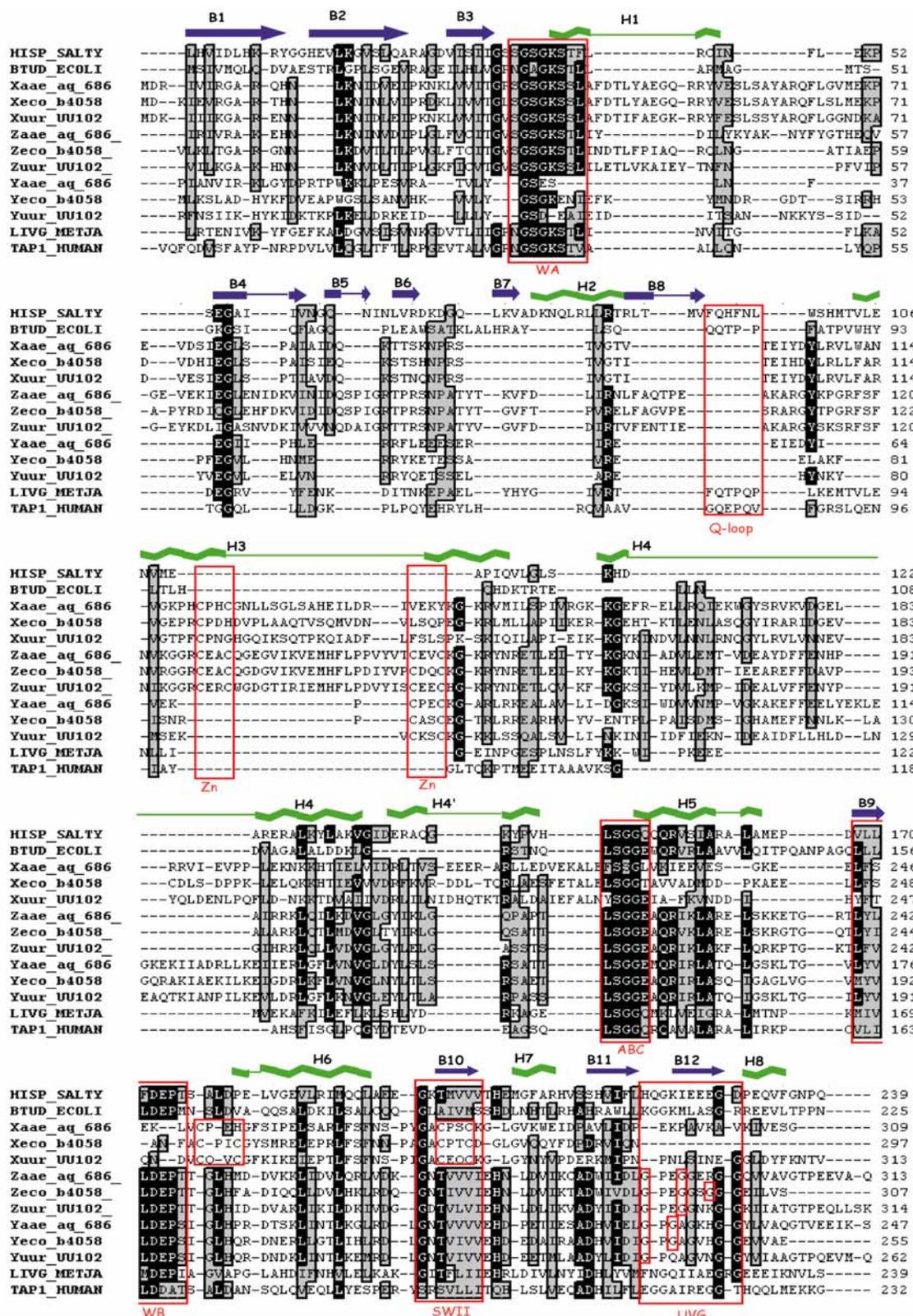


Figura 10. Alineación múltiple de los módulos de unión a nucleótidos ABC de los transportadores ABC a) HisP (*S. typhimurium*), b) BtuD (*E. coli*), c) LivG (*M. jannaschii*), d) Tap1 (*H. sapiens*) con las tres subunidades (X, Y, Z) de las proteínas UvrA de *A. aeolicus*, *U. urealyticum* y *E. coli*. Los recuadros negros indican residuos idénticos, los recuadros grises indican residuos con 60% de identidad. En la parte superior se alinea la estructura secundaria del cristal de HisP (PDB:1BUO). Las flechas indican hojas- $\beta$  plegadas y las espirales representan  $\alpha$ -hélices. Los marcos señalan la posición de cada uno de los motivos de UvrA analizados en el texto. (Preparado con BioEdit).

#### IV. DISCUSIÓN

La presencia de al menos una copia del gen UvrA en la mayoría de los genomas bacterianos analizados sugiere que la ruta UvrABC es uno de los principales mecanismos de reparación de lesiones en el DNA por NER en los sistemas bacterianos, a pesar de la existencia de rutas alternas de escisión de nucleótidos como la UVDE en *D. radiodurans* y *B. subtilis* (endonucleasa UV- $\beta$  o *UvsCDE*) (Bowman *et al.*, 1994; Doetsch, 1995; Yasui, y McCready, 1998).

En eucariontes la vía de escisión de nucleótidos caracterizada (ruta XP/ERCC1/TFIIH), involucra entre 15 y 18 polipéptidos en contraste con los cuatro que necesita la escisión bacteriana (Batty y Wood, 2000), por lo que el no encontrar homólogos a UvrA en eucariontes no resulta tan inesperado. Sin embargo, resulta sorprendente la ausencia de homólogos lejanos a UvrA en los genomas eucariontes, pues se ha reportado que al menos la región C-terminal de la proteína que realiza el corte en la escisión de nucleótidos en mamíferos, ERCC-1, es homóloga a la proteína UvrC bacteriana, que también realiza el corte (van Duin *et al.*, 1988).

##### 1.- PLÁSMIDOS Y ORGANELOS CELULARES

La proteína tampoco fue encontrada en los genomas de mitocondrias y plástidos, por lo que la reparación de lesiones al DNA de éstos organelos debe llevarse a cabo por rutas alternativas, o más probablemente el mecanismo de reparación de DNA por escisión de nucleótidos ha sido reclutado por el genoma nuclear y está a cargo del complejo de escisión de nucleótidos eucarionte que es exportado hacia el interior de los organelos celulares. El

que la proteína no haya sido encontrada en plásmidos accesorios sino exclusivamente en los cromosomas bacterianos centrales indica que la conservación de ésta proteína se debe a que no es un blanco común de los fenómenos de rearrreglo cromosómico a lo largo del genoma bacteriano.

## 2.- GENOMAS BACTERIANOS CARENTES DE UVRA

La ausencia de una copia de UvrA en los cuatro genomas bacterianos de organismos endosimbiontes intracelulares de insectos, *B. aphidicola* APS, *B. aphidicola* Sg, *B. aphidicola* Bp y *Wigglesworthia brevipalpis*, debe considerarse como una excepción resultado de una pérdida secundaria en el genoma. Los genomas de éstos organismos se han visto drásticamente reducidos como consecuencia de sus hábitos de endosimbiontes obligados (Charles & Ishikawa, 1999) y en particular, los genomas de *Buchnera* se encuentran aún en un proceso de reducción genómica (Gil *et al.*, 2002). El genoma de *Buchnera* ha sufrido una severa pérdida de genes involucrados en la reparación y recombinación de DNA, entre los cuales se encuentran los principales genes del mecanismo de recombinación *recA* (Shigenobu *et al.*, 2000) y *recF* (Tamas *et al.* 2002), aunque conservan la ruta *recBCD*. En consecuencia, éstos organismos han reducido sus rearrreglos cromosómicos y el dinamismo genético al mínimo, al punto de alcanzar una estasis genómica imperturbable desde hace 50 millones de años (Tamas *et al.* 2002). Los genes de reparación perdidos no sólo pertenecen a la ruta *uvrABC*, sino también a otras rutas de reparación como las rutas “Direct Damage Reversal” (*ada*, *ogt*), “Base Excision Repair” (*tag*, *mutM*) y “Mismatch Repair” (*mutH*, *recJ*, *uvrD*) (Moran y Mira, 2001).

### 3.- ANÁLISIS DE LA EVOLUCIÓN TEMPRANA DE UVR A

Considerando que la ausencia de UvrA en los cuatro genomas de *Buchnera* son una excepción, los datos obtenidos en éste estudio señalan a la proteína UvrA como una de las proteínas más importantes en los genomas bacterianos, por lo que debe ser considerada como indispensable al reconstruir el genoma mínimo bacteriano. Al estar presente sólo en Bacteria, el gen *uvrA* resulta un mal marcador molecular para rastrear al último ancestro común, sin embargo resulta útil para obtener información sobre la evolución temprana de los principales linajes bacterianos.

La divergencia de los tres dominios puede remontarse a una etapa con un tempo y modo de evolución significativamente diferente al actual, en donde las mutaciones deletéreas o benéficas tenían efecto directamente en la adecuación de los sistemas biológicos por una relación directa entre el fenotipo y el genotipo (Doolittle y Brown, 1994). Si el último ancestro común pertenece al mundo del RNA, entonces es absurdo imaginar un mecanismo de reparación de DNA. Aún más, debido a la ventaja que presentaba el efecto inmediato de las mutaciones en el genotipo-fenotipo de los sistemas ribozimales, tampoco resulta coherente suponer un mecanismo de reparación de RNA por escisión de nucleótidos. Por lo tanto, si el último ancestro común (LCA) data de dicha etapa, la reparación por escisión de nucleótidos debió aparecer en una etapa muy posterior a la divergencia de al menos los dominios Bacteria y Archaea.

Como establece Woese (2002), es muy probable que los primeros sistemas celulares hayan surgido en ésta etapa evolutiva y por lo tanto que la organización celular básica no había completado su evolución en el momento representado por la raíz de árbol filogenético universal. Aunque a decir por la distribución universal de los mecanismos de traducción ésta maquinaria se encontraba totalmente desarrollada en el último ancestro común (Woese

*et al.*, 2000); las maquinarias de transcripción y replicación se encontraban mucho menos desarrolladas. En la maquinaria de transcripción, sólo las subunidades  $\beta$  y  $\beta'$  de la RNAPolimerasa DNA-dependiente poseen una distribución universal (Langer *et al.*, 1995); y no existe evidencia de ortología entre los mecanismos de replicación bacterianos con los mecanismos arqueanos/eucariontes (Olsen y Woese, 1996), y por lo tanto parece ser que los mecanismos de replicación del genoma han surgido independientemente al menos dos veces en la historia evolutiva universal (Olsen y Woese, 1996; Delaye *et al.*, 2001)

Bajo éste modelo, inicialmente los genomas aparecieron en un mundo de RNA y proteínas, por lo que el primer mecanismo biológico de transferencia de información en aparecer fue el mecanismo de traducción con los RNAs de transferencia y las aminoacil-tRNA sintetasas. Después se dio la transición a genomas de DNA, y por lo tanto la siguiente maquinaria en aparecer fue la maquinaria de transcripción con las RNAPolimerasas DNA-dependientes. Todo parece indicar que fue durante ésta transición cuando divergieron los principales linajes procariontes actuales. La estabilidad de los genomas de DNA dependió de la organización de los genes en operones y del establecimiento de redes de regulación génica. Una vez estabilizados, los genomas de DNA desarrollaron una maquinaria de replicación que parece haber sido inventada independientemente en cada dominio.

Además de tener una distribución universal en el dominio Bacteria, los homólogos de UvrA presentan un alto grado de conservación en la estructura primaria a pesar de haber divergido en el momento mismo de la divergencia de los principales grupos bacterianos. Esto solo puede interpretarse como efecto de estar sujetas a grandes presiones de selección

que restringen su divergencia evolutiva y resaltan su importancia en los organismos bacterianos contemporáneos.

#### 4.- RECONSTRUCCIÓN FILOGENÉTICA DE LA FAMILIA UVRA

El árbol filogenético obtenido a partir de la alineación múltiple de todos los homólogos obtenidos (Fig. 5) muestra un clado monofilético que agrupa a una copia de UvrA de cada genoma analizado con dos excepciones importantes: los genomas endosimbiontes intracelulares de insectos analizados anteriormente y el grupo de las *Chlamydia*, que diverge en la base después del grupo externo. En éste clado también se agrupan las cuatro proteínas de los genomas arqueanos encontradas. En otro clado se agrupa el segundo homólogo de algunos genomas de Actinobacteria y Firmicutes, todos junto con las proteínas de resistencia a daunorrubicina DrrC y a noglamicina SnrO. En otro clado más se agrupan homólogos provenientes de genomas de  $\gamma$ -proteobacterias, un  $\alpha$ -proteobacteria y la proteína caracterizada como la segunda copia de *uvrA* en *D. radiodurans*, UvrA2. Finalmente, en los últimos dos clados se agrupan las segundas copias de los genomas de algunas  $\beta$ -proteobacterias y sus duplicaciones subsecuentes.

Las 66 proteínas duplicadas (Tabla 2), no se agrupan con su parálogo, sino que forman clados individuales, sugiriendo eventos de duplicación ancestrales que dieron origen a familias multigénicas. Tanto por los análisis de sus subunidades individuales que se exponen a continuación, como por la topología de la filogenia y las diversas funciones de algunas de las proteínas, cada uno de éstos clados se considerará una subfamilia independiente de la superfamilia UvrA *sensu lato* y se analizarán individualmente a continuación:

**4.1.Subfamilia UvrA *sensu stricto*.** Los árboles inferidos mediante los tres métodos (Neighbour-Joining, Minimum Evolution y Maximum Parsimony) mantienen una topología general en donde pocas ramas divergen en la base del árbol y la mayoría de las proteínas se agrupan en los clados correspondientes a los principales linajes bacterianos. Sin embargo, la mayoría de éstos clados se unen en una misma rama central politómica. El árbol que más difiere del canónico de RNA ribosomal y el menos coherente con las filogenias bacterianas en general es el obtenido por parsimonia.

En los tres árboles, las dos mitades de la única copia de UvrA en los genomas Chlamydiales aparecen en la base del árbol. Desafortunadamente, no podemos considerar esto como evidencia de una divergencia temprana de *Chlamydia*, pues además de que éstos organismos son parásitos y están sujetos a una tasa evolutiva diferente, las proteínas halladas en éstos genomas presentan un gran número de peculiaridades como la presencia de nuevos motivos, seis subunidades fusionadas en el mismo gen, y un gran número de substituciones a lo largo de toda la proteína. En conjunto con el empleo de módulos ABC como grupo externo (Fig 5.), lo anterior señala hacia una alta divergencia, que se traduce en grandes distancias filogenéticas que colocan éstas proteínas en la base del árbol artificialmente. Las filogenias tampoco apoyan una relación cercana entre Planctomycetes y Chlamydiae/Verrucomicrobiae como se propuso anteriormente (Weisburg *et al.*, 1986; Ward-Rainey *et al.*, 1995, 1997). Sin embargo, debido a la evolución taquitética de las proteínas de *Chlamydia*, tampoco es posible apoyar la hipótesis que plantea que los Planctomycetes son tan diferentes del grupo Chlamydiae/Verrucomicrobiae como de cualquier otro linaje bacteriano (Ward *et al.*, 2000).

La topología del árbol filogenético obtenido difiere en varios aspectos de la topología canónica de los árboles obtenidos a partir de filogenias de RNA ribosomal (Pace *et al.*,

1986; Woese, 1987) y otros marcadores moleculares que reproducen ésta topología con filogenias de genes parálogos como los factores de elongación EF-Tu y EF-G (Iwabe *et al.*, 1989; Baldauf *et al.*, 1996). La principal diferencia radica en que en nuestro árbol la divergencia más antigua pertenece a la rama de los Planctomycetes, lo que concuerda con los resultados del análisis filogenético de las terceras posiciones de las proteínas ribosomales (Brochier y Philippe, 2002) y factores de elongación (Jenkins y Fuerst, 2001). Los genomas de las bacterias hipertermófilas *A. aeolicus* y *T. maritima* presentan una sola copia del gen *uvrA*, las cuales se agrupan en un mismo clado monofilético, lo que apoya un origen común de los Thermotogales y Aquificales (Brochier *et al.*, 2002), y aparecen anidados en diferentes posiciones según el método: como una rama relativamente antigua que diverge después de Mollicutes y Planctomyces (MP); como una rama que diverge después de Bacteroides, Planctomyces y  $\epsilon$ -Proteobacteria (ME) o como parte de la multifurcación politómica de la mayoría de los phyla (NJ). De cualquier forma, en ninguna de las topologías obtenidas por los tres métodos aparecen como una rama basal, lo que habla en contra del origen hipertermófilo del dominio Bacteria (Achenbach-Richter *et al.*, 1987; Stetter, 1996). Nuestros concuerdan con lo hallado a partir de otros marcadores moleculares (Brochier *et al.*, 2002), por la el análisis por covarianza (Lopez *et al.*, 1999) y por que la estimación del contenido G+C de rRNA del último ancestro común es incompatible con un estilo de vida por encima de 80°C (54-56%, Galtier *et al.*, 1999). Por lo tanto es posible que la divergencia canónica temprana de los hipertermófilos se deba a (a) el efecto de atracción de ramas largas (Felsenstein, 1978), (b) una tasa evolutiva acelerada (Lopez *et al.*, 1999), (c) un sesgo en la composición de aminoácidos adquirido como adaptación secundaria al estilo de vida hipertermófilo (Cambillau y Claverie, 2000),

y/o (d) a que un gran porcentaje de los genes de éstos organismos han sido adquiridos a través de transferencia horizontal de las arqueas hipertermófilas (Aravind *et al.*, 1998; Deckert *et al.*, 1998; Nelson *et al.*, 1999). Aunque éstos dos órdenes de organismos hipertermófilos no se encuentren en la base de la filogenia, en los tres árboles obtenidos emergen en ramas cercanas a la base y nunca están anidados en los clados terminales. Tomando en cuenta esta localización, es posible que si bien el último ancestro bacteriano no sea de naturaleza hipertermófila, éstos representen una antigua etapa apenas posterior al origen del dominio Bacteria en donde se sufrió una adaptación secundaria a temperaturas elevadas (Islas *et al.*, 2003); o bien, que sean los descendientes directos de otros organismos antiguos originalmente adaptados a condiciones de estrés y que hoy se muestran hipertermófilos (Miller y Lazcano, 1995). De cualquier manera, los órdenes *Thermotogales* y *Aquificales* representan organismos muy antiguos sin que esto signifique que son los primeros linajes en divergir en el dominio Bacteria.

Aunque la antigüedad de los Planctomyces es apoyada con diversos marcadores moleculares como la subunidad 16S del RNA ribosomal (Stackebrandt *et al.*, 1984) y la subunidad  $\beta$  de la ATPasa (Rönnner *et al.*, 1991), ésta divergencia basal también se ha interpretado como un artefacto por atracción de ramas largas resultado de la evolución taquitética de un grupo que presentaría un reloj molecular acelerado (Woese *et al.*, 1985; Liesack *et al.*, 1992; Jenkins y Fuerst, 2001).

Al realizar el análisis de las UvrA bacterianas incorporamos las tres copias halladas en *Rhodopirelulla baltica*, y la copia duplicada fue separada en sus dos componentes. En los tres métodos utilizados, la proteína RB5606, caracterizada como miembro de la subfamilia de las  $\beta$ -Proteobacterias, aparece consistentemente como la rama que diverge

después de las Chlamydiales y fuera del clado principal de UvrA, mientras que en la base del árbol se encuentra la copia sencilla de UvrA RB10876. En los árboles obtenidos por ME y NJ, la siguiente divergencia corresponde al extremo carboxilo de la copia doble RB7959, pero en la reconstrucción por MP las dos mitades de ésta proteína aparecen anidadas al lado de las arqueas *M. thermoautotrophicum* y *Halobacterium* sp. dentro del clado que agrupa a espiroquetas y actinobacterias. Ésta agrupación es claramente un artefacto producto de la atracción de ramas largas, pues es bien conocido que los métodos basados en parsimonia son más vulnerables a éste artefacto filogenético (Felsenstein, 1978; Hendy y Penny, 1989). Si todos los genomas de los Planctomyces estuvieran sujetos a una evolución taquitélica, entonces encontraríamos las copias RB7959 y RB10876 de UvrA ocupando lugares muy diferentes en el árbol, y la copia que se aleja más muy probablemente pertenezca a una subfamilia diferente. Éstas proteínas deben haberse originado en un evento de duplicación previo a la duplicación interna de RB7959. Si realmente presentaran una tasa de evolución acelerada, alguna de ellas aparecería lejos de las demás UvrA y de su parálogo, a menos que se tratara de un evento sumamente reciente, pero no aparecen tampoco formando un solo clado con distancias muy cortas entre ellas. Por lo tanto, debemos suponer que las dos proteínas representan copias de proteínas con una función ortóloga a UvrA.

**4.2. Subfamilia DrrC .** Éste grupo está conformado por proteínas homólogas a UvrA que corresponden a la segunda copia presente en once genomas analizados. La subfamilia contiene los homólogos caracterizados experimentalmente como genes de resistencia a antibióticos, particularmente de resistencia a daunorrubicina/doxorubicina *drrC* en *Streptomyces peucetius* (Lomovskaya *et al.*, 1996), a nogalamicina *SnorO* en *Streptomyces nogalater* (Ylihonko *et al.*, 1996; Palmu *et al.*, sin publicar), y a mitramicina en

*Streptomyces argillaceus* (Fernandez *et al.*, 1996). Ésta última proteína fue eliminada del análisis por presentar una gran divergencia respecto a las demás, por carecer de los motivos W<sub>A</sub>N-, Zn<sub>3</sub>/Zn<sub>4</sub>, y porque no se agrupa en forma robusta con ninguna otra proteína. En el árbol filogenético (Fig. 7) se pueden identificar dos grandes subclados: el primero corresponde a los homólogos relacionados estrechamente con el gen de resistencia a daunorrubicina/doxorubicina DrrC de *S. peucetius*, y que corresponden todos a proteínas de genomas de Firmicutes (*E. faecalis*, *L. plantarum*, *S. epidermidis*, *C. acetobutylicum*, *L. monocytogenes*, *L. innocua*); y el segundo corresponde a los homólogos actinobacterianos estrechamente relacionados al gen de resistencia a nogalamicina SnrO (*S. nogalater*, *S. avermitilis*, *S. coelicolor*, *C. efficiens*, *C. glutamicum*) con dos copias en *S. coelicolor*.

Es poco probable que el mecanismo de acción de DrrC sea la exportación de la droga, pues dentro del mismo grupo de genes en *S. peucetius* se encuentra la ATPasa DrrA, con fuerte similitud al módulo de unión a nucleótidos ABC (Guilfoile y Hutchinson, 1991; Kaur, 1997), y DrrB, el dominio de hélices hidrofóbicas transmembranales (Guilfoile y Hutchinson, 1991; Kaur y Russell, 1998), implicados en el transporte de la daunorrubicina/doxorubicina. En contraste, se ha propuesto que de alguna manera DrrC inhibe o desestabiliza la unión de la daunorrubicina/doxorubicina al DNA del nucleóide, disminuyendo la interferencia con la expresión génica y la dinámica genómica (Lomovskaya *et al.*, 1996).

Aunque la segregación en los grupos taxonómicos Firmicutes y Actinobacteria sugiere un origen común, la naturaleza parafilética de éste grupo puede señalar que el clado se formó a partir de diversos eventos de transferencia horizontal. Además, los componentes de resistencia a drogas y antibióticos son reconocidos como elementos susceptibles a la transferencia horizontal. Es posible que los eventos de transferencia sean eventos antiguos,

pues en el caso de las tres especies de *Corynebacterium*, dos de ellas se agrupan lado a lado mientras que la tercera (*C. diphtheriae*) aparece fuera del clado, junto con las homólogas a UvrA2 utilizadas como grupo hermano. El evento de transferencia pudo ocurrir en el ancestro de las tres especies de *Corynebacterium*, y que en *C. diphtheriae* la copia sufriera una acelerada tasa mutacional y divergiera significativamente de las otras dos.

Por otra parte, en esta subfamilia encontramos dos representantes del genoma de *Streptomyces coelicolor*, uno anidado dentro del clado de las actinobacterias y el otro fuera del clado de la subfamilia. Es probable que la copia original sufriera una duplicación, y la copia resultante se diferenciara del resto de las DrrC.

Es necesario recordar que el mecanismo de acción de las proteínas caracterizadas experimentalmente aún no ha sido dilucidado e incluso se ignora si es el mismo para los dos tipos de drogas (doxorrubicina/daunorrubicina y nogalamicina). Por lo tanto, no es posible determinar si los homólogos encontrados brindan resistencia a alguna de las tres drogas reportadas o a otros compuestos relacionados con la familia de la antraciclina.

Los genomas de los géneros *Corynebacterium* y *Streptomyces* tienen una estrecha relación filogenética, e incluso presentan una antigua sintenia en el núcleo central de genes. La primera copia de DrrC pudo aparecer en una actinobacteria primitiva y conservarse diferencialmente según las presiones ambientales particulares de cada especie de los géneros *Streptomyces* y *Corynebacterium*. La copia en *S. peucetius* fue entonces transferida a Firmicutes, donde tuvo una permanencia diferencial o fue transferida horizontalmente entre los miembros de éste grupo.

*S. coelicolor* ha duplicado éste grupo de proteínas consecutivamente hasta alcanzar cuatro copias genómicas: (a) una perteneciente a la proteína de escisión de nucleótidos UvrA *sensu lato* (SCO1958); (b) otra a la subfamilia de resistencia a daunorrubicina DrrC

(SCO6517); (c) otra más de la subfamilia UvrA2 (SCO0918), y (c) una última copia cercana a la subfamilia DrrC que podría funcionar como proteína de resistencia a algún antibiótico o toxina diferente a la daunorrubicina (SCO6719), lo que resulta coherente con la enorme versatilidad metabólica que presentan los miembros de éste género. Sin embargo, *S. avermitilis* tan sólo presenta dos copias, una de UvrA y otra de DrrC, a pesar de que 70% de los genes de éste organismo poseen ortólogos en *S. coelicolor* (Ikeda *et al.*, 2003).

**4.3. Subfamilia UvrA2.** Se identificaron seis proteínas homólogas a UvrA muy similares a la proteína UvrA2 de *D. radiodurans* (White *et al.*, 1999). Tres proteínas provienen de genomas de  $\gamma$ -proteobacteria (*X. campestris*, *X. axonopodis* y *P. putida*), uno a  $\alpha$ -proteobacteria (*B. japonicum*) y dos más a Actinobacteria (*C. diphtheriae* y *S. coelicolor*).

La proteína DRA0188 se encuentra codificada en el cromosoma II (replicón corto o megaplásmido), a diferencia de la UvrA DR1771 que se encuentra en el cromosoma I. Aunque la actividad y función de DRA0188 no han sido caracterizadas experimentalmente, se ha sugerido que está involucrada en un mecanismo de resistencia a lesiones en el DNA único de *D. radiodurans* consistente en la exportación de los nucleótidos dañados fuera de la célula (Boling y Setlow, 1966), presumiblemente evitando su reincorporación al DNA genómico (Battista, 1997). También se señaló que la función específica de DRA0188 podría ser el transporte activo de los nucleótidos fuera de la célula debido a su similitud con DrrC, que se consideraba un exportador de antibióticos (Eisen, *in* White *et al.*, 1999). Sin embargo, como se analizó anteriormente, la función hipotética de DrrC como exportador de antibióticos no es muy robusta, y el análisis filogenético de las dos subfamilias de proteínas en conjunto señala claramente la existencia de dos clados independientes y filogenéticamente distantes (Fig. 7). Por otra parte, aunque la reducción de la longitud de la

proteína por la casi desaparición de una de las subunidades podría sugerir una reversión a la función de transporte por la similitud con la organización de los módulos ABC, éstas proteínas son aproximadamente 100 residuos más largas que las DrrC, lo que modifica estructuralmente el módulo de transporte ABC al menos como lo conocemos. Por último, no fue posible identificar módulos de transporte transmembranal asociados a éstas proteínas, indispensables para la exportación; ni se identificó el módulo de transporte DrrA, que corresponde efectivamente a la subunidad ABC en *S. peucetius* (Guilfoile y Hutchinson, 1991; Kaur, 1997), en los otros genomas. Por lo tanto, al ser independiente de cualquier otro sistema de transporte ABC, como se sugiere para DrrC, la función de UvrA2 no parece relacionarse con el transporte activo transmembranal.

La distribución taxonómica atípica de ésta subfamilia de proteínas, con algunos miembros de tan sólo un grupo de las proteobacterias y actinobacterias, sugiere que si UvrA2 es un grupo natural deben haber ocurrido al menos dos eventos de transferencia horizontal. Al parecer, la copia original apareció en el genoma de *S. coelicolor*, y de ahí lo transfirió a los demás genomas, todos de organismos bacterianos hallados en suelo (*S. coelicolor*, *D. radiodurans*, *P. putida*, *B. japonicum*, *X. campestris* y *X. axonopodis*). Es posible que la transferencia del gen haya ocurrido antes de la divergencia entre las dos especies de *Xanthomonas*, y que por lo tanto XCC1088 y XAC1186 sean genes ortólogos. Aunque de éstos seis organismos cuatro corresponden a proteobacterias, no suponemos la ortología entre copias proteobacterianas, porque ello implicaría la pérdida secundaria en todos los demás genomas proteobacterianos, lo que resulta difícil de creer cuando hay ausencia de ésta proteína en los otros dos genomas de *X. fastidiosa* (pertenecientes también al grupo de las Xanthomonadales), y en las otras dos Pseudomonadales, *P. aeruginosa* y *P. syringae*. Teniendo en el grupo organismos con tres estilos de vida tan diferentes, a saber,

parásitos de plantas (*Xanthomonas*), de vida libre (*P. putida*, *S.coelicolor*, *D. radiodurans*), y endosimbiontes (*B. japonicum*), no es posible identificar alguna razón por la cual éstos organismos conservaran ésta proteína y se perdieran en otras proteobacterias tanto de vida libre como parásitas de plantas (*X. fastidiosa*, *P. syringae pv tomato*) y endosimbiontes (*M. loti*, *S. meliloti*, *Wolbachia wMel*).

La proteína homóloga del genoma de *C. diphtheriae* debe analizarse con cautela porque 1) pertenece a un organismo parásito sujeto a evolución taquitética, y 2) las segundas copias homólogas a UvrA de las otras dos especies de *Corynebacterium* pertenecen a la subfamilia DrrC. Anteriormente discutimos la monofilia de DrrC al menos dentro de Actinobacteria, donde de todos los genomas de *Streptomyces* y *Corynebacterium* analizados que poseen una segunda proteína homóloga a UvrA, la única proteína que se agrupa fuera de éste clado es la perteneciente a *C. diphtheriae*. Si bien *C. diphtheriae* comparte los hábitos edáficos con el resto de los organismos en la subfamilia UvrA2 y es propenso al transporte horizontal de acuerdo con la reciente adquisición de una amplia gama de genes patogénicos (Cerdeño-Tárraga *et al.*, 2003); es más probable que la proteína DIP2031 que aparece en la subfamilia **UvrA2** en realidad tenga origen en la subfamilia **DrrC**, tras lo cual sufriera evolución taquitética y por lo tanto aparece en UvrA2 artificialmente. Ésto sólo podrá comprobarse evaluando si *C. diphtheriae* presenta efectivamente resistencia a daunorrubicina/doxorrubicina; o caracterizando experimentalmente la función biológica de la subfamilia UvrA2 y analizando si la función es coherente con el estilo de vida del organismo. Hasta ese momento, la subfamilia UvrA2 deberá considerarse como un grupo no natural.

**4.4. Subfamilia Uvr $\beta$ .** En ésta subfamilia se agrupan las proteínas homólogas a UvrA encontradas en los genomas  $\beta$ -proteobacterianos de *R. solanacearum*, patógena de plantas con un amplio rango de hospederos; los patógenos causantes de enfermedades respiratorias en humanos (*B. pertussis*), en humanos y ovejas (*B. parapertussis*), y en un amplio rango de mamíferos y aves (*B. brochiseptica*), y finalmente en la  $\delta$ -Proteobacteria *G. sulfurreducens*.

Todos los miembros de éste grupo poseen una proteína duplicada, es decir con dos copias fusionadas en un sólo gen. La proteína Uvr  $\beta$  original surgió en alguna  $\beta$ -proteobacteria ancestral como resultado de una duplicación completa o parcial de un genoma completo, pues en *R. solanacearum* la copia RS02326 se encuentra en el megaplásmido, mientras que la UvrA RS03388 se encuentra en el cromosoma central.

No es posible determinar si la duplicación ocurrió antes o después de la divergencia del grupo Bordetella-Ralstonia-Nitrosomonas y los dos genomas de *N. meningitidis*, pues éstos últimos son parásitos de mamíferos, mas el género tiene otros representantes de vida libre en el suelo y es posible que la duplicación ocurriese en un ancestro de hábitos edáficos y que dicha duplicación se perdiese en *N. meningitidis* debido a su estilo de vida parásito. Por otra parte, una segunda copia de UvrA no sería seleccionada a favor en *N. meningitidis* cuando su genoma posee dos mecanismos de reparación de lesiones al DNA por radiación UV: el mecanismo de escisión de nucleótidos por UvrA (Black *et al.*, 1997) y por reparación de la recombinación *RecFOR* (Mehr y Seifert, 1998); mientras que ha perdido otros mecanismos de reparación de DNA como la reparación por fotoreactivación (Campbell y Yasbin, 1979) y la reparación de errores por baja fidelidad de la DNA

polimerasa (Campbell y Yasbin, 1984). La proteína debió sufrir un desplazamiento hacia el megaplásmido en *R. solanacearum*.

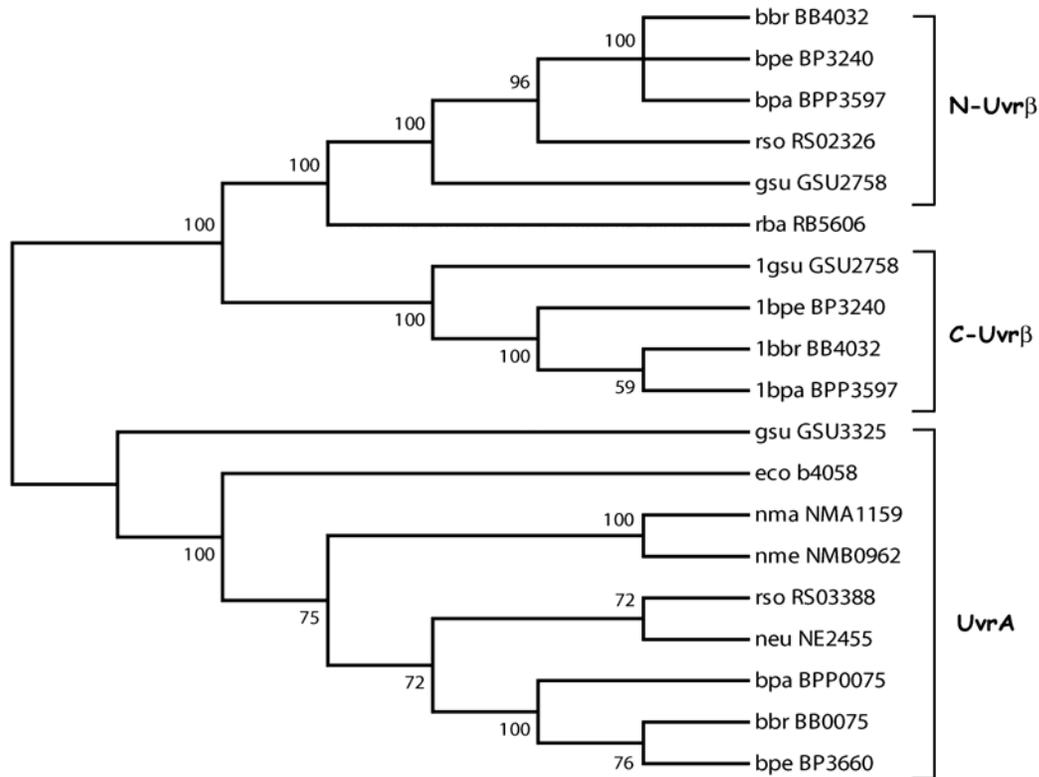


Figura 11. Árbol filogenético de las proteínas homólogas a UvrA encontradas en los genomas de  $\beta$ -proteobacterias, obtenido mediante un NJ con 10,000 réplicas de bootstrap. Los miembros de la subfamilia Uvr $\beta$  se muestran divididos en sus dos mitades para hacer posible su comparación con el resto de las proteínas (Preparado con MEGA2).

La proteína sufrió una duplicación interna originando una sola proteína con dos subunidades funcionales cuyo extremo C-terminal divergió a una tasa evolutiva más acelerada que el N-terminal que conservó la señal filogenética propia de la copia original (Fig. 11). La duplicación ocurrió en un ancestro de *Bordetella*, probablemente alguno relacionado con *B. bronchiseptica*, pues las otras dos especies carecen de mecanismos de sobrevivencia fuera de su único hospedero (Parkhill *et al.*, 2003), mientras que *B. bronchiseptica* posee un amplio rango de hospederos (focas, koalas, leopardos, cerdos, etc.) y se posiciona en la base de los árboles filogenéticos de las tres especies (Cummings *et al.*, 2004). Ello permite sugerir que *B. pertussis* y *B. parapertussis* surgieron a partir de

poblaciones de *B. bronchispetica* como especializaciones en hospederos particulares (van der Zee *et al.*, 1997). Las proteínas de la subfamilia Uvr $\beta$  se caracterizan por poseer un gran número de inserciones en secuencias (IS) y pérdidas de segmentos, cuya presencia robustece la señal filogenética que agrupa a ésta subfamilia. La frecuencia en la ocurrencia de elementos móviles como transposones, IS, inversiones y puntos de alta frecuencia de recombinación han sido reportados en *R. solanacearum* como parte de la propensión de éste genoma a tomar y recombinar DNA exógeno a través de la transformación natural (Salanoubat *et al.*, 2002); y en las especies de *Bordetella* como la consecuencia de relajar las presiones de selección sobre algunos eventos de transposición y reorganización genómica al adoptar un estilo de vida restringido a un hospedero (Andersson y Andersson, 1999; Parkhill *et al.*, 2003; Cummings *et al.*, 2004). Inicialmente consideramos que las pérdidas e IS dentro de éstas proteínas podrían indicar la inactividad del gen, es decir que en realidad el ORF perteneciera a un pseudogen. Sin embargo, las especies del género *Bordetella*, particularmente *B. pertussis* y *B. parapertussis*, se encuentran en un proceso de decaimiento y reducción del genoma (Cummings *et al.*, 2004), y por lo tanto no resultaría coherente conservar una proteína inutilizada de 1927 residuos de aminoácidos (5781pb) si cuenta con una proteína UvrA funcional (BB0075, BP3660, BPP0075).

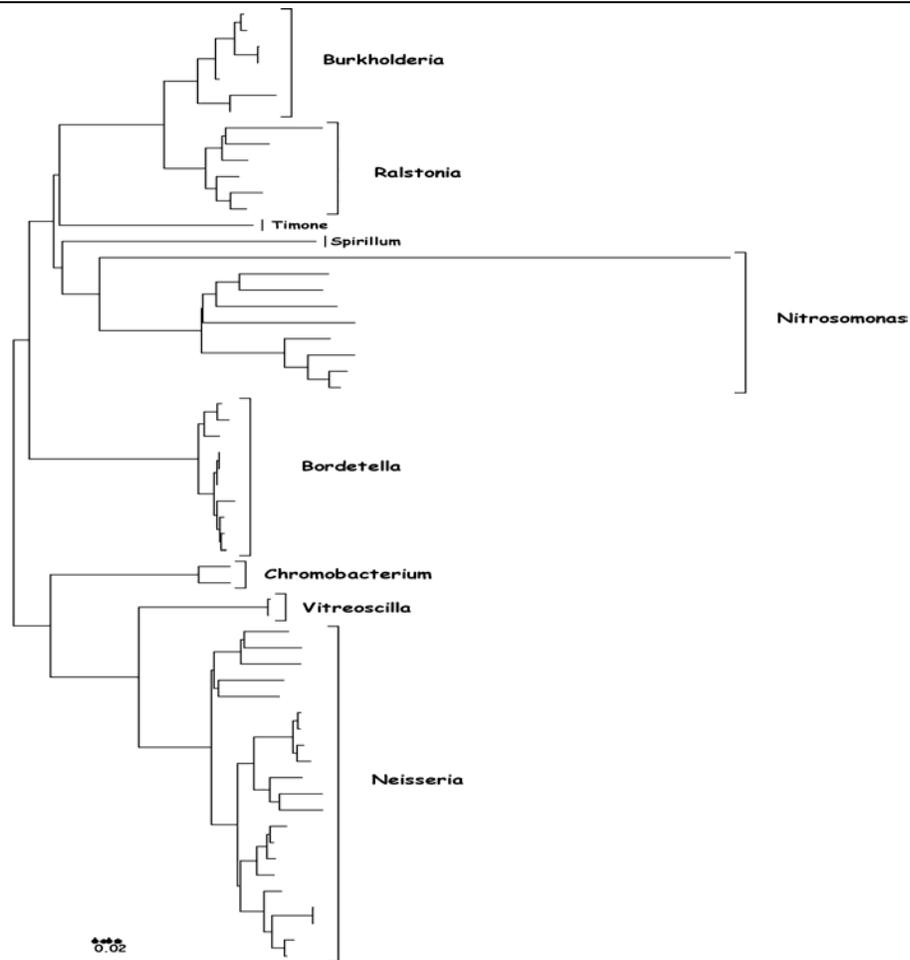


Figura 12. Árbol filogenético de las SSU rRNA de las β-proteobacterias obtenido por el consenso de un análisis de NJ con 1000 réplicas de bootstrap. (Alineación obtenida del Ribosomal Data Base, preparado con MEGA2).

Por último, resulta extraño que la proteína se encuentre ausente en el genoma de *Nitrosomonas europaea*, pues de acuerdo con el árbol filogenético por SSUrRNA de las β-Proteobacterias (Fig. 12), ésta se encuentra en el mismo clado de *Bordetella* -*Ralstonia*, estrechamente relacionada con la última. *Nitrosomonas europea* es una β-proteobacteria terrestre quimiolitioautótrofa obligada que posee un metabolismo basado en la oxidación del amoníaco a nitritos. Al parecer el genoma se encuentra en un proceso de reducción (Tabla 4), a pesar de poseer un gran número de duplicaciones e inserciones (IS), pues de la comparación de los genomas de litótrofos y fotótrofos resulta que basta un genoma muy reducido para mantener un estilo de vida autótrofo (Chain *et al.*, 2003).

Tabla 5. Comparación de los tamaños del genoma de algunas  $\beta$ -proteobacterias con otros organismos fotótrofos y litótrofos.

ORGANISMO	TAMAÑO DEL GENOMA
<i>R. solanacearum</i>	5.81 Gpb
<i>B. bronchiseptica</i>	5.34 Gpb
<i>B. parapertussi</i>	4.77 Gpb
<i>C. violaceum</i>	4.75 Gpb
<i>B. pertussis</i>	4.09 Gpb
<b><i>N. europaea</i></b>	<b>2.81 Gpb</b>
<i>C. tepidum</i>	2.15 Gpb
<i>Prochlorococcus spp.</i>	1.75 Gpb
<i>M. thermoautotrophicum</i>	1.75 Gpb

La proteína de *G. sulfurreducens* fue transferida horizontalmente a éste genoma por un ancestro del grupo *Bordetella* en un evento posterior a la divergencia de *R. solanacearum* y a la duplicación de la proteína. El ancestro pudo ser un organismo cercano a *R. solanacearum* o inclusive a *B. bronchiseptica*, pues ésta puede compartir el nicho edáfico con *G. sulfurreducens*. Sin embargo, para identificar precisamente el genoma de origen será necesario analizar los genomas de otras  $\beta$ -Proteobacterias comprendidas entre *R. solanacearum* y *Bordetella*, un ejemplo son los géneros *Nitrosomonas*, *Chromobacterium*, *Burkholderia* y *Spirillum*, todas ellas bacterias habitantes del suelo. La proteína RBA5606 de *Rhodopirelulla baltica* aparece anidada en el clado Uvr $\beta$ . Sin embargo, un análisis cuidadoso de la secuencia en la alineación múltiple revela que ésta proteína no comparte las pérdidas de segmentos ni las inserciones comunes a los demás miembros de ésta subfamilia (incluyendo a *G. sulfurreducens*), y por lo tanto su localización parece ser causada por un artefacto en el algoritmo de reconstrucción filogenética.

**4.5. Subfamilia UvrChlam.** Las proteínas agrupadas bajo ésta subfamilia pertenecen todas al *phylum* Chlamydiae: eubacterias Gram-negativas parásitas intracelulares obligadas, con un ciclo de vida dual conformado por un Cuerpo Elemental (CE) infeccioso y metabólicamente inactivo, y un Cuerpo Reticulado (CR) metabólicamente activo dentro del hospedero (Corsaro *et al.*, 2003). A pesar de su enorme importancia médica, las Chlamydiae son un grupo poco comprendido en cuanto a su origen y evolución, con un genoma relativamente pequeño resultado de un proceso de evolución reductiva (Read *et al.*, 2000). A pesar de que las dos especies secuenciadas a la fecha (*C. pneumoniae* y *C. trachomatis*) presentan una conservación asombrosa en cuanto al orden y contenido génico, y presentan niveles bajos de transferencia horizontal (Ortutay *et al.*, 2003), muestran un bajo nivel de similitud estructura primaria (Kalman *et al.*, 1999). Los genomas de Chlamydiae son los únicos genomas analizados que no presentan una copia bien conservada de la proteína UvrA dentro del clado central de la subfamilia UvrA. Sólo presentan una copia homóloga por genoma y todas están duplicadas. En la reconstrucción filogenética global es ésta subfamilia la que se desprende en la base, incluso cuando se consideran por separado los dos extremos de la proteína. Ello evidencia una enorme divergencia a nivel de estructura primaria posiblemente debida a la adaptación al estilo de vida intracelular. Por lo tanto, es probable que la posición basal de la subfamilia se deba más al fenómeno de atracción de ramas largas que a una divergencia temprana del grupo. En la atracción de ramas largas, las secuencias más divergentes se agrupan, y éste es el caso de los módulos ABC con la subfamilia UvrChlam. En favor de ésta hipótesis, la divergencia de ésta subfamilia ocurre fuera de cualquier otro clado, cuando es obvio que si reflejara la divergencia temprana de las Chlamydiae entonces la encontraríamos en la base del clado de la subfamilia UvrA *sensu stricto*. La otra posibilidad, que se antoja poco

probable, es que en realidad Chlamydiae represente la primera divergencia basal de las eubacterias, en una etapa anterior incluso a la divergencia del resto de los principales bacterianos. Sin embargo, el complejo ciclo de vida del *phylum* Chlamydiae apoya la hipótesis de las ramas largas.

A pesar de ser tan divergente, ésta proteína debe ser la copia funcional de UvrA involucrada en la escisión de nucleótidos porque, como se analizó anteriormente, éstas proteínas forman parte del núcleo de genes indispensables en los genomas bacterianos. El evento de duplicación no es tan atípico si se considera que UvrA funciona a manera de homodímero *in situ*, explicando la persistencia de una fusión/duplicación en un genoma reducido.

Tanto la reconstrucción filogenética de la familia completa como la inspección visual argumentan en contra de un origen común entre todas las proteínas UvrA *sensu lato* duplicadas ( $\beta$ -Proteobacterias, Planctomycetes y Chlamydia), y en cada caso ocurrió un evento de duplicación independiente.

**4.6. Subfamilia UvrAPlmy.** De las tres proteínas homólogas a UvrA que contiene el genoma de *Rhodopirellula baltica*, dos copias corresponden a una subunidad sencilla (RB10876 y RB5606) y la otra presenta una duplicación interna (RB7959).

RB10876 se posiciona en la base del árbol de las UvrA *sensu stricto* (RB10876) y es la que presenta mayor similitud con esa subfamilia a pesar de la pérdida del WA<sub>x</sub>, y por lo tanto debe corresponder a la subunidad funcional de la proteína de escisión de nucleótidos UvrA. La otra se encuentra artificialmente anidada en la subfamilia Uvr $\beta$ , discutida anteriormente (RB5606). Al analizar las dos mitades de la copia duplicada (RB7959), el extremo N-

terminal guarda una gran similitud a UvrA, inclusive más que RB10876, pues ésta subunidad sí conserva el WA<sub>X</sub>. El extremo C-terminal, por el contrario, aparece fuera de dicho clado, cercana a el extremo C-terminal de las Uvrβ. Aunque en ningún otro genoma bacteriano existen dos copias bien conservadas de una UvrA funcional, es posible que en *R. baltica* exista una copia sencilla funcional bien conservada (RB10876) y una copia funcional duplicada con un extremo bien conservado y el otro erosionado (RB7959), pues la proteína UvrA funciona a manera de homodímero *in situ*. La presencia de dos heterocopias funcionales de UvrA es consistente con que el genoma de *R. baltica* posee diversos mecanismos que reducen el daño por radiación UV: el regulón SOS (*recA*, *lexA*, *uvrA*, *uvrB*, *uvrC*, *inter alia*); un operón que involucra genes de la fotoliasa, fitoeno deshidrogenasa y fitoeno sintasa; la biosíntesis de un carotenoide protector contra UV y varios homólogos a proteínas de estrés de calor (Glöckner *et al.*, 2003).

Se ha sugerido con anterioridad que los Planctomycetes son un grupo antiguo, debido a la ausencia de peptidoglicanos en su pared celular exclusivamente proteínica (König *et al.*, 1984; Liesack *et al.*, 1986) y las evidencias con los marcadores moleculares 16S rRNA (Stackebrandt *et al.*, 1984, Brochier y Philippe, 2002) y la subunidad β de la ATPasa (Rönner *et al.*, 1991) que presentan al grupo como una rama profunda. Sin embargo, el uso de otros marcadores moleculares presentan a los Planctomycetes asociados a las Proteobacteria (23SrRNA, Schleifer y Ludwig, 1989; EF-Tu, Jenkins y Fuerst, 2001) o las Gram-positivas con bajo contenido de G+C (5SrRNA, Van den Eynde *et al.*, 1990). Aunque las dos copias que hipotéticamente están involucradas en escisión de nucleótidos (RB10876 y RB7959<sub>N</sub>.) divergen en la base del árbol de UvrA en Bacteria, no es coherente que un grupo bacteriano con una gran diversidad morfológica y ecológica (Gripenburg *et*

*al.*, 1999; Ward *et al.*, 1995; Wang *et al.*, 2002), una organización celular polar y ciclos de vida complejos (Schlesner, 1994) y una compartimentalización citoplásmica que contiene al nucleoide bacteriano en una membrana interna (Fuerst y Webb, 1991; Lindsay *et al.*, 2001), represente a los organismos bacterianos más antiguos. Además, existe evidencia que relaciona a los Planctomyces con las Chlamydia, como la pérdida secundaria de peptidolicanos en la pared celular (König *et al.*, 1984, Hatch, 1996), la ausencia de los genes *ftsZ* y en consecuencia una división celular desconocida (Glöckner *et al.*, 2003; Brown y Rockey, 2000), la presencia de dos copias genómicas de *dnaA* relacionadas filogenéticamente en ambos grupos (Glöckner *et al.*, 2003), comparten firmas en el 16SrRNA que no se encuentran en otros linajes bacterianos (Fuerst, 1995), y la evidencia filogenética de un clado *Chlamydia-Planctomycetes* con alineaciones concatenadas de proteínas ribosomales, subunidades de la RNAPolimerasa y árboles de genomas completos (Teeling *et al.*, 2004), desplazándolos de su posición basal. Si verdaderamente Chlamydiae y Planctomycetes conforman un clado común, entonces es muy probable que las proteínas duplicadas en *Chlamydia* y en *R. baltica* tengan un mismo origen, y por lo tanto el alto grado de divergencia de la subfamilia UvrChlam por evolución taquitética desplace a las proteínas de *R. báltica* hacia la base de los árboles por atracción de ramas largas. Para comprobar ésta hipótesis, repetimos la reconstrucción filogenética de la subfamilia UvrA *sensu stricto* excluyendo las proteínas de Chlamydiae y RB5606.



las copias homólogas a UvrA presentes en éste genoma, será necesario analizar otros genomas de Planctomyces, como el de *Gemmata obscuriglobus* (en secuenciación por The Institute of Genomic Research, TIGR), y determinar si la presencia de tres copias homólogas y la duplicación en una de ellas es común a otros Planctomycetes o es un evento que ocurrió independientemente en *R. baltica*.

**4.7. Proteínas huérfanas.** Las seis proteínas que no se agrupan en ningún clado evidente tienen una sola cosa en común: que todas han sufrido pérdidas sustanciales en las regiones funcionales. Por lo tanto, existe la posibilidad de que no sean genes verdaderos sino pseudogenes producto de una duplicación y su posterior inactivación. En consecuencia, su agrupación resulta artificial como consecuencia de la exclusión por divergencia. Sin embargo, es posible inferir el origen de algunas agrupaciones:

**UvrLac y UvrBac:** Dos clados de proteínas huérfanas pertenecen a Firmicutes (*L. plantarum*, *L. johnsonii*, *B. halodurans*, *O. iheyensis*) y Actinobacteria (*B. longum*), cuando todas las demás copias secundarias de estos órdenes pertenecen a la subfamilia DrrC. Es posible que las cinco proteínas hayan pertenecido también a DrrC, y ya sea por su inactivación o simple divergencia se segreguen fuera del clado por artefacto del algoritmo. En favor de ésta hipótesis se encuentra el hecho de que ninguno de los cinco genomas poseen otras proteínas homólogas a UvrA en las demás subfamilias (incluyendo la misma DrrC).

**UvrBGS:** Éste clado agrupa tres especies de dos *phyla* particulares, el *phylum* Chlorobia (*Chlorobium tepidum*) de las bacterias verdes sulfurosas, y el *phylum* Bacteroidetes de organismos predominantes en la microbiota intestinal de animales (*Bacteroides thetaiotaomicron*, *Porphyromonas gingivalis*). Tradicionalmente no se considera que Chlorobia pertenezca a una rama profunda en la base del árbol eubacteriano (Woese, 1987, Jenkins y Fuerst, 2001), a diferencia de Bacteroidetes que se reconoce como una divergencia temprana (Xu *et al.*, 2003). Sin embargo, también se ha reconocido que Chlorobia y el *superphylum* Cytophaga-Flavobacteria-Bacteroides (CFB) están relacionados (Woese, 1987; Gruber *et al.*, 1998); que el proteoma de *C. tepidum* resultó poseer la mayor similitud con *P. gingivalis* y que éstos dos organismos se agrupan juntos en todas las filogenias de los genes domésticos (Eisen *et al.*, 2002). El surgimiento de éste clado, idéntico al formado por sus homólogos UvrA, apoya el origen común de Chlorobia-CFB, y además, por la posición de las tres proteínas UvrA en su respectivo árbol, apoya una divergencia temprana de éste conglomerado. Esta subfamilia corresponde entonces a un grupo de proteínas duplicadas antes de la temprana divergencia de los dos *phylum*.

**UvrBba:** La proteína Bd0159 de la  $\delta$ -proteobacteria *Bdellovibrio bacteriovorans* presenta una alta tasa de substitución y pérdida de fragmentos cortos en toda su extensión. Además posee una inserción única rica en lisina de 38 residuos de aminoácidos entre el Zn1 y el ABC<sub>x</sub>.

## 5.- UVRA EN ARCHAEA

Se encontraron homólogos a UvrA en cinco genomas del dominio Archaea de los 18 genomas analizados, determinando la ruta UvrABC como un mecanismo predominantemente bacteriano. Estas proteínas no divergen en la base del árbol, sino que se encuentran al interior del clado de los ortólogos a UvrA *sensu stricto* (Fig. 5). Cuatro de estas proteínas corresponden a genomas de arqueas metanógenas (*M. mazei*, *M. acetivorans*, *M. thermoautotrophicum*, *M. maripaludis*), y la quinta a *Halobacterium* sp. No se encuentran en otras metanógenas como *M. kandleri* ni *M. jannaschii*. Las proteínas de *M. thermoautotrophicum* y *Halobacterium* sp. se segregan independientemente del grupo conformado por *M. maripaludis*, *M. acetivorans* y *M. mazei*, a pesar de que todas las metanógenas son monofiléticas (Slesarev *et al.*, 2001) y que las halófilas son metanógenas derivadas (Kennedy *et al.*, 2001; Forterre *et al.*, 2002); de manera que estas proteínas son más similares a los ortólogos bacterianos de UvrA que entre sí mismas y sugiriendo un origen polifilético de las proteínas arqueanas. La posición filogenética de estas proteínas no se debe a un artificio en el algoritmo de reconstrucción, pues las proteínas no se seplazan hacia la base ni a ramas más profundas al eliminar los artificios de atracción por ramas largas retirando los grupos externos del análisis (Fig. 13). Los cinco organismos arqueanos con UvrA son mesófilos, siendo la mesofilia un carácter secundario derivado porque el último ancestro arqueano parece haber sido hipertermófilo (Galtier *et al.*, 1999; Schouten *et al.*, 2000; Forterre *et al.*, 2000) y no se encontró homólogo alguno en los genomas de *T. volcanium* ni *T. acidophilum*, mesófilos del grupo basal Thermococcales; ni en ningún miembro de Crenarchaeota, lo que sugiere que la proteína se encontraba ausente en el último ancestro arqueano.

Lo anterior sugiere un posible evento de transferencia horizontal del dominio Bacteria hacia Archaea, considerando que el análisis filogenético es robusto al contar con un gran número y diversidad de ortólogos analizados; y que se excluyeron del análisis todos los parálogos que caracterizamos; ambos factores que afectan la sensibilidad de las aproximaciones por similitud (Ochman *et al.*, 2000; Wang, 2001). Ésta suposición está apoyada por el hecho de que en todas las arqueas encontramos las tres proteínas de la ruta (UvrA, UvrB, UvrC) en un sólo operón, y de acuerdo con la teoría del operón egoísta (Lawrence y Roth, 1996; Lawrence, 1999), los operones son evidencia de eventos de transferencia horizontal.

Sin embargo, no se encontraron evidencias de transferencia horizontal utilizando otras aproximaciones como las diferencias composicionales (contenido de G+C%; uso de codones, índice de adaptación de codones CAI) entre las UvrA arqueanas y el resto del genoma respectivo; ni la evidencia de ortología posicional (posición del gen sobre el cromosoma; Lawrence y Ochman, 1998; Koski *et al.*, 2001). No obstante, el uso de codones, contenido G+C% y el CAI resultan malos indicadores de HT por (i) la heterogeneidad propia del genoma en cuanto a éstos índices y (ii) la estandarización y homologación de elementos genéticos (Koski *et al.*, 2001; Wang, 2001; Daubin y Perrière, 2003). Por otra parte, la señal evolutiva de la ortología posicional puede ser borrada por rearrreglos genómicos como son las duplicaciones, transposiciones e inversiones.

En consecuencia, es posible detectar que un evento de HT ocurrió previamente a la divergencia de los géneros *Methanococcus* y *Methanosarcina*, pues las proteínas de éstas tres especies forman un clado monofilético que indica ortología entre ellas. Además, el 21% de los genes reportados en *M. acetivorans* tienen homólogos bacterianos y muestra una gran diversidad fisiológica y metabólica que le ha permitido colonizar una gran

variedad de hábitats (Galagan *et al.*, 2002) manteniéndola en contacto con muchos otros grupos bacterianos y aumentando la probabilidad de HT. También se ha reportado en la secuencia del genoma de *Methanosarcina mazei* la transferencia horizontal de genes bacterianos (Deppenmeier *et al.*, 2002).

Eventos diferentes e independientes debieron ocurrir en los genomas de *Halobacterium* y *M. thermoautotrophicum*, pues no forman un clado monofilético con el resto. La proteína de *Halobacterium* se agrupa con la proteína de *D. radiodurans* al eliminar los grupos externos (Fig. 13), sugiriendo que ésta copia arqueana puede provenir de un organismo del grupo Thermus/Deinococcus. *D. radiodurans* no sólo presenta resistencia a radiación, sino también a un alto nivel de desecación y escasez de nutrientes; y además ha sido aislada de una gran variedad de nichos ecológicos (White *et al.*, 1999).

## 6.- ESTRUCTURA TERCIARIA

La estructura central del módulo ABC se conserva en las tres subunidades UvrA (Fig. 14), y los dominios de zinc ocupan posiciones que no interfieren con la estructura central de la proteína, expuestos hacia su exterior.

La estructura predicha para los motivos Zn5Zn6 concuerda con la presencia de un motivo estructural *treble clef* invertido (Grishin, 2001), con el motivo de hojas- $\beta$  donando los dos primeros residuos de cisteína y el motivo de  $\alpha$ -hélices donando los dos últimos. La subunidad X es la más degenerada, perdiendo incluso una hoja-b del bolso de unión a nucleótidos. Sin embargo, es necesario recordar que para formar un sitio funcional de unión a nucleótidos se necesita la interacción de dos monómeros, y en éste caso el WA<sub>X</sub> se encontraría interactuando con el WB<sub>Y</sub>, que si mantiene éstos elementos bien conservados.

El motivo incompleto Zn1, por otra parte, aunque se encuentra localizado lejos del núcleo de unión a ATP, encara hacia el interior de la proteína.

El que en esta subunidad la pérdida en ésta región se oriente hacia la izquierda y en la subunidad Y se oriente hacia la derecha sugiere la interacción opuesta de los motivos Zn1 y Zn4 para la formación de un sitio de unión a zinc al interior del plegamiento de la subunidad X sobre Y. Por lo tanto, es probable que éste ion se involucre en la interacción entre ambas subunidades, brindándole estabilidad a la conformación XY. La localización de los motivos Zn5Zn6 sugiere que éste dedo de zinc autónomo se sitúa hacia atrás de la proteína, lejos del sitio de interacción con nucleótidos y por lo tanto sin interferir con su función.

Ésto sugiere que el motivo autónomo se encuentra involucrado en la interacción de la proteína con el DNA. La disposición de los motivos Zn5Zn6 es hacia la cara interna de la proteína, y como es ésta subunidad la que forma el homodímero, es probable que éste dedo de zinc esté involucrado en la interacción entre monómeros. Sin embargo, al encarar dos subunidades Z presentamos dos sitios autónomos de unión a zinc, cada uno correspondiente a cada monómero. La región C-terminal de UvrA, equivalente a la subunidad Z, es la que ha demostrado ser indispensable en la reparación por escisión de nucleótidos y que además presenta interacción con el DNA. Por lo tanto, es posible que éste sitio múltiple de unión a zinc permanezca ocupando uno de los motivos al revisar la estructura del DNA, y que cualquier irregularidad en el DNA cause el desplazamiento del ion hacia el otro motivo que consecuentemente traerá un reacomodo en la estructura terciaria de la proteína, inactivando la hidrólisis de ATP y por lo tanto la liberación de la proteína del DNA. La subunidad Y presenta un pérdida significativa en el dominio helicoidal, que aparentemente es la región de señalización, lo que indica que dicha subunidad no participa en la señalización, y por lo

tanto sólo es necesaria para aportar el motivo WB. Ésto explicaría la alta tasa de divergencia global de la subunidad.

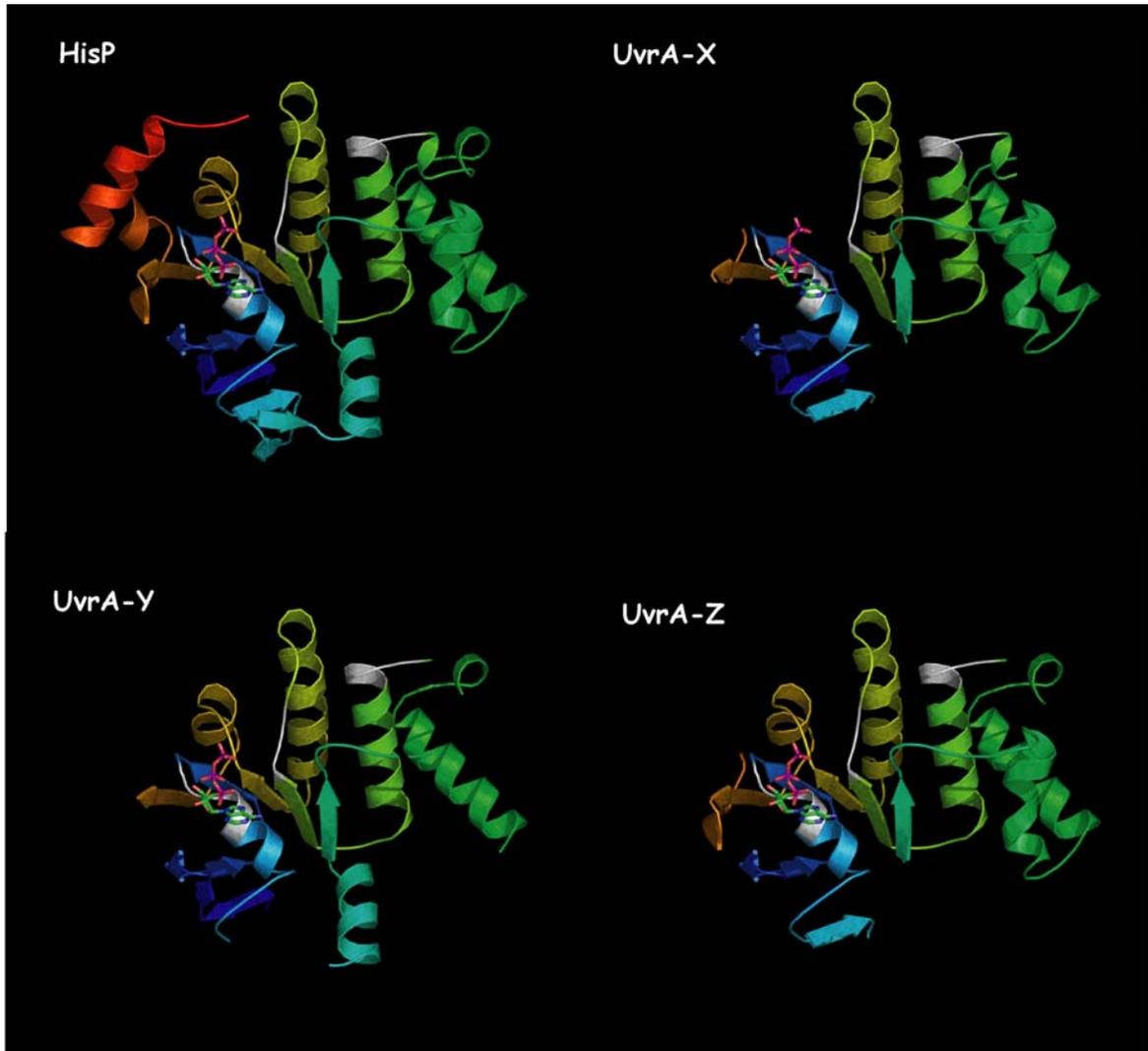


Figura 14. Comparación entre la estructura terciaria del cristal del transportador ABC de histidina HisP (PDB:1BUO) y las estructuras terciarias hipotéticas de las tres subunidades de UvrA calculadas a partir de la alineación de estructura secundaria. Una molécula de ATP se muestra en esquema tubular. En blanco se indican los motivos WA, WB y ABC. (Preparadas con PyMol).

## 7.- MODELO FUNCIONAL DE UVRA

A pesar de que la modalidad tripartita de UvrA fue descubierta desde hace casi veinte años (Doolittle *et al.*, 1986); los estudios experimentales tradicionalmente han considerado a UvrA como una proteína consistente en dos subunidades asimétricas, debido a que las dos primeras subunidades se pliegan sobre si mismas, conformando el equivalente a una sola subunidad estructural (Myles y Sancar, 1991) en donde interaccionan los motivos conservados  $WA_X$  y  $ABC_Y/WB_Y$  (Doolittle *et al.*, 1986). Asimismo, al considerar exclusivamente dos subunidades funcionales cobra sentido la presencia de dos dedos de zinc por subunidad, el  $Zn_2Zn_3$  de la subunidad X, y el  $Zn_5Zn_6$  de la subunidad Z. En consecuencia, se ha subestimado la importancia de los motivos  $Zn_1$  de la subunidad X y  $Zn_4$  de la subunidad Z por considerarlos incompletos (Navaratnam *et al.*, 1989; Gorbalenya y Koonin, 1990; Myles y Sancar, 1991; Wang *et al.*, 1994; Moolenaar *et al.*, 2000; Goosen & Moolenaar, 2001), aún considerando que la modalidad funcional de la proteína completa es como homodímero (Myles *et al.*, 1991) aunque el extremo N-terminal (subunidades X e Y) conserva por sí solo actividad de ATPasa y de unión a DNA (Myles y Sancar, 1991).

Al considerar que las subunidades X e Y se posicionan una sobre otra de tal manera que el motivo  $WA_X$  interactúa con el  $ABC_Y/WB_Y$ , entonces la preservación de los motivos  $Zn_1$  de la subunidad X y  $Zn_4$  de la subunidad Y cobra sentido al indicar la conformación de un sitio de unión a zinc completo  $Zn_1Zn_4$ . Esta teoría se fundamenta en que la estequiometría de la molécula fue calculada en  $1.6 \text{ mol Zn}^{2+} / \text{molécula de UvrA}$ , cifra estimada en 2 átomos de zinc por molécula (Navaratnam *et al.*, 1989), lo cual resulta coherente si se consideran sólo dos sitios de unión a zinc ( $Zn_2Zn_3$  y  $Zn_5Zn_6$ ). Sin embargo la naturaleza dimérica de UvrA presenta mayor consistencia con la extrapolación a  $3.2 \text{ mol Zn}^{2+} / 2 \text{ mol UvrA}$ , es decir aproximadamente tres átomos de zinc por dímero funcional de

UvrA. Por otra parte, el ATP tiene una mayor afinidad por el sitio  $WA_X$  que por  $WA_Z$  la proteína sólo presenta un ATP por dímero (0.5 mol ATP / 1 mol UvrA), lo que ha llevado a proponer un modelo en el cual la hidrólisis de ATP en el  $WA_X$  induce la dimerización de la proteína (Myles y Sancar, 1991) y la unión de ATP al segundo sitio  $WA_Z$  (Myles *et al.*, 1991). El segundo sitio corresponde a la ATPasa conformada por las dos subunidades Z, cuya hidrólisis libera la proteína del DNA cuando no se ha detectado daño. Por último, se ha determinado que si bien el sitio  $Zn5Zn6$  es imprescindible para llevar a cabo tanto la función de escisión de nucleótidos como la de replicación alternativa, el sitio  $Zn2Zn3$  sólo es necesario para la replicación alternativa (Moolenaar *et al.*, 2000a). Con esta información, proponemos el modelo siguiente:

- 1) Un monómero UvrA emplea un átomo de zinc en el sitio  $Zn1Zn4$ , permitiéndole formar el bolsillo necesaria para la unión de nucleótidos al estabilizar la estructura.
- 2) Un ATP se une al monómero, y su hidrólisis induce la dimerización de la proteína (Mazur y Grossman, 1991).
- 3) La dimerización emplea otro átomo de zinc para estabilizar el dímero de las subunidades Z mediante sus respectivos sitios  $Zn5Zn6$ . Como en realidad se confrontan seis residuos de cisteínas, existen al menos dos configuraciones posibles que puede adquirir el complejo  $UvrA_Z-Zn^{2+}$ .
- 4) La dimerización de la proteína induce la unión de ATP al segundo sitio, conformado por el  $WA_Z$  de un monómero y el  $ABC_Z/WB_Z$  del otro (Myles y Sancar, 1991).
- 5) La proteína busca lesiones sobre el DNA. Si no encuentra daño, la hidrólisis del ATP libera el DNA. En caso contrario, la irregularidad en la estructura del DNA induce el movimiento del átomo de zinc a la segunda posición posible, con otra conformación

de Zn5Zn6. Éste cambio separa las subunidades Z, bloqueando la actividad ATPasa y fijando la molécula UvrA sobre la lesión del DNA (Seeberg y Steinum, 1982; Goosen Moolenaar, 2001).

- 6) En el proceso alterno de replicación, el extremo N-terminal (XY) del monómero carente de zinc une el tercer átomo al motivo Zn2Zn3, en donde ayuda a la unión y reconocimiento de la cadena retardada (Moolenaar *et al.*, 2000a). Si no encuentra primers, la hidrólisis de ATP la liberará de la cadena retardada. Si encuentra un híbrido DNA-RNA, el cambio en la posición del zinc Zn5Zn6 bloqueará la liberación de la proteína.
- 7) La proteína UvrB se une covalentemente al segmento dañado o al primer y UvrA es liberada del complejo. En la escisión de nucleótidos, UvrC corta la cadena a ambos lados de la lesión.
- 8) La helicasa UvrD se une al sitio de unión sobre UvrB y desenrolla el segmento dañado o el primer.

## 8.- RELACIONES EVOLUTIVAS CON OTRAS FAMILIAS

**8.1 Transportadores ABC.** La familia UvrA es miembro de la superfamilia ABC, que contiene en su mayoría transportadores transmembranales cuya energía es derivada de la hidrólisis de una molécula de ATP. Ésta superfamilia es la más numerosa que se conoce, con una sorprendente diversidad de sustratos a transportar. También se han encontrado módulos ABC en proteínas con funciones celulares como la resistencia a antibióticos (ART) y el inhibidor L de la RNAsa (RLI). Todas éstas proteínas conservan en diferente grado la estructura primaria del módulo de unión a nucleótidos ABC.

Las diferencias de UvrA con el resto de la superfamilia son: (a) mientras que los ABC poseen dos módulos ABC muy similares, UvrA posee tres subunidades muy diferentes, (b) todos los módulos ABC funcionan como homo/hetero dímeros, y aunque UvrA funciona como homodímero, en realidad su estructura corresponde a un doble trímero, es decir, seis subunidades organizadas por parejas para formar tres subunidades funcionales, (c) UvrA posee motivos de unión a Zinc bien conservados tanto en la mitad de la subunidad como en la región que conecta cada subunidad, que no se hallan en ABC y deben modificar significativamente sus estructuras terciaria y cuaternaria, (d) UvrA posee un motivo rico en glicinas bien conservado hacia el final de las subunidades ABC que no se encuentra en la familia ABC: el motivo UVG, (e) UvrA ha perdido el Q-loop en todas sus subunidades, el cual aparentemente interactúa con el dominio TM en los módulos ABC, (f) UvrA ha perdido el  $WB_X$  y  $SW_X$  y por último, no se ha detectado la presencia de alguna proteína homóloga o equivalente al dominio transmembranal (TM) de los transportadores ABC.

De las alineaciones que se realizaron con los módulos ABC cristalizados, se puede observar que la topología general en cuanto a orden y presencia de la mayoría de los elementos de estructura secundaria se conservan, sugiriendo a su vez la conservación general de la estructura de cada módulo ABC. La estructura de sandwich  $\beta$ - $\alpha$ - $\beta$  que contiene el WA se conserva muy bien. Las inserciones de los motivos Zn1, Zn4, Zn5 y Zn6 se encuentran en la región más variable de la superfamilia, entre el WA y el ABC, que corresponden estructuralmente a las hélices que le siguen al WA, la 5 y 6, y la hoja-B 2. La región helicoidal que precede al ABC se conserva parcialmente, si bien no se puede afirmar que todos los elementos son iguales, sí podemos predecir la presencia de una región helicoidal también. El segmento final de cada subunidad es el menos conservado, que

corresponde a todas las  $\alpha$ -hélices que se encuentran en la parte trasera de la proteína; en ésta región se encuentra el motivo UVG, que aparentemente adquiere una conformación de hojas-b. El Zn<sub>2</sub>Zn<sub>3</sub> de la subunidad X aparece en ésta región.

Los motivos Zn obviamente no pudieron ser alineados, y al realizar la predicción estructural de la proteína sólo, no se hallaron regiones homólogas con cristales. Sin embargo, el JPred logró predecir que el Zn<sub>5</sub> de la subunidad Z debe conformarse por hojas-B, y más vagamente el Zn<sub>6</sub> debe conformarse por  $\alpha$ -hélices. Éste motivo debe presentarse expuesto hacia el frente de la proteína y por lo tanto interactuar con el equivalente en la subunidad Z del homodímero. De la misma manera, los motivos Zn<sub>1</sub> y Zn<sub>4</sub> se muestran hacia el interior de las proteínas, y deben interactuar uno con el otro al formarse el sitio de unión a nucleótidos WA<sub>X</sub>-ABC<sub>Y</sub>/WB<sub>Y</sub>. En contraste, el motivo Zn<sub>5</sub>Zn<sub>6</sub> queda expuestos hacia atrás, lejos de cualquier interacción aparente con las otras subunidades, aunque al desconocer la estructura cuaternaria no deja de ser una simple suposición.

UvrA es claramente un miembro de la superfamilia ABC, pero ha divergido enormemente del resto de los transportadores membranales, perdiendo motivos importantes y adquiriendo otros.

**8.2 Familia SMC.** La familia de mantenimiento estructural de cromosomas SMC (*Structural Maintenance of Chromosomes*) reúne proteínas de 800 a 1500 residuos involucradas en la cohesión de cromátidas hermanas, condensación cromosómica y procesamiento de rupturas de doble hebra en el DNA, caracterizada por poseer los sitios de unión a ATP WA y WB separados por una extensa región de  $\alpha$ -hélices conocida como dominio CC (*coiled coil*) (Cobbe y Heck, 2000). La conservación de éstos dos motivos y la

presencia de la firma de la familia ABC y el dominio SW llevó a Gorbalenya y Koonin (1990) a incluir a Rad50 dentro de su superfamilia de proteínas relacionadas a UvrA.

La proteína procarionte SbcC es una exonucleasa de doble hebra involucrada en la eliminación de estructuras en “pasador” (*hairpin*) del DNA (Connelly *et al.*, 1998). La proteína eucarionte Rad50 es parte del complejo de recombinación homóloga, mantenimiento de telómeros y asociación de cromátidas hermanas Mre11 (Hopfner y Tainer, 2003). Ambas proteínas presentan un motivo de unión a zinc bien conservado (CXXC) a la mitad de la región de  $\alpha$ -hélices conocido como el “gancho” o *hook construct* (Hk) (Sharples y Leach, 1995). Tanto la longitud y función de las proteínas como la presencia y orden de los tres motivos sugerían una relación con las proteínas UvrA. Sin embargo, un análisis más detallado muestra que la longitud alcanzada con un sólo módulo ABC por las subfamilias SbcC y Rad50 es equivalente en UvrA a la repetición de tres módulos ABC. Por lo tanto, mientras que SMC posee sólo un WA, una firma ABC, un WB y un Zn, UvrA posee tres WA, tres ABC, dos WB y seis Zn. Las SMC conservan aún el Q-loop característico de los transportadores ABC, el cual se encuentra ausente en UvrA; y en SMC la mayor parte de las proteínas consiste de  $\alpha$ -hélices sin correspondencia en UvrA. Por otra parte, el dominio UVG característico de las subunidades de UvrA no se encuentra presente en ninguna de las proteínas SMC analizadas (Figs. 15 y 16). Nos encontramos así con que los únicos motivos compartidos entre las dos familias son el WA, el WB y el medio dedo de zinc Zn. Como señalamos anteriormente, aún no existe una clasificación natural y filogenética de los motivos de unión a zinc, que no pueden ser identificados a partir de la estructura primaria (Berg, 1986) e incluso contando con datos estructurales no es sencillo clasificarlos adecuadamente (Grishin, 2001; Krishna *et al.*, 2003) porque sólo

mantienen conservados los residuos de cisteínas (Michael *et al.*, 1992). En la región del motivo Zn sólo se conservan las dos cisteínas, una glicina dos residuos antes de la primera cisteína y parcialmente una prolina después de la primera cisteína; entre las dos familias. El programa JPred no logró predecir la estructura secundaria de uno de los Zn homólogos en UvrA, el Zn1 en la subunidad X que muestra mayor similitud, pero sí fue posible identificar al menos una región de hojas- $\beta$  en el Zn5 y  $\alpha$ -hélices en el Zn6, lo que sugiere una topología de treble-clef. Sin embargo, el cristal del motivo Hk en Rad50 muestra que en ése caso el motivo está conformado por las dos  $\alpha$ -hélices del dominio CC. La simplicidad de los motivos de unión a Zinc sugiere que posiblemente éstos han surgido recurrentemente durante la evolución de diferentes proteínas (Krishna *et al.*, 2003), y por lo tanto es posible que el Hk de Rad50 no sea homólogo a los Zn1 y Zn5 de UvrA, pues aparentemente no conservan tampoco la estructura tridimensional.

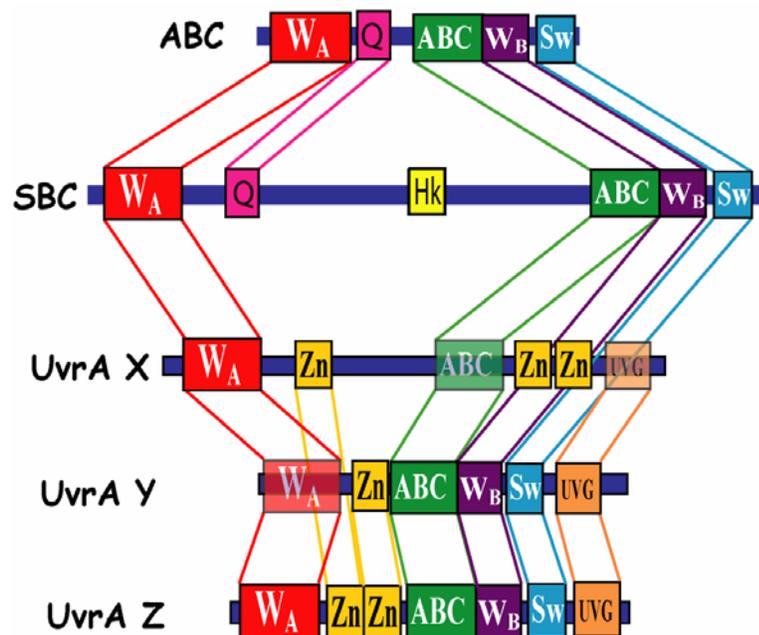


Figura 15. Comparación entre la organización de un módulo de transporte ABC, una proteína de la familia SBC y las tres subunidades de una proteína UvrA.



Figura 16. Alineación múltiple de los dominios característicos de la familia UvrA compartidas por sus tres subunidades (X, Y, Z). En la alineación se les compara con (a) el transportador de histidina HisP de la superfamilia de transportadores ABC; (b) las dos unidades del inhibidor de la RNasa L de la familia RLI; (c) la proteína Rad50 de *P. furiosus* y (d) la proteína SbcC de *D. radiodurans*, de la familia SBC. Se incluye también el motivo Q-loop de la superfamilia de transportadores ABC, compartido con la familia Rad50/SBC pero no con UvrA. (Alineación preparada con BioEdit).

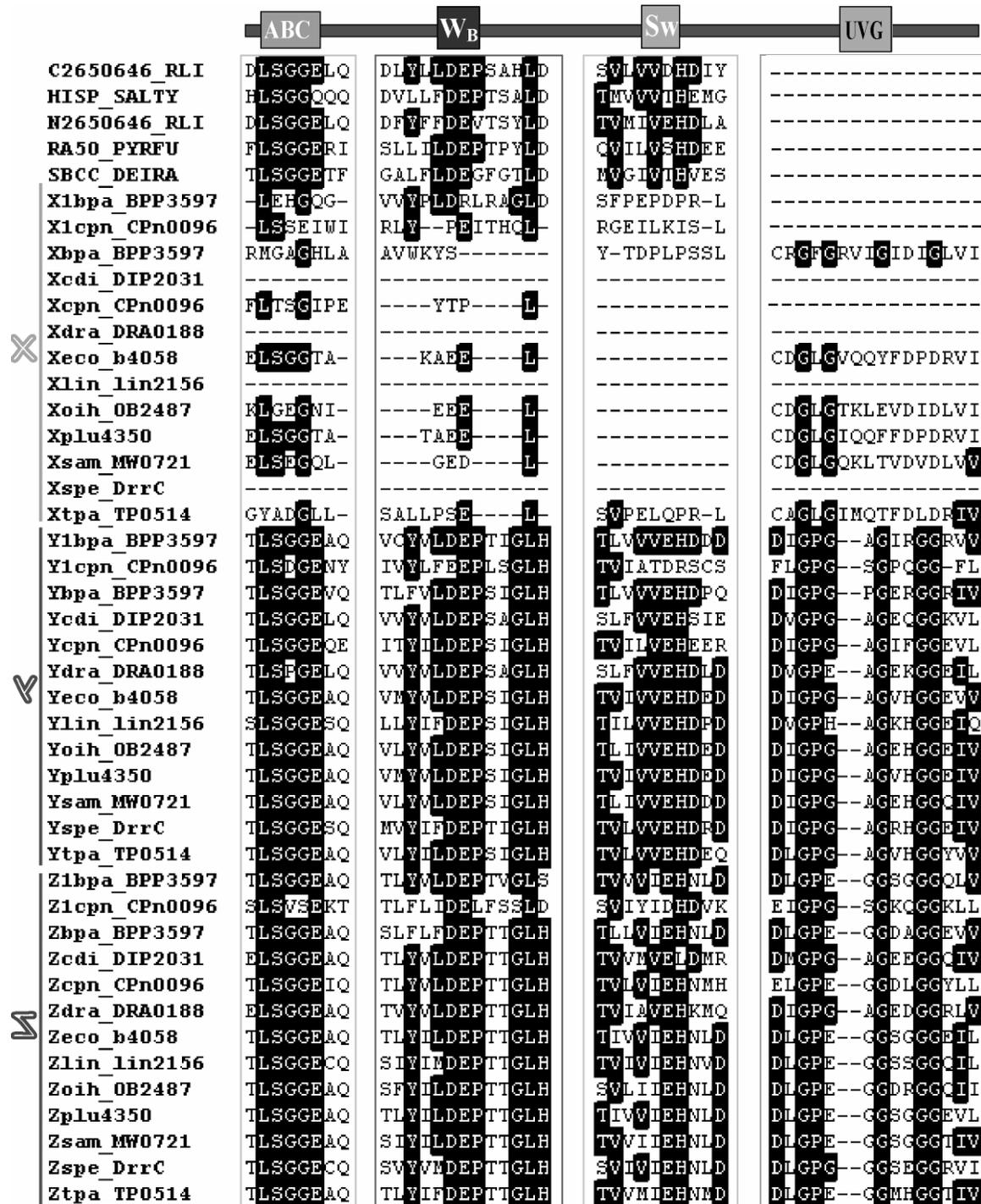


Figura 16. (continuación).

En conclusión, los únicos motivos compartidos por SMC y UvrA son los motivos de unión a nucleótidos WA/WB, la firma ABC y el SW, motivos que ambas familias conservan con la superfamilia ABC de la que son miembros.

Dicho de otra forma, la familia SMC no se parece más a UvrA de lo que se parece al resto de los transportadores ABC. Por lo tanto la superfamilia de proteínas relacionadas con UvrA propuesta por Gorbalenya y Koonin (1990) es una agrupación artificial, pues se basa en la conservación de los motivos WA, WB, ABC y SW que se encuentran conservados en todos los miembros de la superfamilia ABC. (Fig. 15 y Fig. 16).

## 9.- HISTORIA EVOLUTIVA DE LA FAMILIA UVRA

En 1986, Doolittle *et al.* propusieron un modelo de la evolución de los diferentes motivos de UvrA. El modelo mostrado en la Fig. 17 fue construido utilizando exclusivamente la proteína UvrA de *E. coli*, que era la única disponible entonces, y estimando el orden de divergencia individual basado en la similitud de los motivos. Para llevar a cabo el presente trabajo hemos tenido a nuestra disposición las secuencias de las proteínas de 141 genomas bacterianos, de las cuales hemos utilizado las 27 mejor conservadas para evaluar su historia evolutiva. A diferencia del trabajo de Doolittle *et al.* (1986), no consideramos la evolución individual de cada motivo, sino de cada subunidad ABC, pues la presencia del motivo Zn1 sugiere paralogía entre ellas, y por lo tanto que éstas conforman también la subunidad evolutiva.

El proceso de alineación resultó particularmente complicado por dos factores concretos: la aparición del motivo Zn2-Zn3 en la subunidad X exclusivamente y el alto grado de pérdida de aminoácidos en la subunidad Y. Sin embargo, el algoritmo sensible a alineaciones locales del T-COFFEE detectó que los motivos Zn1 de la subunidad X y Zn4

de la subunidad Y son homólogos a los dominios Zn5 y Zn6 de la subunidad Z tanto en su estructura primaria como en la posición correspondiente respecto a los demás motivos. La presencia de dos dedos de zinc entre los motivos WA y ABC debe considerarse entonces como el estado ancestral, y la ausencia del equivalente a Zn5 en X y Zn6 en Y representan pérdidas secundarias como resultado de la interacción entre los dos dominios y la conformación de un dedo de zinc con los motivos Zn1Zn4.

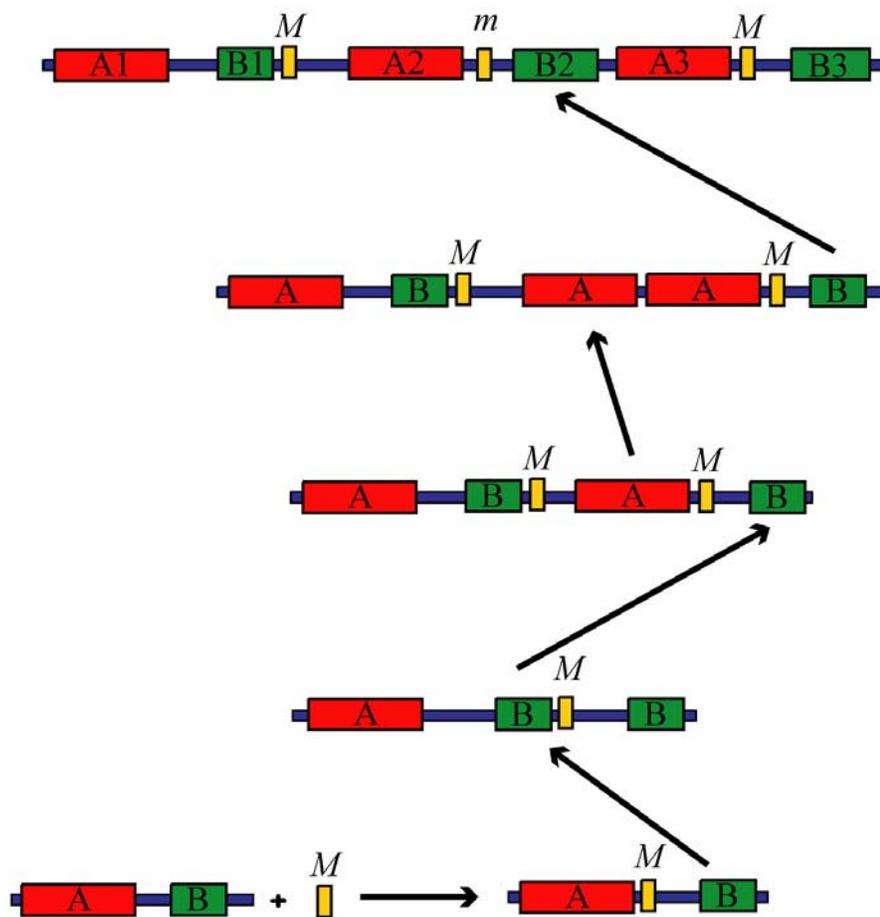
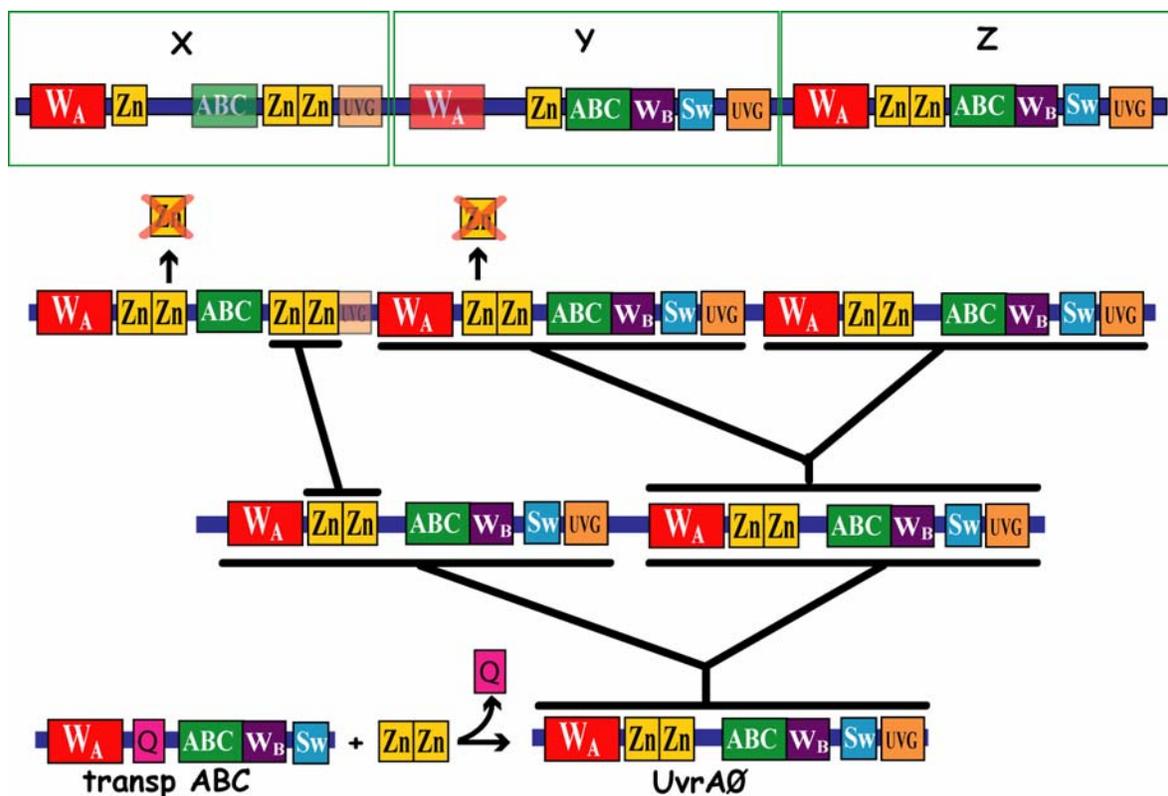


Figura 17. Modelo evolutivo para UvrA propuesto por Doolittle *et al.* (1986). Redibujado del original.

Los motivos Zn<sub>2</sub>Zn<sub>3</sub> surgen de una duplicación posterior y su permanencia muy posiblemente se debe a que al plegarse la proteína para dimerizar las subunidades X e Y, los dedos de zinc originales quedan en la cara interna de la proteína mientras que la posición de los motivos Zn<sub>2</sub>Zn<sub>3</sub> se encuentra en el área de pliegue, orientados hacia afuera y atrás del módulo.

La reconstrucción filogenética de las subunidades revela que la divergencia más reciente corresponde al evento de duplicación que dio origen a las subunidades Y y Z (Fig. 9), apoyado con un valor de bootstrap sorprendentemente alto (81%), si consideramos la antigüedad del evento de duplicación. De esta manera, el estado actual de la proteína UvrA se deriva del siguiente proceso, ilustrado en la Fig. 18:



**Figura 18.** Modelo evolutivo propuesto para las proteínas UvrA a partir de un módulo de unión a nucleótidos ABC sencillo. Las regiones mal conservadas están semitransparentes. Leyenda, **WA**: WalkerA, **Q**: Q-loop, **Zn**: par de cisteínas de unión a Zinc, **ABC**: firma de la familia ABC, **WB**: WalkerB, **SW**: SwitchII, **UVG**: dominio rico en glicinas.

- 1) Un dedo de zinc completo (con cuatro cisteínas) surge o se inserta en la región del dominio helicoidal de un módulo de transporte ABC, específicamente entre el WA y la firma ABC.
- 2) El gen *uvrA* sufre una duplicación interna completa que prevalece gracias a que la función de UvrA es homodimérica, como los módulos de transporte ABC.
- 3) La subunidad del extremo C-terminal sufre una duplicación más, la nueva estructura de la proteína no permite la dimerización pareada de cada dominio y prevalece el plegamiento de X sobre Y con la dimerización de Z.
- 4) El reacomodo espacial de los motivos favorece la aparición de los motivos Zn2Zn3 por recombinación desigual, perdiendo los motivos WB<sub>X</sub>, SW<sub>X</sub> y UVG<sub>X</sub>.
- 5) El surgimiento del dedo de zinc conformado por los motivos Zn1 de X y Zn4 de Y inducen la pérdida de las parejas de Zn correspondientes.
- 6) La conformación de la nueva subunidad ATPasa con los motivos WA<sub>X</sub>-ABC<sub>Y</sub>/WB<sub>Y</sub> liberan la presión de selección sobre el motivo WA<sub>Y</sub> y la firma ABC<sub>X</sub>, degenerando rápidamente.

Éste proceso no sólo es más simple y parsimonioso por involucrar un menor número de eventos evolutivos, sino que además considera a cada módulo UVR como la subunidad evolutiva sin contradecir las similitudes individuales entre motivos identificadas por Doolittle *et al.* (1986), apoyado en la evidencia estructural y filogenética disponible sin apelar a un gran número de eventos de reordenamiento en los motivos al interior del gen.

## V. CONCLUSIONES

1. UvrA es una proteína esencial y bien conservada en los genomas bacterianos secuenciados a la fecha.
2. La proteína UvrA no se encuentra en plásmidos, genomas de organelos celulares, ni endosimbiontes obligados, sugiriendo una pérdida secundaria por la reclutación nuclear de la función durante el establecimiento endosimbiótico.
3. La proteína UvrA se encuentra en cinco genomas de Archaea muy posiblemente debido a eventos de transferencia horizontal desde Bacteria.
4. Las subunidades Y y Z de UvrA poseen un motivo rico en glicinas exclusivo de ésta familia y sin reportar previamente hacia 3' del motivo Switch II.
5. El análisis de secuencias revela la existencia de al menos cuatro subfamilias naturales y al menos dos artificiales de proteínas parálogas. Las subfamilias naturales son:
  - a. **UvrA *sensu stricto***, involucrada en la escisión de nucleótidos.
  - b. **DrrC**, involucrada en la resistencia a daunorrubicina.
  - c. **UvrA2**, sin caracterizar funcionalmente.
  - d. **UvrA $\beta$** , sin caracterizar funcionalmente y presente en algunas  $\beta$ -proteobacterias.

Las familias artificiales son:

- e. **UvrAChlam**, involucrada en la escisión de nucleótidos en Chlamydiae.
- f. **UvrAPlmy**, con proteínas sin caracterizar funcionalmente del genoma de *R. baltica* (Planctomycetes).

6. Las tres unidades de UvrA mantienen la estructura terciaria general característica de los módulos ABC y los motivos de unión a Zinc no se encuentran en regiones que interrumpen ésta estructura.
7. La reconstrucción filogenética de UvrA *sensu stricto* mantiene la presencia de los principales linajes bacterianos, mas parece estar sumamente afectada por el fenómeno de atracción de ramas largas. Las filogenias disponibles apoyan débilmente la existencia del grupo Chlamydiae-Planctomycetes.
8. El funcionamiento homodimérico de UvrA sugiere la interacción de las unidades X e Y, mediada por un ion de zinc unido a los motivos Zn1 de X y Zn4 de Y, que hasta ahora no habían sido considerados en la literatura. Los motivos Zn2 y Zn3 se encuentran expuestos hacia atrás de la proteína, involucrados en la unión a DNA. Los cuatro motivos Zn5 y Zn6 de dos unidades Z estabilizan la dimerización al tiempo que revisan irregularidades en el DNA.
9. La familia UvrA no comparte motivos únicos con las proteínas de las familias SbcC ni Rad50, sólo comparte los motivos comunes a todos los miembros de la superfamilia ABC, a saber, WalkerA, WalkerB y la firma ABC. Por lo tanto, no hay evidencia de la existencia de una superfamilia de proteínas de unión a nucleótidos relacionadas con UvrA que incluyan proteínas de las familias SbcC y Rad50.
10. La disponibilidad de un gran número de secuencias y nuevas herramientas bioinformáticas nos permite proponer un modelo evolutivo más simple que el propuesto Doolittle et al. (1986), sin oponerse a la evidencia presentada originalmente en ese trabajo.

## VI. APÉNDICE

**Código en PerL (*Practical extraction and report Language*) del programa escrito para la automatización en la búsqueda en los genomas completos, construcción de perfiles para búsquedas reiterativas, corte de secuencias según el valor estipulado, obtención de la secuencia completa de cada proteína encontrada y anotación y corte de los dominios particulares de cada secuencia. El programa fue diseñado para el análisis de la superfamilia multigénica de los transportadores ABC (nuestro laboratorio, sin publicar), y por lo tanto los parámetros expuestos a continuación son los que corresponden al análisis de una sólo familia.**

```
#!/usr/local/bin/perl -s

#
#          PMASSYELv.2.0
#          Programa de automatización
#          búsqueda y corte de familias de proteínas en genomas completos.

# Opciones desde línea de comando:
# -b = baklizb; -m = microbio; -P = progenote; -W = watson;

$val = 1; #valor de la resta

if ($b) {
    print ("\n CORRIENDO EN baklizb \n");
    $dirhmm = ("/usr/inv/abb/hmmer-2.3.1/bin");
    $dirgenoma = ("/tmpu/abb");
    $dirlocal = ("/tmpu/abb/abc");
}
if ($m) {
    print ("\n CORRIENDO EN microbio.fciencias.unam.mx \n");
    $dirhmm = ("/usr/bin");
    $dirgenoma = ("/home/genomas");
    $dirlocal = ("/home/german/tst");
}
if ($P) {
    print ("\n CORRIENDO EN progenote.fciencias.unam.mx \n");
    $dirhmm = ("/hmmer-2.3.1/bin");
    $dirgenoma = ("/bio/genomas"); $dirlocal = ("/home/fernis/tst");
}
if ($W) {
    print ("\n CORRIENDO EN Watson \n");
    $dirhmm = ("/bio/hmmer/hmmer-2.2g/binaries");
    $dirgenoma = ("/bio/genomas");
    $dirlocal = ("/home/demian/tst");
}

print ("\nXXXXXXXXXXXXXXXXXXXXX I N I C I O XXXXXXXXXXXXXXXXXXXXXXXX\n\n");
%genomas = qw(
bsu      b.subtilis.pep      atd      a.tumefaciens_d.pep
eco      e.coli.pep              ana      anabaena.pep
pab      p.abyssi.pep           ban      b.anthraxis.pep
chl      plastid.pep        bab      b.aphidicola_bp.pep
mt       mitochondrion.pep  bas      b.aphidicola_sg.pep
ape      a.pernix.pep         bbu      b.burgdorferi.pep
pl       plasmid.pep       bce      b.cereus.pep
aae      a.aeolicus.pep   bha      b.halodurans.pep
afu      a.fulgidus.pep   bja      b.japonicum.pep
ath      a.thaliana.pep   blo      b.longum.pep
atu      a.tumefaciens.pep bme      b.melitensis.pep
atc      a.tumefaciens_c.pep bsu      b.subtilis.pep
```

```

bms      b.suis.pep
bth      b.thetaiotaomicron.pep
buc      buchnera.pep
cac      c.acetobutylicum.pep
cal      c.albicans.pep
cbu      c.burnetii.pep
cca      c.caviae.pep
ccr      c.crescentus.pep
cef      c.efficiens.pep
cel      c.elegans.pep
cgl      c.glutamicum.pep
cje      c.jejuni.pep
cmu      c.muridarum.pep
cpe      c.perfringens.pep
cpn      c.pneumoniae.pep
cpa      c.pneumoniae_ar39.pep
cpj      c.pneumoniae_jl38.pep
cpt      c.pneumoniae_tw183.pep
cte      c.tepidum.pep
ctc      c.tetani.pep
ctr      c.trachomatis.pep
ddi      d.discoideum.pep
dme      d.melanogaster.pep
dra      d.radiodurans.pep
dre      d.rerio.pep
eco      e.coli.pep
ecc      e.coli_cft073.pep
ecj      e.coli_j.pep
ece      e.coli_ol57.pep
ecs      e.coli_ol57j.pep
ecu      e.cuniculi.pep
efa      e.faecalis.pep
fnu      f.nucleatum.pep
hdu      h.ducreyi.pep
hhe      h.hepaticus.pep
hin      h.influenzae.pep
hpy      h.pylori.pep
hpj      h.pylori_j79.pep
hsa      h.sapiens.pep
hal      halobacterium.pep
lin      l.innocua.pep
lil      l.interrogans.pep
lla      l.lactis.pep
lmo      l.monocytogenes.pep
lpl      l.plantarum.pep
mac      m.acetivorans.pep
mbo      m.bovis.pep
mga      m.gallisepticum.pep
mge      m.genitalium.pep
mja      m.jannaschii.pep
mka      m.kandleri.pep
mle      m.leprae.pep
mlo      m.lotii.pep
mma      m.mazei.pep
mmu      m.musculus.pep
mpe      m.penetrans.pep
mpn      m.pneumoniae.pep
mpu      m.pulmonis.pep
mth      m.thermoautotrophicum.pep
mtu      m.tuberculosis.pep
mtc      m.tuberculosis_cdc1551.pep
neu      n.europaea.pep
nme      n.meningitidis.pep
nma      n.meningitidis_a.pep
oih      o.ihyensis.pep
osa      o.sativa.pep

pab
pai
pae
pfa
pfu
pho
pma
pmu
ppu
pst
rba
rco
rno
rpr
rso
sag
san
sau
sav
sam
sau
sma
sce
sco
sep
sfl
sfx
sme
smu
son
spn
spr
spo
spy
spm
spg
sps
sso
sto
sty
stt
stm
syn
tac
tbr
tel
tma
tpa
tte
tvo
twh
tws
uur
vch
vpa
vvu
vrl
wbr
xac
xcc
xfa
xft
ype
ypk
zma
);

p.abysssi.pep
p.aerophilum.pep
p.aeruginosa.pep
p.falciparum.pep
p.furiosus.pep
p.horikoshii.pep
p.marinus.pep
p.multocida.pep
p.putida.pep
p.syringae.pep
r.baltica.pep
r.conorii.pep
r.norvegicus.pep
r.prowazekii.pep
r.solanacearum.pep
s.agalactiae.pep
s.agalactiae_nem316.pep
s.aureus.pep
s.aureus_mu50.pep
s.aureus_mu2.pep
s.aureus_n315.pep
s.avermitilis.pep
s.cerevisiae.pep
s.coelicolor.pep
s.epidermidis.pep
s.flexneri.pep
s.flexneri_2457t.pep
s.meliloti.pep
s.mutans.pep
s.oneidensis.pep
s.pneumoniae.pep
s.pneumoniae_r6.pep
s.pombe.pep
s.pyogenes.pep
s.pyogenes_m18.pep
s.pyogenes_m3.pep
s.pyogenes_ss11.pep
s.solfataricus.pep
s.tokodaii.pep
s.typhi.pep
s.typhi_ty2.pep
s.typhimurium.pep
synechocystis.pep
t.acidophilum.pep
t.brucei.pep
t.elongatus.pep
t.maritima.pep
t.pallidum.pep
t.tengcongensis.pep
t.volcanium.pep
t.whipplei.pep
t.whipplei_s.pep
u.urealyticum.pep
v.cholerae.pep
v.parahaemolyticus.pep
v.vulnificus.pep
viruses.pep
w.brevipalpis.pep
x.axonopodis.pep
x.campestris.pep
x.fastidiosa.pep
x.fastidiosa_t.pep
y.pestis.pep
y.pestis_kim.pep
z.mays.pep
);

```

```

%familias = qw(
UVRx      q
UVRy      q
UVRz      q
);

```

```
# 0 - ALINEAR SECUENCIAS Y CREAR PROFILE
```

```
FAM :foreach $fam (keys (%familias)) {
```

```
    &hmmbuild;
```

```
    #construir el profile de cada familia
```

```

chdir ("${fam}") || die "cannot open $fam directory";
system ("cp ../${fam}.aln ./");

GEN:foreach $key (keys (%genomas)) {
    $genoma = $genomas[$key];
    print ("\n Buscando en el genoma de ".$genomas[$key]."\n");
    FOR:for ($count = 1; $count != 0; $count++) {#iterar b queda hasta count = 0.
        print ("\nIteraci n $count\n");

# 1 - B SQUEDA DEL PROFILE EN EL GENOMA.

        &hmmsearch;

# 2 - CORTE SUPERIOR E INFERIOR
        open (OUT, "${key}$count.out") || die "can't open hmm file\n";
        open (RAWLST, ">${key}$count.raw") || die "can't open raw file\n";
        &corte;
        close (OUT);
        close (RAWLST);

# 3 - B SQUEDA DE DOMINIOS DUPLICADOS (DUPLICACIONES INTERNAS)
        open (RAWLST, "${key}$count.raw") || die "can't open raw file\n";
        &duplic;
        close (RAWLST);

# 4 - IDENTIFICAR REPETICIONES Y COLETAR POR ENCIMA DEL E-value
        open (LISTA, ">${key}$count.lst") || die "can't create san pedro";
        open (RAWLST, "${key}$count.raw") || die "can't read raw file\n";
        if ($count == 1) {
            &primerait;
            &saving;
            &finito
        }

        if ($count > 1) {
            &itsubsec;
            &saving;
            &finito
        }

        close (RAWLST);

# 5 - OBTENCION DE NUEVAS SECUENCIAS
        open (FASTA, ">${key}$count.fas") || die "can't open fasta";
        open (LOCAL, "${key}$count.loc") || die "can't open lista";
        &getseq;
        close (LOCAL);
        close (FASTA);

# 6 - CORTAR DOMINIOS EXCLUSIVAMENTE
        open (OUT, "${key}$count.out") || die "can't read outfile";
        open (MIENTRAS, ">${key}$count.mie") || die "can't create mientras";

        $r = 0;
        while (<OUT>) {
            if (/Histogram/) {
                $r=0;
            }
            if ($r==1) {
                print MIENTRAS "$_";
            }

            if (/Domain/ && /Sequence/) {
                $r =1;
            }
        }
        close (OUT);
        close (MIENTRAS);

        open (MIENTRAS, "${key}$count.mie") || die "can't read mientras file";
        open (PARSED, ">${key}$count.par") || die "can't create parsed";

        while (<MIENTRAS>) {
            if (/Alignments/) {
                last
            }
        }
    }
}

```

```

    }
    if (/#key/) {
        print PARSED $_;
    }
}
close (MIENTRAS);
close (PARSED);

#system ("rm *.mie");

open (PARSED, "#key$count.par") || die "can't read parsed";
open (FASTA, "#key$count.fas") || (warn "can't read fasta" && last);
open (DOMAIN, ">#key$count.dom") || die "can't create domain";

@parsed = <PARSED>;
@fasta = <FASTA>;
@fastaID = grep (/>/, @fasta);

foreach $desc (@fastaID) {
    $desc =~ s/>//;
    @desc = split (/s+/, $desc);
    @grepars = grep (/#desc[0]/, @parsed);

    foreach $grepars (@grepars) {
        @pars = split (/s+/, $grepars);
        $domID = $pars[1];
        $from = $pars[2];
        $to = $pars[3];

        $r = 0;
        foreach (@fasta) {
            $fasline = $_;
            $fasline =~ s/\s//g;
            chomp $fasline;

            if (/>/) {
                $r=0;
            }
            if ($r == 1) {
                push (@secuencia, $fasline);
            }
            if (/#desc[0]/) {
                $r = 1;
            }
        }

        $secuencia = join ("", @secuencia);
        $domain = substr ($secuencia, $from, $to);
        @domain = $domain;
        chomp @domain;
        $domain = join ("", @domain);
        $clave = $desc[0];
        chomp $clave;
        substr ($clave,0,4) = "";
        print DOMAIN ">$clave\-$domID\-$desc\n\n $domain\n\n";

        @secuencia = undef;
        @pars = undef;
        @domain = undef;
    }
    @grepars = undef;
    @desc = undef;
}
close (FASTA);
close (PARSED);
close (DOMAIN);

# 7 - ALINEACION, CONSTRUCCION Y CALIBRACION DEL NUEVO PROFILE
&newprofile;

# 8 - BORRAR ARCHIVOS DE PAJA.

# system ("rm *.hmm");
# system ("rm *.raw");
# system ("rm *.lst");

# 9 - ANALISIS DE DUPLICACIONES INTERNAS

```

```

# &duplisis;
# LD - FIN

}
}
chdir ("../");
}

sub hmmbuild {
  my ($fam) = $fam;
  print ("Analizando familia $fam\n");
  system ("$dirhmm/hmmbuild $fam.hmm $fam.aln");
  system ("hmmcalibrate $fam.hmm");
  #hits locales a la secuencia,
  #globales al profile.

  mkdir ("$fam",0770);
#  chmod (0777,"$fam");
}

sub hmmsearch{
  my ($fam) = $fam;

  if ($count == 1) {
    system ("cp ../$fam.hmm ./");
    system ("$dirhmm/hmmsearch $fam.hmm $dirgenoma/$genoma > $key$count.out");
    print ("\nB queda de $key$count.hmm en $genoma terminada.\n\n");
  } else {
    chdir ("$dirlocal/$fam");
    print ("$dirhmm/hmmsearch -AD $key$count.hmm $dirgenoma/$genoma > $key$count.out \n");
    system ("$dirhmm/hmmsearch $key$count.hmm $dirgenoma/$genoma > $key$count.out" );
    print ("\n B queda de $key$count.hmm en $genoma terminada.\n\n");
  }
}

sub corte {
  my ($linea, $r);
  $r = 0;
  while (<OUT>) {
    $linea = $_;

    if (($r == 1 && /\s+$/ || (/E threshold/)) {
      # ltima linea en
      blanco antes de Parsed for Domains.
      last;
    }

    if ($r == 1) {
      print RAWLST $linea;
    } elsif (/ - - - \n /) {
      $r = 1;
      #si encuentra --- graba
      #la secuencia.
    }
  }
}

sub duplic {
  my ($key) = $key;

  foreach (<RAWLST>) {
    $string = $_;
    @actual = split;
    @reverse = reverse (@actual);

    if (/threshold/) {
      last
    }

    if ($reverse[0] > 1) {
      #busca duplicaciones internas.
      open (REPETAB, ">>$key.rep") || die "can't open repetabs";
      print REPETAB $string;
      close (REPETAB);
      next;
      #next foreach
    }
  }
}

sub finito {
  if (-z "$key$count.lst") {
    $count = -1;
    print ("no mas secuencias!\n");
    last FOR;
  }
}

```

```

}
sub saving {
  open (LISTA, "$key$count.lst") || die "can't read spd";
  open (FLIST, ">$key.flst") || die "can't append to finalst";
  open (LOCAL, ">$key$count.loc") || die "can't create local";

  while (<LISTA>) {
    print FLIST $_;
    print LOCAL $_;
  }

  close (LISTA);
  close (FLIST);
  close (LOCAL);
}

sub primerait {
  my ($anter, $string, $testest);

  print ("entrando a primerait\n");

  foreach $string (<RAWLST>) {
    @actual = split (/s+/, $string);
    @reverse = reverse (@actual);

    print ("AL inicio reverse es ".@reverse[1]."\n");
    if (($reverse[0] eq "thresholds") || ($reverse[1] >= 1e-50)) {
      print ("las secuencias no dieron el ancho:".@reverse[1].".y".@reverse[1]."\n\n");
      $anter = undef;
      last
    }

    @t = split (/-/,@reverse[1]);

    if ($anter == undef || $anter == 0) {
      print ("anter es undef o cero\n imprimiendo a lista\n");
      print LISTA $string;
      $anter = $t[1];
      next
    }

    if ($anter != 0) {
      print ("\nanter no es cero : $anter \ntes es: $tes\n");
      $tes = $t[1];
      $testest = ($anter / $tes);
      print ("testest es : $testest\n");
    }

    if (($reverse[1] <= 1e-60) && ($testest <= $eval)) { #que no se distancie en
      m s de e rdenes de magnitud del anterior
      print ("reverse es ".@reverse[1].". y testest es $testest \n
      las secuencias si dieron el ancho\n");
      print LISTA $string;
      $anter = $tes;
    }
  }
}

sub itsubsec {
  my ($anter, $string, $anterior, $testest, @actual, @reverse, @test);

  foreach $string (<RAWLST>) {

    if ($string == /threshold/) {
      last
    }

    @actual = split (/s+/, $string);
    $actual = $actual[0];
    @reverse = reverse (@actual);
    @t = split (/-/,@reverse[1]);

    # para identificar repetidas...

    open (FLIST, "$key.flst") || die "can't create .flst";
    while (<FLIST>) {
      @anterior = $_;

      if ($anterior[0] == /$actual/ ) {

```

```

        print ("\n\nxxxxxx  ya la tengo! : $anterior con $actual x\n\n");
        push (@test, 1);
    }

    if ($reverse[0] > 1) {
        open (REPETAB, ">>$key.rep") || die "can't open repetabs";
        print REPETAB $string;
        close (REPETAB);
        push (@test, 1);
    } else {
        push (@test, "0");
    }

}
@actual = (); @anterior = ();
$joinif = join (" ", @test);

if ($joinif =~ /1/) {
    @test = undef;
    $anter = $t[1];
    @t = undef;
    next;
}

# primerait de nuevo...

print ("\n\nEntrando en Itsubseq aL inicio reverse es ".$reverse[1]."\n");

if (($reverse[0] eq "thresholds") || ($reverse[1] >= 1e-6)) {
    print ("las secuencias no dieron el ancho:".$reverse[1].".y".$reverse[1]."\n\n");
    $anter = undef;
    last
}

if ($anter == undef || $anter == 0) {
    print ("anter es undef o cero\n imprimiendo a lista\n");
    print LISTA $string;
    $anter = $t[1];
    next
}

if ($anter != 0) {
    print ("\n\nanter no es cero : $anter \ntes es: $tes\n");
    $tes = $t[1];
    $testest = ($anter / $tes);
    print ("testest es : $testest\n");
}

if (($reverse[1] <= 1e-6) && ($testest <= $eval)) { #que no se distancie en
m s de e rdenes de magnitud del anterior
    print ("reverse es ".$reverse[1]." y testest es $testest \n
        las secuencias si dieron el ancho\n");
    print LISTA $string;
    $anter = $tes;
}
@reverse = ();
@test = undef;
}

sub getseq {
my ($line, $clave, $r, $protID);
my (%genomas) = %genomas;

while (<LOCAL>) {
    $line = $_;
    if (/thresholds/) {
        next GEN;
    }

    $clave = (substr ($line, 0, 3));

    open (GENOMA, "$dirgenoma/".$genomas{$clave}) || (print "cant open genoma en 613\n" && next
GEN);

```

```

    $r=0;
    @line = split (/s+/, $line);
    $protID = $line[0];
    while (<GENOMA>) {
        $fasta = $_;
        if (/>/) {
            $r=0;
        }
        if ($r==1) {
            print FASTA "$fasta";
        }
        if ((/$protID/) && ($_ ne \n)) {
            print FASTA ("$fasta");
            $r = 1;
        }
    }
    $fasta = undef;
    close (GENOMA);
}

sub newprofile {
    if ($count == 1) {
        system ("$dirhmm/hmmalign -q $fam.hmm $key$count.dom > $key$count.new")
            && print ("Nueva alineaci n en $key$count.new\n");
    }
    # system ("clustalw -profile1=$fam.aln -profile2=$key$count.dom -sequences -outfile=$key$count.new")
    # && print ("Nueva alineaci n en $key$count.new\n");
    }

    if ($count > 1) {
        $prev = ($count - 1);
        system ("$dirhmm/hmmalign -q $key$count.hmm $key$count.dom > $key$count.new")
            && print ("Nueva alineaci n en $key$count.new\n");
    }
    # system ("clustalw -profile1=$key$prev.new -profile2=$key$count.dom -sequences -
    outfile=$key$count.new")
    # && print ("Nueva alineaci n en $key$count.new\n");
    }

    open (NEW, "$key$count.new");
    open (NEWER, "> $key$count.nwr");

    foreach (<NEW>) {
        $nuez = $_;
        if (/^..RF/) {
            print NEWER ("");
        }
        elsif (/^..GC/) {
            print NEWER ("");
        }
    }

else {
    print NEWER $nuez;
}

close NEW;
close NEWER;
system ("more $key$count.nwr > $key$count.new");

$nnext = ($count + 1);

system ("$dirhmm/hmmbuild $key$nnext.hmm $key$count.new")
    && print ("$key$nnext.hmm est construido\n");
system ("hmmcalibrate $key$nnext.hmm")
    && print ("$key$nnext.hmm est calibrado");

chmod (0777, "$key$nnext.hmm");
}

sub duplisis {
    foreach $fam (keys (%familias)) {
        chdir ("$fam") || die "cannot open $fam directory";

        foreach $key (keys (%genomas)) {
            open (REPLST, "$key.rep") || warn "can't read $key.rep\n";
        }
    }
}

```

#hits locales a la secuencia  
#y locales al profile  
#porque las nuevas est n completas.



## LITERATURA CITADA

1. Achenbach-Richter, L., Gupta, R., Stetter, K.O. y Woese, C.R. **1987**. Were the original eubacteria thermophiles? *Syst Appl Microbiol* **9**:34-39
2. Agostini, H.J., Carroll, J.D. y Minton K.W. **1996**. Identification and characterization of *uvrA*, a DNA repair gene of *Deinococcus radiodurans*. *J Bacteriol* **178**:6759-6765.
3. Alberti, M., Li, Y.F., Sancar A. y Hearst J.E. Presentado en Agosto 1992 a las bases de datos EMBL/GenBank/DDBJ.
4. Alberts, I.L., Nadassy, K. y Wodak, S.J. **1998**. Analysis of zinc binding sites in protein crystal structures. *Protein Sci* **7**:1700-1716
5. Ames, G.F., Mimura, C. y Shyamala, V. **1990**. Bacterial periplasmic permeases belong to a family of transport proteins operating from *Escherichia coli*: relationship to eukaryotic multidrug resistance. *Gene* **136**:231-236
6. Ames, G.F. L y Lecar, H. **1992**. ATP-dependent bacterial transporters and cystic fibrosis: analog between channels and transporters. *FASEB J* **6**:2660-2666
7. Andersson, J.O. y Andersson, S.G. **1999**. Insights into the evolutionary process of genome degradation. *Curr Opin Genet Dev* **9**:664-671
8. Aravind, L., Tatusov, R.L., Wolf, Y.I., Walker, D.R. y Koonin, E.V. **1998**. Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet* **14**:442-444
9. Baldauf, S.L., Palmer, J.D. y Doolittle, W.F. **1996**. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc Natl Acad Sci USA* **93**:7749-7754
10. Battista J.R. **1997** Against all odds: the survival strategies of *Deinococcus radiodurans*. *Annu Rev Microbiol.* **51**:203-24
11. Batty, D.P. y Wood, R.D. **2000**. Damage recognition in nucleotide excision repair of DNA. *Gene* **241**:193-204
12. Berg, B.M., **1986**. Potencial metal binding domains in nucleic acid binding proteins. *Science* **232**:485-487
13. Black, C.G., Fyfe, J. A. y Davies, J. K. **1997**. Cloning, nucleotide sequence and transcriptional analysis of the *uvrA* gene from *Neisseria gonorrhoeae*. *Molec Gen Genet*, **254**(5):479-485
14. Boeckman, B., Bairoch, A., Apweiler, R., Blatter, M., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Pham, I., Pilbout, S. y Schneider, M. **2003**. The Swiss-Prot protein knowledgebase and its supplement TrEMBL in 2003. *Nuc Acid Res* **31**:365-370

15. Boling, M.E. y Setlow, J.K. **1966**. The resistance of *Micrococcus radiodurans* to ultraviolet radiation. 3. A repair mechanism. *Biochim Biophys Acta*. **123**:26-33
16. Bouige, P., Laurent, D., Piloyan, L. y Dassa, E. **2002**. Phylogenetic and functional classification of ATP-binding cassette (ABC) systems. *Curr Protein Pept Sci* **3**(5):541-59
17. Bowman, K.K., Sidik, K., Smith, C.A., Taylor, J.S., Doetsch, P.W. y Freyer, G.A. **1994**. A new ATP-independent DNA endonuclease from *Schizosaccharomyces pombe* that recognizes cyclobutane pyrimidine dimers and 6-4 photoproducts. *Nuc Ac Res***22**:3026-3032
18. Brochier, C. y Philippe, H. **2002** A non-hyperthermophilic ancestor for Bacteria. *Nature* **417**:24
19. Brochier, C., Baptiste, E., Moreira, D. y Philippe, H. **2002**. Eubacterial phylogeny based on translational apparatus proteins. *Trends Genet* **18**:1-5.
20. Brown, J.R., Douady, C.J., Italia, M.J., Marshall, W.E. y Stanhope, M.J. **2001**. Universal trees based on large combined protein sequence data sets. *Nat Genet* **28**(3):281-285
21. Brown, W. J. y Rockey, D. **2000**. Identification of an antigen localized to an apparent septum within dividing *Chlamydiae*. *Infect Immun* **68**:708-715
22. Cambillau, C. y Claverie, J.M. **2000**. Structural and genomic correlates of hyperthermostability. *J Biol Chem. Genome Res* **275**:32383-32386
23. Campbell, L. A., and R. E. Yasbin; 1979. Deoxyribonucleic acid repair capacities of *Neisseria gonorrhoeae*: absence of photoreactivation. **J Bacteriol.** **140**:1109–1111
24. Campbell, L. A., and R. E. Yasbin; 1984. Mutagenesis of *Neisseria gonorrhoeae*: Absence of error-prone repair **J Bacteriol.** **160**:288–293
25. Cerdeño-Tárraga, A.M., Efsftratiou, A., Dover, L.G., Holden, M.T., Pallen, M., Bentley, S.D., Besra, G.S., Churcher, C., James, K.D., De Zoysa, A., Chillingworth, T., Rabinowitsch, E., Rutherford, K.M., Thomson, N.R., Unwin, L., Whitehead, S., Barrell, B.G. y Parkhill, J. **2003**. The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nuc Acid Res* **31**(22):6516-6523
26. Chain, P., Lamerdin, J., Larimer, F., Regala, W., Lao, V., Land, M., Hauser, L., Hooper, A., Klotz, M., Norton, J., Sayavedra-Soto, L., Arciero, D., Hommes, N., Whittaker, M. y Arp, Daniel. **2003**. Complete genome sequence of the ammonia-oxidizing bacterium and obligate Chemolithoautotroph *Nitrosomonas europaea*. *J Bacteriol* **185**(9):2759-2773
27. Chang, G. y Roth, C.B. **2001**. Structure of MsbA from *E. coli*: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* **293**(5536):1793-800
28. Charles H. e Ishikawa H., **1999**. Physical and genetic map of the genome of *Buchnera*, the primary endosymbiont of the pea aphid *Acyrtosiphon pisum*. *J Mol Evol* **48**(2):142-50
29. Cobbe, N. y Heck, M. **2000**. Review: SMCs in the world of chromosome biology – from prokaryotes to higher eukaryotes. *J Struct Biol* **129**(2-3):123-143

30. Connelly, J., Kirkham, L. y Leach, D. **1998**. The SbcCD nuclease of *E. coli* is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. *Proc Natl Acad Sci* **95**:7969-7974
31. Corsaro, D., Valassina, M. y Venditti, D. **2003**. Increasing diversity within Chlamyidae. *Crit Rev in Microbiol* **29**(1):37-78
32. Cuff, J.A. y Barton, G.J. **2000**. Application for enhanced multiple sequence alignment profiles to improve protein secondary structure prediction. *PROTEINS: Structure, Function and Genetics* **40**:502-511
33. Cummings, C.A., Brinig, M.M., Lepp, P.W., van de Pas, S., Relman, D.A. **2004**. *Bordetella* species are distinguished by patterns of substantial gene loss and host adaptation. *J Bacteriol* **186**(5):1484-1492
34. Dassa, E. y Bouige, P. **2001**. The ABC of ABC's: a phylogenetic and functional classification of ABC systems in living organisms. *Res Microbiol* **152**(3-4):211-229
35. Daubin, V. y Perrière, G. **2003**. G+C3 structuring along the genome : a common feature in Prokaryotes. *Mol Biol Evol* **20**(4):471-483
36. de Almeida, R., Beltramini, P., Aparecida, L., Proença-Módena, J.L., Seixas, E., Bueno, G. y Brocchi, M. **2004**. Bacteriophages and insertion sequences of *Chromobacterium violaceum* ATCC 12472. *Gen Mol Res* **3**(1):76-84
37. Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A., Graham, D.E., Overbeek, R., Snead, M.A., Heller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olsen, G.J. y Swanson, R.V. **1998**. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**:353-358
38. DeLano, W. **1998-2003**. The PyMOL Molecular Graphics System. DeLano Scientific LLC, San Carlos, California, USA. <http://www.pymol.org>
39. Delaye, L., Vazquez, H. y Lazcano, A. in "First step in the origin of Life in the Universe" ed. Chela-Flores J. pp. 223-230. (Kluwer Academic, Amsterdam, 2001)
40. Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, Martinez-Arias R, Henne A, Wiezer A, Baumer S, Jacobi C, Bruggemann H, Lienard T, Christmann A, Bomeke M, Steckel S, Bhattacharyya A, Lykidis A, Overbeek R, Klenk H.P, Gunsalus R.P, Fritz H.J, Gottschalk G. **2002**. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *J Mol Microbiol Biotechnol* **4**(4):453-61.
41. Doetsch, P.W. **1995**. What's old is new: an alternative DNA excision repair pathway. *Trends in Bioschem Sci* **20**:384-386
42. Doolittle, R.F., Johnson, M.S., Husain, I., Van Houten, B., Thomas, D.C. y Sancar, A. **1986**. Domainal evolution of a prokaryotic DNA repair protein and its relationship to active-transport proteins. *Nature* **323**:451-52

43. Doolittle, W.F. y Brown J.R. **1994**. Tempo, mode, the progenote and the universal root. *Proc Natl Acad Sci USA* **91**:6721-6728
44. Eddy, S.R. **1998**. Profile Hidden Markov Models. *Bioinformatics*. **14**:755-763
45. Eisen, J., Nelson, K., Paulsen, I., Heidelberg, J., Wu, M., Dodson, R., Deboy, R., Gwinn, M., Nelson, W., Haft, D., Hickey, E., Peterson, J., Durkin, S., Kolonay, J., Yang, F., Holt, I., Umayam, L., Mason, T., Brenner, M., Shea, T., Parksey, D., Niermann, W., Feldblyum, T., Hansen, C., Craven, B., Radune, D., Vamathevan, J., Khouri, H., White, O., Gruber, T., Ketchum, K., Venter, C., Tettelin, H., Bryant, D. y Fraser, C. **2002**. The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. *Proc Natl Acad Sci* **99**(14):9509-9514
46. Felsenstein, J. **1978**. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* **27**:401-410
47. Felsenstein, J. **1985**. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**:783-791
48. Fernández, E., Lombo, F., Méndez, C. y Salas, J.A. **1996**. An ABC transporter is essential for resistance to the antitumor agent mithramycin in the producer *Streptomyces argillaceus*. *Mol Gen Genet* **251**(6):692-698
49. Forterre, P. **2002**. A hot story from comparative genomics: Reverse gyrase as the only protein specific for hyperthermophiles. *Trends Genet* **18**:236-238
50. Forterre, P., Bouthier de la Tour, C., Philippe, H. y Duget, M. **2000**. Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from archaea to bacteria. *Trends Genet* **16**:152-154
51. Fraser C.M, Casjens S, Huang W.M, Sutton G.G, Clayton R, Lathigra R, White O, Ketchum K.A, Dodson R, Hickey E.K, Gwinn M, Dougherty B, Tomb J.F, Fleischmann R.D, Richardson D, Peterson J, Kerlavage A.R, Quackenbush J, Salzberg S, Hanson M, van Vugt R, Palmer N, Adams M.D, Gocayne J, Venter J.C, et al. **1997**. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**(6660):580-586.
52. Fraser C.M, Norris S.J, Weinstock G.M, White O, Sutton G.G, Dodson R, Gwinn M, Hickey E.K, Clayton R, Ketchum K.A, Sodergren E, Hardham J.M, McLeod M.P, Salzberg S, Peterson J, Khalak H, Richardson D, Howell J.K, Chidambaram M, Utterback T, McDonald L, Artiach P, Bowman C, Cotton M.D, Venter J.C, et al. **1998**. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281**(5375):375-88.
53. Fuerst, J.A. **1995**. The planctomycetes: Emerging models for microbial ecology, evolution and cell biology. *Microbiology* **141**:1493-1506
54. Fuerst, J.A. y Webb, R.I. **1991**. Membrane-bounded nucleoid in the eubacterium *Gemmata obscuriglobus*. *Proc Natl Acad Sci USA* **88**:8184-8188

55. Futterer O, Angelov A, Liesegang H, Gottschalk G, Schleper C, Schepers B, Dock C, Antranikian G, Liebl W. Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. **2004**. *Proc Natl Acad Sci U S A*. **101**(24):9091-6.
56. Gabriel, W., Lynch, M. y Burger, R. **1993**. Muller's ratchet and mutational meltdowns. *Evolution* **47**:1744-1757
57. Galagan J.E, Nusbaum C, Roy A, Endrizzi M.G, Macdonald P, FitzHugh W, Calvo S, Engels R, Smirnov S, Atnoor D, Brown A, Allen N, Naylor J, Stange-Thomann N, DeArellano K, Johnson R, Linton L, McEwan P, McKernan K, Talamas J, Tirrell A, Ye W, Zimmer A, Barber R.D, Cann I, Graham D.E, Grahame D.A, Guss A.M, Hedderich R, Ingram-Smith C, Kuettner H.C, Krzycki J.A, Leigh J.A, Li W, Liu J, Mukhopadhyay B, Reeve J.N, Smith K, Springer T.A, Umayam L.A, White O, White R.H, Conway de Macario E, Ferry J.G, Jarrell K.F, Jing H, Macario A.J, Paulsen I, Pritchett M, Sowers K.R, Swanson R.V, Zinder S.H, Lander E, Metcalf W.W, Birren B. **2002**. The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res* **12**(4):532-542.
58. Galtier, N., Tourasse, N y Gouy, M. **1999**. A nonhyperthermophilic common ancestor to extant life forms. *Science* **283**:220-221
59. Gil, R., Sabater-Muñoz, B., Latorre, A., Silva, F.J. y Moya, A. **2002**. Extreme genome reduction in *Buchnera* spp.: Toward the minimal genome needed for symbiotic life. *Proc Natl Acad Sci* **99**(7):4454-4458
60. Glöckner, F.O., Kube, M., Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., Gade, D., Beck, A., Borzym, K., Heitmann, K., Rabus, R., Schlesner, H., Amann, R. y Reinhardt, R. **2003**. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc Natl Acad Sci* **100**(14):8298-8303
61. Goosen, N. y Moolenaar, G. **2001**. Role of ATP hydrolysis by UvrA and UvrB during nucleotide excision repair. *Res Microbiol* **152**:401-409
62. Gorbalenya, A. y Koonin, E. **1990**. Superfamily of UvrA-related NTP-binding Proteins. *J Mol Biol* **213**:583-591
63. Gribaldo, S. y Philippe, H. **2002**. Ancient Phylogenetic Relationships. *Theoretical Population Biology*. **61**:391-408
64. Gripenburg, U., Ward-Rainey, N., Mohamed, S., Schlesner, H., Marxsen, H., Rainey, F., Stackebrandt, E. y Auling, G. **1999**. Phylogenetic diversity, polyamine pattern and DNA base composition of members of the Planctomycetales. *Int J Syst Bacteriol* **49**:689-696
65. Grishin, N.V. **2001**. Treble clef finger-a functionally diverse zinc-binding structural motif. *Nucleic Acids Research* **29**(8):1703-1714
66. Gruber, T.M., Eisen, J.A., Gish, J.A. y Bryant, D.A. **1998**. The phylogenetic relationships of Chlorobium. *FEMS Microbiol Lett* **162**:53-60
67. Guilfoile, P.G. y Hutchinson, C.R. **1991**. A bacterial analog of the *mdr* gene of mammalian tumor cells is present in *Streptomyces peucetius*, the producer of daunorubicin and doxorubicin. *Proc Natl Acad Sci USA* **88**:8553-8557

68. Hall, T.A. **1999**. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**:95-98
69. Hatch, T. P. **1996**. Minireview: disulphide cross-linked envelope proteins: the functional equivalent of peptidoglycan in *Chlamydiae*?. *J Bacteriol* **178**:1-5
70. Hendrickson, E.L., Kaul, R., Zhou, Y., Bovee, D., Chapman, P., Chung, J., Conway de Macario, E., Dodsworth, J., Gillett, W., Graham, D.E., Haydock, A.K., Kang, A., Land, M.L., Levy, R., Lie, T.J., Major, T., Moore, B., Porat, I., Overbeek, R., Palmeiri, A., Rouse, G., Saenphimmachak, C., Soll, D., Whitman, W.B., Larimer, F.W., Olson, M.V. y Leigh, J.A. Complete genome sequence of the mesophilic hydrogenotrophic methanogen *Methanococcus maripaludis*. **2004**. Sin publicar, disponible en GenBank, número de acceso : [NC\\_005791](#).
71. Hendy, M. y Penny, D. **1989**. A framework for the quantitative study of evolutionary trees. *Syst Zool* **38**:297-309
72. Higgins, C.F. **1992**. ABC TRANSPORTERS: From Microorganisms to Man. *Ann Rev Cell Biol* **8**:67-113
73. Higgins, C.F., Gallagher, M.P., Mimmack, M.L., Pearce, S.R. **1988**. A family of closely related ATP-binding subunits from prokaryotic and eukaryotic cells. *BioEssays* **8**:189-91
74. Higgins, C.F., Hiles, I.D., Salmond, G.P.C., Gill, D.R., Downie, J.A., Evans, I.J., Holland, I.B., Gray, L., Buckel, S., Bell, A. y Hermodson, M.A. **1986**. A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature* **323**:448-450
75. Hopfner, K.P., Craig, L., Moncalian, G., Zinkel, R.A., Usui, T., Owen, B., Karcher, A., Henderson, B., Bodmer, J.L., McMurray, C., Carner, J.P., Petrini, J.H. y Tainer, J.A. **2002**. The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. *Nature* **418**:562-566
76. Hopfner, K.P. y Tainer, J.A. **2003**. Rad50/SMC proteins and ABC transporters: unifying concepts from high-resolution structures. *Curr Opin Struct Biol* **13**(2):249-255
77. Husain, I., VanHouten, B., Thomas, D. y Sancar, A. **1986**. Sequences of *Escherichia coli* *uvrA* gene and protein reveal two potential ATP binding sites. *J Biol Chem* **261**(11):4895-4901
78. Ikeda, H., Ishikawa, J., Hanamoto, Akiharu, Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. y Omura, S. **2003**. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nature Biotechnology* **21**:526-531
79. Islas, S., Velasco, A.M., Becerra, A., Delaye, L. y Lazcano, A. **2003**. Hyperthermophily and the origin and earliest evolution of life. *Int Microbiol* **6**:87-94

80. Iwabe, N., Kuma, K.I., Hasegawa, M., Osawa, S y Miyata, T. **1989**. Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci USA* **86**:9355-9359
81. Jenkins, C. y Fuerst, J. **2001**. Phylogenetic analysis of evolutionary relationships of the planctomycete division of the domain bacteria based on amino acid sequences of elongation factor Tu. *J Mol Evol* **52**:405-418
82. Junier, T. y Pagni, M. **2000**. Dotlet: diagonal plots in a web browser. *Bioinformatics* **16**(2):178-179
83. Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W., Olinger, L., Grimwood, J., Davis, R.W. y Stephens, R.S. **1999**. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. *Nature Genet* **21**:385-389
84. Kanehisa, M. **1997**. A database for post-genome analysis. *Trends Genet* **13**:375-376
85. Kanehisa, M. y Goto, S. **2000**. Kyoto Encyclopedia of Genes and Genomes. *Nuc Ac Res* **28**:27-30
86. Kaur, P. **1997**. Expression and Characterization of DrrA and DrrB proteins of *Streptomyces peucetius* in *Escherichia coli*: DrrA is an ATP binding protein. *J Bacteriol* **179**(3):569-575
87. Kaur, P. y Russell, J. **1998**. Biochemical coupling between the DrrA and DrrB proteins of the doxorubicin efflux pump of *Streptomyces peucetius*. *J Biol Chem* **273**(28):17933-17939.
88. Kennedy, S.P., Ng, W.V., Salzberg, S.L., Hood, L. y DasSarma, S. **2001**. Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res* **11**:1641-1650
89. Kenyon, C.J., y Walker, G.C. **1981**. Expression of the *E. coli* *uvrA* gene is inducible. *Nature* **289**:808-810
90. Komaki, K. e Ishikawa, H. **1999**. Intracellular bacterial symbionts of aphids posses many genomic copies per bacterium. *J Mol Evol* **48**:717-722
91. Komaki K. e Ishikawa H. **2000**. Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. *Insect Biochem Molecular Biol* **30**:253-258
92. König, H., Schlesner, H. y Hirsch, P. **1984**. Cell wall studies on budding bacteria of the *Planctomyces/Pasteuria* group and on a *Prosthecomicrobium* sp. *Arch Microbiol* **138**:200-205
93. Koski, L.B., Morton, R.A. y Golding, B. **2001**. Codon bias and base composition are poor indicators of horizontally transferred genes. *Mol Biol Evol* **18**(3):404-412
94. Krishna, S.S., Majumdar, I. y Grishin, N.V. **2003**. Structural classification of zinc fingers. *Nuc Acids Res* **31**(2):532-550
95. Kumar, S., Tamura, K., Jakobsen, I.B. y Nei, M. **2001**. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**(12):1244-1245

96. Langer, D., Hain, J., Thuriaux, P. y Zilling, W. **1995**. Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci USA* **92**:5768-5772.
97. Lawrence, J.G. y Ochman, H. **1998**. Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci USA* **95**:9413-9417
98. Liesack, W., König, H., Schlesner, H. y Hirsch, P. **1986**. Chemical composition of the peptidoglycan-free cell envelope of budding bacteria of the *Pirella/ Planctomyces* group. *Arch Microbiol* **145**:361-366
99. Liesack, W., Söller, R., Stewart, R., Haas, H., Giovannoni, S. y Stackebrandt, E. **1992**. The influence of tachytelically (rapidly) evolving sequences on the topology of phylogenetic trees – Intramily relationships and the phylogenetic position of *Planctomycetaceae* as revealed by comparative analysis of 16S ribosomal RNA sequences. *Syst Appl Microbiol* **15**:357-362
100. Lindsay, M.R., Webb, R.I., Strous, M., Jetten, M., Butler, M., Forde, R. y Fuerst, J. **2001**. Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. *Arch Microbiol* **175**:413-429
101. Locher, K.P., Lee, A.T. y Rees, D.C. **2002**. The *E.coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296**:1091-1098
102. Lomovskaya, N., Hong, S.K., Kim, S.U., Fonstein, L., Furuya, K. y Hutchinson, R.C. **1996**. The *Streptomyces peucetius drrC* gene encodes a UvrA-like protein involved in daunorubicin resistance and production. *J Bacteriol* **178**(11):3238-3245
103. Lopez, P., Forterre, P. y Philippe, H. **1999**. The root of the tree of life in the light of the covarion model. *J Mol Evol* **49**:496-508
104. Markham, B.E., Harper, J.E., Mount, D.W., Sancar, G.B., Sancar, A., Rupp, W.D., Kenyon, C.J. y Walker, G.C. **1984**. Analysis of mRNA synthesis following induction of the *Escherichia coli* SOS system. *J Mol Biol* **178**:237-248
105. Mazur, D.J. y Grossman, L. **1991**. Dimerization of *Escherichia coli* UvrA and its binding to undamaged and ultraviolet-light damaged DNA. *Biochem* **30**:4432-4443
106. Matthews, J.M. y Sunde, M. **2002**. Zinc Fingers- Folds for Many Occasions. *IUBMB Life* **54**:351-355
107. Mehr, I. J. y Seifert H. S. **1998**. Differential roles of homologous recombination pathways in *Neisseria gonorrhoeae* pilin antigenic variation, DNA transformation and DNA repair. *Molec Microbiol* **30**(4):697-710
108. Michael, S.F., Kilfoil, V.J., Schmidt, M.H., Amann, B.T. y Berg, J.M. **1992**. Metal-binding and folding properties of a minimalist Cys2His2 zinc finger peptide. *Proc Natl Acad Sci USA* **89**:4796-4800
109. Miller, J., McLachlan, A.D. y Klug, A. **1985**. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus oocytes*. *EMBO J* **4**:1609-1614

110. Miller, S.L. y Lazcano, A. **1995**. The origin of life-did it occur at high temperatures? *J Mol Evol* **41**:689-692
111. Moolenaar, G.F. Moorman, C. y Goosen, N. **2000a**. Role of the *Escherichia coli* nucleotide excision repair proteins in DNA replication. *J Bacteriol* **182**(20):5706-5714
112. Moolenaar, G.F., Peña H., M.F., Monacos, V., van der Marel, G.A., van Boom, J.H., Visse, R. y Goosen, N. **2000b**. The role of ATP binding and hydrolysis by UvrB during nucleotide excision repair. *J Biol Chem* **275**(11):8044-8050
113. Moran, N.A. **1996**. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* **93**:2873-2878
114. Moran, N.A. y Mira, A. **2001**. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biology* **2**(12):research 0054.1-0054.12
115. Mourez, M., Hofnung, M. y Dassa, E. **1997**. Subunit interaction in ABC transporters. A conserved sequence in hydrophobic membrane proteins of periplasmic permeases define sites of interaction with the helical domain of ABC subunits. *EMBO J* **16**:3066-3077
116. Myles, G.M. y Sancar, A. **1991**. Isolation and Characterization of Functional Domains of UvrA. *Biochemistry* **30**:3834-3840
117. Myles, G.M., Hearst, J.E. Y Sancar, A. **1991**. Site-specific mutagenesis of conserved residues within Walker A and B sequences of *Escherichia coli* UvrA protein. *Biochemistry* **30**:3824-3834
118. Navaratnam, S., Myles, G., Strange, R. y Sancar, A. **1989**. Evidence from extended X-ray absorption fine structure and site-specific mutagenesis for zinc fingers in UvrA protein of *Escherichia coli*. *J Biol Chem* **264**(27):16067-16071
119. Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardosn, D., Heidelberg, J., Sutton, G.G., Fleischman, R.D., Eisen, J.A., Fraser, C.M. **1999**. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**:323-329
120. Ng, W.V., Kennedy, S., Mahairas, G., Berquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N.S., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A., Leithauser, B., Keller, K., Cruz, R., Danson, M., Hough, D., Maddocks, D., Jablonski, P., Krebs, M., Angeviner, C., Daler, H., Isenbargen, T., Peck, R., Pohlschroder, M., Spudich, J., Jung, K., Alam, M., Freitas, T., Hou, S., Daniels, C., Dennis, P., Omen, A., Ebhardt, H., Lowe, T., Liang, P., Riley, M., Hood, L., DasSarma, S. Genome sequence of *Halobacterium* species NRC-1. *Proc Natl Acad Sci* **97**(22):12176-12181
121. Notredame, C., Higgins, D. y Heringa, J. **2001**. T-Coffee: A novel method for multiple sequence alignments. *J Mol Biol* **302**:205-217

122. Ochman, H., Lawrence, J.G. y Groisman, E.A. **2000**. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**:299-304
123. Oeda K., Horiuchi T. y Sekiguchi M. **1982**. The uvr D gene of *E. coli* encodes a DNA-dependent ATPase. *Nature* **298**:98
124. Oh EY, Claassen L, Thiagalingam S, Mazur S, Grossman L. **1989** ATPase activity of the UvrA and UvrAB protein complexes of the *Escherichia coli* UvrABC endonuclease. *Nucleic Acids Res.* **17**(11):4145-59.
125. Ohta T. **1987**. Very slightly deleterious mutations and the molecular clock.. *J Mol Evol* **26**(1-2):1-6.
126. Olsen GJ, Woese CR. **1996**. Lessons from an Archaeal genome: what are we learning from *Methanococcus jannaschii*? *Trends Genet* **12**(10):377-9.
127. Ortutay, C., Gáspári, Z., Tóth, G., Jáger, E., Vida, G., Orosz, L. y Vellai, T. **2003**. Speciation in *Chlamydia*: Genomewide phylogenetic analyses identified a reliable set of acquired genes. *J Mol Evol* **57**:672-680
128. Pace, N.R., Olsen, G.J., Woese, C.R. **1986**. Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* **45**:325-326
129. Palmu, K., Hautala, A., Kunnari, T., Hakala, J., Mantsala, P. e Ylihonko, K. Identification and functional analysis of a glycosyltransferase gene (snogZ) and a resistance gene (snorO) from the nogalamycin producer, *Streptomyces nogalater*. GenBank, sin publicar.
130. Parkhill, J., Sebaihia, M., Preston, A., Murphy, L.D., Thomson, N., Harris, D.E., Holden, M., Churcher, C.M., Bentley, S.D., Mungall, K.L., Cerdeño-Tárraga, A.M., Temple, L., James, K., Harris, B., Quail, M.A., Achtman, M., Atkin, R., Baker, S., Basham, D., Bason, N., Cherevach, I., Chillingworth, T., Collins, M., Cronin, A., Davis, P., Doggett, J., Feltwell, T., Goble, A., Hamlin, N., Hauser, H., Holroyd, S., Jagels, K., Leather, S., Moule, S., Norberczak, H., O'Neil, S., Ormond, D., Price, C., Rabinowitsch, E., Rutter, S., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Unwin, L., Whitehead, S., Barrell, B.G. y Maskell, D.J. **2003**. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Gene* **35**:32-40
131. Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J., Fraser, C.M. **2000**. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nuc Ac Res* **28**:1397-1406
132. Rönner, S., Liesack, w., Wolters, J y Stackebrandt, E. **1991**. Cloning and sequencing of a large fragment of the ATPD-gene of *Pirellula marine* – A contribution to the phylogeny of *Planctomycetales*. *Endocyto Cell Res* **7**:219-229
133. Saitou, N. y Nei, M. **1987**. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**:406-425

134. Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L., Chandler, M., Choisne, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Siguier, P., Thébault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J. y Boucher, C.A. **2002**. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* **415**:497-502
135. Sancar A, Wharton RP, Seltzer S, Kacinski BM, Clarke ND, Rupp WD. **1981**. Identification of the *uvrA* gene product. *J Mol Biol* **148**(1):45-62.
136. Sancar, A., Sancar, G.B., Little, J.W., Mount, D.W. y Rupp, W. **1982**. LexA protein inhibits transcription of the *E. coli* *uvrA* gene in vitro. *Nature* **298**(5869):96-98.
137. Saurin, W., Hofnung, M. y Dassa, E. **1999**. Getting in or out : early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J Mol Evol* **48**:22-41
138. Schleifer, K.H. y Ludwig, W. **1989**. Phylogenetic relationships among bacteria. En: *Fernholm, B., Bremer, K., Jörnvall, H. (eds) The Hierarchy of Life*. Elsevier Science, Amsterdam, pp 103-116
139. Schlesner, H. **1994**. The development of media suitable for the microorganisms morphologically resembling *Planctomyces* spp., *Pirellula* spp. and other *Planctomycetales* from various aquatic habitats using dilute media. *System Appl Microbiol* **17**:135-145
140. Schmitt, L. y Tampé, R. **2002**. Structure and mechanism of ABC transporters. *Curr Opin Struct Biol* **12**:754-760.
141. Schneider, E. y Hunke, S. **1998**. ATP-binding cassette (ABC) transport systems: Functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiol Reviews* **22**:1-20
142. Schouten, S., Hopmans, E.C., Pancost, R.D. y Damste, J.S. **2000**. From the cover: widespread occurrence of structurally diverse tetraether membrane lipids: Evidence for the ubiquitous presence of low-temperature relatives of hyperthermophiles. *Proc Natl Acad Sci USA* **97**:14421-14426
143. Seeberg, E., y Steinum, A. **1982**. Purification and properties of the *uvr A* protein from *Escherichia coli*. *Proc Natl Acad Sci USA* **79**:988-992
144. Servant, F., Bru, C., Carrère, S., Courcelle, E., Gouzy, J., Peyruc, D. y Kahn, D. **2002**. ProDom : Automated clustering of homologous domains. *Briefings in Bioinformatics* **3**(3) :246-251
145. Seshadri R, Myers G.S, Tettelin H, Eisen J.A, Heidelberg J.F, Dodson RJ, Davidsen T.M, DeBoy R.T, Fouts D.E, Haft D.H, Selengut J, Ren Q, Brinkac L.M, Madupu R, Kolonay J, Durkin S.A, Daugherty S.C, Shetty J, Shvartsbeyn A, Gebregeorgis E, Geer K, Tsegaye G, Malek J, Ayodeji B, Shatsman S, McLeod M.P, Smajs D, Howell J.K, Pal S, Amin A, Vashisth P, McNeill T.Z, Xiang Q, Sodergren E, Baca E, Weinstock G.M, Norris S.J, Fraser C.M, Paulsen I.T. **2004**. Comparison of the genome of the oral pathogen *Treponema*

- denticola* with other spirochete genomes. *Proc Natl Acad Sci U S A* **101**(15):5646-51.
146. Sharples, G.J. y Leach, D.R. **1995**. Structural and functional similarities between the SbcCD proteins of *E coli* and the RAD50 and MRE11 recombination and repair proteins of yeast. *Mol Microbiol* **17**(6):1215-1217
147. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y e Ishikawa, H. **2000**. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**:81-85
148. Slesarev A.I, Mezhevaya K.V, Makarova K.S, Polushin N.N, Shcherbinina O.V, Shakhova V.V, Belova G.I, Aravind L, Natale D.A, Rogozin I.B, Tatusov R.L, Wolf Y.I, Stetter K.O, Malykh A.G, Koonin E.V, Kozyavkin S.A. **2002**. The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc Natl Acad Sci U S A* **99**(7):4644-9.
149. Smith D.R, Doucette-Stamm L.A, Deloughery C, Lee H, Dubois J, Aldredge T, Bashirzadeh R, Blakely D, Cook R, Gilbert K, Harrison D, Hoang L, Keagle P, Lumm W, Pothier B, Qiu D, Spadafora R, Vicaire R, Wang Y, Wierzbowski J, Gibson R, Jiwani N, Caruso A, Bush D, Reeve J.N. y et. al. **1997**. Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J Bacteriol* **179**(22):7135-7155.
150. Stackebrandt, E., Ludwig, W., Schubert, W., Klink, F., Schlesner, H., Roggentin, T. y Hirsch, P. **1984**. Molecular evidence for early evolutionary origin of budding peptidoglycan-less eubacteria. *Nature* **307**:735-737
151. Stetter, K.O. **1996**. Hyperthermophiles in the history of life. *Ciba Found Symp* **202**:1-10.
152. Tamas, I., Klasson, L., Canbäck, B., Näslund, A.K., Eriksson, A.S., Wernergreen, J.J., Sandström, J.P., Moran, N. y Andersson, S.G. **2002**. 50 Million years of genomic stasis in endosymbiotic bacteria. *Science* **196**:2376-2379
153. Teeling, H., Lombardot, T., Bauer, M., Ludwig, W. y Glöckner, F.O. **2004**. Evaluation of the phylogenetic position of the planctomycete *Rhodopirellula baltica* SH 1 by means of concatenates ribosomal protein sequences, DNA-directed RNA polymerase subunit sequences and whole genome trees. *J Syst Evol Microbiol* **54**(Pt 3):791-801.
154. Thiagalingam, S. y Grossman, L. **1991**. Both ATPase sites of *Escherichia coli* UvrA have unctional roles in nucleotide excision repair. *J Biol Chem* **266**(17):11395-11403
155. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. y Higgins, D.G. **1997**. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nuc Acids Res* **25**:4876-4882
156. Vallee, B.L., Coleman, J.E. y Auld, D.S. **1991**. Zinc fingers, zinc clusters, and zinc twists in DNA-binding protein domains. *Proc Natl Acad Sci USA* **88**:999-1003.
157. Van den Eynde, H., Van de Peer, Y., Perry, J. y De Watcher, R. **1990**. 5S rRNA sequences of representatives of the genera *Chlorobium*, *Prosthecochloris*, *Thermomicrobium*,

- Cytophaga*, *Flavobacterium*, *Flexibacter* and *Saprospira* and a discussion of the evolution of eubacteria in general. *J Gen Microbiol* 136:11-18
158. Van der Zee, A., Mooi, F., Van Embden, J. y Musser, J. **1997**. Molecular evolution and host adaptation of *Bordetella* spp.: phylogenetic analysis using multilocus enzyme electrophoresis and typing with three insertion sequences. *J Bacteriol* **179**:6609-6617
159. Van Duin, M., van den Tol, J., Warmerdam, P., Odijk, H., Meijer, D., Westerveld, A., Bootsma, D. y Hoeijmakers, J.H.J. **1988**. Evolution and mutagenesis of the mammalian excision repair gene ERCC-1. *Nuc Acid Res* **16**(12):5305-5322
160. Verhoeven, E.E.A., van Kesteren, M., Moolenaar, G.F., Visse, R., Goosen, N. **2000**. Catalytic sites for 3'- and 5'-incision of *E.coli* nucleotide excision repair are both located in UvrC. *J Biol Chem* **275**:5120-5123
161. Walker, J.E., Saraste, M., Runswick, M. y Gay, N.J. **1982**. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**:945-951
162. Wang, B. **2001**. Limitations of compositional approach to identifying horizontally transferred genes. *J Mol Evol* **53**:244-250
163. Wang, J., Jenkins, C., Webb, R.I. y Fuerst, J. **2002**. Isolation of *Gemmata*-like and *Isosphaera*-like Planctomycete bacteria from soil and freshwater. *Applied EnvironMicrobiol* **68**(1):417-422
164. Wang, J., Mueller, K.L. y Grossman, L. **1994**. A Mutational study of the C-terminal zinc-finger motif of the *Escherichia coli* UvrA Protein. *J Biol Chem* **269**(14):10771-10775
165. Ward, N., Rainey, F., Stackebrandt, E. y Schlesner, H. **1995**. Unraveling the extent of diversity within the order *Planctomycetales*. *Applied Environmental Microbiol* **61**(6):2270-2275
166. Ward, N., Rainey, F.A., Hedlund, B.P., Staley, J., Ludwig, W. y Stackebrandt, E. **2000**. Comparative phylogenetic analyses of members of the order Planctomycetales and the division Verrucomicrobia: 23S rRNA gene sequence analysis supports the 16S rRNA gene sequence-derived phylogeny. *Int J Syst Evol Microbiol.* **50**:1965-1972
167. Ward-Rainey, N., Rainey, F.A. y Stackebrandt, E. **1997**. The presence of a *dnaK* (HSP70) multigene family in members of the orders *Planctomycetales* and *Verrucomicrobiales*. *J Bacteriol* **179**:6360-6366
168. Ward-Rainey, N., Rainey, F.A., Schlesner, H. y Stackebrandt, E. **1995**. Assignment of hitherto unidentified 16S rDNA species to a main line of descent within the domain Bacteria. *Microbiology* **141**:3247-3250
169. Weisburg, W.G., Hatch, T.P. y Woese, C.R. **1986**. Eubacterial origin of chlamydiae. *J Bacteriol* **167**:570-574

170. White, O., Eisen, J., Heidelberg, J.F., Hickey, E., Peterson, J., Dodson, R., Haft, D., Gwinn, M., Nelson, W., Richardson, D., Moffat, K., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K., Aravind, L., Daly, M., Minton, K., Fleischmann, R., Ketchum, K., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. y Fraser, C.M. **1999**. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* **286**:1571-1577
171. Woese, C.R. **1987**. Bacterial Evolution. *Microbial Rev* **51**:221-271
172. Woese, C.R. **2002**. On the evolution of cells. *Proc Natl Acad Sci* **99**(13):8742-8747
173. Woese, C.R., Olsen, G.J., Ibba, M. y Soll, D. **2000**. Aminoacyl-tRNA synthetases, the genetic code and the evolutionary process. *Microbiol Mol Biol Rev* **64**:202-236
174. Woese, C.R., Stackebrandt, E., Macke, T.J., Fox, G.E. **1985**. A phylogenetic definition of the major eubacterial taxa. *Syst Appl Microbiol* **6**:143-
175. Yasui, A. y McCready, S.J. **1998**. Alternative repair pathways for UV-induced DNA damage. *BioEssays* **20**:291-297
176. Ylihonko, K., Tuikkanen, J., Jussila, S., Cong, L. and Mantsala, P. **1996**. A gene cluster involved in nogalamycin biosynthesis from *Streptomyces nogalater*: sequence analysis and complementation of early-block mutations in the anthracycline pathway. *Mol Gen Genet* **251**(2):113-120
177. Yoshida, M. y Amano, T. **1995**. A common topology of proteins catalyzing ATP-triggered reactions. *FEBS Lett* **359**:1-5