UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

CENTRO DE CIENCIAS GENÓMICAS

PROGRAMA DE DOCTORADO EN CIENCIAS BIOMÉDICAS

LONGITUD DE LOS SEGMENTOS EMPLEADOS PARA CONVERSIÓN GÉNICA ENTRE SECUENCIAS REPETIDAS EN *Rhizobium etli*

T E S I S

QUE PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

GUSTAVO SANTOYO PIZANO

TUTOR:

DR. DAVID ROMERO CAMARENA

CUERNAVACA, MOR. **MAYO, 2005**

DEDICATORIA

A Marycarmen, mi devoción, mi vida, mi absoluto amor.

A José Gustavo y Leonardo, mis dos premios Nobel[2].

AGRADECIMIENTOS

A Marycarmen, por su total apoyo e infinita comprensión a mis deleites.

A José Gustavo y Leonardo, mi felicidad durante tiempos intrincados.

A mi mamá y mi papá†, por el amor a mi existencia.

A Erika e Iván, mis dos maravillosos hermanos.

A mis suegros (y toooda la familia), por su cordial afecto.

A David, por su amistad, excelsa guía y ejemplar pasión por la ciencia.

A la Dra. Paty, mi admiración y entero agradecimiento.

A Jaime, por su tutoría, consejos y ayuda.

A los revisores de esta tesis, por su tiempo, sus comentarios y sugerencias.

A mis amigos de Cuernavaca, por los momentos compartidos.

A las personas del Programa de Ingeniería Genómica, por su grata amistad.

Al CONACYT, México y DGEP, UNAM, mil gracias por financiar mis estudios.

Al Centro de Ciencias Genómicas, por fortalecerme.

A la UNAM, espero retribuirle un poco por la demasía recibida.

Science is a very difficult life. It´s not well paid frequently. It´s very hard work. It can be extraordinarily frustrating. It can be very, very damaging to the psyche. You know, failing to get grants, failing to get things published. Enormous kicks to your ability. So it´s a hard life. So I think that the only reason for doing it is to do it because you really love it ... don´t do it for any other reason.

Peter Little, 2003.

ESTE TRABAJO SE REALIZÓ EN EL PROGRAMA DE INGENIERÍA GENÓMICA DEL CENTRO DE CIENCIAS GENÓMICAS, PERTENECIENTE A LA UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO, BAJO LA TUTORÍA DEL DR. DAVID ROMERO, ASÍ COMO LA COTUTORÍA DE LA DRA. PATRICIA LEÓN (IBT-UNAM) Y EL DR. JAIME MARTÍNEZ (CCG-UNAM).

INDICE 6
RESUMEN 8
ABSTRACT 10
INTRODUCCIÓN 12
DEFINICIÓN DE CONVERSIÓN GÉNICA 12
HISTORIA SOBRE CONVERSIÓN GÉNICA 13
EL MODELO DE HOLLIDAY 14
EL MODELO DE MESELSON-RADDING 15
EL MODELO DE RECOMBINACIÓN Y REPARACIÓN DE CORTES EN DOBLE CADENA 16
DIVERSOS PAPELES DE CONVERSIÓN GÉNICA 19
REPARACIÓN DE CORTES EN DOBLE CADENA 19
CONVERSIÓN GÉNICA COMO CAUSA DE ENFERMEDADES HUMANAS 20
VARIACIÓN ANTIGÉNICA 22
EVOLUCIÓN CONCERTADA 24
EL GENOMA DE *Rhizobium etli* 26
SECUENCIAS REPETIDAS Y REARREGLOS GENÓMICOS 27
CONVERSIÓN GÉNICA EN *Rhizobium etli* 28
ARTÍCULO I 31
¨ GENE CONVERSION AND CONCERTED EVOLUTION IN BACTERIAL GENOMES ¨
ARTÍCULO II 46
¨ GENE CONVERSION TRACTS ASSOCIATED WITH CROSSOVERS IN *Rhizobium etli* ¨
RESULTADOS ADICIONALES 57
ANÁLISIS DE LOS TRACTOS DE CONVERSIÓN GÉNICA EN UNA CEPA DE *Rhizobium etli* CARENTE DEL SISTEMA DE REPARACIÓN DE MISMATCHES (*mutS*) 57
DISCUSIÓN 64
PERSPECTIVAS 71

ANÁLISIS DE ELEMENTOS GENÉTICOS ADICIONALES QUE INTERVIENEN EN LA ANATOMÍA DE LOS TRACTOS DE CONVERSIÓN EN *Rhizobium etli* 71
CONVERSIÓN GÉNICA A NIVEL GENÓMICO 72
MODIFICANDO EL GENOMA DE *Rhizobium etli* POR MEDIO DE CONVERSIÓN GÉNICA 73
CONCLUSIONES 75
EPÍLOGO 77
REFERENCIAS 78

RESUMEN

La conversión génica, ha sido definida como la transferencia de información genética no recíproca entre dos secuencias de DNA, siendo por tanto uno de los posibles resultados de un evento de recombinación. Este mecanismo ha estado involucrado en diversos aspectos que promueven la homogeneidad o variabilidad genética, la estabilidad genómica, así como también es causa de diversas enfermedades en humanos. A pesar de que conversión génica parece ser un evento universal en los seres vivos, en organismos procariotes ha sido difícil de caracterizar debido a la posibilidad de que múltiples eventos de recombinación podrían mimetizar su acción.

En el presente trabajo, hemos diseñado un sistema novedoso en el cual analizamos los eventos de conversión génica asociados a entrecruzamientos (crossovers) en la bacteria fijadora de nitrógeno *Rhizobium etli*. Esto evita la posibilidad de recuperar eventos múltiples de entrecruzamientos en lugar de eventos de conversión. Además, hemos analizado los tractos de conversión génica en una cepa de *R. etli mutS*, la cual carece del sistema de reparación de "mismatches". Para ello, modificamos el gen *nifH* (que codifica para uno de los componentes estructurales de la nitrogenasa) mediante la introducción de sitios de restricción (RFLPs) aproximadamente cada 100 pares de bases (pb) a lo largo del gen. Este gen modificado fue transferido por conjugación en un vector suicida hacia una cepa receptora de *R. etli* que contiene sólo una copia del gen *nifH*. El evento de cointegración nos permitió recuperar convertantes que fueron caracterizadas por medio de PCR y análisis de restricción. Nuestros resultados muestran que (i) los eventos tipo "crossover" están fuertemente asociados a conversión génica; (ii) los tractos de conversión variaron en longitud, desde 150 pb hasta 800 pb;(iii) inesperadamente, los eventos de conversión génica exhibieron una fuerte desviación hacia la ganancia de sitios de restricción, favoreciendo la dirección de conversión de la secuencia entrante (*nifH* con RFLPs) sobre la receptora (*nifH* silvestre); (iv) la introducción de los RFLPs (1.6% de heterología)

tuvieron un efecto sobre la frecuencia de recombinación; (v) la longitud de los tractos de conversión se redujo en un 40% en la cepa *mutS*. La anatomía de los tractos de conversión sugiere que nuestros resultados se explican mejor por eventos de conversión génica verdadera, en lugar de múltiples eventos de recombinación. Así mismo, el análisis de los tractos en la cepa *mutS*, nos indica que este gen esta jugando un papel importante en la estructura de los tractos de conversión génica, aunque no totalmente. Es probable que otros sistemas de reparación de mismatches estén participando.

ABSTRACT

Gene conversion is the result of a recombination event, and has been defined as the non-reciprocal transfer of genetic information between two DNA sequences. This mechanism has been involved in promoting either homogenization or variability of genetic information, genomic stability, as well as in the genesis of diverse human diseases. Gene conversion seems to be universal among living organisms, but unfortunately, this mechanism has been difficult to detect and characterize in bacterial genomes, due to the likelihood to obtain multiple rounds of recombination that should mimic gene conversion.

In the present work, we have designed a novel system to analyze gene conversion events associated with crossovers in the nitrogen-fixing bacterium *Rhizobium etli*. This system avoids the possibility to obtain multiple rounds of recombination instead of having true gene conversions. Besides, we have analyzed the gene conversion tracts in a mismatch repair deficient strain. Thus, we have modified the *nifH* gene by introducing RFLPs approximately 100 bp along the sequence. This mutagenized gene was transferred by conjugation into a strain that contains only one copy of the wild type *nifH* gene. After the cointegration event, the convertants were subjected to PCR and restriction analysis. Our results show that (i) crossover events were almost invariably accompanied by a gene conversion event; (ii) gene conversion events ranged in size from 150 bp up to 800 bp; (iii) gene conversion events displayed a strong bias, favoring the preservation of incoming sequences; (iv) even small amounts of sequence divergence had a strong effect on recombination frequency; (v) length of the gene conversion tracts were roughly reduced 40% in a mismatch repair deficient strain; (vi) as soon as the anatomy of the conversions was modified; (vii) finally, the MutS mismatch repair system plays an important role in determining the length of gene conversion segments.

A detailed analysis of the anatomy of the gene conversion events suggests that multiple crossovers are an unlikely alternative for its generation. Our results

are better explained as the product of true gene conversions, occurring under the double strand break repair model for recombination. Additionally, the analysis of the conversion tracts in the *mutS* mutant suggests that this gene is contributing in the structure and length of the conversions, although not totally. It is probable that other mismatch repair systems are participating.

DEFINICIÓN DE CONVERSIÓN GÉNICA

La recombinación genética es una característica universal de todos los seres vivos. Entender este fenómeno a nivel molecular en los diversos organismos, desde bacterias hasta humanos, es de suma importancia para conocer el impacto que ha tenido en la evolución, función y estructura de sus genomas (Cromie, *et al.*, 2001). La recombinación genética se puede dar a nivel intragenómico, ya sea entre familias de genes o secuencias de DNA que son totalmente idénticas, un proceso llamado recombinación homóloga. La recombinación también puede suceder entre genes que divergen ligeramente en secuencia, por lo que a este proceso se le conoce como recombinación homeóloga. Los eventos de recombinación entre secuencias se pueden dar de manera recíproca, es decir, cada molécula de DNA que interviene en el proceso de recombinación recibe una dotación de información genética (Figura 1). Por otra parte, un resultado interesante de un evento de recombinación es la conversión génica. La conversión génica usualmente se define como la transferencia de información genética de forma no recíproca entre secuencias de DNA. Sin embargo, la conversión génica puede estar asociada a eventos de recombinación tipo crossover (Szostak, *et al.*, 1983), así como también se generarían eventos de conversión

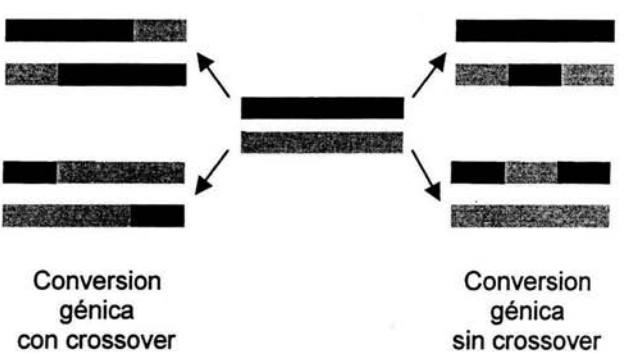

Figura 1. Conversión génica asociada o no a eventos de tipo crossover (Modificado de Santoyo y Romero, 2005).

sin tener intercambio de marcadores aledaños (sin crossover) (Figura 1).

HISTORIA SOBRE CONVERSIÓN GÉNICA

De acuerdo a Whitehouse (1982) y Liu y West (2004), el término conversión génica fue introducido por el científico alemán Hans Winkler en 1930 para describir el rango aberrante 3:1 en tétradas de levadura (*Saccharomyces cerevisiae*). Evidencia posterior fue obtenida por Zickler en 1934 (en Whitehouse, 1982), trabajando con algunas mutantes del hongo *Bombardia lunata* que carecían de color en sus esporas. En la mayoría de sus análisis notaba una segregación en tétradas 2:2 respecto al color de las esporas, dos de color oscuro y otras dos sin color. Todo parecía respetar las leyes Mendelianas de segregación de un carácter, sin embargo, en ocasiones observó proporciones de 3:1 en el color de las esporas. Sus observaciones le llevaron a proponer de nuevo el término de conversión para este fenómeno (Whitehouse, 1982).

Unos años después, Lindegren (1953) utilizó la levadura *Saccharomyces cerevisiae* para obtener evidencia adicional de conversión génica. Análisis de las tétradas haploides, producto de la meiosis, encontró nuevamente una proporción 3:1, cuando se esperaba una proporción 2:2. Estos resultados de conversión génica rápidamente fueron corroborados por Mitchell (1955) en *Neurospora crassa*. Este hongo, en una sola estructura llamada ascus, produce ocho esporas en hilera, por lo que se pueden encontrar proporciones de 6:2 o 5:3, como resultado de un evento de conversión.

Hasta ese momento sólo se había estudiado conversión génica durante la meiosis. Sin embargo, utilizando a *Saccharomyces cerevisiae* como modelo, Roman en 1957 (en Whitehouse, 1982) demostró que la conversión génica también ocurre durante recombinación mitótica. Estudios posteriores de marcadores que flanqueaban la región implicada en los eventos de conversión demostraron que el intercambio de estos marcadores ocurría en el 50% de las ocasiones, un evento que se conoció como crossing over o crossover. Datos que sirvieron para que Robin Holliday (1964) propusiera por primera vez un modelo

molecular que predecía los resultados de conversión génica asociada o no a crossovers.

EL MODELO DE HOLLIDAY

Robin Holliday propuso en 1964 el primer modelo que explicaba los eventos de recombinación genética (Holliday, 1964; Stahl, 1996; Liu y West; 2004), y el cual a su vez, predecía perfectamente la segregación de alelos de forma no Mendeliana en experimentos antes mencionados sobre conversión génica.

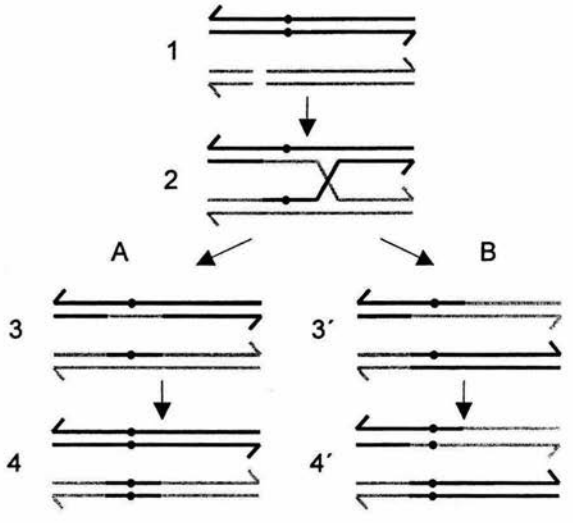

Figura 2. Modelo de Holliday. Los círculos negros indican diferencias en secuencias nucleotídica entre las cadenas de DNA dúplex.Ver el texto para más detalle.

El modelo de Holliday predice que la recombinación inicia con un corte de cadena sencilla en ambas cadenas dúplex participantes (Figura 2, 1). Posteriormente hay intercambio de cadenas sencillas, seguido por ligación (Figura 2, 2). Una vez que se tiene este tipo de unión de cadenas entrecruzadas (al cual se le llama hoy en día la unión o estructura de Holliday), pueden darse dos tipos de resultados, dependiendo de la orientación del corte de las cadenas, ya sea horizontal o vertical. En el caso A (Figura 2, 3), observamos conversión sin crossover o intercambio de marcadores externos. Para el caso B, observamos conversión asociada con crossover (Figura 2, 3´).

Es necesario remarcar que en el modelo de Holliday, la conversión génica solo existe si se forman regiones heteroduplex o híbridas entre las cadenas participantes. De esta manera, el modelo de Holliday predice que conversión génica será el resultado de eventos de reparación de mismatches (Figura 2, 3´ y 4´).

EL MODELO DE MESELSON-RADDING

De acuerdo al modelo de Holliday, la formación de regiones híbridas entre las cadenas parecía ser recíproca, es decir, que los eventos de conversión derivados de la reparación de mismatches se darían en un 50% de los casos para cada una de las cadenas dúplex. Sin embargo, surgieron nuevos datos en la levadura *Saccharomyces cerevisiae* donde no se observó tal reciprocidad de regiones heteroduplex (Stahl,1994). Por lo tanto, Matt Meselson y Charles Radding

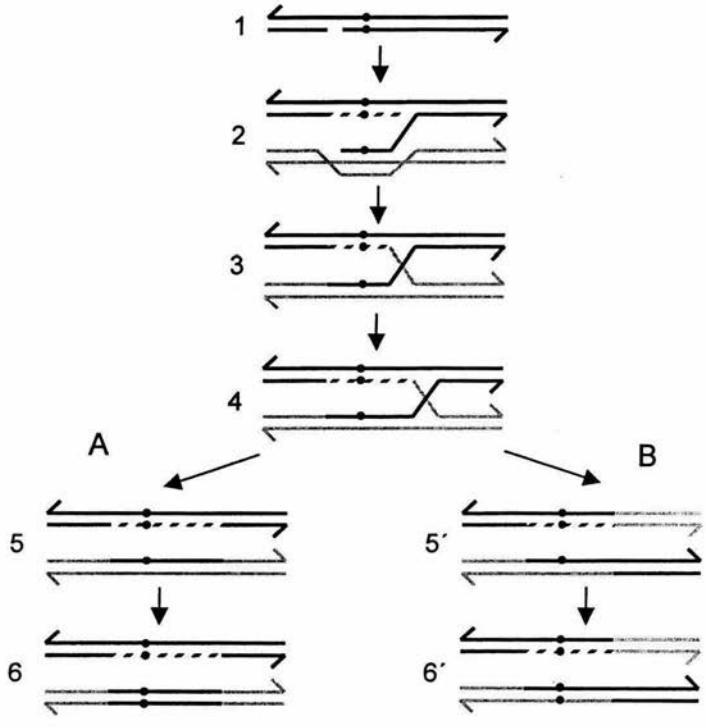

Figura 3. Modelo de recombinación Meselson-Radding. Los círculos negros indican las diferencias en secuencia nucleotídica entre las cadenas de DNA. Resultado de conversión sin crossover (A) y asociado a crossover (B).

(1975), modificaron el modelo de Holliday para adecuar los nuevos resultados de la no reciprocidad en la formación de regiones híbridas (Figura 3).

En su modelo, se conserva el inicio de los eventos de recombinación por un corte en cadena sencilla (Figura 3, 1). Una vez que se da el corte, el extremo 3´ sirve como iniciador para que la DNA polimerasa entre en acción y desplace la cadena, la cual invadirá la región homóloga de otra cadena dúplex, formando una asa en D (Figura 3, 2). Este desplazamiento genera la formación de una región heteroduplex. Posteriormente el asa en D se degrada y las cadenas que quedan se unen para formar una estructura conocida como intermediario o unión de Holliday. Nótese que se genera una región heteroduplex no simétrica (Figura 3, 3). Posteriormente, la unión de Holliday migra y extiende la región heteroduplex de forma simétrica (Figura 3, 4). Nuevamente dependiendo de la orientación del corte, se dará el resultado de conversión sin crossover (Figura 3, A) o con crossover (Figura 3, B). Para que haya un evento de conversión dependerá, como en el modelo de Holliday, de la reparación de mismatches en las regiones heteroduplex (Figura 3, 4-5 y 5´-6´).

El modelo de Meselson-Radding es importante por que en ese momento resolvía el problema de la formación no recíproca de regiones heteroduplex entre las cadenas, además de que innovaba al agregar el nuevo elemento de la polimerización, el cual se conservaría para la propuesta de un nuevo modelo: el modelo de recombinación y reparación de cortes en doble cadena.

EL MODELO DE RECOMBINACIÓN Y REPARACIÓN DE CORTES EN DOBLE CADENA

Los modelos previos de recombinación que explican eventos de conversión génica se iniciaban únicamente con un corte en cadena sencilla. Sin embargo, trabajos realizados por Hicks (1979), Orr-Weaver *et al.* (1981) y Orr-Weaver y Szostak (1983), sugirieron que los cortes en doble cadena podrían dar inicio a procesos de recombinación. En 1981, Orr-Weaver *et al.* construyeron un plásmido que contenía un ¨gap¨ (en doble cadena). Este plásmido se transformó en

Saccharomyces cerevisiae y posteriormente se seleccionaron aquellas cepas que contenían el plásmido cointegrado en el genoma de la levadura. Al cointegrarse este replicón se forzaba la selección para recuperar únicamente eventos tipo crossover. Algo interesante de estos datos es que el gap estaba siendo reparado por recombinación, utilizando una secuencia homóloga como templado del genoma de *Saccharomyces*. Unos años más tarde, Orr-Weaver y Szostak (1983) incorporaron un origen de replicación en el plásmido, de tal manera que ahora el replicón podría sobrevivir sin algún evento de cointegración (crossover). En aproximadamente el 50% de los casos recuperaron eventos crossover y el otro 50% eran eventos de reparación del gap sin crossover. Cabe destacar que en ambos resultados, con o sin crossover se estaba reparando el gap por conversión génica. Este fue un resultado interesante ya que los modelos previos de recombinación postulaban que conversión génica era resultado de reparación de mismatches, y no de reparación de gaps o cortes en doble cadena en el DNA. Por lo tanto, enseguida se propuso el modelo de recombinación y reparación de cortes en doble cadena (Szostak, *et al.*, 1983) (Figura 4).

En este modelo, la recombinación inicia con un corte en doble cadena (Figura 4, 1), seguido de degradación que deja extremos 3´ libres, las cuales pueden invadir una región homóloga en otra cadena dúplex. Al haber invasión del extremo 3´ se forma una asa en D en la cadena receptora, debido al efecto de la polimerización. El asa en D formada, ahora puede servir como templado para que la cadena invasora sea polimerizada (Figura 4, 2). Después de esto se da la ligación de las cadenas y se forman dos uniones de Holliday, las cuales pueden migrar y extender la longitud de la región heteroduplex (Figura 4, 3). Finalmente, dependiendo de la orientación del corte de cada unión de Holliday será el resultado, ya sea de conversión asociada (Figura 4, B) o no a crossover (Figura 4, A). Esto da la posibilidad de tener como resultado la mitad de los eventos de conversión sin crossover y viceversa.

Cabe destacar que en este modelo de recombinación y reparación de cortes en doble cadena ya predice el dato donde conversión puede ser resultado de:

1. Reparación de gap.

2. Formación y reparación de mismatches en regiones heteroduplex.
3. Capacidad de la migración de los intermediarios de Holliday.

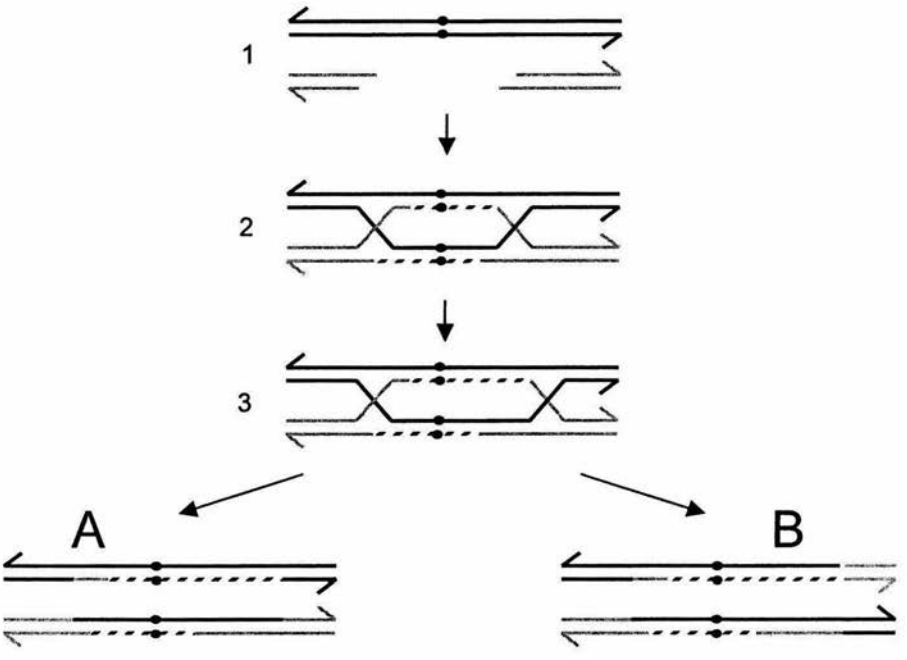

Figura 4. Modelo de reparación de cortes en doble cadena por recombinación. Los círculos negros indican las diferencias en secuencia nucleotídica entre las cadenas de DNA. Resultado de conversión sin crossover (A) y asociado a crossover (B). Ver el texto para mayor detalle.

Es importante mencionar que este modelo mantiene aún algunos elementos de los modelos previos de recombinación, tales como la formación de las uniones de Holliday del modelo de Holliday (que ahora son dos), así como el acto de polimerización del modelo Meselson-Radding.

DIVERSOS PAPELES DE CONVERSIÓN GÉNICA

REPARACIÓN DE CORTES EN DOBLE CADENA

Los cortes en doble cadena en el DNA pueden surgir por una diversidad de factores, incluyendo la radiación ionizante, la exposición a químicos, por factores como la replicación (colapso de la horquilla de la replicación) y de manera programada durante la meiosis (Norbury y Hickson, 2001). Así como también por la acción de enzimas de restricción, transposones, bacteriófagos, algunos antibióticos, entre otras causas (Kobayashi, 2002). De no repararse eficientemente los cortes en doble cadena, se pueden provocar diversos rearreglos cromosomales (Khanna y Jackson, 2001). Por ejemplo, en células de humanos las translocaciones pueden darse por cortes en doble cadena que no fueron reparados debidamente, conduciendo a la generación de diversos tumores (Rabbitts, 1994; Richardson y Jasin, 2000). Una vía que es importante, y de hecho la principal por la cual se reparan cortes en doble cadena en la levadura *Saccharomyces cerevisiae* y en células de humanos (Liang *et al.*, 1998), es a través de recombinación homóloga, siendo la conversión génica el evento que predomina en la reparación. Así, la reparación de cortes en doble cadena a través de conversión génica es la más eficiente, ya que evita la generación de rearreglos como deleciones, traslocaciones o amplificaciones. De hecho, en un trabajo realizado por Wiese *et al.* (2002) observaron que los cortes en doble cadena en células humanas inducen la reparación de los cortes por conversión génica, manteniendo así, la estabilidad genómica (Wiese *et al.*, 2002) y evitando la generación de diversas enfermedades.

LA CONVERSIÓN GÉNICA COMO CAUSA DE ENFERMEDADES EN HUMANOS

La conversión génica puede ser un arma de doble filo. Así como mencionábamos anteriormente que puede reparar cortes en doble cadena de manera eficiente, evitando la generación de rearreglos cromosomales, puede a su vez ser el factor que cause enfermedades en humanos (Hurles, 2002).

Cuando un gen se duplica puede acumular mutaciones en cada una de las copias, lo cual puede resultar en que alguna de ellas termine por ser un pseudogen o adoptar una nueva función. Supongamos que uno de ellos se convierte en pseudogen y la otra copia aún retiene su función inicial, ya que es indispensable para el organismo. En teoría, si ambas copias conservan una identidad en secuencia razonablemente alta como para recombinar entre ellas, puede haber eventos de conversión en ambas direcciones. El problema sucede cuando el pseudogen es el donador de información, ya que podría afectar la función del gen esencial, lo cual deriva en el desarrollo de alguna enfermedad genética. Es por ello que se han detectado algunos desórdenes genéticos en humanos, donde el mecanismo de conversión génica de este tipo es el responsable de varias enfermedades (Hurles, 2002). En la tabla 1 mencionamos algunos ejemplos.

Tabla 1. Enfermedades y genes afectados por conversión génica.

Enfermedad	Gen afectado	Longitud de los segmentos de conversion génica	Ref.
Hiperplasia adrenal congénita	*CYP21A2*	De 200-500 pb	37, 76
Atrofia espinal muscular II y III	*SMN1*		5, 6, 21
Shwachman Diamond	*SBDS*	240 pb	2
Granulomatosa	*p47-phox*,	377 pb	63
Gaucher	*GBA*	De 190 1193 pb	15, 75
Riñon Poliquística	*PKD1*		32, 66
Catarata	*CRYBB2*	De 9 a 104 pb	78
Pancreatitis	*PRSS1*		8

Existen algunos otros desórdenes genéticos en humanos, donde a la expansión de tripletes (por ejemplo: CTG-CAG, CGG-CCG o GAA-TTC) se le ha asociado con más de 14 enfermedades (Jakupciak y Wells, 2000a). Un ejemplo de ello es la distrofia miotónica, donde se ha sugerido que eventos de conversión génica pueden ser los responsables de la expansión de tripletes, y estos a su vez, son los causales de la enfermedad (O´Hoy, *et al.*, 1993). En este mismo sentido, en la bacteria *Escherichia coli* se trató de elucidar el origen de la expansión de

tripletes (Jakupciak y Wells, 2000b). En este modelo, la conversión génica fue el principal mecanismo que incrementó considerablemente el número de tripletes, además de que otros elementos como replicación y reparación pudieran contribuir. De esta manera, en algunas otras enfermedades neurodegenerativas como Huntington, el síndrome X frágil o la ataxia de Friedreich, que sus causas son también la expansión excesiva de tripletes, es probable que la conversión génica esté jugando un papel importante en su etiología.

VARIACIÓN ANTIGÉNICA

El éxito de muchos patógenos bacterianos para infectar a sus huéspedes se debe, entre varios aspectos, a que pueden escapar del sistema de inmune del huésped. El mecanismo conocido como variación antigénica, es el responsable por el cual diversas bacterias patógenas pueden evitar el sistema inmune de sus huéspedes (Deitsch, *et al.*, 1997). De esta manera, utilizan varios mecanismos genéticos para tener diversas combinaciones en sus proteínas de membrana (antígenos). Algunos de ellos pueden ser la regulación a nivel transcripcional, generación de rearreglos, altas tasas de mutación puntual y conversión génica (Deitsch, *et al.*, 1997). Respecto a este último, la conversión ha sido reconocida como uno de los mecanismos más usados por diversos patógenos. De hecho, a través de conversión génica en algunas bacterias se pueden generar millones de variantes en las proteínas de membrana, lo cual hace que este repertorio le antígenos le permita iescapar del sistema inmunológico del huésped (Serkin y Seifert, 1998).

Por lo general, los sistemas de variación antigénica en bacterias patógenas se caracterizan por tener un gen funcional y varios pseudogenes. Cada uno de estos pseudogenes es un donador potencial de información genética a través de conversión génica hacia el gen funcional (Figura 5). Además de ello, los eventos

de conversión o donación de información, pueden ser por segmentos génicos, lo que significa que un solo pseudogene podría generar varias combinaciones antigénicas (Zhang y Norris, 1998).

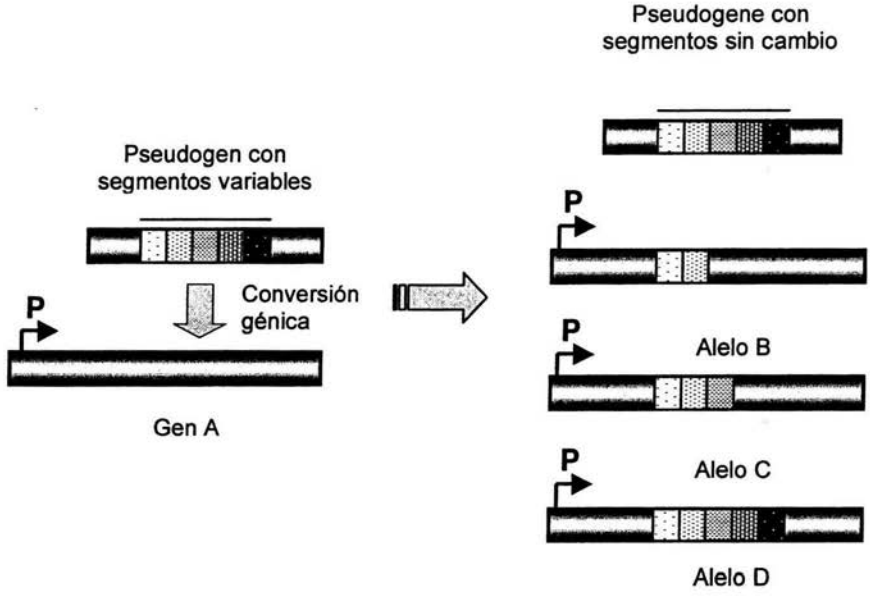

Figura 5. Mecanismo de variación antigénica por medio de conversión génica (Modificado de Santoyo y Romero, 2005).

Uno de los sistemas más estudiados ha sido la bacteria *Neisseria gonorrhoeae*, agente causal de la gonorrea, una enfermedad de transmisión sexual en humanos (Haas y Meyer, 1986; Haas, *et al.*, 1992). En esta bacteria se ha reportado que su genoma contiene un gen funcional *pilE* y hasta 19 pseudogenes *pilS*, dependiendo de la cepa. El gen *pilS* codifica para el pilus de la bacteria y se ha visto que está involucrado en la capacidad para infectar a su huésped. En diversos trabajos se han detectado eventos de conversión entre las secuencias *pilS* y el gen *pilE*, así como eventos segmentales donde se podrían generar hasta 47 millones de variantes de la proteína que compone el pili (Serkin y Seifert, 1998). Otros han reportado que entre estos mismos elementos genéticos, *pilS* y *pilE*, las frecuencias de conversión son de las más altas comparadas con

otros sistemas de recombinación, ya que por medio de RT-PCR se detectaron eventos de conversión génica de 3.3×10^{-2} (Serkin y Seifert, 1998).

Otro ejemplo interesante de variación antigénica en patógenos se reportó en la bacteria *Treponema pallidum*, el agente etiológico de la sífilis en humanos (Centurión-Lara, *et al.*, 2004). En esta bacteria, el gen *tprK es* el responsable de generar variaciones antigénicas, además de otros 47 pseudogenes que actúan como donadores de información genética. Adicionalmente, dentro de cada secuencia de los pseudogenes se reconocen hasta siete regiones variables, lo cual hace que se puedan generar hasta 420,000 variantes. Aún cuando no se reportaron las frecuencias de conversión entre el gen *tprK* y los pseudogenes, el tener este número de variantes le permite al patógeno tener una gran flexibilidad para poder evitar y escapar al sistema inmune del huésped. Algunos otros ejemplos de variación antigénica y conversión se han reportado en patógenos como *Campylobacter jejuni* (Harrington, *et al.*, 1997), *Borrelia burgdorferi* (Zhang y Norris, 1998), *Borrelia hermsii* (Restrepo y Barbour, 1994), *Anaplasma marginale* (Brayton, *et al.*, 2002), entre otros.

Los ejemplos anteriores nos muestran que la conversión génica puede ser un aliado muy importante para patógenos como *Neisseria gonorrhoeae* y *Treponema pallidum*, entre otros, ya que de ello depende en gran parte el éxito que tengan durante la infección al huésped, además de que representan un grave probable de salud en humanos.

EVOLUCIÓN CONCERTADA

La duplicación de genes conduce a la generación de familias multigénicas, generando de dos a varias copias del mismo gen (Ohno, 1970). Cada copia puede adquirir mutaciones puntuales a través del tiempo y adquirir nuevas funciones. En algunos casos más drásticos pueden derivar en pseudogenes, sin tener alguna función aparente. Por otra parte, comúnmente se le ha asociado a mecanismos de conversión génica con la homogenización de familias multigénicas, un proceso llamado evolución concertada (Dover, 1992).

Un ejemplo de evolución concertada en bacterias y arqueobacterias lo representan la familia de genes ribosomales (rRNA), los cuales se encuentran repetidos de dos a siete copias dependiendo de la especie (Liao, 2000). En este trabajo se realizaron análisis filogenéticos de los genes ribosomales de 19 genomas completos de bacterias y arqueobacterias. Los resultados mostraron que las diversas repeticiones son muy similares a nivel de especie, pero muy diferentes si se comparan con otros géneros. Esto sugiere que un mecanismo como conversión génica podría estar jugando un papel importante en la evolución de los genes ribosomales, manteniendo su alta identidad a través de procesos de recombinación no recíproca (Liao, 2000).

Otro dato interesante de evolución concertada, derivado de análisis filogenéticos, se da en los genes *tuf* (que codifican para factores de elongación), en bacterias como *Escherichia coli*, *Haemophilus influenzae*, *Vibrio cholerae*, *Pseudomonas aeruginosa*, *Neisseria meningitidis* y *Salmonella typhimurium* (Lathe III y Bork, 2001). Derivado de este trabajo se reportó que la secuencia nucleotídica de los genes *tuf* es más similar a nivel comparativo de especie que entre géneros.

Utilizando la bacteria patógena *Salmonella typhimurium* se detectaron diversos eventos de conversión génica entre repeticiones de los genes *tuf*, las cuales se encuentran en orientación inversa en el cromosoma y separadas por 700 kilobases. Las tasas de conversión entre estos elementos genéticos fueron dependientes de genes para recombinación como *recA* y *recB*. La orientación inversa de los genes *tuf* podría dar lugar a inversiones, por lo que Hughes (2000) en un estudio adicional demostró que la conversión puede estar asociada a eventos crossover, tal y como sucede con las inversiones de aproximadamente 700 kilobases dentro del cromosoma de *Salmonella typhimurium*.

Respecto a ejemplos de evolución concertada en organismos eucariotes, en un trabajo reciente Drouin (2002) detectó eventos de conversión génica en el genoma completo de la levadura *Saccharomyces cerevisiae*. Varias familias multigénicas mostraron eventos de conversión entre sus miembros, encontrándose además que existía una preferencia por la región 3´ para participar en conversión. Este resultado sugirió que un mecanismo de conversión mediado

por moléculas de cDNA incompletas podría estar dando esta preferencia por convertir la región 3´ de las familias génicas. Esto podría ser posible ya que datos publicados por Derr y Strathern (1993), demostraron que conversión génica puede estar mediada por transcriptos reversos en la levadura *Saccharomyces cerevisiae*.

El desarrollo de la genómica, junto con algunas pruebas estadísticas (Sawyer, 1986) han permitido investigar eventos de conversión en familias multigénicas de genomas completos, ya sean procariotes o eucariotes. Es por ello que conforme se vayan descifrando más genomas, además de secuenciar varias cepas de una misma especie, sin duda que ayudará a detectar nuevos eventos de conversión en diversos organismos, además de tener un mejor conocimiento de este mecanismo y su impacto en la evolución y estructura de los genomas.

EL GENOMA DE *Rhizobium etli*

Rhizobium etli es una bacteria que pertenece a la familia de las Rhizobiaceae, demás de que tiene la capacidad de reducir el nitrógeno atmosférico a amonio, una actividad conocida como la fijación biológica de nitrógeno. Para realizar la fijación de nitrógeno, *R. etli* es capaz de infectar las raíces de plantas de frijol, formando unas estructuras globulares llamadas nódulos. En los nódulos es donde *R. etli* fija el nitrógeno y lo dona a la planta en forma de amonio para poderlo asimilar, lo cual mejora considerablemente su desarrollo (Segovia, *et al.*, 1993: Dávila, *et al.*, 2000).

De esta manera, *Rhizobium etli* es un modelo importante de estudio desde el punto de vista agronómico, ya que por su capacidad de fijar nitrógeno puede mejorar considerablemente el crecimiento de plantas de frijol. En algunos estudios recientes su genoma ha sido modificado genéticamente para incrementar la fijación de nitrógeno en experimentos de invernadero y de campo, obteniendo resultados favorables sobre un incremento en la fijación de nitrógeno y el contenido de nitrógeno de las semillas (Peralta, *et al.*, 2003).

Por otra parte, los genes involucrados en el proceso de nodulación y fijación de nitrógeno en *R. etli*, así como en la mayoría de las rhizobia, se encuentran

localizados en plásmidos llamados simbióticos o pSim (Brom y Romero, 2004). Recientemente, se publicó la secuencia del plásmido simbiótico (pSim) de *Rhizobium etli*, el cual tiene un tamaño de 371,255 pb (González, *et al.*, 2003). Respecto al resto de su genoma, desde hace tiempo se sabe que se encuentra dividido en varios replicones, un cromosoma y seis plásmidos (González, *et al.*, resultados no publicados). El cromosoma tiene un tamaño de 4,381,611 pb, mientras que los plásmidos son considerados de gran tamaño, varían desde los 184,338 pb (p42b) hasta los 642,334 pb (p42f) (González, *et al.*, resultados no publicados).

SECUENCIAS REPETIDAS Y REARREGLOS GENÓMICOS

La determinación de la secuencia del plásmido simbiótico mostró diversos aspectos sobre su origen evolutivo. Uno de ellos, es que su estructura compleja ha sido parte de una dinámica genética que involucra eventos de transposición, transferencia horizontal y recombinación (González, *et al.*, 2003). Respecto a este último, para llevar a cabo eventos de recombinación se requiere la presencia de secuencias repetidas como parte de la arquitectura de este replicón. Previamente se habían descrito algunas familias multigénicas en el pSim (Quinto, *et al.*, 1982; Flores, *et al.*, 1987: Girard, et al., 1991). Actualmente, con la secuencia completa del plásmido simbiótico se conocen en total 12 familias multigénicas, las cuales tienen de dos a tres miembros cada una. Para determinar cada una de estas familias se requirió que mantuvieran al menos 300 pb de identidad continua entre las diversas copias (González, *et al.*, 2003).

Las secuencias reiteradas o las familias multigénicas representan material disponible para recombinar, lo cual puede generar rearreglos genómicos. En un trabajo realizado por Romero *et al.* (1991) se determinó que la familia multigénica de *nifH*, la cual codifica para uno de los componentes de la nitrogenasa, puede llevar a cabo eventos de amplificación y deleción de una zona de aproximadamente 120 kb de el pSim, plásmido donde residen las tres copias de *nifH*. En un trabajo posterior se observó que algunas otras regiones repetidas del

pSim eran al igual susceptibles de recombinar y generar diferentes rearreglos (Romero, *et al.*, 1995). Así mismo, algunos otros eventos naturales como rondas adicionales de replicación, pueden incrementar la frecuencia de recombinación entre las copias de *nifH* del plásmido simbiótico (Valencia-Morales y Romero, 2000).

Con el avance de la genómica y el conocimiento de la secuencia completa de los genomas de diversos organismos, hoy en día existe la posibilidad de predecir rearreglos genómicos de diversos replicones, todo ello previo a observarlos experimentalmente. Flores y colaboradores (2000) publicaron un trabajo donde basados en el diseño de oligonucleótidos para PCR; bordeando las regiones repetidas, pudieron predecir diversos rearreglos, incluyendo deleciones o amplificaciones en el plásmido simbiótico de la cepa de *Rhizobium* NGR234.

En este mismo sentido, la presencia de secuencias repetidas también permite la recombinación interplasmídica o plásmidica-cromosomal, tal y como lo demuestran Mavingui *et al.* (2002) y Guo *et al.* (2003). En tales trabajos, se obtienen nuevas estructuras genómicas que involucran eventos de recombinación entre secuencias repetidas, obteniéndose la cointegración de plásmidos y cromosomas como un solo replicón. Por lo tanto, la presencia de secuencias repetidas ha permitido generar nuevas arquitecturas genómicas como parte de un nuevo diseño genómico natural (Guo, *et al.*, 2003).

CONVERSIÓN GÉNICA EN *Rhizobium etli*

La recombinación entre elementos genéticos repetidos que generan rearreglos genómicos representan eventos de intercambio recíproco de información o tipo crossover. En el caso de *Rhizobium etli*, la presencia de familias multigénicas en sus diferentes replicones representa también la posibilidad de llevar a cabo eventos de conversión entre sus miembros, y que estos a su vez evolucionen de manera concertada. Tal es el caso de la familia *nifH*, en la cual Sepúlveda y Romero (resultados no publicados) aislaron cada una de las copias de *nifH* de 11 cepas de *R. etli* que tienen diversos orígenes geográficos, como

Brasil, Belice, E. U. A., Colombia, Hawai y México. Una vez que se conoció la secuencia completa de cada una de sus copias, se alinearon y por un análisis filogenético se buscó un patrón que mostrara rastros de evolución concertada. En sus resultados, aún cuando fueron pocas las diferencias nucleotídicas encontradas entre los miembros de la familia *nifH*, se pudo observar que en algunas cepas las reiteraciones eran más similares a nivel intracepa que intercepa. Esto sugiere que existe evolución concertada entre las copias de *nifH* de varias cepas de *Rhizobium*. Adicionalmente, este trabajo revela que las tres copias que forman parte de la familia multigénica *nifH*, es muy improbable que se hayan originado recientemente, ya que se encontró la triplicación de sus miembros en las mayoría de las 11 cepas de *R. etli* (Sepúlveda y Romero, manuscrito en preparación).

Tratando de encontrar evidencia experimental de que mecanismos como conversión génica podrían estar homogenizando las copias de la familia multigénica *nifH*, Rodríguez y Romero (1998) describieron diversos eventos de recombinación entre sus tres copias, incluyendo conversión génica. En dicho trabajo, los eventos de conversión representaron el 14% de las recombinantes analizadas, con una frecuencia de 8×10^{-5}. Tal frecuencia fue mayor que la mutación espontánea. Algunos de los eventos de conversión pudieron ser parte de un evento de recombinación recíproca, pero que no pudieron ser detectados, por lo que se le llamó conversión génica aparente. Sin embargo, los eventos de recombinación, incluyendo conversión aparente, necesitaron de la función del gen *recA*, así como de la presencia de más de una copia del gen *nifH*, además de la corrección precisa de la mutación que se insertó en una de las copias. Todo ello sugiere que los eventos de conversión aparente, junto con otros eventos de recombinación, podrían estar siendo parte del mecanismo que impera en la evolución concertada de la familia *nifH* de *Rhizobium etli*.

Durante la realización de este proyecto de doctorado, se profundizó más allá del trabajo experimental sobre el tema de conversión génica y evolución concertada en organismos bacterianos, por lo que se escribió una revisión que fue aceptada para su publicación en la revista *FEMS Microbiology Reviews*. La

referencia es la siguiente: Santoyo, G. and D. Romero. 2005. Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiol. Rev.* 29, 169-183. Dicha revisión se incluye en esta tesis. (Artículo I).

En el presente proyecto de investigación, se trató de obtener un mejor conocimiento sobre el impacto de conversión génica en la evolución concertada de la familia multigénica *nifH* de *Rhizobium etli*. Para ello, uno de los objetivos fue analizar la longitud de los tractos de conversión entre los genes *nifH*, además de tratar de evitar tener eventos aparentes de conversión génica, o al menos, que su contribución fuera mínima. Para tal fin, se diseñó un nuevo sistema para analizar los eventos de conversión génica asociados a crossovers. Así mismo, como parte de los resultados adicionales de este proyecto, se investigó el efecto de una mutación en el gen *mutS* de *Rhizobium etli* y cómo podría participar en la formación de los tractos de conversión, además de analizar si se modifica su anatomía. Los resultados de este proyecto se presentan en el artículo aceptado para su publicación en la revista *Journal of Bacteriology*. La referencia es la siguiente: Santoyo, G., Martínez-Salazar, J. M., Rodríguez, C. and D. Romero. 2005. Gene conversion tracts associated with crossovers in *Rhizobium etli*. *J. Bacteriol. Aceptado para su publicación.* (Artículo II).

ELSEVIER

FEMS Microbiology Reviews 29 (2005) 169–183

FEMS MICROBIOLOGY Reviews

www.fems-microbiology.org

Gene conversion and concerted evolution in bacterial genomes ☆

Gustavo Santoyo, David Romero *

Programa de Ingeniería Genómica, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Apartado Postal 565-A, Cuernavaca, Morelos, Mexico

Received 7 July 2004; received in revised form 21 October 2004; accepted 27 October 2004

First published online 19 November 2004

Abstract

Gene conversion is defined as the non-reciprocal transfer of information between homologous sequences. Despite methodological problems to establish non-reciprocity, gene conversion has been demonstrated in a wide variety of bacteria. Besides examples of high-frequency reversion of mutations in repeated genes, gene conversion in bacterial genomes has been implicated in concerted evolution of multigene families. Gene conversion also has a prime importance in the generation of antigenic variation, an interesting mechanism whereby some bacterial pathogens are able to avoid the host immune system. In this review, we analyze examples of bacterial gene conversion (some of them spawned from the current genomic revolution), as well as the molecular models that explain gene conversion and its association with crossovers.

Keywords: Gene conversion; Homologous recombination; Concerted evolution; Antigenic variation

Contents

1. Introduction 169
2. Molecular models that explain recombination and gene conversion 170
3. Experimental verification of gene conversion in bacteria 172
4. Molecular evolutionary inference as an aid to detect gene conversion 175
5. Role of gene conversion for the generation of antigenic variation 177
6. Concluding remarks and perspectives 179
Acknowledgments 180
References 181

1. Introduction

Homologous recombination is crucial for the long-term survival and evolution of bacterial cells. Although recombination is frequently analyzed in the context of lateral gene transfer (an intergenomic event), most of the recombination occurring in bacterial cells is an intragenomic event. This kind of recombination may help the repair of collapsed replication forks, a rather frequent event in bacteria [1–3]. Moreover, intragenomic recombination between resident repeated sequences (either between insertion sequences or with members

☆ Edited by Martin Kupiec.

* Corresponding author. Tel.: +52 777 3175867/3291691; fax: +52 777 3175581.
E-mail address: dromero@cifn.unam.mx (D. Romero).

0168-6445/$22.00 © 2004 Federation of European Microbiological Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.femsre.2004.10.004

of multigene families) also occur with a high frequency. Multigene families are an inherent part of eukaryotic, archeal and prokaryotic genomes [4–6]. Commonly, members of these families share high sequence similarity, thereby they could be potential targets for exchange of genetic material by homologous recombination. Besides exchange of variant genomic information, other possible outcomes of intragenomic recombination could be genomic rearrangements, such as translocations, deletions, duplications and inversions with diverse biological implications [7–9]. Another result of a recombination event is the non-reciprocal transfer of genetic information between two or more gene copies, a process called gene conversion.

Several studies have shown that gene conversion may play an important role in the evolution of multigene families. For example, some bacterial pathogens have the capacity to produce a high diversity of transmembrane proteins, which help to avoid the host immune system. The common theme in this case is that variant gene sequences are transferred from unexpressed genes (termed pseudogenes or gene cassettes) into an expressed sequence, through gene conversion. This process has the consequence that different expressed genes are now generated, provoking antigenic diversity [10]. The degree of variability achieved is conditioned by the number of unexpressed cassettes and their sequence diversity. The concerted evolution of multigene families (i.e. the spread of identical mutations between its members) is another consequence of probable gene conversion events [11–13]. Furthermore, mutations that confer a weak biological advantage when present in a single member, can be transferred by gene conversion to all the copies of the family, thus maximizing their effect on fitness [14].

As shown in Fig. 1, gene conversion is associated, half of the time, with exchange of flanking markers. When this occurs, a sector of a sequence is transferred into its homologous zone. Note that in this case there is exchange of flanking markers (Fig. 1(a)). This kind of recombination may affect the genome stability, provoking genomic rearrangements. On the other hand, if gene conversion occurs without associated exchanges,

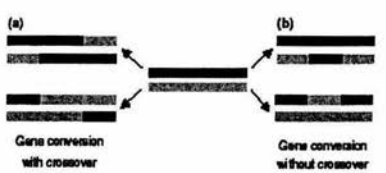

Fig. 1. Association between crossovers and gene conversion. As shown in (a), gene conversion can be associated, half of the time, to crossover or marker flanking exchange. For the rest of the cases, gene conversion occurs without association with crossovers (b).

only a sector is transferred (Fig. 1(b)). This opens the possibility for concerted evolution of a multigene family without affecting the genome architecture. In this review, we will analyze the current experimental and phylogenetic evidence for the occurrence of gene conversion in bacteria (see Table 1). This analysis also includes a study of antigenic variation mediated by gene conversion. Other specialized instances of gene conversion, such as intron homing and marker exclusion in phages [15–17], are out of the scope of this review. Excellent texts on homologous recombination are available [18,19].

2. Molecular models that explain recombination and gene conversion

The first molecular model that explained gene conversion was initially proposed by Holliday in 1964 ([20], for a historical account see [21]). His pioneering proposal (based solely in experiments with lower fungi) leads to the model that bears his name (Fig. 2). In this model, recombination inititiates with a single-strand nick made in both of the DNA participating molecules, followed by unwinding and strand exchange. Then, a four-strand structure is formed (Holliday junction), which can be resolved to give two different results: crossing over or gene conversion. Of course, if the DNA molecules are not identical in sequence, some mismatches will arise in the heteroduplex DNA, and so, the repair system will recognize and probably correct them. Depending of the repair preference, there will be gene conversion or only reciprocal exchange. The length of the gene conversion tract will depend on both the migration of the Holliday structure and the capacity to repair the mismatches before replication. Finally, orientation of the cuts needed to resolve the Holliday junction will ultimately determine the end result, being gene conversion events, crossover events, or both. Therefore, in the Holliday model, gene conversion will be esentially a consequence of DNA heteroduplex formation and the role of the repair system.

The great flexibility and richness of the Holliday proposal, coupled to its heuristic and predictive power, made this model to stay unchallenged for a decade. However, Meselson and Radding [22] (Fig. 2) proposed some important modifications to the Holliday model, trying to fit in some new data in *Saccharomyces cerevisiae*; in this organism little or no reciprocal heteroduplex DNA could be detected [21,23], a feature that departs from the expectatives of the Holliday model. In the Meselson–Radding model, one single-strand cut is made in one of the chains, which is then displaced by the action of a DNA polymerase. The displaced chain invades the homolog sequence; ligation of the newly synthesized strand with a strand of the same polarity in the other homolog generates the Holliday junction (Fig. 2). Reso-

Table 1
Gene conversion in bacterial species and its associated role

Species	Gene	Evidence	Role	Ref.
Acinetobacter calcoaceticus	*pcaJ*	Experimental	Reversion	[63]
Anaplasma marginale	*msp2*	Experimental	Antigenic variation	[11]
Borrelia burgdorferi	*vls*	Experimental	Antigenic variation	[81]
Borrelia hermsii	*vmp*	Experimental	Antigenic variation	[82]
Campylobacter jejuni	*fla*	Phylogenetic	Concerted evolution	[83]
Chlamydia pneumoniae	*hap*	Experimental	Concerted evolution	[68]
Deinococcus radiodurans	*tuf*	Phylogenetic	Concerted evolution	[72]
Escherichia coli	*neo*	Experimental	Gap repair	[38]
	rRNA	Phylogenetic	Concerted evolution	[71]
	tuf	Phylogenetic	Concerted evolution	[72]
Haemophilus influenzae	*tuf*	Phylogenetic	Concerted evolution	[72]
Helicobacter pylori	*hcb*	Phylogenetic	Concerted evolution	[73]
	hsp	Phylogenetic	Concerted evolution	[68]
Mycobacterium smegmatis	*16S*	Experimental	Antibiotic resistance	[56]
Mycoplasma synoviae	*vlhA*	Experimental	Antigenic variation	[80]
Neisseria gonorrhoeae	*pilE*	Experimental	Antigenic variation	[85]
Neisseria meningitidis	*pilE*	Experimental	Antigenic variation	[84]
	tuf	Phylogenetic	Concerted evolution	[72]
Pseudomonas aeruginosa	*tuf*	Phylogenetic	Concerted evolution	[72]
Rhizobium etli	*nifH*	Experimental	Concerted evolution	[61]
	nifH	Phylogenetic	Concerted evolution	Unpubl.
Salmonella typhi	*rrl*	Phylogenetic	Concerted evolution	[78]
Salmonella typhimurium	*tuf*	Experimental	Concerted evolution	[53]
	rrl	Phylogenetic	Concerted evolution	[75]
Treponema pallidum	*tprK*	Experimental	Antigenic variation	[103]
Vibrio cholerae	*tuf*	Phylogenetic	Concerted evolution	[72]

lution of this intermediate occurs in the same form as in the Holliday model. In this model, the invading sequence can give origin to a non-reciprocal DNA heteroduplex, that later on becomes a reciprocal heteroduplex region, upon migration of the Holliday junction; repair in any of these heteroduplexes may generate gene conversion.

A more recent model that has gained wide acceptance is the double-strand break repair model (DSBR) proposed by Szostak et al. [24] (Fig. 2). This DSBR model emerged as the canonical model because of the huge amount of genetic evidence in fungi that indicate that double-strand breaks (rather than single-strand breaks) can act as initiators of recombination [25,26]. In this model, recombination initiates with a double-strand break, continued by extensive single chain degradation in the 5′ to 3′ direction, thus generating a gap with exposed 3′ overhangs. One of these 3′ overhangs can invade the uncut homolog, thus displacing a D-loop that can pair with the remaining 3′ overhang; the paired D-loop can act as a template for DNA synthesis, primed by the 3′ overhang. Of course, the other invading 3′ end should be also a primer for a DNA polymerase. These events of DNA synthesis repair the gap formed during initiation with information from the uncut homolog. Upon strand ligation, two Holliday junctions are created, which are able to migrate and extend the heteroduplex segment. As in the previous models, orientation of cutting of the Holliday junctions will result in crossover, gene conversion or both. According to the DSBR model, segments of gene conversion can be generated in two different ways. One alternative is through repair, by DNA synthesis, of the gap produced during initiation. The other is through mismatch repair of heteroduplex DNA. Thus, gene conversion segment length will depend on the extent of the gap, size of the heteroduplex region and the migration capacity of the Holliday structures.

A specific prediction of all these models is that gene conversion is associated, half of the time, with crossover. However, a wealth of data from yeast, both from meiotic [25–29] as well as mitotic [30–32] recombination, shows that gene conversion may occur at significant proportions without an associated crossover. A recent modification of the DSBR model, termed the synthesis-dependent strand annealing model (SDSA, Fig. 2) [33], deals nicely with this result. In this model, a double-strand cut is made in one DNA duplex; this double-strand break is then processed by degradation to generate protruding 3′ ends. One of these 3′ ends invades a homologous region and starts DNA synthesis using as template the homologous strand. So, a D-loop is formed as a consequence of strand displacing and DNA synthesis. The model then postulates that the newly synthesized strand may dislodge from the invaded duplex (conceivably by the action of helicases), making it available to pair with the other 3′ end in its original duplex. A full duplex is then restored by limited DNA synthesis. Gene

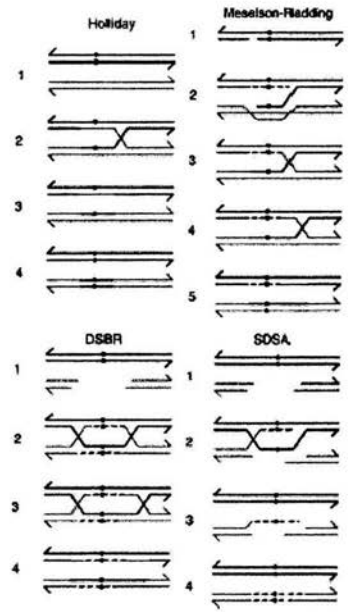

Fig. 2. Molecular models of recombination that explain gene conversion. Abbreviated versions of the Holliday, Meselson-Radding, DSBR (double-strand break repair) and SDSA (synthesis-dependent strand annealing) models are given. Numbers at the left of each block indicate successive steps in the corresponding model. Different shadings indicate each interacting homologs, with the arrowhead marking the 3′ end for each strand. DNA synthesis is indicated by broken lines. Black dots mark the position of sequence differences between the homologs.

conversion may arise in this model through DNA synthesis or through mismatch repair; however, in both cases, gene conversion is not associated with crossover.

Support for these models, especially the DSBR model, have been obtained through detailed analysis of meiotic recombination in ascomycete fungi, where recovery of all the products of a recombination event in a non-selective way is possible. Molecular analysis of these systems, plus the finding of enzymes that promote double-strand breaks on recombination hotspots, validate the general applicability of the DSBR model in eukaryotes [25–29,34,35]. However, the DSBR model is also a useful framework to explain recombination and gene conversion in prokaryotes. In fact, most of the proteins that play essential roles in recombination, such as for annealing and strand exchange (RecA), migration and resolution of the Holliday junction (RuvA, RuvB, RuvC, and RecG), to mention but a few, were first discovered in *Escherichia coli* [36].

Moreover, there is ample evidence of the role of double-strand breaks in promoting recombination in prokaryotes, including phage systems [35,37] and bacterial chromosomes as well [35,38–41]. These double-strand breaks are thought to be generated by progression of the replication fork over regions on the template that harbor DNA nicks as a product of DNA damage [42].

3. Experimental verification of gene conversion in bacteria

One particular caveat in working with gene conversion in prokaryotes is that, in general, it is difficult to ensure the recovery of the two products of a recombination event. This is an important limitation, because a possible gene conversion event may originate either from gene conversion or from two crossover events flanking the "converted" region. This last event is possible to occur during replication. Double crossover events occurring between sister strands may create a product similar to gene conversion. For this reason, many plausible gene conversion events have been dubbed as "apparent gene conversion events" [43]. In fact, the only group of organisms in which recovery of the products of a single recombination event is possible is the ascomycete fungi (such as yeast, *Neurospora* or *Sordaria*), where all the products of a single meiosis are enclosed in an ascus. In these, scoring if recombination occurs in a reciprocal or non-reciprocal way, as well as an evaluation of the number of crossover events is possible. Given the absence of such a nice biological trick as the ascus in most organisms, we have to resort to indirect strategies to make likely that the two products of a recombination event are recovered.

One of these, used frequently for analysis of gene conversion in prokaryotes, is based on the use of special substrates that undergo intragenomic recombination. Molecules harboring inverted repeats (either in small plasmids or in the chromosome) are preferred for this purpose, because it is easier to ascertain if a product arise through gene conversion and its association with crossover [43]. The common theme under this approach is that specific mutations present in only one repeat will appear in the two copies whenever a gene conversion event has occurred; if this event is associated with a reciprocal crossover, this would lead to the generation of an inversion of the intervening segment. Note that under the scenario of reciprocal recombination, the finding of an inversion makes more likely the recovery of the two products of a recombination event. Even in this case, the issue of gene conversion in bacterial genomes is far from settled. As we will see in this review, there are sev-

eral reports in which gene conversion has been adequately demonstrated. In other cases, it is impossible to ascertain if the reported event arose through gene conversion or by double recombination between sister strands.

Kobayashi and his colleagues made a clever use of molecules harboring inverted repeats in their studies on gene conversion in *E. coli*. They employed a plasmid system harboring two inverted, inactive copies of the neomycin/kanamycin resistance gene [38,44]. One of the copies has a deletion, 283 bp long, towards the 5′ end of the gene, while the other is inactivated by a 248 bp deletion near its 3′ end. Thus, recombination between the two inactive copies (either as crossover, gene conversion or both) is needed to reconstruct the *neo* gene, giving resistance to kanamycin. During the analysis of the results obtained with this system, it must be kept in mind that *E. coli* has three different pathways for recombination (RecBCD, RecF and RecE) that differ in the mode of initiation [45]. Two of these pathways (RecBCD and RecF) are active in wild-type strains, while expression of the RecE pathway requires of special activating mutations.

Upon introduction of this plasmid system into a wild-type *E. coli* strain, kanamycin-resistant derivatives were readily obtained. The participation of gene conversion in their generation was initially surmised because plasmids isolated from neo^r derivatives have replaced the deletion in one of the copies with wild-type information from the other copy, in some cases with an associated crossover leading to inversion. However, detailed analysis revealed that these products originated from multiple rounds of recombination, rather than from gene conversion. In this scenario, two plasmid molecules engage in an intermolecular crossover, thus generating a dimer. Intramolecular recombination in this dimer then generates monomers, which resemble gene conversion products because they are the result of *two* crossover events. Conclusive evidence for the existence of the dimer and the generation of apparent gene conversion monomers has been obtained [44]. Although the occurrence of gene conversion in a wild-type strain remained possible, its presence is obscured by the high recombination rate found in this strain.

Convincing evidence for gene conversion was found, however, when this plasmid system was introduced into mutant strains that have only a functional RecE pathway (*recBC sbcA*) or a RecF pathway (*recBC sbcBC*) [46]. In both cases, gene conversion events were obtained. Dimer formation was not observed in neither of these backgrounds and, in fact, an artificially formed dimer introduced in these strains was not resolved into apparent gene convertants [46]. Thus, gene conversion is generated by both the RecE and the RecF recombination pathways.

These two pathways, however, do not produce gene convertants in the same way. For the RecE pathway, in almost half of the products that displayed gene conversion, there was a crossover nearby [44,46], as expected according to the DSBR model. Similar results were also observed for gene convertants obtained through the action of the phage λ Red pathway for recombination [37]. Surprisingly, convertants arising through the RecF pathway were rarely, if ever, associated with crossovers [46]. Although the gene convertants associated with crossovers are clearly explained through the DSBR model, a majority of gene convertants not associated with crossovers was an unexpected finding.

To explain this finding, Kobayashi has argued, rather persuasively, for a novel kind of recombination model, called the half-crossing over model [46,47]. According to this model, recombination through the RecF pathway occurs most of the time as a half-crossover (that is, generating one recombinant DNA molecule out of two parental DNA molecules). The gene convertants without an associated crossover can be explained as a result of two successive half-crossover events [46,47].

This model makes the explicit prediction that recombination through the RecF pathway has to happen without the generation of two products from a recombination event (termed conservative recombination by Kobayashi), but generating only a single recombinant product (non-conservative recombination). Extensive studies using this system have shown that while recombination through the RecF pathway is non-conservative [39,46,48,49], recombination through both the RecE and the phage λ Red pathways proceed in a conservative (i.e. reciprocal) way [39,40,49]. There is conflicting evidence regarding the reciprocity of recombination in wild-type *E. coli* cells. While data from Kobayashi suggest that there is an appreciable amount of non-conservative recombination [39], other groups have found, using both chromosomal [50,51] as well as plasmid systems [52], that recombination proceeds in a reciprocal way. In fact, it has been suggested that recombination in wild-type cells occurs in a two-step fashion. In the first step (RecBCD independent), recombination proceeds in a non-conservative way, but in a second step, the presence of RecBCD stimulate the production of reciprocal crossovers [51].

These studies have shown that recombination through the RecF pathway generates gene convertants in a radically different way, i.e., through the formation of half-crossovers. In our opinion, gene convertants generated through both the RecE and the λ Red pathways, and perhaps also in wild-type cells, can be adequately explained through the DSBR model. Under this view, gene convertants not associated with crossovers may occur by normal resolution of the two Holliday intermediates or through the SDSA model, a variation of the DSBR model that deals adequately with this finding

[33]. More experiments are needed to ascertain the validity of the SDSA model in *E. coli*.

Another well studied example is concerned with the concerted evolution of duplicated *tufA* and *tufB* genes (which encode for the translation factor EF-Tu) in *Salmonella typhimurium* [53]. These two genes diverge 1% in nucleotide sequence and are located 700 Kb apart on the chromosome of *S. typhimurium*, in an inverted orientation. Detection of gene conversion is possible using special mutant alleles (*tufA8* or *tufB103*, both producing a change Ala375Thr); any of these two alleles give a kirromycin sensitive phenotype in the presence of a wild-type copy of the other *tuf* gene. A kirromycin-resistant phenotype is produced, however, upon sequence homogenization to a mutant allele (i.e. a gene conversion event) in both *tufA* and *tufB*. A similar rationale allows isolation of putative gene conversion events by differences in growth rate. Extent of sequence transfer by gene conversion can be evaluated by scoring the cotransfer of unselected sequence differences between *tufA* and *tufB*. Using this system, Abdulkarim and Hughes [53] detected transfer of sequence information between these two genes at rates of 10^{-8}–10^{-9} per cell division, depending of which copy acts as the donor. Mutations in recombination and repair genes affect gene conversion rates, either lowering (*recA*, *recB*) or increasing (*mutSLH*) them. The length of gene conversion tracts is variable, but can be almost the entire gene length (1182 bp). Such gene conversion events are closely associated with the generation of chromosomal inversions, linking, as expected from molecular models of recombination, gene conversion with crossover [54].

For the *tufA–tufB* model, it was also possible to provide compelling evidence for the occurrence of gene conversion rather than double crossovers (apparent gene convertants) as the cause for homogenization. In this system, homogenization of the two genes to the mutant sequence leads to a kirromycin (Kr)-resistant phenotype, while homogenization toward the wild-type sequence leads to a streptomycin (Sm)-resistant phenotype [55]. Thus, if reciprocal crossovers between sister molecules are responsible for homogenization, both Kr- and Sm-resistant derivatives should be found, at similar magnitudes, in individual tubes of fluctuation experiments aimed to detect homogenization. In contrast, Arwidsson and Hughes [55] found conclusive evidence in favor of a non-reciprocal transfer of these mutations, because the frequencies of Kr- and Sm-resistant derivatives were different in individual tubes of a fluctuation experiment by about one order of magnitude. Therefore, reciprocal recombination between both *tuf* genes in sister molecules is not the major cause for homogenization. Additionally, they sequenced both *tuf* genes and analyzed the gene conversion tracts and observed that the tracts were not similar in Kr- and Sm-resistant strains. Such experiments support gene conversion events as a mechanism of homogenization of the *tuf* genes in *Salmonella* [55].

Another plausible evidence for the occurrence of gene conversion has been reported for *Mycobacterium smegmatis* [56]. In this organism, resistance to aminoglycoside antibiotics (such as amikacin, gentamicin and tobramycin) is due to a recessive mutation in the 16S rRNA gene (at position 1408, A → G); since this organism contains two complete rRNA operons, aminoglycoside resistance is seen only at a very low frequency (10^{-11} per viable cell). To investigate the mechanisms involved in homogenization of this allele, an integrative vector carrying a mutated rRNA operon was introduced by transformation into a sensitive strain (1408^{wt}) that harbors a single copy of the rRNA operon. The vector integrates as a single copy into an ectopic site (*attB*) of the *M. smegmatis* genome. Initial integrants were sensitive to aminoglycosides, due to its heterozygotic nature ($1408^{wt}/1408^{mut}$). However, aminoglycoside-resistant derivatives arose at a very high frequency (10^{-4} per viable cell); these resistant derivatives were due to homogenization of the 1408^{mut} allele into the original rRNA operon. Interestingly, the homogenization process does not occur in a *recA* mutant strain, indicating the participation of homologous recombination in the conversion of this allele. These results are consistent with a non-reciprocal transfer of this recessive mutation by gene conversion between the two rRNA copies [56]. Unfortunately, the possible association of gene conversion with crossover was not explored in this work.

Artificial merodiploid strains have been also used in *Helicobacter pylori* to detect gene conversion. In this case, a truncated, silent copy of the *rpsL* gene (encoding the S12 ribosomal protein) was inserted into an ectopic site [57]. This silent copy also harbors a mutation that, when incorporated into the wild-type gene, confers streptomycin resistance. Derivatives resistant to streptomycin were isolated readily (at a frequency of 10^{-3} per viable cell) from this merodiploid; as expected, their isolation frequency is strongly reduced in a *recA* mutant background. Sequencing of both copies in streptomycin-resistant derivatives revealed homogenization of the mutation between the two copies. Since no hybrid genes, product of a putative reciprocal recombination event were found, these results are more consistent with a *recA*-dependent gene conversion event [57].

The genome of the nitrogen-fixing bacteria *Rhizobium etli* is an excellent model to study concerted evolution and gene conversion events because of the presence of many reiterated elements and multigene families [58]. One interesting example is the nitrogenase (*nifH*) family, comprised by three non-contiguous copies located on the symbiotic plasmid [59]. Nucleotide sequence of the three *nifH* copies is identical [60]. The nature of events that homogenize sequence differences was explored by introducing a 28 bp insertion into the *nifH* gene. This

insertion abolishes, by polarity, the expression of a *nifD*::kan gene fusion inserted downstream. Thus, events that lead to the loss of the 28 bp insertion can be easily scored by selection for resistance to kanamycin. Diverse recombination events, including gene conversion, maintain the sequence identity of the members of the family [61]. These events require of the existence of additional, wild-type copies of the *nifH* gene, as well as of a functional *recA* gene. However, at least in some cases, apparent gene conversion events might result from repeated reciprocal exchanges, because it was not possible to recover all the products of the recombination event.

To circumvent this difficulty, G. Santoyo and D. Romero (unpublished results) designed a two-plasmid crossover system, which allows the recovery of all the recombination products, making feasible the analysis of *bona fide* gene conversion events. Single-nucleotide RFLPs introduced every 100 bp along the *nifH* gene allow an evaluation of the size of the segment transferred by gene conversion. Interestingly, most of the crossovers were strongly associated with gene conversion (98%). The length of the tracts transferred varies from 150 to 800 bp, with a mean of 500 bp (half of the gene).

Gene conversion has also been found between paralogous genes (i.e. generated by an ancient duplication, see below) in the naturally transformable bacterium *Acinetobacter calcoaceticus* [62,63]. In this species, both the *catJ* and the *pcaJ* genes (located 20-kb apart on the chromosome) encode β-ketoadipate:succinyl CoA transferase; the *catJ* gene is required for growth on the xenobiotic compound catechol, while *pcaJ* is needed for growth on protocatechuate. These genes respond to different metabolic inducers, so inactivation of *pcaJ* precludes growth on protocatechuate, despite the presence of a wild-type *catJ* gene. As expected for genes that are the product of a duplication, they share 99% of identity. Interestingly, high-frequency reversion (10^{-5}) of a *pcaJ3125* mutation was observed only in the presence of a *catJ* gene [64]. Reversion frequency is reduced nearly 100-fold by deletion of the *catJ* gene [64] or by a mutation in *recA* [62]. High-frequency reversion is unaffected by deoxyribonuclease treatment, thus ruling out natural transformation as the cause of this phenomenon. Gene conversion in *pcaJ* revertants was ascertained by tracking the transfer of single-nucleotide differences from *catJ* into *pcaJ*; this revealed that apparent gene conversion tracts may range from less than 315 bp to more than 881 bp [63].

Data reviewed thus far describes evidence for intragenomic gene conversion among repeated zones on the one-kilobase size range. This repeat size, or even shorter, is perhaps the most frequent size for repeated sequences in bacterial genomes. However, extended repeat regions may appear as a product of the action of transposons or through the generation of long tandem duplications [7,8]. Roth and Segall [43] have studied the characteristics of gene conversion in *S. typhimurium* on instances when the recombining regions are large (either 5 or 40 kb) and located in an inverted orientation. The mutations to be converted were, in both cases, insertions of transposable elements.

Interestingly, the phenomenology of gene conversion departs from the one observed with one-kb repeats. Crossover between long repeated sequences (generating an inversion) is frequently accompanied by gene conversion (38–88% of the time); in contrast, crossover between shorter (5 kb repeats) is accompanied by gene conversion only in a minority of the cases (1–4%). This effect was seen upon a variety of chromosomal intervals. The formation of gene conversion associated with crossovers is clearly dependent on a functional *recA* gene. While inactivation of *recB* reduces inversion formation and associated gene conversion, it preserves a class of convertants, dubbed "apparent gene convertants" in which conversion is not associated with crossovers [43].

These data are clearly different to those obtained in previous assays that analyze gene conversion using sequences in the one-kb range. Although these differences might be explained by assuming an enhanced frequency of multiple recombination events in repeats in the 5–40 kb range, there are other, perhaps more interesting alternatives. One of these is that this kind of substrates, by virtue of the long heteroduplex sequences needed to cover these insertions, have different enzymatic requirements for its processing. Another alternative is that these long substrates are differentially sensitive to local variations in conformation, imposed by loops on the bacterial chromosome. More work is needed to establish the generality of these findings, as well as to try to correlate this with possible landmarks in genome architecture.

4. Molecular evolutionary inference as an aid to detect gene conversion

The accessibility to complete genome sequences of over 100 different bacteria is paving the way to novel approaches to study bacterial genome evolution. Analysis of whole genome sequences has revealed an unexpected abundance of repeated sequences and multigene families in prokaryotes [4–6]. Specially, short, identical repeated sequences (aprox. 300 bp long) are very common of some bacterial genomes [6]. More germane to the purpose of this review, studying gene conversion at the genomic level is now feasible, due to the availability of powerful computer programs and statistical tests able to detect the occurrence of gene conversion events in multigene families [65–67].

All the programs designed to detect gene conversion rely on an analysis on the extent of similarity between

orthologs and paralogs. The distinction between orthologs and paralogs is crucial to this end. Orthologous genes are linked by descent, having shared a common ancestor in the course of its evolution. Paralogous genes, in contrast, are commonly generated by duplication, sharing a common ancestor only at the time of the duplication [68,69] (Fig. 3(a)). A different, perhaps less frequent route for the generation of a paralog gene, would be by horizontal acquisition of an ortholog from another species. These sequences are sometimes called xenologs, and pose special problems for their detection and characterization [69]. Sequence comparisons between the members of a given multigene family within a strain are mostly comparisons between paralogs. Ortholog comparisons are only those that compare the sequence of a given gene (X) with the one that is linked by descent (X′).

In principle, for a multigene family generated by ancient duplication events, paralogous comparisons should exhibit the same overall degree of divergence than orthologous comparisons. This would be the most likely result if evolution of a given copy occurs independently of the other copies (Fig. 3). However, for many cases of multigene families in eukaryotes, paralogous comparisons tend to be more similar than orthologous comparisons (Fig. 3). This is tantamount to say that copies of a multigene family within an organism are evolving in a non-independent way, a process that it is known as concerted evolution. Although there are several molecular mechanisms that can explain the process of concerted evolution (such as repeated transposition and unequal crossing over), gene conversion has gained wide acceptance because only requires a high sequence similarity between the sequences to be homogenized. Moreover, homogenized sequences can be generated without the production of rearrangements, a by-product of homogenization via unequal crossover.

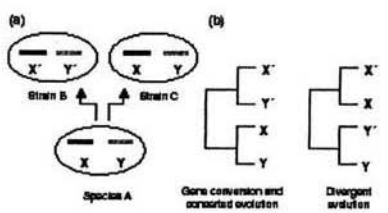

Fig. 3. Phylogenetic analysis can reveal concerted vs. divergent evolution. In strain B, X′ and Y′ genes are paralogous, the same situation applies for strain C. Genes (in both strains B and C) with the same shading are orthologous (a). Paralogous genes can take two different routes through time, concerted or divergent evolution (b). These are distinguishable because, under concerted evolution, paralogs within a given strain retain a high identity. Under divergent evolution, paralogs within a strain become diverse in sequence.

Up to now, there are many examples of concerted evolution of multigene families in prokaryotes. Repeated sequences, such as the REP sequences in *E. coli* [70], were initially proposed to keep homogeneity in sequence through gene conversion. Multigene families encoding essential cell components, however, have been prime targets for this type of analysis, since they are widespread in bacteria. One typical example refers to the ribosomal genes (rRNA) in bacteria and archaea [71]. Although all the ribosomal genes of bacteria and archaea are similar between genera, a higher similarity is seen for comparisons between copies within a given species (paralogous comparisons), a situation strongly suggestive of concerted evolution. Since frequent transposition has not been reported for rRNA genes, and given their scattered location in the genomes analyzed, gene conversion emerges as the most likely mechanism for sequence homogenization. Based on the distribution of similar regions between these copies, gene conversion tracts are short (less than 500 bp); moreover, a patchy distribution of the conversion segments was commonplace in these cases [71].

Another interesting example pertains to the evolution of the elongation factor protein EF-Tu, which is encoded by the *tuf* genes. The *tuf* genes are duplicated in several bacteria, including *E. coli*, *Haemophilus influenzae*, *Vibrio cholerae*, *Pseudomonas aeruginosa*, *Neisseria meningitidis*, *Deinococcus radiodurans* [72] and *S. typhimurium* [53]. Although the presence of duplicate *tuf* genes is not universal among bacteria, genomic neighborhood analysis supports the interpretation that these repeats are the product of an ancient duplication, followed by differential loss or maintenance of the duplicates [72]. The high identity between the duplicates within a species, coupled with the divergence seen in orthologous comparisons, indicates the occurrence of concerted evolution through gene conversion for this family [72]. In fact, as mentioned before, the action of gene conversion between members of this family has been supported in *S. typhimurium* [53].

Molecular evolutionary inference is also useful to detect patterns of concerted evolution at the intraspecies level. Analyses of this kind have unraveled a complex evolutionary pattern for the *babA* and *babB* genes, which belong to a family of outer membrane proteins in *H. pylori* [57]. Upon sequencing the *babA* and *babB* genes from 23 *H. pylori* strains from different origins worldwide, a segmental pattern of concerted evolution was found for the 3′ region of these genes, but not in the 5′ or middle zones. Additionally, it was found that the nucleotide substitution frequencies (NSF) is different for each gene segment, with the NSF for the 3′ segment being significantly lower than the overall mean substitution frequency. Moreover, the NSFs for this segment

were also lower for comparisons within the same strain than between strains. All these observations are consistent with a non-independent evolution of this gene family, most likely through gene conversion. The experimental demonstration of the occurrence of gene conversion in this species supports this interpretation.

Although it is clear that the 3′ segment is evolving through concerted evolution, a recent article [73] gives an unexpected turn to this story. Upon experimental infection of rhesus monkeys with *H. pylori* J166, it was found that most of the strains have lost the ability to express the BabA protein. Most of these strains have modifications that inactivate the promoter region of *babA*, but other show a pattern indicative of gene conversion in the *babA* locus, using *babB* information as the donor. Interestingly, the segment transferred from *babB* corresponds to the middle region, not to the 3′ region [73]. It is thought that the increased expression of *babB*, as a consequence of gene conversion, increases adherence to the gastric epithelium, thus favoring chronic infection. Thus, even if this particular gene conversion event is rare in the population, a strong positive selection within the host may allow its predominance.

An analysis of the distribution of conserved segments in multigene families has also been done for genes encoding outer membrane proteins in *H. pylori* and *Chlamydia pneumoniae* [68]. In this work, the authors took advantage of the existence of complete genome sequences for two strains of each species; this allowed the possibility of clear ortholog and paralog comparisons. Phylogenetic analysis revealed multiple gene conversion events in several members of the Hop family of outer membrane proteins in *H. pylori*. However, for the *C. pneumoniae* strains, convincing evidence for gene conversion was only found for two paralog genes encoding predicted outer membrane proteins.

Another instance of concerted evolution at the intraspecies level has been reported for *Campylobacter jejuni* [74]. Strains of this species are the ethiological agent of important diarrhoeal diseases; this species harbor two highly similar copies of the flagellin gene (*flaA* and *flaB*). Upon sequencing and phylogenetic analysis of these genes in 16 strains from *Campylobacter*, clear evidence was found for concerted evolution. Interestingly, regions of similarity between these genes are segmental [74], an observation clearly expected from the phenomenology of gene conversion.

In the nitrogen-fixing bacterium *R. etli*, it has been suggested that the identity between the members of the *nifH* multigene family may be due to concerted evolution. Besides the evidence mentioned above for gene conversion among the members of this family, the pattern of similarity among the three *nifH* members from 11 strains isolated from diverse world regions support the existence of concerted evolution (E. Sepúlveda and D. Romero, unpublished results).

All the instances of concerted evolution analyzed here share the implicit assumption that the homogenization process occurs by interaction of endogenous gene copies. It is possible, however, that lateral gene transfer plays a role in this process. In this view, a variant gene copy is acquired through gene transfer; this variant information may be spread then to the endogenous copies through gene conversion. An interesting example in this regard is posed by the evolution of the genes for 23S rRNA (*rrl*) in *S. typhimurium*. In the type strain of this organism (LT2), there are seven rrl genes; these genes are interrupted by at least an intervening sequence (IVS), that is excised from the transcript by RNase III during rRNA maturation [75]. Despite the fact that this process results in a fragmented 23S rRNA, this is still functional. Presence of IVSs in genes for 23S rRNA has been reported in fourteen different bacterial genera [76].

For *S. typhimurium* LT2, two of the genes (*rrlG* and *rrlH*) share a 110 bp IVS called helix 25; these shared an identity of 56%. Six of the genes (*rrlA-E* and *rrlH*) possess a 90 bp IVS (helix 45) that is identical among all the copies. An analysis of the *rrl* genes in 21 natural isolates showed that, although many isolates were identical to *S. typhimurium* LT2, some have the 90 bp IVS in all the seven genes, and lack the 110 bp IVS [75]. This 90 bp IVS is at least 80% identical to the IVS from *S. arizonae* and *Yersinia enterocolytica* group 2 [75], suggesting that these have a common evolutionary origin [77]. These data suggest that *S. typhimurium* acquired the 90 bp IVS (and perhaps the 110 IVS) through lateral transfer; the IVS was then spread to the remaining *rrl* copies by gene conversion, unequal crossover with a sister molecule or both [75]. A similar scenario has been proposed for the IVS of *S. typhi* [78].

5. Role of gene conversion for the generation of antigenic variation

Many bacterial pathogens have developed antigenic variation systems to avoid or escape from host immune systems [10]. *Neisseria gonorrhoeae* and *N. meningitidis* [79], *Mycoplasma synoviae* [80], *Anaplasma marginale* [11], *Borrelia burgdorferi* [81], *Borrelia hermsii* [82] and *C. jejuni* [83] are some examples of pathogens that avoid clearance by the immune system through the use of antigenic variation. Survival in all the cases studied relies on a rapid switching in the type of outer membrane antigens produced, thus avoiding elimination. Diverse genetic mechanisms have been invoked to explain the molecular basis of antigenic variation, including modifications of transcriptional levels, genomic rearrangements and high rates of point mutation [10]. Although all these mechanisms participate in variation, the greatest richness in alternatives is reached through the use of

combinatorial gene conversion (Fig. 4). In most of the cases, there is a single gene that encodes an outer membrane antigen (the expressed sequence); variation in the kind of antigen expressed is achieved through the unidirectional transfer of information (i.e., gene conversion) from a large series of variant, unexpressed gene cassettes into the expressed site. Use of combinatorial gene conversion allows a successful balancing between the need of variation and the need of preserving some parts of the gene invariant, in order to permit proper biological function.

Pilin variation in the neisseriales, such as *N. gonorrhoeae* and *N. meningitidis*, was one of the first systems for antigenic variation described (see [84] for a review). For *N. gonorrhoeae*, a functional pilin protein is encoded by the *pilE* locus; this expressed locus occurs in one or two copies depending of the strain [85–87]. The *pilE* locus has a 5′ region that is constant among isolates; in contrast, the central part is semivariable, while the 3′ region is hypervariable [85–87]. Besides the expressed loci, *N. gonorrhoeae* has a repertoire of 16–19 silent pilin sequences called *pilS* cassettes; these loci have a complex structure, each including six semivariable or hypervariable regions, termed minicassettes. Antigenic variation occurs by gene conversion from the minicassettes in the *pilS* loci into the *pilE* locus; since antigenic variation was resistant to DNase in the culture medium, alternative interpretations such as intercell recombination mediated by transformation were excluded [79,85]. Interestingly, since each minicassette locus can participate in up to six gene conversion events, the number of potential antigen variants is enormous (19^6 or roughly 47 million variants) [88].

Although not all of these variants may represent functional proteins, the potentiality of this combinatorial strategy is vast. In fact, during studies of experimental human infection with *N. gonorrhoeae*, 11 novel antigenic variants were detected as early as two days after infection [87]. In most of these cases, the *pilE* sequences represented chimeras of minicassettes derived from up to three different *pilS* loci [87]. The fast variability observed is possible due to the high frequency of gene conversion even during in vitro conditions, which has been determined, by quantitative RT-PCR, at 3.3×10^{-2} [88]. Although a clear environmental control of antigenic variation has been difficult to demonstrate, the gene conversion frequency is 2-fold higher upon transition into late log/stationary phase [88], and is enhanced 10-fold upon growth in iron-depleted media [89], by unknown mechanisms.

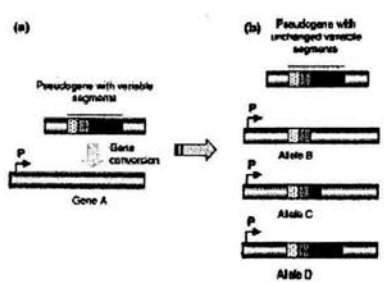

Fig. 4. Combinatorial gene conversion participates in antigenic variation. A pseudogene may have several variable cassettes, represented by squares with different stippling patterns. Cassettes can be transferred by gene conversion into an expressed gene copy (a). Alleles with different combinations of variable cassettes are generated depending on the strain, leaving the pseudogene unchanged (b).

Genes participating in homologous recombination play an important role in gonococcal antigen variation. Mutations affecting strand exchange (either in *recA* or in its modulator, *recX*, [90,91]), invasion through the RecF pathway (*recJ*, *recO*, *recQ* or *recR* [84,92,93]), or branch migration (*ruvA* or *recG*, [84]) all reduce the frequency of antigenic variation. Enhancements in antigenic variation have been detected, for *N. gonorrhoeae*, upon mutation of *recD* [94] or in *recB* (for *N. meningitidis* [95]). Although the last findings have been difficult to reproduce [84], these data are consistent with antigenic variation occurring through the RecF pathway of recombination, while the RecBCD pathway may be removing substrates used for antigenic variation [84].

Homologous recombination through the RecF pathway, however, does not suffice to explain the phenomenology of antigenic variation in the neisseriales. The high frequency of antigenic variation, coupled with the fact that gene conversion events may start with as few as 11 bp of fully homologous sequences, suggests that there may be specialized systems responsible for initiation. So far, these systems have resisted detailed characterization.

The detailed mechanism of gene conversion is also another potential avenue for research. It has been reported that recombination between repeated *pilS* sequences may generate circular structures that harbor a hybrid *pilS* locus [13]. These circles may promote transfer via gene conversion into the *pilE* locus. However, as noted before [84], the frequency of circle formation is rather low compared to the frequency of antigenic variation (10^{-5} vs. 10^{-2}), making circles unlikely intermediates for this process. Moreover, the antigenic variation system in the neisseriales shows a clear bias towards gene conversion, as opposed to gene conversion with crossover. This bias has previously been observed for other instances of gene conversion through the RecF pathway [46]. As noted previously, this bias may be explained by alternative models, such as the half-crossing over model or the SDSA model [33]. Presently, a distinction about which model is operating is still lacking.

A similar mechanism of antigenic variation is seen for the spirochaetes *B. burgdorferi* and *B. hermsii*. In *B. burgdorferi* (the causal agent of Lyme disease), the system responsible for the diversity in antigenic types observed for the surface-exposed lipoprotein is VlsE. Production of this lipoprotein is dependent on a single expressed locus, *vlsE*, located in a 28 kb linear plasmid; upstream of this locus, there are 15 silent *vls* cassettes [81,96]. Antigenic variants of the VlsE protein are readily produced during mice infection; most of the variations are restricted to the central part of the protein. Gene conversion events between the silent copies and the expressed *vlsE* locus are responsible for this variability. As in the case of the neisseriales, gene conversion events occur in a combinatorial way, where a given antigenic variant may be the product of 6–11 segmental conversion events [81,96]. Previous work has also shown the occurrence of combinatorial gene conversion for the *vmp* genes in *B. hermsii* [82,97].

Another interesting case occurs in *A. marginale*, which is a tick-transmitted ehrlichial pathogen responsible for anaplasmosis in mammalian hosts around the world [98]. Successful avoidance of the immune system is achieved by antigenic variation in the outer membrane proteins encoded by the *msp2* family. In this organism, there is a single expressed gene, termed *msp2* as well as 9 *msp2* pseudogenes, equivalent to the silent sequences in the neisseriales. Variation is achieved by combinatorial gene conversion from the silent pseudogenes into the expressed sequence [11,12,99]. Up to four sequential changes have been detected in the expressed gene; therefore, the potential of variation in this system is quite high (9^4 or 6561 potential variants [11]). This number greatly exceeds the number of variants (500) thought to be needed to allow a persistent, lifelong infection [100]. Interestingly, in *A. marginale* there is a separate system, termed *msp3*, which also engages in gene conversion leading to antigenic variation [99]. The role of *msp3* variation in immune system avoidance is currently unknown.

Combinatorial gene conversion has also been identified as one of the mechanisms participating in antigenic variation in *M. synoviae*, a pathogen for poultry. In this organism, MSPB (a lipoprotein) and MSPA (a haemagglutinin) are the main surface antigens; these display a high antigenic diversity [101]. Both antigens are encoded by a single gene (*vlhA1*), which undergoes post-translational cleavage and modification to yield MSPB and MSPA [102]. Although both the lipoprotein and the haemagglutinin sectors of the gene experience antigenic variation, combinatorial gene conversion is restricted to the haemagglutinin part. In this case, the *vlhA1* locus functions as the expressed gene; there are at least other eight pseudogenes (*vlh2*–*vlh9*) that participate as donors in combinatorial gene conversion towards the expression locus [80]. This process may generate 120 potential MSPA variants [80]; this is a conservative estimate, since the size of the full complement of *vlh* pseudogenes has not been determined yet. Although the exact frequency of antigenic variation for this protein has not been determined, it is frequent enough to be detected as sectored colonies that synthesize a variable antigen [101].

The most recent example of combinatorial gene conversion as a source of heterogeneity was found in *Treponema pallidum*, the ethiological agent of human syphilis [103]. In this organism, there is a large variation in the structure of the major antigen encoded in the *tprK* gene. The expressed *tprK* gene has seven variable regions; there are also 47 pseudogene sequences, corresponding to these variable regions, which may function as donors in combinatorial gene conversion. Although it is not possible to evaluate the frequency of variation, the dynamics of the process has been followed during experimental infection in rabbits. TprK variants were readily isolated; structural analyses of these clearly show that variants were generated by gene conversion. Assuming only a one-to one interaction between the pseudogenes and variable sites in *tprK*, this process would generate at least 420 000 variants; of course the number would be higher if a combinatorial rule is used.

Data mentioned before illustrate the role of gene conversion on antigenic variation of multiple gene families in some bacterial pathogens. Additionally, it is shown that repeats are used as a reservoir to produce new gene variants, important to survive and stay within the host. Besides this, gene conversion is also playing an important role on the evolution gene families and genome architecture, essential for adaptation of variable environments.

6. Concluding remarks and perspectives

As reviewed above, experimental data from a variety systems have succedeed in demonstrating the feasibility of gene conversion in bacteria. This was achieved, despite the important limitation of being unable to recover both products of a recombination event, by exploring recombination between repeated sequences, especially those in an inverted orientation. Although the most detailed studies are still concentrated on *E. coli* and *S. typhimurium*, it can be anticipated that experimental verification of gene conversion will extend to representatives of other bacterial groups.

One important aspect that remains to be clarified regards the exact mechanism for gene conversion. Even if the current evidence favors the DSBR model, several instances of gene conversion, especially those that do not show association with crossovers, could be better explained through the SDSA model. Although the enzymatic requirements for the SDSA model are entirely likely to be fullfilled in bacteria, there are no

demonstrations, to our knowledge, of its operation in bacteria. This demonstration is going to be crucial to understand the operation of the RecF pathway, prevalent in most bacteria, where gene conversion apparently proceeds with a weak association with crossovers.

More work is also needed to understand the basis of the differences in gene conversion for long vs. short repeats. In the only case analyzed thus far, gene conversion proceeds through a long insertion (5 kb) immersed in the context of an even larger repeat (40 kb). This might demand the generation of a very long heteroduplex region; although this may occur in bacteria [104] their requirements are still poorly understood.

On evolutionary terms, gene conversion fulfills two opposing roles in bacteria: concerted evolution and gene diversification (antigenic variation). These seemingly opposed roles in fact reflect the operation of gene conversion at different stages of evolution in a multigene family. In the first stage, after the generation of a duplicated sequence (either through intragenomic events or by horizontal transfer) homogenization through gene conversion is likely to predominate. This is entirely possible because, in general, gene conversion proceeds at higher frequency than spontaneous mutation. At this stage, the low or null divergence between the repeats allows the efficient operation of the homologous recombination system. So, the general tendency of duplicate sequences would be concerted variation.

The system, however, is not foolproof. Despite the existence of efficient systems for gene conversion, some of the duplicates would start to diverge in nucleotide sequence, due to random genetic drift. Even small proportions of sequence divergence would reduce severely the rate of homologous recombination, through the action of the *mutS* system, the most efficient "editor" of homologous recombination in bacteria [105,106]. Reducing the rate of gene conversion then enhances the diversification rate of the duplicates, thus increasing the probability of functional diversification or even of extinction of the duplicate.

From this point of view, events that restore a high rate of gene conversion even with divergent duplicates would ultimate reset the state of affairs to one of concerted variation. One of these events would be the inactivation of the *mutS* function. One of the phenotypes of this kind of mutant, besides a general increase in mutability, is an enhancement in the rate of gene conversion, even with divergent duplicates [53,105,106]. Natural mutants in *mutS* may constitute a high proportion of the natural isolates in *E. coli* [107,108] *Salmonella* [107], *P. aeruginosa* [109] and *Staphylococcus aureus* [110]. The natural occurrence of these isolates has been rationalized by proposing that their enhanced mutability allows the bacteria to cope with variant environments [111,112], or as ideal recipients for horizontal transfer events [113]. However, as a collateral effect, these mutants would allow the restoration of concerted variation between diverging duplicates. Restoration of concerted variation does not require, however, the permanent inactivation of *mutS*. Environmental conditions, such as treatment with a high mutagen dose would generate *mutS* phenocopies [114], that should be the origin of lineages displaying concerted variation.

Even faced with extensive divergence between duplicates, functional diversification or extinction are not inescapable fates for members of a multigene family. High-frequency gene conversion interactions are commonly seen among divergent copies, as happens in all systems for antigenic variation. Although the nature of the enzymatic systems that allow selectivity and directionality towards the expressed sequences remains obscure, these systems appear to have evolved by a combination of site-specific and homologous recombination steps. In this scenario, the systems that allow specific interaction among a expressed sequence and their cognate pseudogene sequences should be the product of recruitment of site-specific recombination systems for a novel purpose, namely antigenic variation. Extensive molecular diversity among these systems can be anticipated, in order to fulfill the need of a more specific interaction. The nature of these systems is eagerly awaited in the field.

The current revolution in bacterial genomics is generating a treasure trove from where data mining for novel examples of concerted evolution would be possible. The current emphasis, as analyzed in this review, is to look for examples of repeated genes that maintain a high degree of sequence identity. These studies are greatly aided by the existence of more than 150 completely sequenced bacterial genomes. In several cases, there are complete sequences of two or more isolates from a single species, thus facilitating the study of concerted variation at the intraspecific and interspecific levels. These studies need to be complemented by careful screening of examples of a "complete gene-divergent pseudogenes" arrangement, likely candidates for systems analogous to antigenic variation. So far, these systems appear to predominate in pathogenic bacteria, where increased variability gives a much needed leverage to survive in the complex environment of a mammalian host. This need for increased variability is not restricted to pathogens; in fact, is a desirable characteristic to thrive in a variety of ecological niches. Thus, it is not far-fetched to expect the discovery of similar systems in saprophytic or even symbiotic bacteria.

Acknowledgments

We are grateful to Rafael Camacho-Carranza and two anonymous reviewers for helpful comments on the manuscript. Work in our laboratory is partially sup-

ported by Grant No. 31753-N (CONACyT, México). G.S. was supported by scholarships from CONACyT, México and Dirección General de Estudios de Posgrado, UNAM.

References

[1] Marians, K.J. (2004) Mechanisms of replication fork restart in *Escherichia coli*. Philos. Trans. R. Soc. Lond. B 29, 71–77.

[2] Kuzminov, A. (2001) DNA replication meets genetic exchange: chromosomal damage and its repair by homologous recombination. Proc. Natl. Acad. Sci. USA 98, 8461–8468.

[3] Michel, B., Flores, M.J., Viguera, E., Grompone, G., Seigneur, M. and Bidnenko, V. (2001) Rescue of arrested replication forks by homologous recombination. Proc. Natl. Acad. Sci. USA 98, 8181–8188.

[4] Rocha, E.P., Danchin, A. and Viari, A. (1999) Functional and evolutionary roles of long repeats in prokaryotes. Res. Microbiol. 150, 725–733.

[5] Rocha, E.P., Danchin, A. and Viari, A. (1999) Analysis of long repeats in bacterial genomes reveals alternative evolutionary mechanisms in *Bacillus subtilis* and other competent prokaryotes. Mol. Biol. Evol. 16, 1219–1230.

[6] Romero, D., Martínez-Salazar, J., Ortiz, E., Rodríguez, C. and Valencia-Morales, E. (1999) Repeated sequences in bacterial chromosomes and plasmids: a glimpse from sequenced genomes. Res. Microbiol. 150, 735–743.

[7] Romero, D. and Palacios, R. (1997) Gene amplification and genomic plasticity in prokaryotes. Annu. Rev. Genet. 31, 91–111.

[8] Hughes, D. (2000) Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. Genome Biol. 1, reviews 0006.1–0006.8.

[9] Flores, M., Mavingui, P., Perret, X., Broughton, W.J., Romero, D., Hernández, G., Dávila, G. and Palacios, R. (2000) Prediction, identification, and artificial selection of DNA rearrangements in *Rhizobium*: toward a natural genomic design. Proc. Natl. Acad. Sci. USA 97, 9138–9143.

[10] Deitsch, K.W., Moxon, E.R. and Wellems, T.E. (1997) Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. Microbiol. Mol. Biol. Rev. 61, 281–293.

[11] Brayton, K.A., Palmer, G.H., Lundgren, A., Yi, J. and Barbet, A.F. (2002) Antigenic variation of *Anaplasma marginale msp2* occurs by combinatorial gene conversion. Mol. Microbiol. 43, 1151–1159.

[12] Barbet, A.F., Lundgren, A., Yi, J., Rurangirwa, F.R. and Palmer, G.H. (2000) Antigenic variation of *Anaplasma marginale* by expression of MSP2 mosaics. Infect. Immun. 68, 6133–6138.

[13] Howell-Adams, B. and Seifert, H.S. (2000) Molecular models accounting for the gene conversion reactions mediating gonococcal pilin antigenic variation. Mol. Microbiol. 37, 1146–1158.

[14] Dover, G. (2002) Molecular drive. Trends Genet. 18, 587–589.

[15] Belle, A., Landthaler, M. and Shub, D.A. (2002) Intronless homing: site-specific endonuclease Seg P of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. Genes Dev. 16, 351–362.

[16] Liu, Q., Belle, A., Shub, D.A., Belfort, M. and Edgell, D.R. (2003) SegG endonuclease promotes marker exclusion and mediates co-conversion from a distant cleavage site. J. Mol. Biol. 334, 13–23.

[17] Landthaler, M., Lau, M.C. and Shub, D.A. (2004) Group I intron homing in *Bacillus* phages SPO1 and SPβ: a gene conversion event initiated by a nicking homing endonuclease. J. Bacteriol. 186, 4307–4314.

[18] Stahl, F.W. (1979) Genetic Recombination: Thinking About it in Phage and Fungi. Freeman, New York, 333 pp..

[19] Leach, D.R.F. (1996) Genetic Recombination. Blackwell Science, London, 192pp.

[20] Holliday, R. (1964) A mechanism for gene conversion. Genet. Res. 5, 282–304.

[21] Stahl, F. (1994) The Holliday junction on its thirtieth anniversary. Genetics 138, 241–246.

[22] Meselson, M.S. and Radding, C.M. (1975) A general model for genetic recombination. Proc. Natl. Acad. Sci. USA 72, 358–361.

[23] Alani, E., Reenan, R.A. and Kolodner, R.D. (1994) Interaction between mismatch repair and genetic recombination in *Saccharomyces cerevisiae*. Genetics 137, 19–39.

[24] Szostak, J., Orr-Weaver, T., Rothstein, R. and Stahl, F. (1983) The double-strand-break repair model for recombination. Cell 33, 25–35.

[25] Stahl, F. (1996) Meiotic recombination in yeast: coronation of the double-strand-break repair model. Cell 87, 965–968.

[26] Petes, T.D. (2001) Meiotic recombination hotspots and coldspots. Nat Rev. Gen. 2, 360–369.

[27] Keeney, A., Ben-Aroya, S. and Kupiec, M. (2002) Control of meiotic recombination initiation: a role for the environment? Curr. Genet. 42, 129–139.

[28] Wolf-Dietrich, H. (2004) Recombination: Holliday junction resolution and crossover formation. Curr. Biol. 14, R56–R58.

[29] Bishop, D. and Zickler, D. (2004) Early decision: meiotic crossover interference prior to stable strand exchange and synapsis. Cell 117, 9–15.

[30] Aguilera, A., Chavez, S. and Malagón, F. (2000) Mitotic recombination in yeast: elements controlling its incidence. Yeast 16, 731–754.

[31] Prado, F., Cortes-Ledesma, F., Huertas, P. and Aguilera, A. (2003) Mitotic recombination in *Saccharomyces cerevisiae*. Curr. Genet. 42, 185–198.

[32] Ira, G., Malkova, A., Liberi, G., Foiani, M. and Haber, J.E. (2003) Srs2 and Sgs1-Top3 suppress crossovers during double-strand break repair in yeast. Cell 115, 401–411.

[33] Allers, T. and Lichten, M. (2001) Differential timing and control of noncrossover and crossover recombination during meiosis. Cell 106, 47–57.

[34] West, S.C. (2003) Molecular views of recombination proteins and their control. Nat. Rev. Mol. Cell Biol. 4, 435–445.

[35] Cromie, G.A., Connelly, J.C. and Leach, D.R. (2001) Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. Mol. Cell 8, 1163–1174.

[36] Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D. and Rehrauer, W.M. (1994) Biochemistry of homologous recombination in *Escherichia coli*. Microbiol. Rev. 58, 401–465.

[37] Takahashi, N. and Kobayashi, I. (1990) Evidence for the double-strand break repair model of bacteriophage lambda recombination. Proc. Natl. Acad. Sci. USA 87, 2790–2794.

[38] Kobayashi, I. and Takahashi, N. (1983) Double-stranded gap repair of DNA by gene conversion in *Escherichia coli*. Genetics 119, 751–757.

[39] Takahashi, N.K., Kusano, K., Yokochi, T., Kitamura, Y., Yoshikura, H. and Kobayashi, I. (1993) Genetic analysis of double-strand break repair in *Escherichia coli*. J. Bacteriol. 175, 5176–5185.

[40] Yokochi, T., Kusano, K. and Kobayashi, I. (1995) Evidence for conservative (two-progeny) DNA double-strand break repair. Genetics 139, 5–17.

[41] Handa, N. and Kobayashi, I. (2003) Accumulation of large non-circular forms of the chromosome in recombination-defective mutants of *Escherichia coli*. BMC Mol. Biol. 4, 5.

[42] Kuzminov, A. (2001) Single-strand interruptions in replicating chromosomes cause double-strand breaks. Proc. Natl. Acad. Sci. USA 98, 8241–8246.
[43] Segall, A.M. and Roth, J.R. (1994) Approaches to half-tetrad analysis in bacteria: recombination between repeated, inverse-order chromosomal sequences. Genetics 136, 27–39.
[44] Yamamoto, K., Yoshikura, H., Takahashi, N. and Kobayashi, I. (1983) Apparent gene conversion in an *Escherichia coli rec*⁺ strain is explained by multiple rounds of reciprocal crossing-over. Mol. Gen. Genet. 212, 393–404.
[45] Clark, A.J. (1991) rec Genes and homologous recombination proteins in *Escherichia coli*. Biochimie 73, 523–532.
[46] Yamamoto, K., Kusano, K., Takahashi, N.K., Yoshikura, H. and Kobayashi, I. (1992) Gene conversion in the *Escherichia coli* RecF pathway: a successive half crossing-over model. Mol. Gen. Genet. 234, 1–13.
[47] Kobayashi, I. (1992) Mechanisms for gene conversion and homologous recombination: the double-strand gap and break repair model and the successive half crossing-over model. Adv. Biophys. 28, 81–133.
[48] Takahashi, N.K., Yamamoto, K., Kitamura, Y., Luo, S.Q., Yoshikura, H. and Kobayashi, I. (1992) Nonconservative recombination in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 89, 5912–5916.
[49] Yamamoto, K., Takahashi, N., Fujitani, Y., Yoshikura, H. and Kobayashi, I. (1996) Orientation dependence in homologous recombination. Genetics 143, 27–36.
[50] Mahan, M.J. and Roth, J.R. (1988) Reciprocality of recombination events that rearrange the chromosome. Genetics 120, 23–35.
[51] Mahan, M.J. and Roth, J.R. (1989) Role of recBC function in formation of chromosomal rearrangements: a two-step model for recombination. Genetics 121, 433–443.
[52] Lovett, S.T., Hurley, R.L., Sutera Jr., V.A., Aubuchon, R.H. and Lebedeva, M.A. (2002) Crossing over between regions of limited homology in *Escherichia coli*: RecA-dependent and RecA-independent pathways. Genetics 160, 851–859.
[53] Abdullarim, F. and Hughes, D. (1996) Homologous recombination between the *tuf* genes of *Salmonella typhimurium*. J. Mol. Biol. 260, 506–522.
[54] Hughes, D. (2000) Co-evolution of the *tuf* genes links gene conversion with the generation of chromosomal inversions. J. Mol. Biol. 297, 355–364.
[55] Arvidson, O. and Hughes, D. (2004) Evidence against reciprocal recombination as the basis for *tuf* gene conversion in *Salmonella enterica* serovar *typhimurium*. J. Mol. Biol. 338, 463–467.
[56] Prammananan, T., Sander, P., Springer, B. and Bottger, E.C. (1999) RecA-mediated gene conversion and aminoglycoside resistance in strains heterozygous for rRNA. Antimicrob. Agents Chemother. 43, 447–453.
[57] Pride, D.T. and Bluser, M.J. (2002) Concerted evolution between duplicated genetic elements in *Helicobacter pylori*. J. Mol. Biol. 316, 629–642.
[58] Flores, M., Gonzalez, V., Brom, S., Martinez, E., Piñero, D., Romero, D., Dávila, G. and Palacios, R. (1987) Reiterated DNA sequences in *Rhizobium* and *Agrobacterium* spp. J. Bacteriol. 169, 5782–5788.
[59] Quinto, C., de la Vega, H., Flores, M., Fernández, L., Ballado, T., Soberón, G. and Palacios, R. (1982) Reiteration of nitrogen fixation gene sequences in *Rhizobium phaseoli*. Nature 299, 724–726.
[60] Quinto, C., de la Vega, H., Flores, M., Leemans, J., Cevallos, M.A., Pardo, M.A., Azpiroz, R., Girard, M.L., Calva, E. and Palacios, R. (1985) Nitrogenase reductase: a functional multigene family in *Rhizobium phaseoli*. Proc. Natl. Acad. Sci. USA 82, 1170–1174.
[61] Rodriguez, C. and Romero, D. (1998) Multiple recombination events maintain sequence identity among members of the nitrogenase multigene family in *Rhizobium etli*. Genetics 149, 785–794.
[62] Gregg-Jolly, L.A. and Ornston, L.N. (1994) Properties of *Acinetobacter calcoaceticus recA* and its contribution to intracellular gene conversion. Mol. Microbiol. 12, 985–992.
[63] Kowalchuk, G.A., Gregg-Jolly, L.A. and Ornston, L.N. (1995) Nucleotide sequences transferred by gene conversion in the bacterium *Acinetobacter calcoaceticus*. Gene 153, 111–115.
[64] Doten, R.C., Gregg, L.A. and Ornston, L.N. (1987) Influence of the *catBCE* sequence on the phenotypic reversion of a *catE* mutation in *Acinetobacter calcoaceticus*. J. Bacteriol. 169, 3175–3180.
[65] Sawyer, S. (1989) Statistical tests for detecting gene conversion. Mol. Biol. Evol. 6, 526–538.
[66] Drouin, G., Prat, F., Ell, M. and Clarke, G.D. (1999) Detecting and characterizing gene conversions between multigene family members. Mol. Biol. Evol. 16, 1369–1390.
[67] Awadalla, P. (2003) The evolutionary genomics of pathogen recombination. Nat. Rev. Genet. 4, 50–60.
[68] Jordan, I.K., Makarova, K.S., Wolf, Y.I. and Koonin, E.V. (2001) Gene conversions in genes encoding outer-membrane proteins in *H. pylori* and *C. pneumoniae*. Trends Genet. 17, 7–10.
[69] Gogarten, J.P. and Olendzenski, L. (1999) Orthologs, paralogs and genome comparisons. Curr. Opin. Genet. Dev. 9, 630–636.
[70] Higgins, C.F., McLaren, R.S. and Newbury, S.F. (1988) Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? A review. Gene 72, 3–14.
[71] Liao, D. (2000) Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. J. Mol. Evol. 51, 305–317.
[72] Lathe III, W.C. and Bork, P. (2001) Evolution of *saf* genes: ancient duplication, differential loss and gene conversion. FEBS Lett. 502, 113–116.
[73] Solnick, J.V., Hansen, L.M., Salama, N.R., Boonjakuakul, J.K. and Syvanen, M. (2004) Modification of *Helicobacter pylori* outer membrane protein expression during experimental infection of rhesus macaques. Proc. Natl. Acad. Sci. USA 101, 2106–2111.
[74] Meinersmann, R.J. and Hiett, K.L. (2000) Concerted evolution of duplicate *fla* genes in *Campylobacter*. Microbiology 146, 2283–2290.
[75] Mattatall, N.R. and Sanderson, K.E. (1996) *Salmonella typhimurium* LT2 possesses three distinct 23S rRNA intervening sequences. J. Bacteriol. 178, 2272–2278.
[76] Pronk, L.M. and Sanderson, K.E. (2001) Intervening sequences in *rrl* genes and fragmentation of 23S rRNA in genera of the family *Enterobacteriaceae*. J. Bacteriol. 183, 5782–5787.
[77] Skurnik, M. and Toivanen, P. (1991) Intervening sequences (IVSs), in the 23S ribosomal RNA genes of pathogenic *Yersinia enterocolitica* strains. The IVSs in *Y. enterocolitica* and *Salmonella typhimurium* have a common origin. Mol. Microbiol. 5, 585–593.
[78] Mattatall, N.R., Daines, D.A., Liu, S.-L. and Sanderson, K.E. (1996) *Salmonella typhi* contains identical intervening sequences in all seven *rrl* genes. J. Bacteriol. 178, 5323–5326.
[79] Zhang, Q.Y., De Ryckere, D., Lauer, P. and Koomey, M. (1992) Gene conversion in *Neisseria gonorrhoeae*: evidence for its role in pilus antigenic variation. Proc. Natl. Acad. Sci. USA 89, 5366–5370.
[80] Noormohammadi, A.H., Markham, P.F., Kanci, A., Whithear, K.G. and Browning, G.F. (2000) A novel mechanism for control of antigenic variation in the haemagglutinin gene family of *Mycoplasma synoviae*. Mol. Microbiol. 35, 911–923.

[81] Zhang, J.R. and Norris, S.J. (1998) Genetic variation of the *Borrelia burgdorferi* gene *elpE* involves cassette-specific segmental gene conversion. Infect. Immun. 66, 3698–3704.
[82] Restrepo, B.I. and Barbour, A.G. (1994) Antigenic diversity in the bacterium *B. hermsii* through "somatic" mutations in rearranged vsp genes. Cell 78, 867–876.
[83] Harrington, C.S., Thomson-Carter, F.M. and Carter, P.E. (1997) Evidence for recombination in the flagellin locus of *Campylobacter jejuni*: implications for the flagellin gene typing scheme. J. Clin. Microbiol. 35, 2386–2392.
[84] Kline, K.A., Sechman, E.V., Skaar, E.P. and Seifert, H.S. (2003) Recombination, repair and replication in the pathogenic *Neisseriae*: the 3 R's of molecular genetics of two human-specific bacterial pathogens. Mol. Microbiol. 50, 3–13.
[85] Haas, R. and Meyer, T.F. (1986) The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: evidence for gene conversion. Cell 44, 107–115.
[86] Haas, R., Veit, S. and Meyer, T.F. (1992) Silent pilin genes of *Neisseria gonorrhoeae* MS11 and the occurrence of related hypervariant sequences among other gonococcal isolates. Mol. Microbiol. 6, 197–208.
[87] Hamrick, T.S., Dempsey, J.A., Cohen, M.S. and Cannon, J.G. (2001) Antigenic variation of gonococcal pilin expression in vivo: analysis of the strain PA1090 pilin repertoire and identification of the *pilS* gene copies recombining with *pilE* during experimental human infection. Microbiology 147, 839–849.
[88] Serkin, C.D. and Seifert, H.S. (1998) Frequency of pilin antigenic variation in *Neisseria gonorrhoeae*. J. Bacteriol. 180, 1955–1958.
[89] Serkin, C.D. and Seifert, H.S. (2000) Iron availability regulates DNA recombination in *Neisseria gonorrhoeae*. Mol. Microbiol. 37, 1075–1086.
[90] Koomey, M., Gotschlich, E.C., Robbins, K., Bergstrom, S. and Swanson, J. (1987) Effects of *recA* mutations on pilus antigenic variation and phase transitions in *Neisseria gonorrhoeae*. Genetics 117, 391–398.
[91] Stohl, E.A. and Seifert, H.S. (2001) The *recX* gene potentiates homologous recombination in *Neisseria gonorrhoeae*. Mol. Microbiol. 40, 1301–1310.
[92] Mehr, I.J. and Seifert, H.S. (1998) Differential roles of homologous recombination pathways in *Neisseria gonorrhoeae* pilin antigenic variation, DNA transformation and DNA repair. Mol. Microbiol. 30, 697–710.
[93] Skaar, E.P., Lazio, M.P. and Seifert, H.S. (2002) Roles of the *recJ* and *recN* genes in homologous recombination and DNA repair pathways of *Neisseria gonorrhoeae*. J. Bacteriol. 184, 919–927.
[94] Chaussee, M.S., Wilson, J. and Hill, S.A. (1999) Characterization of the *recD* gene of *Neisseria gonorrhoeae* MS11 and the effect of *recD* inactivation on pilin variation and DNA transformation. Microbiology 145, 389–400.
[95] Salvatore, P. et al. (2002) Phenotypes of a naturally defective *recB* allele in *Neisseria meningitidis* clinical isolates. Infect. Immun. 70, 4185–4195.
[96] Zhang, J.R., Hardham, J.M., Barbour, A.G. and Norris, S.J. (1997) Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. Cell 89, 275–285.
[97] Donelson, J.E. (1995) Mechanisms of antigenic variation in *Borrelia hermsii* and african trypanosomes. J. Biol. Chem. 270, 7783–7786.
[98] Dumler, J.S. and Bakken, J.S. (1998) Human ehrlichioses: newly recognized infections transmitted by ticks. Annu. Rev. Med. 49, 201–213.
[99] Brayton, K.A., Knowles, D.P., McGuire, T.C. and Palmer, G.H. (2001) Efficient use of a small genome to generate antigenic diversity in tick-borne ehrlichial pathogens. Proc. Natl. Acad. Sci. USA 98, 4130–4135.
[100] French, D.M., Brown, W.C. and Palmer, G.H. (1999) Emergence of *Anaplasma marginale* antigenic variants during persistent rickettsemia. Infect. Immun. 67, 5834–5840.
[101] Noormohammadi, A.H., Markham, P.F., Whithear, K.G., Walker, I.D., Gerovich, V.A., Ley, D.H. and Browning, G.F. (1997) *Mycoplasma synoviae* has two distinct phase-variable major membrane antigens, one of which is a putative hemagglutinin. Infect. Immun. 65, 2542–2547.
[102] Noormohammadi, A.H., Markham, P.F., Duffy, M.F., Whithear, K.G. and Browning, G.F. (1998) Multigene families encoding the major hemagglutinins in phylogenetically distinct mycoplasmas. Infect. Immun. 66, 3470–3745.
[103] Centurion-Lara, A., LaFond, R.E., Hevner, K., Godornes, C., Molini, B.J., Van Voorhis, W.C. and Lukehart, S.A. (2004) Gene conversion: a mechanism for generation of heterogeneity in the *tprK* gene of *Treponema pallidum* during infection. Mol. Microbiol. 52, 1579–1596.
[104] Smith, G.R. (2001) Homologous recombination near and far from DNA breaks: alternative roles and contrasting views. Annu. Rev. Genet. 35, 243–274.
[105] Worth Jr., L., Clark, S., Radman, M. and Modrich, P. (1994) Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. Proc. Natl. Acad. Sci. USA 91, 3238–3241.
[106] Zahrt, T.C., Mora, G.C. and Maloy, S. (1994) Inactivation of mismatch repair overcomes the barrier to transduction between *Salmonella typhimurium* and *Salmonella typhi*. J. Bacteriol. 176, 1527–1529.
[107] LeClerc, J.E., Li, B., Payne, W.L. and Cebula, T.A. (1996) High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. Science 274, 1208–1211.
[108] Matic, I., Radman, M., Taddei, F., Picard, B., Doit, C., Bingen, E., Denamur, E. and Elion, J. (1997) Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. Science 277, 1833–1834.
[109] Oliver, A., Cantón, R., Campo, P., Baquero, F. and Blázquez, J. (2000) High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. Science 288, 1251–1253.
[110] Prunier, A.L., Malbruny, B., Laumain, M., Brouard, J., Duhamel, J.F. and Leclercq, R. (2003) High rate of macrolide resistance in *Staphylococcus aureus* strains from patients with cystic fibrosis reveals high proportions of hypermutable strains. J. Infect. Dis. 187, 1709–1716.
[111] Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P.H. and Godelle, B. (1997) Role of mutator alleles in adaptive evolution. Nature 387, 700–702.
[112] Giraud, A., Matic, I., Tenaillon, O., Clara, A., Radman, M., Fons, M. and Taddei, F. (2001) Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. Science 291, 2606–2608.
[113] Townsend, J.P., Nielsen, K.M., Fisher, D.S. and Hartl, D.L. (2003) Horizontal acquisition of divergent chromosomal DNA in bacteria: effects of mutator phenotypes. Genetics 164, 13–21.
[114] Matic, I., Rubic, A. and Radman, M. (2003) 2-aminopurine allows interspecies recombination by a reversible inactivation of the *Escherichia coli* mismatch repair system. J. Bacteriol. 185, 1459–1461.

JOURNAL OF BACTERIOLOGY, June 2005, p. 001
0021-9193/05/$08.00+0 doi:10.1128/JB.187.12.000-000.2005

Vol. 187, No. 12

Gene Conversion Tracts Associated with Crossovers in *Rhizobium etli*

Gustavo Santoyo, Jaime M. Martínez-Salazar, César Rodríguez, and David Romero*

Programa de Ingeniería Genómica, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Apartado Postal 565-A, Cuernavaca, Morelos, México

Received 8 December 2004/Accepted 11 March 2005

Gene conversion has been defined as the nonreciprocal transfer of information between homologous sequences. Despite its broad interest for genome evolution, the occurrence of this mechanism in bacteria has been difficult to ascertain due to the possible occurrence of multiple crossover events that would mimic gene conversion. In this work, we employ a novel system, based on cointegrate formation, to isolate gene conversion events associated with crossovers in the nitrogen-fixing bacterium *Rhizobium etli*. In this system, selection is applied only for cointegrate formation, with gene conversions being detected as unselected events. This minimizes the likelihood of multiple crossovers. To track the extent and architecture of gene conversions, evenly spaced nucleotide changes were made in one of the nitrogenase structural genes (*nifH*), introducing unique sites for different restriction endonucleases. Our results show that (i) crossover events were almost invariably accompanied by a gene conversion event occurring nearby; (ii) gene conversion events ranged in size from 150 bp to 800 bp; (iii) gene conversion events displayed a strong bias, favoring the preservation of incoming sequences; (iv) even small amounts of sequence divergence had a strong effect on recombination frequency; and (v) the MutS mismatch repair system plays an important role in determining the length of gene conversion segments. A detailed analysis of the architecture of the conversion events suggests that multiple crossovers are an unlikely alternative for their generation. Our results are better explained as the product of true gene conversions occurring under the double-strand break repair model for recombination.

Gene duplication is the main mechanism that gives rise to gene families in both eukaryotic and prokaryotic genomes (28). One common observation is that members of a multigene family tend to maintain a higher degree of sequence conservation at the intraspecific level than that seen in interspecific comparisons. This is a clear indication that members of a multigene family are evolving in a concerted way, a process called concerted evolution (5). Concerted evolution of multigene families occurs not only in eukaryotic organisms but in prokaryotes as well (39). Evidence for its occurrence has been reported for the rRNA genes (16S rRNA) in Bacteria and Archaea (17), the 23S rRNA intervening sequences in *Salmonella enterica* serovar Typhimurium and *Salmonella enterica* serovar Typhi (21, 22), the genes coding for the translation factor EF-Tu (*tuf*) in bacteria (1, 16), the flagellin genes (*fla*) in *Campylobacter* (23), the family of outer membrane proteins (*bab*) in *Helicobacter pylori* (30), and the genes coding for the nitrogenase enzyme (*nifH*) in *Rhizobium etli* (34; E. Sepúlveda and D. Romero, unpublished data). Although several molecular processes may participate in concerted evolution of gene families, it is generally thought that gene conversion among repeated genes is responsible for this evolutionary trend. Specialized systems for gene conversion also participate in the generation of antigenic variation in several pathogenic bacteria (reviewed in reference 39).

Gene conversion is one of the possible outcomes of a recombination event and has been defined as the nonreciprocal transfer of genetic information from one DNA duplex to another. This process was initially demonstrated in ascomycete fungi (18), and the yeast *Saccharomyces cerevisiae* is still the preferred organism for its study due to the ability to recover all the products of a meiotic recombination event. This ability facilitates the demonstration of nonreciprocal transfer events. The need to understand gene conversion and its association with crossovers was the main motivation for the development of the Holliday model for recombination (18); the possibility for gene conversion has been retained in successive models of recombination, including the double-strand break repair model (18, 43, 44).

Gene conversion has been more difficult to study in other organisms, including bacteria. The main problem is that, since the recovery of all the products of a recombination event is not possible, the characteristic nonreciprocity of gene conversion events cannot be ensured. In fact, it has been argued that several possible examples of gene conversion in bacteria may be due to selection for rare double crossovers rather than to gene conversion (40). Despite this limitation, some groups have provided convincing evidence for gene conversion in bacteria, including *Escherichia coli* (13, 45) and *Salmonella enterica* serovar Typhimurium (1, 3, 11), using substrates harboring repeated sequences in an inverted orientation. In these reports, selection is applied for gene conversion events occurring between inverted repeats, afterwards exploring their relation with crossover. In at least one case, there is convincing albeit indirect evidence for the nonreciprocal origin of these events (3).

Our group has been studying *Rhizobium etli*, an α-proteobacterium that is able to form nitrogen-fixing symbiotic associations with bean plants. Besides its symbiotic capabilities, *R. etli* is also interesting because of the presence of reiterated gene elements (8), which can play important roles in shaping geno-

* Corresponding author. Mailing address: Programa de Ingeniería Genómica, Centro de Ciencias Genómicas-Universidad Nacional Autónoma de México, Apartado Postal 565-A, 62210 Cuernavaca, Mor., México. Phone: 52 (777) 3175867. Fax: 52 (777) 3175581. E-mail: dromero@ccg.unam.mx.

TABLE 1. PCR primers used in this work[a]

Primer	Sequence[b]	Source/position[c]	GenBank accession no.
HindIII	AGGAAGCTTATATGTCAGATTTGC	*nifHa*, (−)11–13	M10587
BamHI	GGATCCGACCCGAAAGCC	*nifHa*, 115–132	M10587
ApaI	GTGCACATGACGATGTCGACT	*nifHa*, 344–364	M10587
NarI	GCGCCCGTTACAGATCAG	*nifHa*, 564–547	M10587
MluI	TCCGACGCGTTACTGGATCA	*nifHa*, 701–683	M10587
BclI	CTTGATCATGCCGAAGTCGAG	*nifHa*, 831–811	M10587
XbaI	TCCTCTAGACAGCGGCAGTTAT	*nifHa*, 912–891	M10587
1	CTGAAACCCAACAAAAG	*nifHa*, (−)135–(−)119	M10587
2	GCAAGGCGATTAAGTTG	pIC20R, 385–369	L08913
3	AGTCGGCAAATAATGTC	fITc, 2543–2559	U35135
4	AAAACGCTGTCATTCTC	*nifHa*, 1033–1017	U80928

[a] All the oligonucleotides are shown in the 5′ to 3′ direction.
[b] Nucleotides that were modified to generate the corresponding restriction site (underlined) are shown in boldface.
[c] Positions correspond to the start codon of the indicated sequence (*nifHa*) or to the initial nucleotide in the reported sequence (pIC20R and fITc).

mic structure. One important multigene family is the *nifH* family (encoding one component of the nitrogenase enzyme), comprised of three identical members (31, 32) located in a large plasmid (371 kb) called the symbiotic plasmid or pSym (9, 35). Homologous recombination among the members of this family promotes different genomic rearrangements in pSym, having important symbiotic consequences (36, 37, 47). Phylogenetic evidence indicates the existence of concerted evolution among members of this family (E. Sepúlveda and D. Romero, unpublished data), perhaps generated through gene conversion (34). A previous evaluation of the occurrence of gene conversion in *R. etli* relied on the introduction of a 28-bp insertion into one of the *nifH* copies, followed by its elimination by recombinational interactions with either of the other two *nifH* copies (34). Although products consistent with the occurrence of gene conversion were isolated, it is formally possible that at least some of these arose from repeated reciprocal exchanges rather than true gene conversion (34).

In this work, we employ a novel approach to evaluate gene conversion, using a genetic system based on the cointegration between sequences harboring planned sequence alterations, or restriction fragment length polymorphism (RFLPs). In this approach, selection is only applied for cointegration; any gene conversion arises as an unselected event. This avoids weaknesses present in previous approaches, where direct selection for gene convertants raises the possibility of multiple crossovers to explain its generation. Characterization of the gene convertants obtained through this approach allowed us to evaluate the association of recombination with gene conversion, the length of converted tracts, and the role of sequence heterology. Our results show that (i) crossover events are frequently accompanied (98%) by a gene conversion event occurring nearby; (ii) gene conversion events frequently encompass more than half of the length of this gene; (iii) gene conversion events display a strong polarity, favoring the preservation of incoming sequences; (iv) even small amounts (1.6%) of sequence divergence have a strong effect on recombination frequency; and (v) the MutS mismatch repair system plays an important role in determining the length of gene conversion segments.

MATERIALS AND METHODS

Bacterial strains and media. *Escherichia coli* strains were grown in LB medium (25) at 37°C. *Rhizobium etli* strains were grown in PY medium (27) at 30°C. Antibiotics were added to the media when needed at the following concentrations (in micrograms per milliliter): carbenicillin, 100; chloramphenicol, 15; kanamycin, 15; nalidixic acid, 20; spectinomycin 100; and tetracycline, 10 (*E. coli*) or 2 (*R. etli*). For selection in cloning experiments, 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) was added to LB plates at 30 μg ml^{-1}.

General DNA manipulations and mutagenesis of the *nifH* gene. All DNA manipulations were done using standard procedures (38). Most of the plasmid transformations employed *Escherichia coli* DH5α as a host (10). PCRs were done in a Techgene thermocycler using Platinum *Taq* High Fidelity DNA polymerase (Invitrogen) for mutagenesis and conventional *Taq* DNA polymerase for analytical characterization. For ligations, T4 polynucleotide ligase (Amersham Biosciences) was used. Restriction enzymes were purchased from diverse companies and used according to the recommendations of the supplier. Custom oligonucleotides were synthesized at the Unidad de Síntesis de Oligonucleótidos (Instituto de Biotecnología, Universidad Nacional Autónoma de México, México).

Introduction of specific restriction sites into the *nifH* gene was done by a variation of published PCR mutagenesis procedures (26). Specific oligonucleotide primers (see Table 1) containing one or more modified nucleotides to introduce a restriction site at the time of polymerization were designed. Only single base changes (either transitions or transversions) were used, avoiding the introduction of stop or otherwise rare codons. To introduce mutations, two PCR products were generated, using either the HindIII/NarI primer combination (product size, 567 bp) or the ApaI/XbaI combination (size 557 bp), employing as a template DNA from *R. etli* CFN42. Both products were gel purified, using a GeneClean II kit (Bio 101). The purified PCR products were mixed, heat denatured (at 90°C for 2 min), and annealed (at 60°C for 1 min), taking advantage of a 200-bp overlap between both products. A mixture of the four deoxyribonucleotides as well as Platinum *Taq* DNA polymerase was added to the annealing mixture and incubated at 72°C for 6 min to allow the generation of a complete *nifH* sequence. After this, the HindIII and XbaI oligonucleotides were added to the reaction and subjected to PCR (30 cycles with denaturation at 92°C for 1 min, annealing at 56°C for 1 min, and extension at 72°C for 2 min). The PCR products were cloned into pUC19 (48) and analyzed to verify the introduction of the restriction sites in the *nifH* sequence. Two full-sized *nifH* clones were obtained, one containing the HindIII, ApaI, I, and XbaI restriction sites and the other with the HindIII, NarI, and XbaI sites.

The whole process was repeated cyclically to introduce new mutations, using as templates the product of the preceding step and primer pairs BamHI/XbaI (to introduce mutations in the 5′ end of the gene) or HindIII/MluI and HindIII/BclI (to mutagenize the 3′ end of *nifH*). This process generates two clones, one harboring the BamHI and ApaLI sites in the 5′ end of *nifH* and another with the NarI, MluI, and BclI sites in the 3′ end; both clones share a native BglII site and are flanked by HindIII and XbaI sites. The mutations were combined in a single gene by digesting the first clone with HindIII and BglII and the second with BglII and XbaI and ligating the desired fragments into pUC19. Introduction of these mutations was verified by manually sequencing both strands of the modified *nifH* gene by the Sanger dideoxy chain termination method, using a thermosequenase cycle sequencing kit (Amersham Biosciences). Oligonucleotides were labeled by kinasing with [γ-^{32}P]ATP and T4 DNA kinase (Amersham Biosciences). Sequencing reactions were electrophoresed in 6% polyacrylamide–8 M urea gels. Besides the planned mutations, spontaneous sequence changes, perhaps occurring during the successive PCR steps, generated a novel MaeIII site. A map of the relevant restriction sites in the *nifH* gene is shown in Fig. 1.

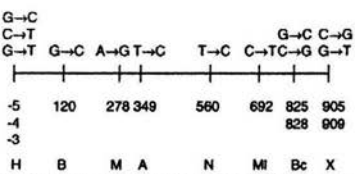

FIG. 1. Locations of the nucleotide substitutions introduced on the *nifH* gene. Numbers below the horizontal line indicate the position of the base substitutions (showed above the line) with respect to the starting nucleotide of the gene. These substitutions generate unique sites for different restriction enzymes, which are indicated by a one-letter code (H, HindIII; B, BamHI; M, MaeIII; A, ApaLI; N, NarI; MI, MluI; Bc, BclI; X, XbaI).

Construction of integrative plasmids. For construction of integrative vectors, pIC20R (20) was modified by introduction of an RK2 *oriT* sequence (29, 46). This was done by PCR amplification from pEYM1 (47), using custom-made oligonucleotides that introduce a XhoI site. The PCR product was cloned into the single XhoI site in pIC20R. This was used as a vector for introduction, by HindIII-XbaI digestion and ligation, of the wild-type *nifH* sequence, the *nifH* gene harboring the eight RFLPs (Fig. 1), and other variants, such as one having the mutations HindIII, BamHI, MaeIII, and ApaLI in the 3′ end of *nifH* and the other with the NarI, MluI, BclI, and XbaI mutations in the 5′ zone. Finally, a 1.8-kb HindIII-HindIII tetracycline resistance cassette from pRSL193 (2) was introduced into each of these plasmids, giving rise to pMC0, pMC11, pMC52, and pMC63, respectively.

To simplify cointegrate selection in the *nucS* background, a kanamycin-resistant derivative from pMC11 was constructed by substitution of the HindIII-HindIII tetracycline resistance cassette with a 2.2-kb HindIII-HindIII kanamycin resistance cassette from pHP45Ω-Km (7), giving rise to pJGus2R.

Molecular characterization of transconjugants. *Escherichia coli* S17-1 (F⁻ *pro-82 thi-1 endA1 hsdR17 supE44 recA13*, chromosomally integrated RP-4-2 [Tc::Mu, Km::Tn*7*]) was used as a host for conjugative transfer of integrative plasmids (42). To that end, biparental matings were set up on solid media between *Escherichia coli* S17-1 harboring the desired plasmid and *R. etli* as described previously (47); transconjugants were selected by its resistance to nalidixic acid and tetracycline. In most cases, *R. etli* CFNX55 (36) (harboring a large deletion that removes two of the three *nifH* genes in pSym) was used as a recipient. To ensure the independence of the observed events, 10 separate conjugation experiments were set up, retaining not more than five single-colony isolates from each experiment. Total DNA was isolated from each transconjugant and analyzed by PCR with specific primers (left PCR with primers 1 and 2, right PCR with primers 3 and 4, Fig. 2) to amplify both *nifH* products. All the PCR products were purified by using CentriSep spin columns (Applied Biosystems) before digestion with restriction enzymes. Determinations of conjugation frequency were repeated at least 10 times and are expressed as number of transconjugants per recipient cell ± standard deviation.

Generation of strain CFNX704. To isolate an *R. etli* strain harboring the desired RFLPs on the *nifH* copy in pSym, plasmid pMC11 was transferred by conjugation to *R. etli* CFNX55, selecting integrants by their resistance to tetracycline. An integrant harboring a gene conversion event encompassing the BamHI, MaeIII, ApaLI, and NarI markers was identified by PCR. From this strain, loss of pMC11 by excision was screened by checking single-colony isolates for a tetracycline-sensitive phenotype; these were found at a frequency of 10^{-4}. The excisant was analyzed by PCR and restriction analysis to verify the retention of the BamHI, MaeIII, ApaLI, and NarI markers. This strain, called CFNX704, was then used in crosses with the pMC0 plasmid (containing the *nifH* wild-type sequence). As before, we analyzed 50 transconjugants coming from 10 independent experiments.

Construction of a *nucS* derivative. To evaluate the participation of the *nucS* repair system in gene conversion, we employed strain CFNX706, a *nucS::kanP*Sp derivative from *R. etli* (J. M. Martínez-Salazar, J. Zuñiga-Castillo, and D. Romero, unpublished data). This strain harbors a *kanP*Sp insertion in the *nucS* gene, interrupting codon 292. Strain CFNX706 was modified by generating a large deletion on pSym that eliminates two of the three *nifH* genes, using the recombination enhancement by replication system (47). In this system, activation of a supernumerary replication origin on pSym leads to the high-frequency generation of a deletion on pSym identical to the one in strain CFNX55 (47).

To apply this system, plasmid pEYM13, harboring *oriV* from RK2 (47) was inserted by single-crossover recombination into one of the *nifH* genes of strain CFNX706, selecting integrants by their resistance to kanamycin. To activate replication from the supernumerary origin, plasmid pEYM5, encoding the replication initiator protein *trfA* from RK2, was introduced by conjugation with several integrants; transconjugants were selected by their resistance to chloramphenicol. Over 25% of the chloramphenicol-resistant transconjugants also displayed the loss of the kanamycin resistance marker, indicating the presence of possible deletions on pSym. The presence of the desired deletion was verified by analyses of plasmid profiles (revealing a 107-kb deletion on pSym) as well as by Southern hybridization against a *nifH* probe (data not shown). Spontaneous loss of pEYM5 from the *nucS::kanP*Sp derivative harboring the desired deletion was screened by checking single-colony isolates for a chloramphenicol-sensitive phenotype, giving rise to strain CFNX712. This strain was then used as a recipient in crosses with plasmid pJGus2R. Fifty transconjugants, coming from 10 independent experiments, were analyzed as described before.

RESULTS

Experimental strategy. The *nifH* multigene family has three completely identical members located on pSym. To evaluate the length and distribution of gene conversion tracts in this family, one of the *nifH* genes was modified by introducing single-base-pair changes approximately every 100 bp along the gene sequence, as described in Materials and Methods. Twelve different changes were made, generating unique recognition sites for eight different restriction enzymes (Fig. 1). Such modifications serve as convenient landmarks to evaluate the extension of gene conversion tracts (see below).

This modified *nifH* gene was inserted into a plasmid (pMC11)

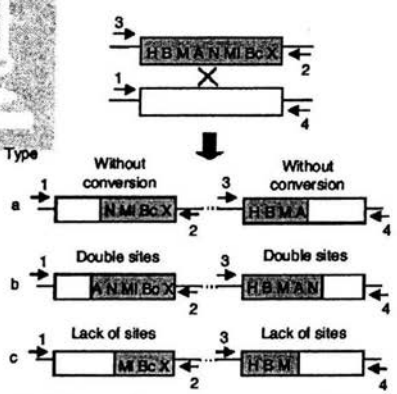

FIG. 2. Experimental strategy to detect gene conversion associated to cointegration. A shaded rectangle represents the *nifH* gene harboring different RFLPs (Fig. 1); the open rectangle corresponds to the wild-type *nifH* gene. Cointegration between circular molecules bearing these genes (indicated by crossed lines on the top part of the figure) generates three different types of cointegrate, depending on its possible association with gene conversion. In type a, a single crossover is depicted, without gene conversion. Gain of sites on both sides of the cointegrate (type b, sites A and N) or loss of sites (type c, sites A and N) is interpreted as evidence of gene conversion. The primers used to amplify each side of the cointegrate are indicated by numbered arrowheads.

that can be transferred by conjugation from *Escherichia coli* to *Rhizobium etli*, but it is unable to replicate in this latter host. The recipient for such crosses was *Rhizobium etli* CFNX55 (36), a derivative that harbors only one copy of the wild-type *nifH* gene. Upon single-crossover recombination in this strain, three types of cointegrates are possible (Fig. 2). The first type entails cointegrate formation without associated gene conversion (Fig. 2, type a), while the other two involve cointegrate formation associated with gene conversion, favoring either incoming sequences (type b) or endogenous sequences (type c). As shown in Fig. 2, these types can be easily distinguished by looking at the distribution of restriction sites in both sides of the cointegrate. Type a recombinants have only a redistribution of restriction sites at the crossover point, while types b and c display an increase in the number of restriction sites (double gain, type b) or a reduction in restriction sites (double loss, type c). Thus, the number and position of restriction sites modified in the gene conversion events allows an evaluation of the length and position of gene conversion tracts along the *nifH* gene.

In this experimental approach, selection is applied only for cointegration of the plasmid; no selection whatsoever was applied for recovery of gene conversion events. This is an important difference with previous studies, because it minimizes the likelihood that gene conversion events arise through selection for rare double crossover events, which would mimic gene convertants.

Independent isolation of both sides of the cointegrate is possible through PCR amplification, using primer pairs that amplify either the left side (primers 1 and 2) or the right side (primers 3 and 4) of the cointegrate (Fig. 2). These PCR products were then subjected to restriction analysis, searching for instances in which cutting with a specific enzyme occurred in both sides or in neither side of the cointegrate; these were interpreted as examples of gene conversion.

Crossover events are strongly associated with gene conversion. To evaluate the association of crossover formation to gene conversion, cointegrates between pMC11 (harboring eight different RFLPs) and pSym were selected, as described in Materials and Methods. In *R. etli*, formation of cointegrates of this kind is strictly dependent on *recA* (50). Interestingly, when plasmid pMC11 was used as a donor, integrants were obtained at a very low frequency ($1.63 \times 10^{-7} \pm 0.52 \times 10^{-7}$). Higher integration frequencies were seen when plasmids lacking RFLPs (pMC0, integration frequency $1.6 \times 10^{-5} \pm 0.05 \times 10^{-5}$) or with RFLPs only on the 5′ half (pMC32, $1.18 \times 10^{-5} \pm 0.12 \times 10^{-5}$) or on the 3′ half (pMC63, $6.6 \times 10^{-6} \pm 3.1 \times 10^{-6}$) of the gene were used. Thus, the low integration frequency observed with pMC11 may be attributed to degree of sequence divergence (1.6%) between the recombining sequences. A similar sensitivity of recombination frequency to degree of heterology has been observed previously (33, 49).

DNA was purified from 50 independent cointegrates with pMC11 and subjected to separate PCRs to amplify the left- and right-hand sides of each cointegrate, which were then subjected to restriction analysis. An example of these analyses is shown in Fig. 3. In this case, digestion with BamHI, MaeIII, and ApaLI was observed for both sides of the cointegrate, while restriction on only one side was observed for the remaining enzymes. This indicates that a continuous gene conversion

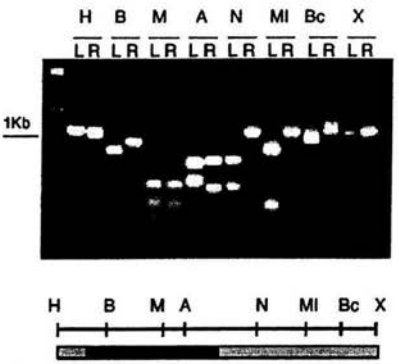

FIG. 3. Detection of gene conversion in cointegrates. The left (L) and right (R) parts of a specific cointegrate were amplified by PCR (Fig. 2 and Material and Methods). These products were digested with different restriction enzymes (H, HindIII; B, BamHI; M, MaeIII; A, ApaLI; N, NarI; MI, MluI; Bc, BclI; X, XbaI) and analyzed by agarose gel electrophoresis (top). Note that both sides of the cointegrate were cut with BamHI, MaeIII, and ApaLI, indicating a gene conversion event encompassing these markers. This is summarized in the lower part of the figure (shaded bar, regions of *nifH* not subjected to gene conversion; black bar, region undergoing gene conversion towards gain of the markers).

event encompassing these three sites had occurred in this particular cointegrate. Since it is impossible to determine the exact endpoint of conversions occurring between two markers, the middle zone between two restriction sites was chosen to register the end of every conversion event. Thus, this particular conversion event was roughly 400 bp in size.

Figure 4 show the results obtained for the 50 different cointegrates analyzed. Four classes of events were observed. Classes A to C represent different kinds of gene conversion events, while class D comprises crossover events not associated with gene conversion. Interestingly, 98% of the events (49 out of 50) fall in the conversion classes (A to C), while only a single event was located in the no-conversion class (class D, 2%). Thus, crossover formation is frequently accompanied by a gene conversion event occurring nearby.

Structure of gene conversion events. Class A events (continuous gene conversion), such as the one shown in Fig. 3, were more frequent in this sample. Fifty-eight percent of the isolates (29 out of 50) belonged to this class. As shown in Fig. 4, continuous gene conversion tracts ranged in size from 150 bp up to 800 bp; more than half of the members in this class (17 out of 29 events) displayed continuous gene conversion tracts at least 600 bp in size.

The second most frequent class corresponds to discontinuous gene conversion events (class B). Members of this class, encompassing 26% of the isolates, display two tracts of continuous gene conversion, with a marker or two between these tracts that do not display gene conversion. The most complex class corresponds to the bipolar conversions (class C), which

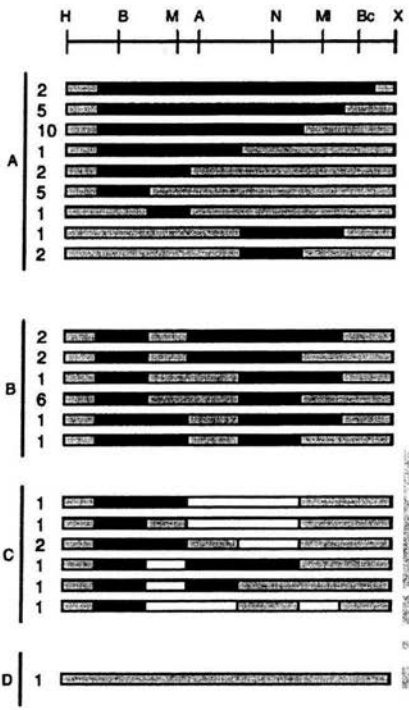

FIG. 4. Structure of gene conversion tracts obtained upon introduction of pMC11 into *R. etli* CFNX55. The RFLP map of the *nifH* gene is shown on top as a reference. Letters at the left side represent the four cointegrate classes found (A, continuous conversion; B, discontinuous conversion; C, bipolar conversions; and D, single crossover with no evidence of gene conversion). Values indicate the number of isolates with the corresponding conversion tract. Black bars represent the extent of gene conversion tracts towards marker gain; white bars indicate gene conversion tracts showing marker loss; shaded bars are regions not subjected to gene conversion. Note that gene conversion is biased towards marker gain.

represent 14% of the isolates characterized. Members of this class display at least two conversion tracts; these tracts are clearly discernible, because one of these displays gain of sites in the conversion tract (double gain, Fig. 2), while in the other tract the restriction sites were absent (double loss, Fig. 2). In a single isolate, the contrasting conversion tracts are contiguous, while in the rest these tracts are separated by an intervening marker that does not display conversion. Thus, class C conversions are a mixed class that contains both continuous and discontinuous events.

Length and distribution of gene conversion tracts. To evaluate the length distribution of gene conversion tracts, continuous and discontinuous classes were analyzed separately; bipolar conversions were included in either the continuous or discontinuous class depending on architecture (Fig. 5). The F5 size of the conversion tract was evaluated for each class; isolates harboring a continuous conversion contribute only with a single tract to the total, but isolates in classes B and C contribute with two or three tracts, depending on structure. Therefore, the data in Fig. 5 are based on the analysis of 72 conversion tracts. Gene conversion tract length for the continuous class reveals a bimodal distribution, centered at 150 bp and 600 bp. The discontinuous tracts, in contrast, show a single unimodal distribution, centered at 150 bp. This is consistent with an interpretation that continuous tracts are formed through the cooperation of two separate processes, such as gap filling and heteroduplex correction (see Discussion).

To evaluate if all the sites are equally likely to participate in a gene conversion event, representation of each site in conversion tracts was counted separately for both continuous and discontinuous events. Figure 5 show that for continuous events the sectors containing the BamHI, MaeIII, ApaLI, and NarI

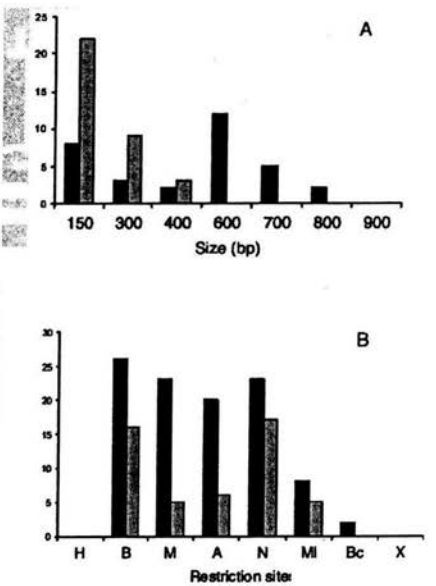

FIG. 5. Size distribution (A) and positions (B) involved in gene conversion events. Data are derived from Fig. 4. For both panels, vertical black bars represent continuous tracts, while shaded bars correspond to discontinuous tracts. The letters in panel B represent the different RFLPs, as shown in Fig. 1.

markers are equally likely to participate in gene conversion. Participation of the MluI and BclI markers is somewhat reduced, while the HindIII and XbaI markers were not included in the continuous conversion events analyzed. The lack of participation of these two terminal markers may be artifactual, due to the reduced homology available at the end of the gene (Fig. 4). In contrast, for the discontinuous events, the BamHI and NarI markers are included preferentially in conversion events; markers flanking these regions, such as MaeIII, ApaLI, MluI, and BclI, were poorly represented in discontinuous conversion events. This suggests that continuous conversion tracts encompassing the BamHI or NarI marker have a significant probability of terminating in the markers surrounding them.

Biased transfer of genetic information by gene conversion. In the system described thus far, we have taken advantage of gain or loss of markers in both sides of each cointegrate (double gain and double loss, respectively, Fig. 2) to detect a gene conversion event. In principle, both kinds of convertants should be observed in the same proportions. Interestingly, we found that conversion is strongly biased towards the double gain class. From Fig. 4, it is clear that almost 90% of the tracts observed (64 out of 72) showed gain of sites. In fact, the few tracts displaying loss of sites come exclusively from class C (bipolar) convertants.

This lack of marker parity may be explained under two contrasting hypotheses. One alternative is that all the markers employed might have an intrinsic repair preference, favoring their use as templates for heteroduplex repair over the wild-type sequence of the *nifH* gene. A second possibility is that the observed bias may arise as a consequence of the way in which the recombining sequences are brought together. In all our experiments, the *nifH* copy harboring the RFLP markers is introduced by conjugation from *Escherichia coli* into an *R. etli* strain bearing a wild-type *nifH* gene. In this case, the observed bias is favoring conversion towards double gain (i.e., towards the incoming sequence) rather than its restoration to a wild-type sequence. This may be explained, under the double-strand break repair model of recombination by saying that the resident copy preferentially receives a double-strand cut, thus being a receptor of information (see Discussion).

These hypotheses may be distinguished by exchanging the configuration of markers participating in conversion, putting the RFLP markers in the resident copy. If the bias is due to preferential repair, convertants should be still biased towards double gain; if the bias is due to preferential cutting of the resident sequence, the bias should be reversed towards the double loss class. To that end, the BamHI, MaeIII, ApaLI, and NarI markers were transferred to the *nifH* gene present in pSym of *R. etli*, generating strain CFNX704 (see Materials and Methods). Plasmid pMCO, harboring a wild-type *nifH* gene, was introduced by conjugation into strain CFNX704 to generate 50 independent cointegrates, which were screened for conversion as before. The results of this experiment are shown in Fig. 6.

Exchanging the configuration of markers does not greatly affect the proportion of the different types of convertants; 58% of the isolates were still class A convertants (continuous gene conversion). However, there were slight differences in the proportion of classes B (discontinuous conversion, 16% versus 26%) and C (bipolar convertants, 26% versus 14%). More importantly, conversion bias is now reversed towards loss of markers. Nearly 82% of the tracts observed (68 out of 83, Fig. 6) showed double marker loss. Again, the few cases displaying double marker gain come from the bipolar class. These results indicate that, in this system, gene conversion is strongly biased towards the acquisition of markers present in the incoming sequence.

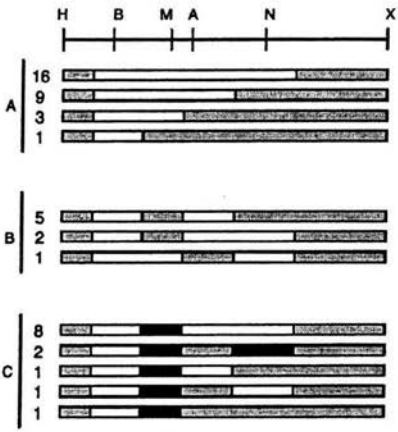

FIG. 6. Structure of gene conversion tracts obtained upon introduction of pMC0 into *R. etli* CFNX704. Letters at the left side represent the three cointegrate classes found (A, continuous conversion; B, discontinuous conversion; and C, bipolar conversions). Black bars represent the extent of gene conversion tracts towards marker gain; white bars indicate gene conversion tracts showing marker loss; shaded bars are regions not subjected to gene conversion. Note that gene conversion is biased towards marker loss. (H, HindIII; B, BamHI; M, MaeIII; A, ApaLI; N, NarI; Ml, MluI; Bc, BclI; X, XbaI.)

MutS mismatch repair system is an important determinant for length of gene conversion segments. The data presented here suggest that continuous conversion segments may be formed through the cooperation of two separate processes, such as gap filling and heteroduplex correction (see Discussion). If that were the case, inactivation of the MutS system, one of the main systems for mismatch correction in bacteria (33, 49), would instigate a marked reduction in the length of gene conversion segments. To evaluate this possibility, a kanamycin-resistant derivative of pMC11 (pJGus28, containing the eight different RFLPs) was introduced by conjugation into an *R. etli mutS::lacP*Sp derivative that harbors only one copy of the wild-type *nifH* gene on pSym (strain CFNX712, see Materials and Methods). As expected for knocking out one of the main barriers for recombination between divergent sequences, integrants were obtained readily in this mutant background (at a frequency of $2.8 \times 10^{-5} \pm 1.49 \times 10^{-5}$). This frequency is 30-fold higher than the one obtained upon transfer of pJGus28 into CFNX55 ($7.54 \times 10^{-7} \pm 3.84 \times 10^{-7}$).

The analysis of 50 cointegrates obtained in the *mutS::lacP*Sp

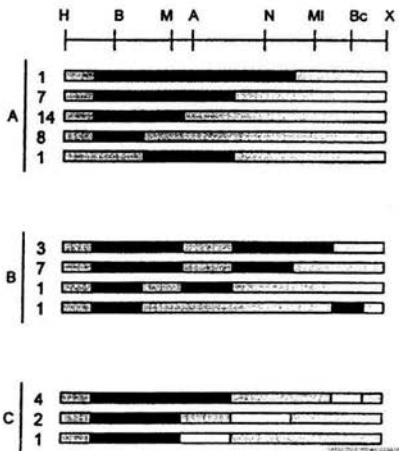

FIG. 7. Structure of gene conversion tracts obtained upon introduction of pJGus28 into *R. etli* CFNX712 (*mutS::lacZPSp*). Letters at the left side represent the three cointegrate classes found (A, continuous conversion; B, discontinuous conversion; and C, bipolar conversions). Numbers indicate the amount of isolates with the corresponding conversion tract. Black bars represent the extent of gene conversion tracts towards marker gain; white bars indicate gene conversion tracts showing marker loss; shaded bars are regions not subjected to gene conversion. (H, HindIII; B, BamHI; M, MaeIII; A, ApaLI; N, NarI; MI, MluI; Bc, BclI; X, XbaI.)

derivative is shown in Fig. 7. In this mutant background, no effect was seen either on the proportion of the different conversion classes (A, 62%; B 24%; C, 14%) or on the bias towards the acquisition of markers present in the incoming sequence (62 out of 68 conversion tracts displayed double marker gain). Striking differences were detected, however, for both the length of gene conversion tracts and the sectors covered by these tracts.

As shown in Fig. 8, both the continuous and discontinuous tracts obtained in a *mutS* background displayed unimodal distributions, centered at 300 bp and 150 bp, respectively. In contrast, continuous conversion tracts in the wild-type strain displayed a bimodal distribution, centered at 150 and 600 bp (Fig. 5). Moreover, while for continuous tracts in the wild-type strain, sectors encompassing the BamHI, MaeIII, ApaLI, and NarI markers are equally likely to participate in gene conversion (Fig. 5), in the *mutS* background the BamHI and MaeIII sectors participate preferentially (Fig. 8). Both the reduction in gene conversion tract length and the preferential use of two of the markers in the *mutS* background support the interpretation that the MutS mismatch repair system participates in the generation of gene conversion.

DISCUSSION

The data reported here show that, when two homologous sequences recombine to form a cointegrate, the majority of the products have a structure consistent with a gene conversion event occurring nearby. These events may encompass most of the recombining sequences. Roughly one half of the events are represented by continuous gene conversion events, while the other half show discontinuous or even bipolar events. Moreover, there is a clear bias in information transfer, favoring the conversion toward the markers present in the incoming sequence.

There are two alternatives to explain these data, one based on sister exchanges and the other postulating gene conversion. As argued before (40), apparent gene conversion events may be formed through the chance formation of double crossover events between dissimilar alleles located on sister molecules; these products, upon segregation, would be scored as convertants. Thus, these products have been dubbed apparent gene convertants, to reflect the possibility that they may have formed through reciprocal events rather than the nonreciprocal events that are the hallmark of gene conversion.

The most likely scenario for the formation of these double crossover events would be after formation of the cointegrate. In Fig. 9 we present the sequence of events required to explain

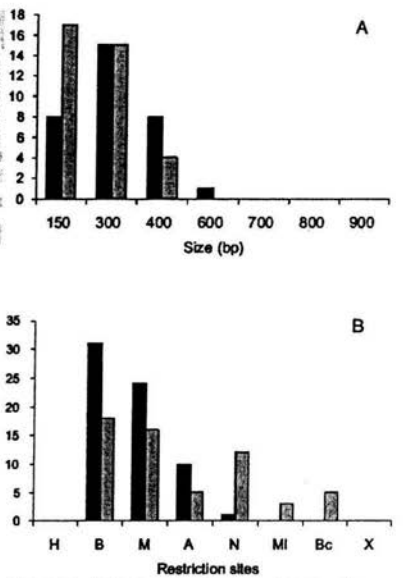

FIG. 8. Size distribution (A) and positions (B) involved in gene conversion events in *R. etli* CFNX712 (*mutS::lacZPSp*). Data are derived from Fig. 7. For both panels, vertical black bars represent the continuous tracts, while shaded bars correspond to discontinuous tracts. The letters in panel B represent the different RFLPs, as shown in Fig. 1.

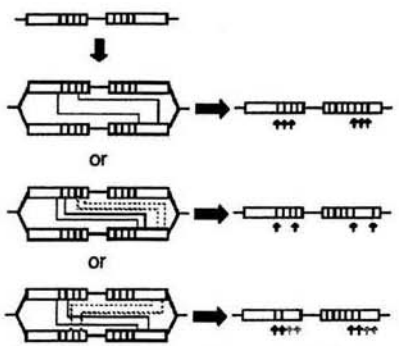

FIG. 9. Alternative model to explain gene conversion through unequal crossovers. After a cointegration event, a partially replicated molecule would generate four *nifH* sequences (open rectangles). Vertical lines within the open rectangles represent RFLPs. Pairs of continuous or broken lines joining the rectangles indicate the region of the unequal crossover. The events needed to generate continuous conversion tracts (top), discontinuous conversions (middle), and bipolar conversions (bottom) are shown. Arrows indicate regions transferred in an apparent gene conversion event, leading to marker gain (black) or marker loss (shaded). Note that generation of continuous conversions can be explained by two unequal crossover events, but both discontinuous and bipolar conversions require the participation of four unequal crossover events.

our results under this hypothesis. The initial crossover event required to form the cointegrate might generate a simple redistribution of markers. Upon replication of the cointegrate structure, unequal double crossovers may generate an apparent gene conversion segment. This might be a tenable explanation for the continuous gene conversion class (class A in Fig. 4), which represents 58% of the events observed. However, more complex rationales have to be used to explain the discontinuous (class B) and bipolar (class C) events, representing 40% of the events observed here. For these, unequal double crossovers do not suffice. As shown in Fig. 9, under this hypothesis, four crossovers are needed to generate these classes. In the case of discontinuous conversions, two crossovers would be needed to generate the first conversion tract, followed by two additional crossovers farther away from the first pair. A similar situation has to be posed for most of the cases of bipolar conversions, but in this case crossovers have to involve the four copies present in the replicated structure.

We consider the hypothesis of unequal multiple crossovers an unlikely alternative to explain our results. First, all the conversion events observed here were isolated in the absence of selection for the conversion event itself. The only selection applied was for the integration event; thus, selection for the conversion event cannot be invoked as the reason to observe multiple crossovers. Second, putative double crossover events (explaining class A convertants) were found at roughly the same frequency as the quadruple crossovers needed to explain convertants belonging to classes B and C; it is hard to envisage, in the absence of selection for conversion, why this has to occur at such a high frequency. Third, we have checked specifically for the occurrence of additional crossovers in our sample. In particular, we have evaluated the frequency of an additional unequal crossover leading to the formation of triplications. Its presence can be detected using PCR amplification with primers 3 and 2 (Fig. 2); only two out of 50 isolates revealed triplications (data not shown). Thus, unequal crossovers are infrequent enough to explain our results. Fourth, the unequal crossover model cannot explain the bias toward acquisition of the incoming sequence observed among the convertants. In fact, the unequal crossover model predicts that the double gain and double loss classes should appear at the same frequency. Moreover, the reversal of transfer bias observed upon exchanging the configuration of markers is also an unexpected feature under this model.

We think that our data are better explained by invoking the occurrence of gene conversion, perhaps generated under the double-strand break repair model (44). This model for recombination is now widely accepted for both prokaryotic and eukaryotic organisms (4). Variations of this model have been used to explain recombinational repair of collapsed replication forks, a rather frequent event in bacteria (15, 19, 24). As shown in Fig. 10, this model explains all our data in an economical way. In this model, a double-strand gap made on the resident *nifH* sequence may be repaired by the modified *nifH* sequence present in the incoming plasmid. The DNA synthesis associated with gap repair generates, in this case, a short conversion tract; migration of the Holliday junction generates heteroduplex DNA. If the mismatches in the heteroduplex segment are corrected using the strands containing the information for the RFLPs, a long continuous conversion tract will ensue (Fig. 10, class A). Correction favoring the strand containing the RFLP information in one heteroduplex and the wild-type information in the other would generate a short continuous tract (Fig. 10, class A). Thus, in this model, continuous tracts are formed in two ways: by gap repair and also by heteroduplex correction. The bimodal distribution observed for the size of continuous conversion tracts (Fig. 5) is consistent with this interpretation. Similar sizes for converted tracts have been observed in other systems, such as *Escherichia coli* (45) *Salmonella enterica* serovar Typhimurium (1, 11), and *Acinetobacter calcoaceticus* (14). F10

This model also predicts that variations in the way in which the heteroduplexes are corrected should generate additional classes. For instance, correction of an intervening marker in one heteroduplex toward the mutated sequence and to wild-type sequence in the other should generate a discontinuous conversion segment (Fig. 10, class B). Also, correction of both heteroduplexes to wild-type sequence should generate a bipolar conversion tract (Fig. 10, class C). As expected under this model, both classes were found in our data, at a cumulative frequency similar to the one for the continuous class.

The presence of continuous and discontinuous conversion segments has also been observed in other studies of gene conversion in bacteria, particularly with the *tufA-tufB* genes in *S. enterica* serovar Typhimurium (1, 11). In that case, most of the converted segments belong to the continuous class, with a minority of discontinuous events. An important difference with our work is that in the case of *S. enterica* serovar Typhimurium (1, 11), selection was applied for isolation of the conversion

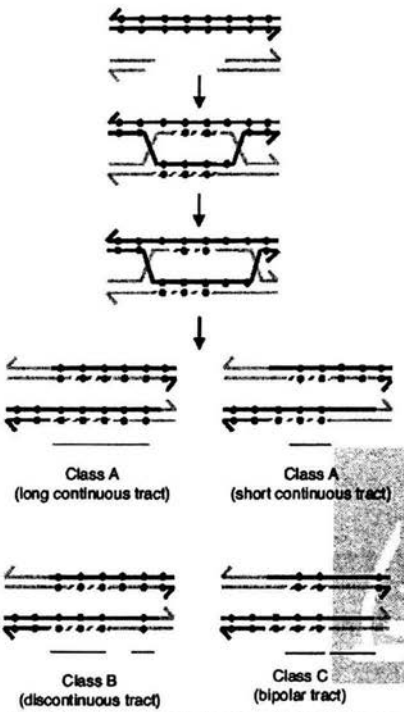

FIG. 10. Model for gene conversion of one *nifH* gene to the other initiated by a double-stranad break. Black dots in the black double-strand represent differences in nucleotide sequence (RFLPs). According to the double-strand break repair model, a double-strand cut on the recipient molecule is enlarged by degradation to a gap, followed by strand exchange and gap DNA resynthesis (discontinuous lines). After that, the DNA heteroduplex could appear because of the migration of the Holliday junction. Arrows indicate the orientation of the cuts needed to generate a cointegrate. In class A, a long continuous tract could be the result of the action of gap DNA resynthesis and the repair of both DNA heteroduplexes, favoring retention of the RFLPs. For a short continuous tract (class A), we suggest that both DNA heteroduplexes may be repaired using the wild-type sequence as a template. Thus, the short tracts may arise only by the action of gap repair. In class B, a discontinuous tract could be the result of gap resynthesis and, in some mismatches, their repair favoring the wild-type or the modified sequence. For bipolar tracts (class C), gap repair results in tracts biased toward marker gain, while mismatch repair using the wild-type sequence as a template would result in marker loss.

events; this would reduce the representation of the discontinuous class.

The reduction in the length of conversion tracts in the *mutS* background presented here lends further support to the interpretation that these tracts appear through the operation of gap filling and heteroduplex correction. These data suggest that the MutS mismatch correction system is one of the major players in mismatch correction during gene conversion. This system, however, may not be the only one to participate in heteroduplex correction. Even in its absence, classes that should be reduced in abundance, such as the discontinuous and bipolar classes, are unabated. Thus, other mismatch repair systems, such as the very short repair system, are likely to participate in heteroduplex correction. Similar conclusions were also reached for the *S. enterica* serovar Typhmurium *tufA-tufB* system (1, 11), although in that case the characterization of the conversion tracts obtained in the *mutS* background was not presented.

This model also explains the close association between crossover formation (a selected event) and gene conversion (an unselected event). Under the double-strand break repair model, the strong association observed here should be the result of a preference to start a crossover in regions with a gap at least 100 bp in size. This will frequently include at least one marker, thus forcing the repair of that gap and the conversion of the restriction site. Association between crossover formation and conversion has been observed for *S. enterica* serovar Typhimurium (1, 11).

The fact that conversion is biased towards the incoming sequence is an unexpected characteristic from our data. To our knowledge, such a strong bias has not been reported previously, with the possible exception of natural transformation in *Bacillus subtilis*. In this organism, a weak preference to incorporate incoming markers has been observed (12). According to the double-strand break repair model, the molecule that receives the double-strand break will be the one to receive information through gap repair (i.e., the one to be converted). To explain the observed bias, we have to postulate that the resident molecule, not the incoming molecule, is the one that frequently receives a double-strand cut. This preference should arise in different ways, including generation of a double-strand break by collapse of a replication fork in the resident molecule and through the operation of endonucleases that preferentially cut resident molecules, to mention but two.

The proposed mechanisms for the bias towards the donor sequence may also help to illuminate the way in which cointegrates are generated in this organism. It is commonly thought that during conjugation, DNA is transferred as a linear concatemeric array that provides flanking homology for the selected marker (6, 41). A double crossover on such a direct duplication substrate could give rise to an integrant. If that were the case, there must be a strong bias favoring the retention of the resident sequence, because the discontinuity on the linear array should be detected (and corrected) by the mismatch repair system. The fact that the observed bias is toward the donor sequences militates against this view. Thus, we favor an alternative view in which the donor sequence is first circularized in the recipient cell and then integrated, using existing discontinuities on the resident sequence.

Although the specific mechanism involved in the observed bias remains to be clarified, these findings suggest an easy way to introduce specific mutations into the *R. etli* genome using gene conversion. Moreover, if the observed bias applies to other, more natural ways of transfer, this would make *R. etli* a

rather permissive host to incorporate variations arising in a different host. Work in progress will clarify if this is the case.

In summary, we have provided evidence consistent with the operation of gene conversion in *R. etli*. Since the observed conversion tracts may frequently encompass more than half of the *nifH* gene, this process would be a good way to explain the concerted evolution among the members of this family. Moreover, our interpretation of these data lead to the prediction that both the size of the converted segment and the classes observed should be modified in backgrounds deficient in migration of the Holliday intermediates (*ruvB*, *recG*, and *radA*). Experiments are under way to test these hypotheses.

ACKNOWLEDGMENTS

We are grateful to Patricia León and Edgardo Sepúlveda for helpful discussions, Laura Cervantes and Javier Rivera for technical help and advice, Araceli Dávalos for plasmid pIC20R *oriT*, and Paul Gaytán and Eugenio López for oligonucleotide synthesis. We are also indebted to the anonymous reviewers for helpful comments on the manuscript.

Work in our laboratory is partially supported by grant 31753-N (CONACyT, México). G.S. was supported by scholarships from CONACyT, México, and Dirección General de Estudios de Posgrado, Universidad Nacional Autónoma de México.

REFERENCES

1. **Abdulkarim, F., and D. Hughes.** 1996. Homologous recombination between the *nif* genes of *Salmonella typhimurium*. J. Mol. Biol. 260:506–522.
2. **Alexeyev, M. F., I. N. Shokolenko, and T. P. Croughan.** 1995. Improved antibiotic-resistance gene cassettes and omega elements for *Escherichia coli* vector construction and in vitro deletion/insertion mutagenesis. Gene 160: 63–67.
3. **Arwidsson, O., and D. Hughes.** 2004. Evidence against reciprocal recombination as the basis for *nif* gene conversion in *Salmonella enterica* serovar *typhimurium*. J. Mol. Biol. 338:463–467.
4. **Cromie, G. A., J. C. Connelly, and D. R. Leach.** 2001. Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. Mol. Cell 8:1163–1174.
5. **Dower, G.** 2002. Molecular drive. Trends Genet. 18:587–589.
6. **Erickson, M. J., and R. J. Meyer.** 1993. The origin of greater-than-unit-length plasmids generated during bacterial conjugation. Mol. Microbiol. 7: 289–298.
7. **Felsay, R., J. Frey, and H. Krisch.** 1987. Interposon mutagenesis of soil and water bacteria: a family of DNA fragments designed for in vitro insertional mutagenesis of gram-negative bacteria. Gene 52:147–154.
8. **Flores, M., V. González, S. Brom, E. Martínez, D. Piñero, D. Romero, G. Dávila, and R. Palacios.** 1987. Reiterated DNA sequences in *Rhizobium* and *Agrobacterium* spp. J. Bacteriol. 169:5782–5788.
9. **González, V., P. Bustos, M. A. Ramírez-Romero, A. Medrano-Soto, H. Salgado, I. Hernández-González, J. C. Hernández-Celis, V. Quintero, G. Moreno-Hagelsieb, L. Girard, O. Rodríguez, M. Flores, M. A. Cevallos, J. Collado-Vides, D. Romero, and G. Dávila.** 2003. The mosaic structure of the symbiotic plasmid of *Rhizobium etli* and its relation with other symbiotic genome compartments. Genome Biol. 4:R36.
10. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166:557–580.
11. **Hughes, D.** 2000. Co-evolution of the *nif* genes links gene conversion with the generation of chromosomal inversions. J. Mol. Biol. 297:355–364.
12. **Iglesias, A., and T. A. Trautner.** 1983. Plasmid transformation in *Bacillus subtilis*: symmetry of gene conversion in transformation with a hybrid plasmid containing chromosomal DNA. Mol. Gen. Genet. 189:73–76.
13. **Kobayashi, I.** 1992. Mechanisms for gene conversion and homologous recombination: the double-strand end break repair model and the successive half crossing-over model. Adv. Biophys. 28:81–133.
14. **Kowalchuk, G. A., L. A. Gregg-Jolly, and L. N. Ornston.** 1995. Nucleotide sequences transferred by gene conversion in the bacterium *Acinetobacter calcoaceticus*. Gene 153:111–115.
15. **Kuzminov, A.** 2001. Single-strand interruptions in replicating chromosomes cause double-strand breaks. Proc. Natl. Acad. Sci. USA 98:8241–8246.
16. **Lathe, W. C., III, and P. Bork.** 2001. Evolution of *nif* genes: ancient duplication, differential loss and gene conversion. FEBS Lett. 502:113–116.
17. **Liao, D.** 2000. Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. J. Mol. Evol. 51: 305–317.
18. **Lilie, Y., and S. C. West.** 2004. Happy Holidays: 40th anniversary of the Holliday junction. Nat. Rev. Mol. Cell. Biol. 5:937–946.
19. **Marians, K. J.** 2004. Mechanisms of replication fork restart in *Escherichia coli*. Phil. Trans. R. Soc. Lond. B Biol. Sci. 29:71–77.
20. **Marsh, J. L., M. Erfle, and E. J. Wykes.** 1984. The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. Gene 32:481–485.
21. **Marshall, N. R., D. A. Daines, S. L. Lin, and K. E. Sanderson.** 1996. *Salmonella typhi* contains identical intervening sequences in all seven *rrl* genes. J. Bacteriol. 178:5323–5326.
22. **Marshall, N. R., and K. E. Sanderson.** 1996. *Salmonella typhimurium* LT2 possesses three distinct 23S rRNA intervening sequences. J. Bacteriol. 178: 2272–2278.
23. **Meinersmann, R. J., and K. L. Hiett.** 2000. Concerted evolution of duplicate *flg* genes in *Campylobacter*. Microbiology 146:2283–2290.
24. **Michel, B., M. J. Flores, E. Viguera, G. Grompone, M. Seigneur, and V. Bidnenko.** 2001. Rescue of arrested replication forks by homologous recombination. Proc. Natl. Acad. Sci. USA 98:8181–8188.
25. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
26. **Mingfu, M., and B. Robinson.** 1997. Approaches to DNA mutagenesis: an overview. Anal. Biochem. 254:157–178.
27. **Noel, K. D., A. Sánchez, L. Fernández, J. Leemans, and M. A. Cevallos.** 1984. *Rhizobium phaseoli* symbiotic mutants with transposon Tn5 insertions. J. Bacteriol. 158:148–155.
28. **Ohno, S.** 1970. Evolution by gene duplication. Springer-Verlag, Berlin, Germany.
29. **Passegrue, W., E. Lanka, P. T. Barth, D. H. Figurski, D. G. Guiney, D. Haas, D. R. Helinski, H. Schwab, V. A. Stanisich, and C. M. Thomas.** 1994. Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. J. Mol. Biol. 239:623–663.
30. **Pride, D. T., and M. J. Blaser.** 2002. Concerted evolution between duplicated genetic elements in *Helicobacter pylori*. J. Mol. Biol. 316:629–642.
31. **Quinto, C., H. de la Vega, M. Flores, L. Fernández, T. Ballado, G. Soberón, and R. Palacios.** 1982. Reiteration of nitrogen fixation gene sequences in *Rhizobium phaseoli*. Nature 299:724–726.
32. **Quinto, C., H. de la Vega, M. Flores, J. Leemans, M. A. Cevallos, M. A. Pardo, R. Azpiroz, M. L. Girard, E. Calva, and R. Palacios.** 1985. Nitrogenase reductase: a functional multigene family in *Rhizobium phaseoli*. Proc. Natl. Acad. Sci. USA 82:1170–1174.
33. **Rayssiguier, C., D. Thaler, and M. Radman.** 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. Nature 342:396–401.
34. **Rodríguez, C., and D. Romero.** 1998. Multiple recombination events maintain sequence identity among members of the nitrogenase multigene family in *Rhizobium etli*. Genetics 149:785–794.
35. **Romero, D., and S. Brom.** 2004. The symbiotic plasmids of the *Rhizobiaceae*, p. 271–290. *In* B. Funnell and G. Phillips (ed.), Plasmid biology. American Society for Microbiology, Washington, D.C.
36. **Romero, D., S. Brom, J. Martínez-Salazar, M. L. Girard, R. Palacios, and G. Dávila.** 1991. Amplification and deletion of a *nod-nif* region in the symbiotic plasmid of *Rhizobium phaseoli*. J. Bacteriol. 173:2435–2441.
37. **Romero, D., J. Martínez-Salazar, L. Girard, S. Brom, G. Dávila, R. Palacios, M. Flores, and C. Rodríguez.** 1995. Discrete amplifiable regions (amplicons) in the symbiotic plasmid of *Rhizobium etli* CFN42. J. Bacteriol. 177:973–980.
38. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
39. **Santoyo, G., and D. Romero.** Gene conversion and concerted evolution in bacterial genomes. FEMS Microbiol. Rev., in press. AQ: A
40. **Segall, A. M., and J. R. Roth.** 1994. Approaches to half-tetrad analysis in bacteria: recombination between repeated, inverse-order chromosomal sequences. Genetics 136:27–39.
41. **Silberstein, Z., and A. Cohen.** 1987. Synthesis of linear multimers of *oriC* and pBR322 derivatives in *Escherichia coli* K-12: role of recombination and replication functions. J. Bacteriol. 169:3131–3137.
42. **Simon, R.** 1984. High frequency mobilization of gram-negative bacterial replicons by the *in vivo* constructed Tn*5*-Mob transposon. Mol. Gen. Genet. 196:413–416.
43. **Stahl, F.** 1994. The Holliday junction on its thirtieth anniversary. Genetics 138:241–246.
44. **Szostak, J., T. Orr-Weaver, R. Rothstein, and F. Stahl.** 1983. The double-strand-break repair model for recombination. Cell 33:25–35.
45. **Takahashi, N. K., K. Kusano, T. Yokochi, Y. Kitamura, H. Yoshikura, and I. Kobayashi.** 1993. Genetic analysis of double-strand break repair in *Escherichia coli*. J. Bacteriol. 175:5176–5185.
46. **Thomas, C. M., and C. A. Smith.** 1987. Incompatibility group P plasmids: genetics, evolution, and use in genetic manipulation. Annu. Rev. Microbiol. 41:77–101.

47. **Valencia-Morales, E., and D. Romero.** 2000. Recombination enhancement by replication (RER) in *Rhizobium etli*. Genetics 154:971–983.
48. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268.
49. **Zahrt, T. C., G. C. Mora, and S. Maloy.** 1994. Inactivation of mismatch repair overcomes the barrier to transduction between *Salmonella typhimurium* and *Salmonella typhi*. J. Bacteriol. 176:1527–1529.
50. **Zañiga-Castillo, J., D. Romero, and J. M. Martínez-Salazar.** 2004. The recombination genes *addAB* are not restricted to Gram-positive bacteria: genetic analysis of the recombination initiation enzymes RecF and AddAB in *Rhizobium etli*. J. Bacteriol. 186:7905–7913.

ANÁLISIS DE LOS TRACTOS DE CONVERSIÓN GÉNICA EN UNA CEPA DE *Rhizobium etli* CARENTE DEL SISTEMA DE REPARACIÓN DE MISMATCHES (*mutS*)

El sistema de reparación de mismatches en el DNA (Mismatch Repair System, MMRS por sus siglas en inglés) es parte estratégica de cualquier tipo de célula, ya que forma parte importante de los mecanismos que conservan la integridad de los diversos genomas. Una de sus funciones especificas es el incremento en la fidelidad de la replicación de aproximadamente 1000 veces. Además, participa en la reparación, junto con otros sistemas, de lesiones al DNA que son provocadas por medio de rayos UV, especies reactivas de oxígeno y otro tipo de agentes (Schofield, 2003). Así mismo, juega diversos papeles en la recombinación genética, ya que se ha reportado que inhibe el intercambio genético entre especies (Rayssiguier, *et al.*, 1989). De esta manera, el MMRS por medio de la reparación de malos apareamientos entre las bases, puede mantener viable funciones tan básicas como la replicación y recombinación en el genoma de cualquier organismo.

En las bacterias, El MMRS lo conforman un grupo de proteínas que participan, desde el reconocimiento del mismatch, hasta dejar íntegra la parte reparada del DNA (Tabla 2). Una proteína indispensable que participa en la fase inicial de reparación, mediante el reconocimiento del mismatch, es MutS, la cual es codificada por el gen *mutS*.

Como habíamos mencionado anteriormente, de acuerdo al modelo de recombinación y reparación por cortes en doble cadena, se predice que durante el evento de recombinación se pueden formar regiones heteroduplex, las cuales se forman por malos apareamientos entre las bases. Así, estos malos apareamientos pueden ser reconocidos por el MMRS para ser corregidos, por lo que pueden influir en la estructura de los tractos de conversión génica. Por lo tanto, y como en

nuestro sistema de recombinación de dos plásmidos estamos utilizando dos versiones de las secuencias del gen *nifH*, las cuales varían en secuencia en un 1.6%, es probable que se formen regiones heteroduplex como intermediarios de la recombinación y conversión génica. De esta manera, decidimos explorar la función del gen *mutS* de *Rhizobium etli* en la formación de tractos de conversión génica.

Tabla 2. Proteínas que participan en la reparación de malos apareamientos en el DNA.

PROCESO DONDE ACTÚA	PROTEÍNA	ACTIVIDAD QUE REALIZA
Inciación	MutS	Reconoce y se une al mismatch
Inciación	MutL	Interacciona con MutS para activar la actividad nucleasa de MutH
Inciación	MutH	Corta la cadena no metilada en el sitio GATC hemimetilado
Escisión	Helicasa II	(UvrD) Desenrolla el DNA previo a la escisión
Escisión	RecJ ExoVII	Son requeridos para la escisión 3´-5´ entre el corte y el mismatch
Escisión	ExoI ExoX	Son requeridos para la escisión 5´-3´ entre el corte y el mismatch
Resíntesis	pol III	DNA polimerasa requerida para la reparación del corte por medio síntesis del DNA
Resíntesis	SSB	Proteínas de unión a DNA, ayudan a la escisión y resíntesis
Resíntesis	Ligasa	Sella los cortes en el DNA

(Modificado de Schofield y Hsieh, 2003).

Para ello, utilizamos una cepa de *Rhizobium etli* CFN42 mutada en el gen *mutS* (Martínez-Salazar, *et al.*, 2004), proporcionada por el Dr. Jaime Martínez. Posteriormente se generó en ella una deleción de la región simbiótica del pSim, desde *nifHDK*a hasta *nifHDK*b, por medio del sistema RER (Recombination enhancement by replication) (Valencia-Morales y Romero, 2000). Tal deleción la realizó el Dr. Cesar Rodríguez. Esta cepa fue nombrada CFNX712.

Para realizar las cruzas con la cepa CFNX712, se modificó el plásmido pMC11, ya que se cambio la resistencia de tetraciclina por kanamicina, derivando en el plásmido pJGus28.

Como se describe previamente en el artículo, se realizaron 10 cruzas independientes entre el plásmido pJGus28 y la cepa CFNX712, aislando nuevamente 50 recombinantes. El análisis se dio por medio de PCR, seguido por el análisis con enzimas de restricción de ambos productos (*nifH* izquierdo y derecho) del evento de cointegración. Esto nos llevó a buscar un patrón de conversión mediante la pérdida o ganancia de sitios, tal y como se describe en el articulo.

Con este análisis se observó que la frecuencia de recombinación (cointegración) se incremento aproximadamente 30 veces en la cepa *mutS* (CFNX712), comparado con la cepa CFNX55 (Figura 6), donde el MMRS es funcional. Esto debido a que el MMRS disminuye la frecuencia de recombinación entre secuencias que no son completamente homólogas (Rayssiguier, *et al.*, 1989), tema que se discutirá con más detalle en la discusión.

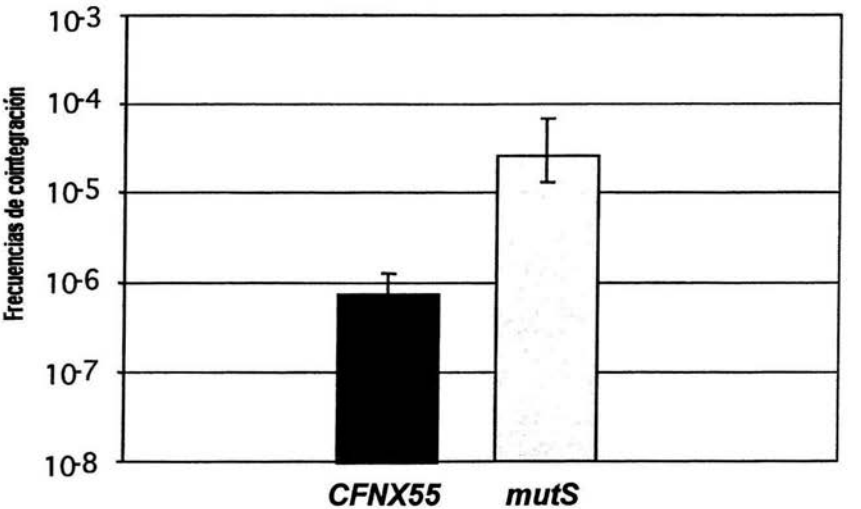

Figura 6. Frecuencias de cointegración del plásmido pJGus28 y la cepa *mutS* (CFNX712). Los datos son el promedio de 10 repeticiones independientes.

Respecto a la estructura de los tractos, nuevamente encontramos regiones continuas de conversión (62%) (Figura 7, A). Así mismo, se observaron tractos discontinuos (24%) (Figura 7, B) y bipolares (14%) (Figura 7, C) en proporciones similares con respecto de la cepa CFNX55 (Figura 7). En este mismo sentido encontramos que en cada evento de cointegración (crossover) analizado, había algún tracto de conversión génica (100% de los casos), por lo que se corroboró la fuerte asociación entre estos dos eventos de recombinación. En este caso no se detectó evento alguno de cointegración sin la asociación de conversión génica (Figura 7).

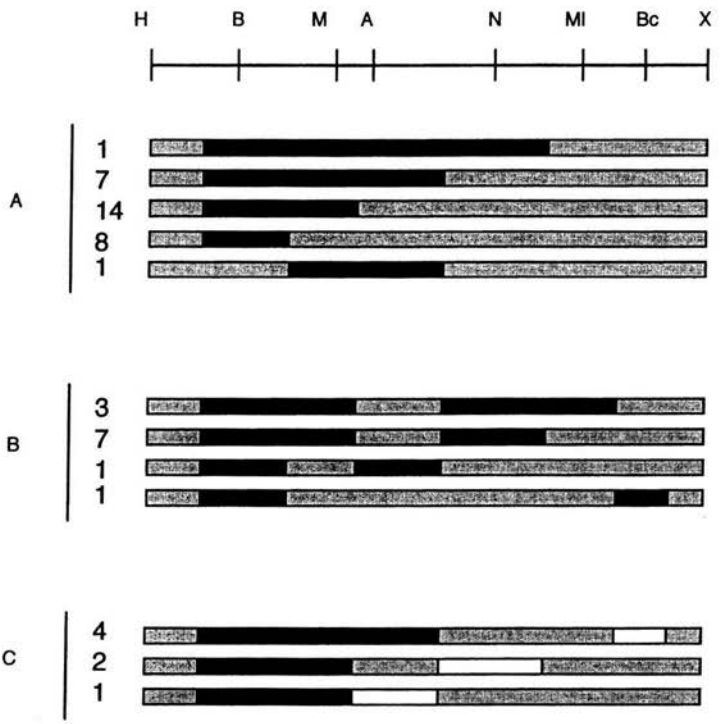

Figura 7. Anatomía de los tractos de conversión en la cepa *mutS* (CFNX712). Arriba se muestra el mapa de *nifH* con los RFLPs como referencia. Las letras de la izquierda muestran los tractos continuos (A), discontinuos (B) y bipolares (C). Los números indican la cantidad de recombinantes aisladas con el tracto correspondiente. Las barras negras indican la extensión del tracto con ganancia de sitios; las barras blancas la doble pérdida de sitios; mientras que en los segmentos grises no ocurrió conversión.

Interesantemente, se observó que los tractos fueron mucho más cortos en la cepa *mutS*, comparado con cepa CFNX55, ya que hubo una reducción del 40% en la longitud. Así, la longitud media de los tractos en la cepa silvestre fue de 473 pb, mientras que para la cepa *mutS* fue de 290 pb (Figura 8). Para este análisis comparativo se tomó en cuenta sólo la longitud de los tractos continuos, sin incluir los discontinuos y bipolares.

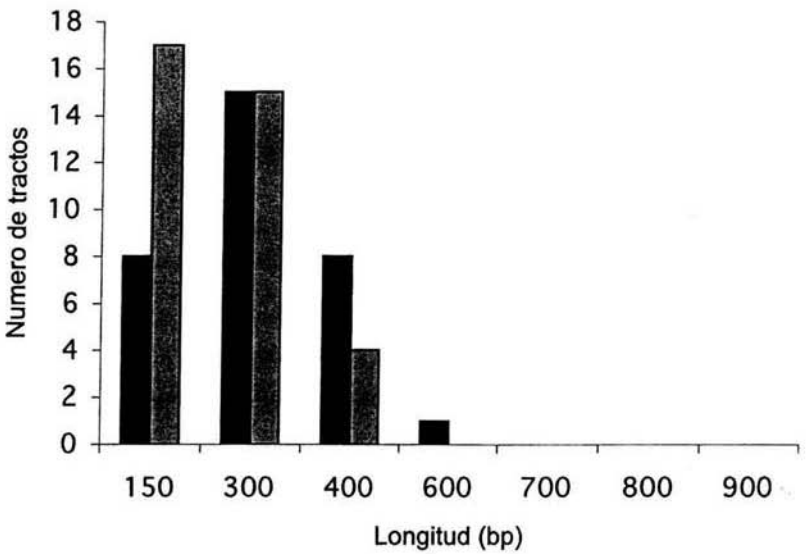

Figura 8. Longitud de los tractos de conversión génica en la cepa *mutS* (CFNX712). Las barras negras indican los tractos discontinuos; mientras que las barras grises los tractos continuos.

Como consecuencia de la disminución en la longitud de los tractos, se observó que las regiones que participaron en los tractos de conversión también fueron modificadas en la cepa *mutS* (Figura 9). Sin embargo, algunos RFLPs participaron mucho más activamente en los eventos de conversión, tal es el caso de BamHI y MaeIII.

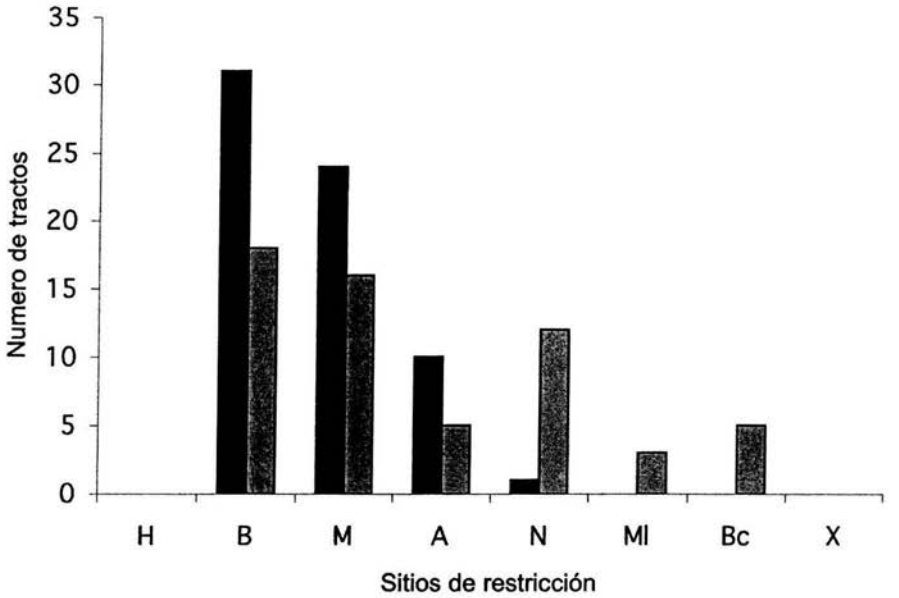

Figura 9. RFLPs (sitios de restricción) que participaron en conversión génica en la cepa *mutS* (CFNX712).

DISCUSIÓN

Para conocer la longitud de los tractos de conversión en la bacteria *Rhizobium etli* hemos diseñado un sistema de recombinación de dos plásmidos. Este sistema consiste en un evento de recombinación (cointegración o crossover) entre dos secuencias (*nifH* wt y *nifH* con RFLPs) que contienen 1.6% de heterología. Esto nos permitió observar tractos con anatomías diversas; tractos continuos, discontinuos y bipolares. Una cantidad muy pequeña (1 de 100) de los tractos no tuvo asociación alguna con eventos de conversión. Esto nos demuestra que conversión génica está muy asociada a eventos de recombinación tipo crossover.

Desde hace algún tiempo se argumenta que los eventos de conversión génica detectados en bacterias son producto de múltiples eventos de recombinación (Segall y Roth, 1994). Por tal motivo, se ha sugerido que se denominen " conversión génica aparente ". Los resultados que nosotros obtuvimos mediante los eventos de recombinación en nuestro sistema no apoyan esta hipótesis, o, al menos no podrían explicar la gran mayoría de nuestros tractos de conversión. En primer lugar, todos los eventos de conversión que observamos no tienen algún tipo de selección alguna, para ello se ha aplicado selección únicamente para recuperar cointegrados, no conversión asociada. Por lo tanto, no se puede decir que exista algún tipo de selección para forzar la recombinación de eventos múltiples, y así obtener rastros de conversión génica. Segundo, para obtener conversiones de clase A se requieren al menos dobles entrecruzamientos, mientras que para los tractos de clase B (discontínuos) y C (bipolares), se necesitarían cuádruples entrecruzamiento. Por ello, es difícil imaginar que dobles y cuádruples entrecruzamientos se puedan dar a una misma frecuencia. Tercero, por medio de oligonucleótidos específicos se podrían detectar entrecruzamientos adicionales que provocaran duplicaciones (ver artículo). Esto se llevó a cabo y únicamente se detectaron eventos de duplicación en sólo 2 de 50 convertantes, por lo que este tipo de entrecruzamientos adicionales son muy poco frecuentes de

que se lleven a acabo en nuestro sistema, y que ello explique el origen de conversión. Cuarto, los entrecruzamientos adicionales no pueden explicar la desviación que observamos sobre la ganancia de sitios de restricción, ya que este tipo de predicciones dice que se tendría la misma cantidad de ganancia versus pérdida de sitios.

Debido a los motivos anteriores, nosotros proponemos que nuestros datos son más fácil de explicar si nos basamos en el modelo de recombinación y reparación de cortes en doble cadena (DSBRM, pos sus siglas en inglés) (Szostak *et al*, 1983).

De acuerdo a la figura 10, proponemos que los tractos de conversión continuos y largos (clase A) se pueden originar con reparación del corte y degradación de doble cadena (gap), además de la probable acción del sistema de reparación de mismatches, favoreciendo la extensión de la conversión. Asimismo, los tractos cortos continuos (clase A) se explicarían por la acción única de reparación del gap formado. Para la clase B, se daría nuevamente la reparación del gap, pero al repararse los mismatches en una dirección y viceversa, se formarían tractos discontinuos. En este mismo sentido, si los

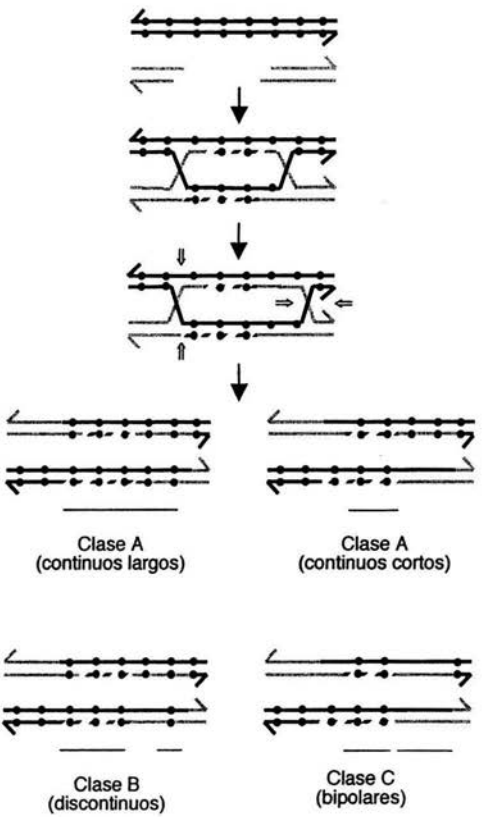

Figura 10. Modelo basado en el DSBRM que predice el origen de los tractos de conversión génica asociados a crossovers en *Rhizobium etli*. (Modificado de Santoyo y Romero, 2005)

mismatches se repararán en una sola dirección en ambas cadenas, favoreciendo la pérdida de sitios, se crearía un tracto bipolar.

Por lo tanto, la diversidad de tractos encontrados en nuestro análisis se explica de una manera más simple mediante el modelo de recombinación y reparación de cortes en doble cadena, así como la polaridad hacia la ganancia de sitios, lo cual se explicará en detalle más adelante. Finalmente, este modelo es ampliamente estudiado y aceptado en organismos procariotes y eucariotes (Cromie, *et al.*, 2001; Santoyo y Romero, 2005).

En nuestro proyecto de investigación existía la posibilidad de obtener paridad en la ganancia y pérdida de nuestros marcadores (sitios de restricción). Sin embargo, encontramos un dato inesperado: una fuerte tendencia hacia la ganancia de sitios de restricción. De hecho, de 50 recombinantes analizadas encontramos que 49 (98%) tenían algún rastro de doble ganancia de sitios. Este dato podía ser explicado por dos hipótesis; una es que al introducir los sitios de restricción, estos, por sus características intrínsecas serían reparados favoreciendo su presencia; dos, que el gen *nifH* residente en el plásmido simbiótico es preferentemente cortado, y por lo tanto, susceptible de ser reparado a través de la recombinación con el gen *nifH* entrante con los RFLPs. Así, favoreciendo la transferencia de información de la secuencia entrante (RFLPs). Es interesante notar que al intercambiar los marcadores, ahora se tuvo una preferencia hacia la pérdida de sitios (ver artículo). En la figura 11 se ilustran un poco mejor estos datos.

Esto resultados corroboran la hipótesis dos, donde no existe nada particular en la introducción de sitios, sino que es probable que el *nifH* residente sea el receptor de información por medio de conversión, sin importar el tipo de secuencia donadora. Como mencionamos anteriormente, el modelos de recombinación y reparación de cortes en doble cadena puede explicar la desviación hacia la ganancia y pérdida de sitios en ambas cepas (CFNX55 y CFNX704).

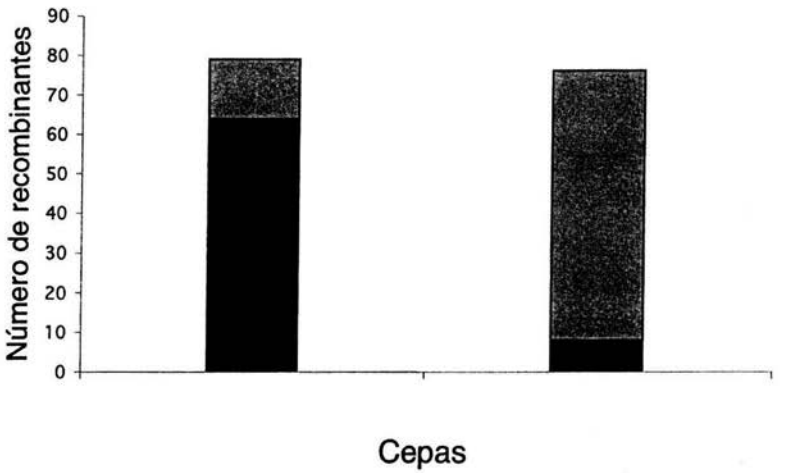

Figura 11. Relación pérdida versus ganancia de sitios de restricción en las cepas CFNX55 (Barra izquierda) y CFNX704 (Barra derecha). En color negro se representa en número de recombinantes que ganaron sitios. En color gris es el número de recombinantes que perdieron sitios.

Al parecer, todos los seres vivos contienen en sus genomas mecanismos que mantienen la integridad de su DNA, heredándolo con el menor número de errores posibles. Tal es el caso de los sistemas de reparación de mismatches, que se encuentra conservado desde bacterias hasta humanos, aunque cada especie conserva ciertas particularidades (Cromie, *et al.*, 2001). Sin embargo, el MMRS no sólo participa en la corrección de malos apareamientos y la corrección de otro tipo de daños en el DNA, sino que también se ha visto como una barrera para el intercambio genético entre distintas especies debido a sus diferencias en la secuencia nucleotídica de sus genomas. Tal es el caso de *Escherichia coli* y *Salmonella typhimurium* (Rayssiguier, *et al.*, 1989).

En nuestro sistema para detectar eventos de conversión asociados a crossovers utilizamos secuencias que varían 1.6%, por lo que la frecuencia de

recombinación se ve disminuida considerablemente (ver artículo). Sin embargo, cuando anulamos la función del gen *mutS* en *R. etli*, la frecuencia de recombinación entre estas secuencias se incrementó 30 veces (Figura 6). Para detallar el mecanismo por el cual el MMRS puede inhibir el intercambio genético entre secuencias de DNA heterólogas, se han propuesto varías hipótesis. (i) De acuerdo a Evans y Alani (2000), durante el intercambio de información genética entre dos cadenas de DNA, las cuales contienen algunas diferencias en secuencia nucleotídica, se forma el DNA heteroduplex, y por lo tanto los malos apareamientos entre las bases. Estos mismatches son reconocidos por el sistema de reparación, seguido de un corte en cadena sencilla. Posteriormente pueden escindirse desde unos cuantos nucleótidos hasta cientos de pares de bases hasta encontrar el próximo GATC hemimetilado (Evans y Alani, 2000). Esto podría destruir el intermediario y la unión de ambas cadenas que participan en el proceso de recombinación (Rayssiguier, *et al.*, 1989). (ii) Una segunda hipótesis es que las helicasas que participan comúnmente en la reparación de mismatches podrían desenrollar y desaparear las cadenas alineadas, destruyendo así el intermediario de recombinación. (iii) La unión de las diversas proteínas que participan en el MMRS podrían bloquear la formación del DNA heteroduplex, posiblemente a través de la interacción con la maquinaría de recombinación (Evans y Alani, 2000). Esta última hipótesis se ve respaldada por un trabajo hecho por Worth y colaboradores (1994) donde observaron que proteínas como MutS y MutL inhiben la acción catalizadora de RecA, la cual participa en el reconocimiento y transferencia de la punta 3´ con la región homóloga, un paso inicial y fundamental para el inicio de la recombinación.

Robin Holliday propuso un modelo de recombinación donde conversión génica se explicaba por medio de la reparación de regiones heteroduplex. Desde entonces, esta característica ha sido conservada en posteriores modelos de recombinación, como el de reparación de cortes en doble cadena. De esta manera, es posible que nuestros tractos de conversión se generen por reparación de cortes en doble cadena y/o de la corrección de mismatches en zonas heteroduplex. Por lo tanto, al analizar nosotros los tractos de conversión en una

cepa de *R. etli mutS* (CFNX712), los resultados nos sugieren que el sistema de reparación de mismatches está participando en la estructura y longitud de los segmentos de conversión (Figura 7 y 8). Se sugiere que los tractos son más largos en una cepa silvestre versus la mutante *mutS*, ya que al no estar activo el MMRS los probables mismatches que eran reparados favoreciendo la cadena entrante con los RFLPs ahora no son corregidos. Por lo tanto, la longitud de los tractos, de acuerdo al modelos de ruptura y doble cadena podrían ser únicamente el resultado de la reparación del gap, y no de corrección de mismatches.

Es interesante recalcar que los sitios de BamHI y MaeIII incrementaron su participación en los tractos de conversión, mientras que el resto disminuyeron considerablemente (Figura 7). Esto nos puede sugerir que la región 5′ del gen *nifH* podría ser el sitio de inicio del proceso de recombinación, seguido por la migración del intermediario de Holliday hasta la zona 3′ del gen. Por lo tanto, al no tener activo el sistema de reparación de mismatches, las regiones heteroduplex formadas por la migración del intermediario de Holliday no serían reparadas, disminuyendo su participación de estos sitios (Figura 9).

Al anular la actividad de *mutS* en *R. etli*, esperábamos que los tractos discontinuos y bipolares hubieran disminuido en proporción respecto a los tractos continuos. Sin embargo esto no fue así, ya que observamos una proporción muy similar con respecto de la cepa silvestre. Así mismo, es necesario recalcar que se han detectado tractos discontinuos y complejos (bipolares) de conversión génica en cepas carentes del MMRS de la levadura *Saccharomyces cerevisiae*, y que aún este tipo de tractos son más frecuentes en este tipo de mutantes (Chen y Jinks-Robertson, 1998).

Por otra parte, se podría proponer la posibilidad de que los tractos discontinuos y bipolares observados en la cepa *R. etli mutS*, sean el resultado de segundos eventos de recombinación o entrecruzamientos desiguales. Sin embargo, esta hipótesis es muy remota, ya que para generar la gran mayoría de los tractos discontinuos y bipolares, se requerirían al menos de cuatro entrecruzamientos desiguales. Estos eventos tendrían que suceder posterior al evento de selección del cointegrado, por lo que las frecuencias de tener un solo

evento de recombinación versus dos o cuatro, tendrían que darse a la misma frecuencia. Esto es sumamente improbable, aún cuando se trate de una cepa *mutS*. Adicionalmente, todo ello tendría que darse sin algún elemento de selección hacia los tractos de conversión. Finalmente, este modelo no predice de manera clara la desviación observada hacia la ganancia de sitios, algo que se observó de igual manera en la cepa silvestre (ver artículo).

Por lo tanto, nosotros proponemos que la generación de tractos discontinuos y bipolares en una cepa *mutS*, podrían ser el resultado de la actividad de otros sistemas de reparación de mismatches (Modrich, 1989). Una vía de reparación de mismatches podría ser por medio de parches cortos (Abdulkarim and Hughes, 1996), donde el gen *mutY* (una adenina glicosilasa) puede reconocer y reparar mismatches A-G y C-A (Lu, 2000). Adicionalmente, la timina glicosilasa repara mismatches G-T (Norbury y Hickson, 2001). Cabe destacar que al introducir los RFLPs en el gen *nifH*, en los sitios BamH, MaeIII y MluI se generarían mismatches G-T al momento de recombinar (apareamiento de cadenas), los cuales podrían ser reparados por la timina glicosilasa. Además, los cambios que realizamos para generar los sitios NarI y MluI, podrían crear mismatches C-A con la secuencia silvestre, por lo que podrían ser reparados por el producto del gen *mutY*. Estas vías de reparación de mismatches han sido discutidas por otros autores para proponer el origen de tractos discontinuos en los genes *tuf* de *Salmonella typhimurium*, aunque sus datos no distinguen la acción entre la vía de reparación de parches largos (*mutSHL*) y parches cortos (*mutB* en *S. typhimurium*) (Abdulkarim y Hughes, 1996).

ANÁLISIS DE ELEMENTOS GENÉTICOS ADICIONALES QUE INTERVIENEN EN LA ANATOMÍA DE LOS TRACTOS DE CONVERSIÓN EN *Rhizobium etli*

En nuestros resultados adicionales analizamos el papel del gen *mutS* en la estructura de los tractos de conversión, observando una disminución en su longitud de un 40%. Eso nos sugiere que el gen *mutS* de *R. etli* podría estar jugando un papel en la reparación del DNA heteroduplex (mismatches). Sin embargo, para que se forme este tipo de DNA híbrido, además del papel que esta jugando el sistema de reparación de malos apareamientos de *Rhizobium etli* (*mutS* y otros genes), es probable que el intermediario de Holliday haya migrado con la ayuda de proteínas como RuvA, RuvB, RecG y RadA. Así, la migración del intermediario de Holliday puede provocar la generación de tractos cortos o largos, dependiendo de la orientación hacia donde migre a partir del corte en doble cadena. Es por ello interesante demostrar el papel de estas otras proteínas que participan en la migración del intermediario de Holliday, así como su papel en los tractos de conversión génica. La manera que se podrían caracterizar es por medio de la transferencia del plásmido pJGus28, el cual contiene el *nifH* mutagenizado, hacia cepas mutantes en genes *ruvB*, *recG* y *radA* de *Rhizobium etli*. Así mismo el análisis de dobles y una triple mutante de igual manera serían muy interesantes de analizar, además de muy informativas sobre la función de estos genes. Es necesario mencionar que hasta donde conocemos, en ningún otro organismo se ha analizado el papel de estos genes (*ruvB*, *recG* y *radA*) y su papel en la formación de tractos de conversión génica.

Además de la migración de la unión de Holliday y la reparación de DNA heteroduplex, existen algunas otras características que participan en la anatomía de los tractos de conversión génica. Proteínas que participan en el inicio del proceso de recombinación, tales como *recBCD* o *recF* (que en el primer caso en *Rhizobium etli* su homólogo es *addAB*) podrían influir en la formación y extensión

del gap. Probablemente su actividad influya en la estructura de los eventos de conversión. Estos genes que participan en el inicio del proceso de recombinación han sido ya aislados y caracterizados en *Rhizobium etli* (Zúñiga, *et al.*, 2004), por lo que un análisis comparativo sobre su participación en la generación de tractos conversión asociados a crossovers sería igualmente interesante.

CONVERSIÓN GÉNICA A NIVEL GENÓMICO

El desarrollo de la genómica en la última década del siglo pasado cambió por completo el estudio de los seres vivos. Hasta este momento hemos sido testigos de la revolución que ha causado en el mundo biológico mediante el estudio de genomas completos y como ha cambiado nuestra perspectiva de los sistemas biológicos que estudiamos. Respecto a nuestro tema, es casi imposible no preguntarse si se pueden analizar eventos de conversión génica a nivel genómico. La respuesta es que si se puede lograr mediante el análisis de familias multigénicas en genomas completos. Por una lado, se podrían caracterizar en un primer caso el total de las familias multigénicas en el genoma de *Rhizobium etli*, incluyendo plásmidos y el cromosoma, el cual ha sido recientemente descifrado (González, *et al.*, resultados no publicados). Algo que es indispensable es que las familias contengan elementos con un alto porcentaje de similitud en secuencia para que existan posibilidades altas de que se hayan sometido a eventos de conversión. Una vez que se haya realizado este tipo de análisis, se podría analizar la información a programas computacionales que detectan eventos de conversión entre secuencias repetidas, tales como GENECONV (Sawyer, 1989). Este tipo de programa ha sido utilizado para conocer eventos de conversión en familias multigénicas de la levadura *Saccharomyces cerevisiae* (Drouin, 2002) y cepas de *Escherichia coli* (Morris y Drouin, 2003). Si el programa no puede detectar algún tipo de conversión, se podría hacer semi-manualmente la comparación mediante programas de comparación de secuencias nucleotídicas de diversas cepas o especies.

Por otra parte, para realizar análisis de conversión génica genómica, el análisis filogenético (o de inferencia molecular, Santoyo y Romero, 2005) de familias multigénicas de varias cepas (cuyos genomas hayan sido completamente secuenciados) de una misma especie podría de igual manera darnos un indicio de eventos de conversión génica. Hasta el momento existen aproximadamente 22 especies de eubacterias y arqueobacterias en las cuales se ha reportado el genoma completo de más de una cepa, en algunos casos hasta 5 cepas de la misma especie (datos del *National Center for Biotechnology Information*). De esta manera, comparando los genes parálogos de una cepa con los ortólogos de otras, nos podría indicar si sus miembros están sometidos a evolución concertada, lo que podría ser indicativo de que conversión génica estuviera generando tal identidad entre las diversas copias.

MODIFICANDO EL GENOMA DE *Rhizobium etli* POR MEDIO DE CONVERSIÓN GÉNICA

Como habíamos mencionado anteriormente, el fenómeno de polaridad de conversión génica en nuestro sistema nos sugiere que el DNA entrante por medio de conjugación es el que preferentemente convierte al residente. Esto lo corroboramos en los experimentos donde intercambiamos los marcadores (cepa CFNX704).

Esta característica que observamos en nuestro trabajo podría ser de utilidad para generar cualquier tipo de modificación genética que deseemos en bacterias no transformables como *R. etli*. Para ello se diseñaría un vector suicida con cierta homología en los extremos de la secuencia a modificar. Este vector se transferiría por conjugación a *R. etli* y se seleccionarían aquellas recombinantes sencillas. Posteriormente se corroboraría la modificación deseada por algún método (por ejemplo PCR) para posteriormente seleccionar la escisión del vector, sin dejar huella alguna de marcadores.

Mediante esta metodología se podrían crear diversas modificaciones, tales como deleciones en genes para conocer su función, inserciones para crear

fusiones de genes con promotores transcripcionales. Así mismo se intercambiarían secuencias promotoras, las cuales tuvieran algunas diferencias en sus bases para conocer la función de cada una de ellas o generar modificaciones genéticas en promotores que incrementen su actividad transcripcional, entre muchos otros ejemplos. El intercambiar cualquier tipo de secuencia sin dejar alguna huella de marcadores podría ser favorable en la generación de organismos modificados genéticamente, ya que no tendrían rastro de algún tipo de antibiótico o marcador no deseado en su genoma. Es importante mencionar que actualmente ya se han desarrollado sistemas similares para crear modificaciones genéticas en otros organismos (Court, *et al.*, 2002; Oppenheim, *et al.*, 2004). Sin embargo, en *R. etli* su principal limitante es que no es transformable por medios químicos y por medio de electroporación es muy baja su eficiencia (Dávalos y Romero, com. pers.).

Adicionalmente durante la realización de este trabajo de Doctorado se colaboró en un proyecto con la Dra. Lourdes Girard para estudiar la región regulatoria del gen *fixNd* de *R. etli*. Para ello se generó una deleción de 234 pb en la zona regulatoria del gen *fixNd*. Hasta este momento podemos decir que sí es posible deletar regiones específicas de algún replicón por medio de conversión génica, en particular en bacterias no transformables como *R. etli*. Sin embargo, sería interesante investigar que otros tamaños de deleciones o inserciones se pueden generar, así como determinar cuál es la longitud mínima indispensable de homología para recombinar en las orillas de la región de interés por modificar. Conocer este parámetro facilitaría posteriores modificaciones y se tendría un dato adicional de algo que hasta el momento es desconocido en bacterias como *Rhizobium etli*. Así mismo, este parámetro sería sumamente útil si en el futuro se genera un método de transformación eficiente en *R. etli*, ya que se podría modificar por medio de fragmentos de PCR (Court, *et al.*, 2002; Oppenheim, *et al.*, 2004), pero sería necesario de cualquier manera conocer el fragmento mínimo eficiente para recombinar. De ser posible esto, se ahorraría mucho tiempo y trabajo en la generación de mutantes, además de que abriría nuevas posibilidades para el análisis genómico funcional de *R. etli*.

CONCLUSIONES

1. La recombinación tipo crossover está fuertemente ligada a eventos de conversión génica.

2. La longitud de los tractos de conversión génica de la familia multigénica *nifH* de *Rhizobium etli* varían desde150 pb hasta 800 pb.

3. Los eventos de conversión génica mostraron una fuerte desviación hacia la ganancia de RFLPs o sitios de restricción, en lugar de perderlos.

4. La heterología de 1.6% entre las dos secuencias de *nifH* redujo la frecuencias de recombinación (cointegración) aproximadamente 100 veces.

5. Los tractos de conversión génica observados en nuestro sistema de cointegración se explican mejor por medio de conversión génica verdadera y no por múltiples eventos de recombinación.

6. Los tractos de conversión génica analizados en la cepa de *R. etli mutS* se redujeron un 40% en su longitud respecto a la cepa silvestre.

7. El sistema de reparación de mismatches en *R. etli* esta participando en la anatomía de los tractos de conversión, aunque no se excluye la participación de otros sistemas.

8. Las frecuencias de recombinación entre secuencias que divergen 1.6% se incrementan considerablemente en una cepa de *R. etli mutS*.

9. Nuestro sistema para detectar eventos de conversión génica asociados a crossovers, abre la posibilidad de analizar otros elementos genéticos que participan en la estructura de los tractos de conversión génica, así como la implementación de un sistema que permita modificar sistemáticamente el genoma de *Rhizobium etli*.

EPÍLOGO

Reflecting on the prospects for further advance, one may be tempted to take an attitude of romantic pessimism: all that remains is either applications or epistemological disquisition. Such was the mood in physics around 1900-after Maxwell and Boltzmann, and just before Curie, Planck, Rutherford, Einstein and Bohr entered the picture. ***And so there is hope for the young biologists who dream of discovery***.

Salvador Luria, 1986

REFERENCIAS

1. Abdulkarim, F. y Hughes, D. (1996). Homologous recombination between the *tuf* genes of *Salmonella typhimurium*. *J. Mol. Biol.* 260, 506-522.

2. Boocock, G., Morrison, J., Popovic, M., Richards, N., Ellis, L., Durie, P. y Rommens, J. (2003). Mutations in *SBDS* are associated with Shwachman-Diamond syndrome. *Nat. Genet.* 33, 97-101.

3. Brayton, K.A., Palmer, G.H., Lundgren, A., Yi, J. y Barbet, A.F. (2002). Antigenic variation of *Anaplasma marginale msp2* occurs by combinatorial gene conversion. *Mol. Microbiol.* 43, 1151-1159.

4. Brom, S. y Romero, D. (2004). The symbiotic plasmids of the *Rhizobiaceae*. Plasmid Biology. (Funnell, B.E. y Phillips, G.J., eds.) ASM Press.

5. Burghes, A.H. (1997). When is a deletion not a deletion? When it is converted. *Am. J. Hum. Genet.* 61, 9-15.

6. Campbell, L., Potter, A., Ignatius, J., Dubowitz, V. y Davies, K. (1997). Genomic variation and gene conversion in spinal muscular atrophy: implications for disease process and clinical phenotype. *Am. J. Hum. Genet.* 61, 40-50.

7. Centurion-Lara. A., LaFond, R.E., Hevner, K., Godornes, C., Molini, B.J., Van Boris, W.C. y Lukehart, S.A. (2004). Gene conversion: a mechanism for generation of heterogeneity in the *tprK* gene of *Treponema pallidum* during infection. *Mol. Microbiol.* 52,1579-96.

8. Chen, J.M. y Ferec, C. (2000). Gene conversion-like missense mutations in the human cationic trypsinogen gene and insights into the molecular evolution of the human trypsinogen family. *Mol. Genet. Metabol.* 71, 463-469.

9. Chen, W. y Jinks-Robertson, S. (1998). Mismatch repair proteins regulate heteroduplex formation during mitotic recombination in yeast. *Mol. Cell. Biol.* 18, 6525-6537.

10. Court, D.L., Sawitzke, J.A. y Thomason, L.C. (2002). Genetic engineering using homologous recombination. *Annu. Rev. Genet.* 36, 361-88.

11. Cromie, G.A., Connelly, J.C. y Leach, D.R. (2001). Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. *Mol. Cell* 8, 1163–1174.

12. Dávila, G., Brom, S., Collado, J., Hernández, G., Mora, J., Palacios, R. y Romero, D. (2000). Genomics of *Rhizobium etli*, Prokaryotic nitrogen fixation: a model system for the analysis of a biological process. (Triplett, E.W. ed.). Horizon Scientific Press.

13. Deitsch, K.W., Moxon, E.R. y Wellems, T.E. (1997). Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol. Mol. Biol. Rev.* 61, 281-93.

14. Derr, L.K. y Strathern, J. (1993). A role for reverse transcripts in gene conversion. *Nature* 361, 170-173.

15. Diaz-Font, A., Cormand, B., Blanco, M., Chamoles, N., Chabas, A., Grinberg. D. y Vilageliu, L. (2003). Gene rearrangements in the glucocerebrosidase-metaxin region giving rise to disease-causing

mutations and polymorphisms. Analysis of 25 Rec NciI alleles in Gaucher disease patients. *Hum. Genet.* 112, 426-9.

16. Dover, G. (2002). Molecular drive. *Trends Genet.* 18, 587-589.

17. Drouin, G. (2002). Characterization of the gene conversions between the multigene family members of the yeast genome. *J. Mol. Evol.* 55, 14-23.

18. Evans, E. y Alani, E. (2000). Roles for mismatch repair factors in regulating genetic recombination. *Mol. Cell. Biol.* 20, 7839-7844.

19. Flores, M., Mavingui, P., Perret, X., Broughton, W.J., Romero, D., Hernández, G., Dávila, G. y Palacios, R. (2000). Prediction, identification, and artificial selection of DNA rearrangements in *Rhizobium*: toward a natural genomic design. *Proc. Natl. Acad. Sci. U. S. A.* 97, 9138-9143.

20. Flores, M., Gonzalez, V., Brom, S., Martinez, E., Pinero, D., Romero, D., Davila, G. y Palacios, R. (1987). Reiterated DNA sequences in *Rhizobium* and *Agrobacterium* spp. *J. Bacteriol.* 169, 5782-5788.

21. Frugier, T., Nicole, S., Cifuentes-Diaz, C. y Melki, J. (2002). The molecular bases of spinal muscular atrophy. *Curr. Opin. Genet. Dev.* 12, 294-8.

22. Girard, M.L., Flores, M., Brom, S,. Romero, D., Palacios, R. y Davila, G. (1991). Structural complexity of the symbiotic plasmid of *Rhizobium leguminosarum* bv. *phaseoli*. *J. Bacteriol.* 173, 2411-9.

23. González, V., Bustos, P., Ramírez-Romero, M.A., Medrano-Soto, A., Salgado, H., Hernández-Gonzalez, I., Hernández-Celis, J.C., Quintero, V., Moreno-Hagelsieb, G., Girard, L., Rodríguez, O., Flores, M., Cevallos, M.A., Collado-Vides, J., Romero, D. y Dávila, G. (2003). The mosaic

structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. *Genome Biol.* 4, R36.

24. Guo, X., Flores, M., Mavingui, P., Fuentes, S.I., Hernández, G., Dávila, G. y Palacios, R. (2003). Natural genomic design in *Sinorhizobium meliloti*: novel genomic architectures. *Genome Res.* 13, 1810-7.

25. Haas, R. y Meyer, T.F. (1986). The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: evidence for gene conversion. *Cell* 44, 107-15.

26. Haas, R., Veit, S. y Meyer, T.F. (1992). Silent pilin genes of *Neisseria gonorrhoeae* MS11 and the occurrence of related hypervariant sequences among other gonococcal isolates. *Mol. Microbiol.* 6,197-208.

27. Harrington, C.S., Thomson-Carter, F.M. y Carter, P.E. (1997). Evidence for recombination in the flagellin locus of *Campylobacter jejuni*: implications for the flagellin gene typing scheme. *J. Clin. Microbiol.* 35, 2386-2392.

28. Hicks, J.B., Hinnen, A. y Fink, G.R. (1979). Properties of yeast transformation. *Cold Spring Harbor Symp. Quant. Biol.* 43, 1305-1313.

29. Holliday, R. (1964). A mechanism for gene conversion. *Genet. Res.* 5, 282-304.

30. Hughes, D. (2000). Co-evolution of the *tuf* genes links gene conversion with the generation of chromosomal inversions. *J. Mol. Biol.* 297, 355-364.

31. Hurles, M.E. (2002). Gene conversion. *Encyclopedia of life sciences.* Nature Publishing Group.

32. Inoue, S., Inoue, K., Utsunomiya, M., Nozaki, J., Yamada, Y., Iwasa, T., Mori, E., Yoshinaga, T. y Koizumi, A. (2002). Mutation analysis in *PKD1* of Japanese autosomal dominant polycystic kidney disease patients. *Hum. Mutat.* 19, 622-8.

33. Jakupciak, J.P. y Wells, R.D. (2000a). Genetic instabilities of triplet repeat sequences by recombination. *IUBMB Life* 50, 355-9.

34. Jakupciak, J.P. y Wells, R.D. (2000b). Gene conversion (recombination) mediates expansions of CTG-CAG repeats. *J. Biol. Chem.* 275, 40003-13.

35. Khanna, K.K. y Jackson, S.P. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. *Nat. Genet.* 27, 247-54.

36. Kobayashi, I. (2002). DNA double-strand break repair in bacteria. *Encyclopedia of life sciences.* Nature Publishing Group.

37. Koppens, P.F., Hoogenboezem, T. y Degenhart, H.J. (2002). Carriership of a defective tenascin-X gene in steroid 21-hydroxylase deficiency patients: TNXB -TNXA hybrids in apparent large-scale gene conversions. *Hum. Mol. Genet.* 11, 2581-90.

38. Lathe III, W.C. y Bork, P. (2001). Evolution of *tuf* genes: ancient duplication, differential loss and gene conversion. *FEBS Lett.* 502, 113-116.

39. Liao, D. (2000). Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. *J. Mol. Evol.* 51, 305-17.

40. Liang, F., Han, M., Romanienko, P.J. y Jasin, M. (1998). Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5172-7.

41. Lindegren, C.C. (1953). Gene conversion in *Saccharomyces. J. Genet.* 51, 625-637.

42. Liu, Y. y West, S.C. (2004). Happy Hollidays: 40th anniversary of the Holliday junction. *Nature Revs. Mol. Cell Biol.* 5, 937-946.

43. Lu, A.L. (2000). Repair of A/G and A/8-oxoG mismatches by MutY adenine DNA glycosylase. *Methods Mol. Biol.* 152, 3-16.

44. Martínez-Salazar, J.M., Zúñiga-Castillo, J. y Romero, D. (2004) Stimulation of recombination by inactivation of *recG* or *mutS* gene in *Rhizobium etli. Manuscrito en preparación.*

45. Mavingui, P., Flores, M., Guo, X., Dávila, G., Perret, X., Broughton, W.J. y Palacios, R. (2002). Dynamics of genome architecture in *Rhizobium* sp. strain NGR234. *J. Bacteriol.* 184, 171-176

46. Meselson, M.S. y Radding, C.M. (1975). A general model for genetic recombination. *Proc. Natl. Acad. Sci. U. S. A.* 72, 358-361.

47. Mitchell, M.B. (1955). Aberrant recombination of pyridoxine mutants of *Neurospora. Proc. Natl. Acad. Sci. U. S. A.* 41, 215-220.

48. Modrich, P. (1989). Methyl-directed DNA mismatch repair correction. *J. Biol. Chem.* 264, 6597-6600.

49. Morris, R.T. y Drouin, G. (2004). Ectopic gene conversions in four *Escherichia coli* genomes: increased recombination in pathogenic strains. *J. Mol. Evol.* 58:596-605.

50. Norbury, C. J. y Hickson, I. D. (2001). Cellular responses to DNA damage. *Annu. Rev. Pharmacol. Toxicol.* 41, 367-401.

51. Ohno, S. (1970). *Evolution by gene duplication*. pp. 160. Springer-Verlag, Berlin.

52. O'Hoy, K.L., Tsilfidis, C., Mahadevan, M.S., Neville, C.E., Barcelo, J., Hunter, A.G. y Korneluk, R.G. (1993). Reduction in size of the myotonic dystrophy trinucleotide repeat mutation during transmission. *Science* 259, 809-12.

53. Oppenheim, A.B., Rattray, A.J., Bubunenko, M., Thomason, L.C. y Court, D.L. (2004). In vivo recombineering of bacteriophage lambda by PCR fragments and single-strand oligonucleotides. *Virology* 319, 185-189.

54. Orr-Weaver, T.L., Szostak, J.W. y Rothstein, R.J. (1981). Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci. U. S. A.* 78, 6354-6358.

55. Orr-Weaver, T. L. y Szostak, J. W. (1983). Yeast recombination: the association between double strand gap repair and crossing over. *Proc. Natl. Acad. Sci. U. S. A.* 80, 4417-4421.

56. Peralta, H., Mora, Y., Salazar, E., Encarnación, S., Palacios, R. y Mora, J. (2004). Engineering the *nifH* promoter region and abolishing poly-beta-hydroxybutyrate accumulation in *Rhizobium etli* enhance nitrogen fixation in symbiosis with *Phaseolus vulgaris*. *Appl. Environ. Microbiol.* 70, 3272-81.

57. Quinto, C., de la Vega, H., Flores, M., Fernández, L., Ballado, T., Soberón, G. y Palacios, R. (1982) Reiteration of nitrogen fixation gene sequences in *Rhizobium phaseoli. Nature* 299, 724-726.

58. Rabbitts, T.H. (1994). Chromosomal translocations in human cancer. *Nature* 372,143-9.

59. Rayssiguier, C., Thaler, D.S. y Radman, M. (1989). The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* 342, 396-401.

60. Restrepo, B.I. y Barbour, A.G. (1994). Antigen diversity in the bacterium *B. hermsii* through "somatic" mutations in rearranged *vmp* genes. *Cell* 78, 867-76.

61. Richardson, C. y Jasin, M. (2000). Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* 405, 697-700.

62. Rodríguez, C. y Romero, D. (1998). Multiple recombination events maintain sequence identity among members of the nitrogenase multigene family in *Rhizobium etli. Genetics* 149, 785-794.

63. Roesler, J., Curnutte, J.T., Rae, J., Barrett, D., Patino, P., Chanock, S.J. y Goerlach, A. (2000). Recombination events between the p47-phox gene and its highly homologous pseudogenes are the main cause of autosomal recessive chronic granulomatous disease. *Blood* 95, 2150-6.

64. Romero, D., Brom, S., Martínez-Salazar, J., Girard, M. L., Palacios, R. y Dávila, G. (1991). Amplification and deletion of a *nod-nif* region in the symbiotic plasmid of *Rhizobium phaseoli. J. Bacteriol.* 173, 2435-2441.

65. Romero, D., Martínez-Salazar, J., Girard, L., Brom, S., Dávila, G., Palacios, R., Flores, M. y Rodríguez, C. (1995). Discrete amplifiable regions (amplicons) in the symbiotic plasmid of *Rhizobium etli* CFN42. *J. Bacteriol.* 177, 973-980.

66. Rossetti, S,, Strmecki, L., Gamble, V., Burton, S., Sneddon, V., Peral, B., Roy, S., Bakkaloglu, A., Komel, R., Winearls, C.G. y Harris, P.C. (2001). Mutation analysis of the entire *PKD1* gene: genetic and diagnostic implications. *Am. J. Hum. Genet.* 68, 46-63.

67. Santoyo, G. y Romero, D. (2005). Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiol. Rev.* 29, 169-183.

68. Sawyer, S. (1989) Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* 6, 526-538.

69. Schofield, M. J. y Hsieh, P. (2003). DNA mismatch repair: molecular mechanisms and biological functions. *Annu. Rev. Microbiol.* 57, 579-608.

70. Segall, A.M., y Roth, J.R. (1994). Approaches to half-tetrad analysis in bacteria: recombination between repeated, inverse-order chromosomal sequences. *Genetics* 136, 27-39.

71. Segovia, L., Young, J.P. y Martinez-Romero, E. (1993) Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* type I strains as *Rhizobium etli* sp. nov. *Int. J. Syst. Bacteriol.* 43, 374-7.

72. Serkin, C.D. y Seifert, H.S. (1998). Frequency of pilin antigenic variation in *Neisseria gonorrhoeae*. *J. Bacteriol.* 180, 1955-1958.

73. Stahl, F. (1994) The Holliday junction on its thirtieth anniversary. *Genetics* 138, 241-246.

74. Szostak, J., Orr-Weaver, T., Rothstein, R. y Stahl, F. (1983). The double-strand-break repair model for recombination. *Cell* 33, 25-35.

75. Tayebi, N., Stubblefield, B.K., Park, J.K., Orvisky, E., Walker, J.M., LaMarca, M.E. y Sidransky, E. (2003). Reciprocal and nonreciprocal recombination at the glucocerebrosidase gene region: implications for complexity in Gaucher disease. *Am. J. Hum. Genet.* 72, 519-34.

76. Tusie-Luna, M.T. y White, P.C. (1995). Gene conversions and unequal crossovers between *CYP21* (steroid 21-hydroxylase gene) and *CYP21P* involve different mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* 92, 10796-10800.

77. Valencia-Morales, E., y Romero, D. (2000). Recombination enhancement by replication (RER) in *Rhizobium etli. Genetics,* 154, 971-983.

78. Vanita, Sarhadi V., Reis, A., Jung, M., Singh, D., Sperling, K., Singh, J.R. y Burger, J. (2001). A unique form of autosomal dominant cataract explained by gene conversion between beta-crystallin B2 and its pseudogene. *J. Med. Genet.* 38, 392-6.

79. Whitehouse, H.L.K. (1982). Genetic recombination: understanding the mechanisms. John Wiley & Sons.

80. Wiese, C., Pierce. A.J., Gauny, S.S., Jasin, M. y Kronenberg, A. (2002). Gene conversion is strongly induced in human cells by double-strand breaks and is modulated by the expression of BCL-x(L). *Cancer Res.* 62, 1279-83.

81. Zhang, J.R. y Norris, S.J. (1998). Genetic variation of the *Borrelia burgdorferi* gene *vlsE* involves cassette-specific, segmental gene conversion. *Infect. Immun.* 66, 3698-704.

82. Zúñiga-Castillo, J., Romero, D. y Martínez-Salazar, J.M. (2004). The recombination genes *addAB* are not restricted to gram-positive bacteria: genetic analysis of the recombination initiation enzymes RecF and AddAB in *Rhizobium etli*. *J. Bacteriol.* 186, 7905-7913.