

**UNIVERSIDAD MICHOACANA DE SAN  
NICOLÁS DE HIDALGO**

**INSTITUTO DE INVESTIGACIONES  
QUÍMICO-BIOLÓGICAS**



**“ANÁLISIS BIOINFORMÁTICO DE SECUENCIAS REITERADAS Y  
CONVERSIÓN GÉNICA EN GENOMAS DE RHIZOBIAS”**

TESIS PARA OBTENER EL GRADO DE:

**MAESTRA EN CIENCIAS EN BIOLOGÍA EXPERIMENTAL**

PRESENTA:

**Biol. Julie Eliene Hernández Salmerón**

TUTOR:

**D.C. Gustavo Santoyo Pizano**

COTUTOR:

**D.C. Agustino Martínez Antonio**

Morelia, Mich., México.

Noviembre de 2011

ESTE TRABAJO FUE REALIZADO EN EL LABORATORIO DE RECOMBINACIÓN Y DIVERSIDAD GENÓMICA DEL INSTITUTO DE INVESTIGACIONES QUÍMICO-BIOLÓGICAS DE LA UNIVERSIDAD MICHOACANA DE SAN NICOLAS DE HIDALGO, BAJO LA TUTORÍA DEL D.C. GUSTAVO SANTOYO PIZANO Y LA COASESORÍA DEL D.C. AGUSTINO MARTÍNEZ ANTONIO EN EL CINVESTAV-IPN, EN IRAPUATO, GTO.



TESIS APOYADA POR CIC-UMSNH Y CONACYT BECA # 232463

## ÍNDICE GENERAL

ÍNDICE DE TABLAS Y FIGURAS	III
ABREVIATURAS	IV
RESUMEN	VI
ABSTRACT	VII
1. INTRODUCCIÓN	
1.1 Fijación biológica del nitrógeno	1
1.2 Bacterias fijadoras de nitrógeno	2
1.2.1 <i>Rhizobium etli</i>	4
1.2.2 <i>Rhizobium leguminosarum</i>	4
1.2.3 <i>Sinorhizobium meliloti</i>	4
1.2.4 <i>Mesorhizobium loti</i>	5
1.2.5 <i>Bradyrhizobium japonicum</i>	5
1.3 Recombinación homóloga	6
1.4 Conversión génica	9
1.5 Secuencias reiteradas y familias multigénicas	12
1.6 Rearreglos genómicos	14
2. JUSTIFICACIÓN	16
3. HIPÓTESIS	17
4. OBJETIVOS	17
4.1 Objetivo general	17
4.2 Objetivos específicos	17
5. MATERIALES Y MÉTODOS	18
5.1 Búsqueda de familias multigénicas en rizobias	18
5.2 Análisis filogenéticos para inferir eventos de conversión génica	19
5.3 Análisis de secuencias reiteradas en genomas	20

6. RESULTADOS	22
6.1 Familias multigénicas y conversión génica en rizobias	22
6.2 Estructura genómica de las rizobias analizadas en este trabajo	31
6.3 Secuencias reiteradas en los genomas de rizobias	32
6.3.1. Reiteraciones por número de copias en genomas de rizobias	34
6.3.2 Longitud de las reiteraciones localizadas en genomas de rizobias	35
6.3.3 Densidades de las reiteraciones por genoma	36
6.3.4 Densidad de las reiteraciones por localización genómica	37
6.3.5 Funciones de las reiteraciones en genomas de rizobias	38
7. DISCUSIÓN	42
8. RESUMEN DE RESULTADOS	50
9. CONCLUSIONES	52
10. RESULTADOS EXPERIMENTALES ADICIONALES	53
11. PUBLICACIONES	59
12. LITERATURA CITADA	61

## ÍNDICE DE TABLAS Y FIGURAS

<b>Tabla 1.</b> Familias multigénicas encontradas en algunas especies de rhizobias representativas	23
<b>Tabla 2.</b> Arquitectura genómica de las rhizobias	31
<b>Tabla 3.</b> Densidad de las secuencias reiteradas a nivel genómico	36
<b>Tabla 4.</b> Densidad de las reiteraciones de acuerdo a su localización en el genoma	38
<b>Figura 1.</b> Modelos moleculares de recombinación homóloga	7
<b>Figura 2.</b> Rearreglos genómicos producidos por recombinación homóloga	9
<b>Figura 3.</b> Conversión génica	10
<b>Figura 4.</b> Relaciones filogenéticas entre los miembros de una familia multigénica	12
<b>Figura 5.</b> Metodología general	20
<b>Figura 6.</b> Árbol filogenético de la familia multigénica <i>nifD</i>	25
<b>Figura 7.</b> Alineamiento de los genes <i>nifD</i> en <i>Rhizobium etli</i>	26
<b>Figura 8.</b> Árbol filogenético de la familia multigénica <i>nifK</i>	27
<b>Figura 9.</b> Alineamiento de los genes <i>nifK</i> en <i>Rhizobium etli</i>	28
<b>Figura 10.</b> Árbol filogenético de la familia multigénica <i>groEL</i>	30
<b>Figura 11.</b> Secuencias reiteradas localizadas en genomas de rhizobias	33
<b>Figura 12.</b> Reiteraciones por número de copias en genomas de rhizobias	34
<b>Figura 13.</b> Longitud de las reiteraciones localizadas en genomas de rhizobias	35
<b>Figura 14.</b> Funciones de las secuencias reiteradas en <i>R. etli</i>	39
<b>Figura 15.</b> Funciones de las secuencias reiteradas en	

<i>R. leguminosarum</i>	39
<b>Figura 16.</b> Funciones de las secuencias reiteradas en <i>M. loti</i>	40
<b>Figura 17.</b> Funciones de las secuencias reiteradas en <i>S. meliloti</i>	40
<b>Figura 18.</b> Funciones de las secuencias reiteradas en <i>B. japonicum</i>	41
<b>Figura 19.</b> Técnica RAPD-PCR	54
<b>Figura 20.</b> Polimorfismos por RAPD-PCR en los aislados de la rizósfera	57
<b>Figura 21.</b> Dendrograma de la diversidad genética de la rizósfera de <i>M. truncatula</i>	58

## ABREVIATURAS

ADN	Ácido Desoxirribonucleico
ARDRA	Análisis de restricción de ADN ribosomal (por sus siglas en inglés Amplified rDNA Restriction Analysis)
ARNr	Ácido ribonucleico ribosomal
BLAST	Herramienta de búsqueda de un alineamiento local básico (por sus siglas en inglés Basic Local Alignment Search Tool)
DL	Densidad en longitud
DN	Densidad en número
DSBR	Reparación de cortes en doble cadena (por sus siglas en inglés Double-Strand Break Repair)
FBN	Fijación biológica del nitrógeno
IS	Secuencias de inserción (por sus siglas en inglés Insertion sequences)
Kb	Kilobases

Mb	Megabases
NCBI	Centro Nacional de Información Biotecnológica (por sus siglas en inglés National Center for Biotechnology Information)
nt	nucleótidos
pb	pares de bases
PCR	Reacción en cadena de la polimerasa (por sus siglas en inglés Polymerase Chain Reaction)
RFLP	Polimorfismos de longitud de fragmentos de restricción (por sus siglas en inglés Restriction Fragment Length Polymorfism)
RAPD	ADN polimórfico amplificado al azar (por sus siglas en inglés Random Amplified Polymorphic DNA)

## RESUMEN

En el presente trabajo se realizó una búsqueda de familias multigénicas en cinco genomas de bacterias fijadoras de nitrógeno, mejor conocidas como rizobias, incluyendo: *Rhizobium etli* cepas CFN42 y CIAT652, *Rhizobium leguminosarum* bv. *viciae* 3841, *Bradyrhizobium japonicum* USDA110, *Sinorhizobium meliloti* 1021 y *Mesorhizobium loti* MAFF3030. Dicha búsqueda se llevó a cabo en la literatura reportada y en las bases de datos del NCBI, encontrando 11 familias multigénicas en *R. etli*, algunas de las cuales se encontraron también presentes en otras rizobias. Las familias multigénicas presentes en al menos tres genomas se sometieron a un análisis filogenético en busca de eventos de conversión génica, los cuales se detectaron en las familias *nifD* y *nifK* de la bacteria *R. etli* CFN42 y CIAT652. Asimismo, se realizó una búsqueda *in silico* de secuencias repetidas mayores a 300 pb en los cinco genomas de rizobias anteriormente mencionados. Nuestros resultados muestran que la bacteria *B. japonicum* contiene el mayor número de reiteraciones con 410, seguido de *S. meliloti* con 276, *R. etli* con 155, así como 129 en *R. leguminosarum* y finalmente *M. loti* con 79. También se encontró una alta frecuencia de reiteraciones de bajo número de copias (2-7) en todos los genomas, mientras que su longitud fue variable, desde 300 hasta 1000 pb. En el análisis de densidad por número de reiteración se encontró una mayor densidad en el genoma de *B. japonicum* (45.1) y *M. loti* mostró el valor más bajo (9.6), mientras que en la densidad por longitud, *S. meliloti* resultó ser el más alto (3.6) y *M. loti* el más bajo (1.0). De acuerdo a las densidades obtenidas en plásmidos y cromosoma por separado, se encontró una mayor densidad en plásmidos tanto de densidad por longitud como por número de copias, además no se observó una relación directa entre un mayor tamaño de genoma con una mayor



densidad de reiteraciones. Las principales funciones del total de las reiteraciones localizados en los presentes genomas están relacionadas con elementos móviles como transposasas y secuencias de inserción, siendo relevantes además, las funciones importantes para la sobrevivencia de este grupo de organismos, tales como: metabolismo, transcripción, replicación, traducción, conjugación y transporte. Finalmente, nuestros resultados sugieren que la longitud y presencia de múltiples reiteraciones génicas en genomas de rizobias, distribuidas en diferentes replicones, podrían participar en una frecuente dinámica de rearrreglos genómicos con consecuencias biológicas aún por descubrir.

#### ABSTRACT

In the present study, we conducted a search of multigene families in five genomes of nitrogen-fixing bacteria, better known as rhizobia, including: *Rhizobium etli* CFN42 and CIAT652 strains, *Rhizobium leguminosarum* bv. *viciae* 3841, USDA110 *Bradyrhizobium japonicum*, *Sinorhizobium meliloti* 1021 and *Mesorhizobium loti* MAFF3030. We found 11 multigene families in *R. etli*, some of which were also present in other rhizobia. Multigene families present in at least three genomes were subjected to phylogenetic analysis looking for gene conversion events, which were detected in the *nifD* and *nifK* gene families of the bacterium *R. etli* CFN42 and CIAT652. In addition, we searched repeated sequences longer than 300 bp in the five genomes of rhizobia. Our results showed that *B. japonicum* contains the largest number of repetitions with 410, followed by *S. meliloti* with 276, *R. etli* with 155, 129 in *R. leguminosarum* and finally *M. loti* with 79. We also found a high frequency of repetition of low copy

number (2-7) in all the genomes, while its length was variable, from 300 to 1000 bp. In the analysis of density in number of copies, a higher density was found in the genome of *B. japonicum* (45.1) while *M. loti* showed the lowest value (9.6); the density in length analysis showed that *S. meliloti* appears to be the highest (3.6) and *M. loti* the lowest (1.0). However, according to the densities obtained in separate plasmids and chromosome, we found a higher density in plasmids for both, length density and number of copies. No possible direct relationship between genome size and density of repetition was found. The main functions of all the reiterations located in these genomes belong to mobile elements such as transposases and insertion sequences. We also detected important functions encoded in the repeated sequences for the survival of this group of organisms, such as metabolism, transcription, replication, translation, conjugation and transport. Finally, our results suggest that the length and presence of multiple reiterations in rhizobia genomes, distributed in different replicons may participate in common mechanism of genomic rearrangements whose dynamics and biological consequences are yet to be discovered.

## 1. INTRODUCCIÓN

---

### 1.1. Fijación biológica de nitrógeno

La restringida disponibilidad de compuestos de nitrógeno en el suelo es con frecuencia un factor limitante para el crecimiento y productividad de las plantas (Russelle y Birr, 2004). Las plantas leguminosas han sobrellevado este problema estableciendo una simbiosis con algunas bacterias del suelo llamadas rizobia, las cuales fijan el nitrógeno atmosférico para ser usado por la planta.

La fijación biológica de nitrógeno (FBN) es un proceso que surgió hace aproximadamente 3,700 millones de años (m.a), cuando todos los procesos eran llevados a cabo dentro de la atmósfera primitiva y se desarrollaron en condiciones desprovistas de oxígeno (Awramik *et al.*, 1983). Fue probablemente en este periodo donde algunos procesos metabólicos aparecieron: fermentación, glucólisis, metanogénesis, ciclo de Krebs, la fotosíntesis anaerobia y la fijación del dióxido de carbono y del nitrógeno (Margulis, 1993). La capacidad de fijar nitrógeno (diazotrofía) es exclusiva de los procariotes (Raymond *et al.*, 2004) y se cree que su aparición fue también anterior a la acumulación del oxígeno molecular. Sin embargo, la simbiosis de los diazótrofos con las plantas debió darse mucho tiempo después (Young, 1992) recorriendo un largo camino evolutivo desde la aparición de las primeras plantas terrestres hace aproximadamente 400 millones de años.

Los diazótrofos fijan el nitrógeno mediante la reducción del nitrógeno atmosférico en amonio, la cual se lleva a cabo mediante el complejo enzimático de la nitrogenasa en condiciones anaerobias propiciadas por la formación de nódulos en las raíces de las plantas. El complejo de la nitrogenasa está formado

por dos componentes: la dinitrogenasa heterotetramérica con dos subunidades,  $\alpha$  y  $\beta$ , donde la subunidad  $\alpha$  es codificada por el gen *nifD* y la subunidad  $\beta$  por el gen *nifK*. El otro componente es la dinitrogenasa reductasa, un homodímero codificado por el gen *nifH* (Raymond *et al.*, 2004). La nitrogenasa ( $\alpha$ ) contiene el sitio activo para la reducción del  $N_2$ , usando el cofactor Fe-Mo, mediante la siguiente reacción:



Así, las poblaciones microbianas favorecen la asimilación de nutrientes para las plantas, que a su vez brindan un lugar propicio a las bacterias para sobrevivir formando una importante interacción y una de las asociaciones más eficientes entre las plantas y los microorganismos de millones de años de coevolución.

## 1.2. Bacterias fijadoras de nitrógeno

Las bacterias simbióticas conocidas como rhizobias tienen la capacidad de fijar el nitrógeno atmosférico en estructuras globulares llamadas nódulos, los cuales se forman en las raíces de las plantas, especialmente en leguminosas (Rivas *et al.*, 2009). Con base en la filogenia de los genes 16S ARNr se ha agrupado a estos organismos dentro de la división Proteobacteria (Woese *et al.*, 1984), principalmente en la subdivisión  $\alpha$ -Proteobacteria en diferentes órdenes y familias; sin embargo, las más representativas pertenecen al orden Rhizobiales, en las familias Rhizobiaceae, Phyllobacteriaceae y Bradyrhizobiaceae. Un grupo muy reducido de rhizobias, recientemente descritas, pertenecen al grupo de las  $\beta$ -Proteobacteria (Moulin *et al.*, 2001).

Hasta ahora, se han descrito más de 50 especies de rhizobias en  $\alpha$  y  $\beta$ -Proteobacteria (Lloret *et al.*, 2005), todas ellas son bacilos aerobios gram-negativos que no forman esporas, con formas irregulares llamadas bacteroides (formas diferenciadas que fijan nitrógeno) (Garrity *et al.*, 2001); la movilidad se lleva a cabo por un flagelo polar o subpolar o entre 2 a 6 flagelos peritricos (Werner, 1992). Este grupo bacteriano contiene en sus genomas de uno a dos cromosomas y varios plásmidos de tamaño que va de entre 100 kilobases (Kb) a más de 2000 kilobases (Kb). Una característica importante de sus genomas es que los genes involucrados en la simbiosis, o genes simbióticos, se pueden localizar en el cromosoma formando islas simbióticas, o bien en los plásmidos (González *et al.*, 2003); esto ha causado dificultades en la taxonomía de las rhizobias debido a su capacidad de ser transferidos tanto de plásmidos a islas cromosómicas, de la bacteria a la planta y entre las mismas bacterias (Finan, 2002).

La capacidad de las rhizobias para nodular plantas es una característica muy estudiada. El primer contacto lo inician las bacterias de la rizósfera que se aproximan a la raíz bajo la influencia de flavonoides secretados por la planta activando en la bacteria la expresión de genes *nod*, los cuales son los se encargan de sintetizar las sustancias responsables de la inducción de la formación de los nódulos y son llamados factores de nodulación o factores nod (Downie, 1994). Posteriormente, los factores nod son percibidos por la planta, desencadenando una vía de señalización dependiente de calcio que dirige el encurvamiento de los pelos de la raíz y la división de células meristemáticas que conducirán a la formación del nódulo (Oldroyd *et al.*, 2007). Así, la rhizobacteria penetra la pared celular infectando el nódulo en desarrollo que le propiciará un ambiente anaerobio adecuado para el buen funcionamiento de la nitrogenasa encargada de la fijación de nitrógeno.

Si bien se ha mencionado la amplia diversidad de rizobias en la naturaleza, a continuación se describen algunas especies representativas en las que se enfocó el presente trabajo.

#### 1.2.1. *Rhizobium etli*

Es un bacilo Gram-negativo, aerobio, de crecimiento rápido y producción de ácido en el medio; pertenece a la familia Rhizobiaceae y nodula específicamente a la leguminosa *Phaseolus vulgaris* (Segovia *et al.*, 1993). Actualmente se cuenta con su genoma completamente secuenciado (González *et al.*, 2006) el cual consiste de un cromosoma circular y seis plásmidos; uno de ellos, el plásmido simbiótico, contiene los genes encargados de la simbiosis con la planta. Además, se han localizado varias familias de genes con diferente número de copias, incluyendo aquellos que son importantes para la nodulación y fijación de nitrógeno (González *et al.*, 2003).

#### 1.2.2. *Rhizobium leguminosarum*

Al igual que *R. etli*, pertenece a la familia Rhizobiaceae y sus características morfológicas y bioquímicas son similares. Este microorganismo tiene la característica de que presenta tres biovars: bv. *viciae*, que forma nódulos en *Pisum*, *Vicia*, *Lathyrus* y *Lens* ; bv *trifolii* (*Trifolium* ) y bv. *phaseoli* (*Phaseolus*) (Young, 1996). En el año 2006 fue presentado su genoma completo (Young *et al.*, 2006), el cual consiste en un cromosoma circular y seis plásmidos que cuentan también con la mayoría de los genes de nodulación.

#### 1.2.3. *Sinorhizobium meliloti*

Son bacterias de rápido crecimiento, bacilos que acidifican el medio y se agrupan en la misma familia que *R. etli* y *R.leguminosarum*; sin embargo, esta

especie nodula géneros como *Medicago*, *Melilotus* y *Trigonella* (de Lajudie *et al.*, 1994). Su genoma está organizado en un cromosoma circular y dos megaplásmidos (>1Mb) que se ha sugerido son de reciente origen (Galibert *et al.*, 2001).

#### 1.2.4. *Mesorhizobium loti*

Estas bacterias son bacilos de crecimiento lento o moderado; forman nódulos con los géneros *Lotus*, *Lupinus* y *Antillus*. Su genoma completamente secuenciado está conformado por un cromosoma circular y dos plásmidos (Kaneko *et al.*, 2000). A diferencia de las especies anteriores, *M. loti*, contiene los genes de nodulación y fijación de nitrógeno en su cromosoma y no en los plásmidos, en una región de aproximadamente 611 Kb llamada isla simbiótica (Kaneko *et al.*, 2000).

#### 1.2.5. *Bradyrhizobium japonicum*

Son células aerobias, Gram-negativas, de crecimiento muy lento y son productoras de álcali. Esta especie forma nódulos efectivos solamente en *Glicine max* (Jordan, 1982). Su genoma está formado por un solo cromosoma circular de gran tamaño y no cuenta con algún plásmido en su genoma (Kaneko *et al.*, 2002).

Es importante notar la reciente publicación de las secuencias completas de los genomas anteriores (Kaneko *et al.*, 2000; Galibert *et al.*, 2001; Kaneko *et al.*, 2002; González *et al.*, 2006; Young *et al.*, 2006), lo cual abre un campo de estudio muy amplio de estos organismos desde diversas perspectivas, incluyendo el área de la ciencia aplicada, es decir, aquellos aspectos relacionados con el mejoramiento de las capacidades fijadoras de nitrógeno y

promoción del crecimiento vegetal, algo que es de suma importancia en la agrobiotecnología. Sin embargo, dados los objetivos del presente trabajo, el conocimiento de la secuencia completa de estas rizobias abre la posibilidad de conocer mejor su genoma respecto a su grado de reiteración génica. Lo anterior brinda nuevas formas de estudio *in silico* para conocer las familias multigenéticas, su localización y posible predicción de rearrreglos cromosomales, los cuales podrían estar implicados en capacidades de adaptación en un ambiente tan cambiante como lo es la rizósfera.

### **1.3. Recombinación homóloga**

La recombinación homóloga se define comúnmente como el proceso por el cual dos secuencias idénticas o altamente idénticas de ADN intercambian información genética. Este mecanismo tiene diferentes papeles en la reparación del ADN, la segregación de cromosomas y la generación de variación genética, entre otras (West, 2003).

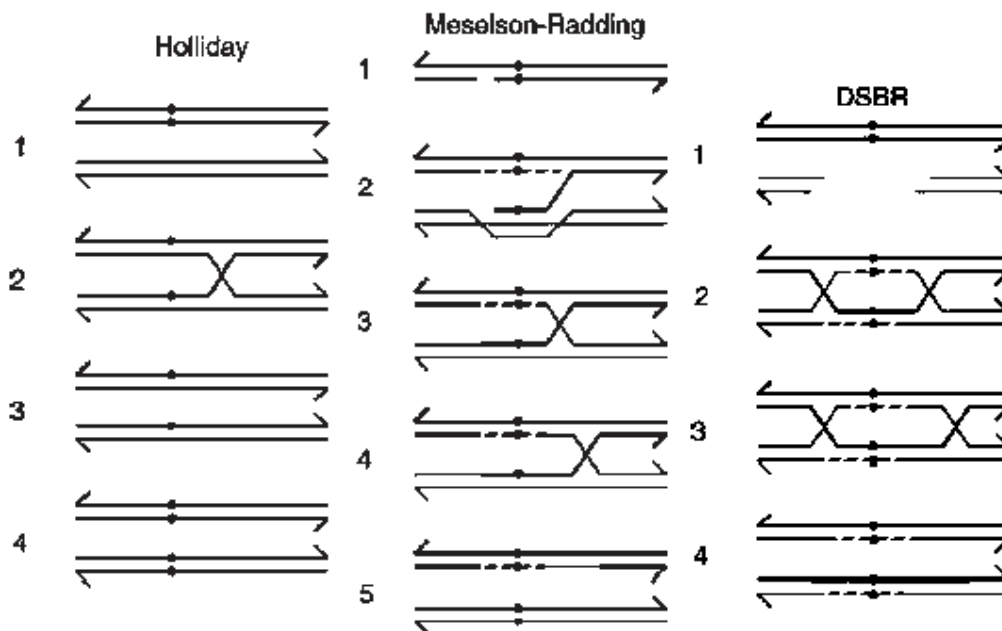
La recombinación homóloga es una fuerza potente para la evolución de los genomas (Rocha *et al.*, 2005). Todos los organismos en la tierra tienen probablemente algún mecanismo de recombinación homóloga; en el caso de bacterias, representa un evento crucial para la sobrevivencia y evolución de las células (Marians, 2004).

Existen varios modelos que tratan de explicar cómo se llevan a cabo los eventos de recombinación. El primero de ellos fue el descrito por Robin Holliday en 1964 (Holliday, 1964), el cual lleva su nombre. En el modelo de Holliday (Fig. 1), la recombinación inicia con un corte de cadena sencilla en ambas moléculas participantes, enseguida se da un intercambio de cadena sencilla seguido por ligación. Una vez entrecruzadas las moléculas se forma una estructura de cuatro cadenas llamada unión de Holliday, la cual puede tener finalmente dos



resultados dependiendo de la orientación del corte, ya sea horizontal o vertical. En un primer caso puede resultar un entrecruzamiento (“crossover”) o bien, conversión génica.

Posteriormente, Meselson y Radding propusieron algunas modificaciones importantes para el modelo de Holliday (Meselson y Radding, 1975). Este modelo (Fig. 1), también llamado de invasión de una sola cadena, inicia con un corte de cadena sencilla en una de las moléculas; el extremo 3’ sirve como iniciador para que la ADN polimerasa entre en acción y desplace la cadena, la cual invadirá la secuencia homóloga formando un asa. Finalmente la ligación genera la unión de Holliday, la cual será resuelta de la misma manera que el modelo anterior.

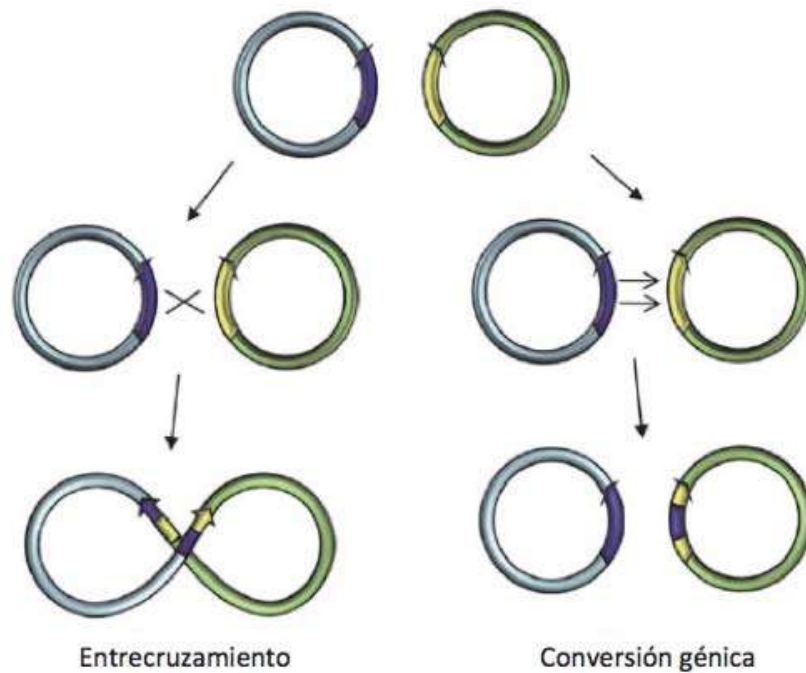


**Fig. 1.** Modelos moleculares de recombinación homóloga. Modelo de Holliday, Meselson-Radding y reparación de cortes en doble cadena (DSBR). Los números de la izquierda indican los pasos sucesivos de cada modelo, los diferentes sombreados de las líneas indican las moléculas que interactúan, mientras que las líneas punteadas indican la síntesis de ADN y los puntos negros marcan la posición de diferencias en la secuencia. Tomada de Santoyo y Romero.

(2005).

Un modelo más recientemente propuesto es el de reparación de cortes en doble cadena (DSBR), propuesto por Szostak *et al.* (1983). En sus trabajos, ellos sugieren que la recombinación inicia con cortes en doble cadena (Fig. 1), que continúa por una degradación extensiva de una cadena en dirección 5' a 3' generando un hueco con el extremo 3' terminal el cual invade la cadena homóloga. Al invadir se forma una especie de cadena doble que sirve como templado para que se inicie un proceso de síntesis de ADN. Posteriormente se da la ligación para formar las uniones de Holliday y la resolución dependerá de la orientación del corte.

El intercambio de segmentos de moléculas de ADN, como resultado de la recombinación homóloga, induce a diversos arreglos genómicos. Por un lado, dirige a la diversificación de las secuencias provocando duplicaciones o deleciones (Fig. 2 A y B); o bien, puede resultar en una conversión génica (Fig. 2 C), volviendo idénticas las secuencias.



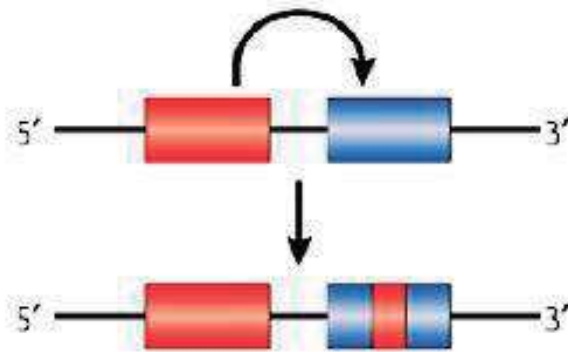
**Fig. 2.** Rearreglos genómicos producidos por recombinación homóloga. Dos posibles resultados: entrecruzamiento y conversión génica. Un evento de entrecruzamiento involucra el intercambio recíproco de material genético, mientras que en una conversión génica el intercambio se da de manera unidireccional entre las dos moléculas que interactúan.

Tomado de Orozco-Mosqueda *et al.* (2009).

#### 1.4. Conversión génica

Tal como se mencionó anteriormente, un resultado interesante de un evento de recombinación es la conversión génica (Fig. 3). La conversión génica se define como la donación unidireccional de información genética entre las dos moléculas que interactúan (Szostak *et al.*, 1983). Está asociada con la evolución de familias multigénicas, las cuales son parte inherente de genomas procariontes, eucariontes y de arqueas (Romero *et al.*, 1999). Comúnmente, los miembros de estas familias comparten alta similitud en sus secuencias lo cual

los hace blancos potenciales para intercambiar información genética a través de recombinación homóloga.



**Figura 3.** Conversión génica. Mecanismo de transferencia unidireccional de material genético entre segmentos de ADN. Las cajas roja y azul representan secuencias altamente similares o idénticas y las flechas indican la transferencia de material genético de una secuencia a la otra.

Una consecuencia importante de la conversión génica es la evolución concertada; es decir, la homogenización de mutaciones entre las secuencias reiteradas o familias génicas (Howell-Adams y Seifert, 2000). Una mutación presente en un miembro que le confiere cierta ventaja biológica puede ser transferida por conversión a otro miembro y así maximizar su efecto para la sobrevivencia. A su vez, una mutación que altere o dañe a un miembro puede ser eliminada por conversión. Sin embargo, algunos problemas pueden surgir si la secuencia donadora de material genético no es funcional, entonces la conversión génica puede dirigir a algunas enfermedades y desórdenes congénitos (Lawson *et al.*, 2009).

Por otro lado, la conversión génica juega también un papel importante en la variación antigénica. Las bacterias patógenas tienen la capacidad de producir una alta diversidad de proteínas transmembranales, las cuales le ayudarán a

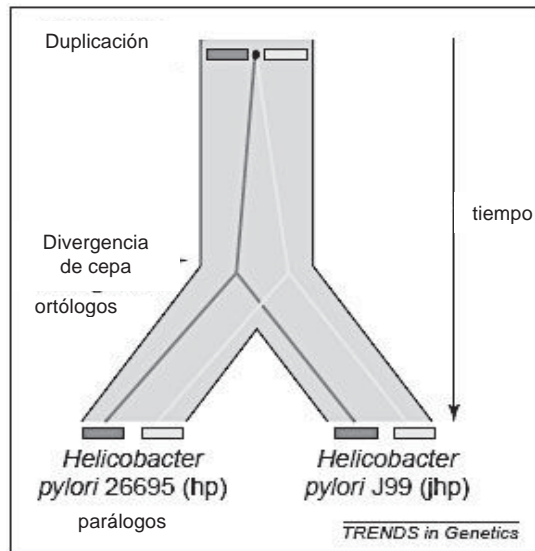
evadir las defensas del hospedero; esto puede ocurrir debido a que las secuencias génicas variantes son transferidas de genes no expresados a genes expresados, mediante conversión génica (Deitsch *et al.*, 1997).

La conversión génica ha sido ampliamente estudiada en sistemas eucariontes y procariontes como en *Saccharomyces cerevisiae* (Kuzminov, 2001), *Escherichia coli* (Takahashi y Kobayashi, 1990), *Salmonella typhimurium* (Alani *et al.*, 1994) y *Bacillus subtilis* (Belle *et al.*, 2004), en distintas familias de genes (Santoyo y Romero, 2005), pero pocos estudios se han realizado en otros grupos bacterianos, tales como las bacterias fijadoras de nitrógeno o rizobias, donde solo se ha estudiado la familia génica *nifH*, que codifica para la nitrogenasa en la fijación de nitrógeno (Santoyo *et al.*, 2005), donde además fue posible detectar de manera experimental el posible segmento de conversión, el cual varía entre 150-800 pb.

Una gran limitante se presenta al estudiar conversión génica particularmente en los procariontes, principalmente debido a la dificultad de recuperar los dos productos de la recombinación, lo cual ha dificultado este tipo de análisis de manera experimental (Orozco-Mosqueda *et al.*, 2009). Para ello, recientemente se ha propuesto detectar conversión génica mediante evidencia filogenética, tomando en cuenta la hipótesis de los genes ortólogos y parálogos.

La hipótesis establece que los genes ortólogos presentes en dos cepas o especies cercanas, se encuentran más cercanamente relacionados que los genes parálogos (Jordan *et al.*, 2001); esto se debe a que los ortólogos comparten un ancestro común más cercano en el momento de divergencia de cepa o especiación, mientras que los parálogos, comparten su último ancestro común en el momento de la duplicación génica, antes de la divergencia (Fig. 4). En un árbol filogenético será posible detectar un evento de conversión cuando

los genes parálogos se relacionen más entre ellos, como efecto de un proceso de homogenización en sus secuencias.



**Fig. 4.** Relaciones filogenéticas entre los miembros de una familia multigénica. Se muestran dos cepas, cada una con dos copias de una familia génica. Los genes ortólogos se indican con cajas del mismo color y los parálogos con distintos colores. Los genes parálogos comparten su último ancestro común en el momento de la duplicación génica, mientras que los ortólogos en el momento de la divergencia de cepas. Tomado de Jordan *et al.* (2001).

### 1.5. Secuencias reiteradas y familias multigénicas

Como se ha mencionado anteriormente, para que ocurran eventos de conversión génica es necesaria la presencia de secuencias repetidas, ya sea codificantes o no; las rhizobias han recibido poca atención en ésta área ya que se tenía la noción de que sus genomas contaban con pocos elementos reiterados o repetidos. En el presente, existen varios estudios que muestran la presencia de secuencias reiteradas y familias multigénicas en este grupo de organismos (Achaz *et al.*, 2002; Hernández-Salmerón y Santoyo, 2011).

Una secuencia reiterada se define como un segmento de ADN que comparte una extensa similitud en cuanto a su secuencia de nucleótidos con otro segmento del mismo genoma (Achaz *et al.*, 2002). Las reiteraciones pueden estar presentes tanto en cromosoma como en los plásmidos dando lugar a diferentes rearrreglos genómicos como resultado de la recombinación homóloga.

Hasta ahora las secuencias reiteradas se han clasificado en repeticiones cortas y largas. Las repeticiones cortas incluyen repeticiones de pocos nucleótidos, arreglados en tándem, es decir, uno seguido del otro repetido un número variable de veces (Katti *et al.*, 2001). La segunda categoría, en la que se centra este trabajo, está formada por reiteraciones largas que incluyen elementos móviles, como secuencias de inserción (IS), repeticiones en tándem largas y repeticiones espaciadas o dispersas.

Se conocen algunos estudios de reiteraciones en eucariontes, como *Saccharomyces cerevisiae* (Warner, 1989) y *Homo sapiens* (Sakai *et al.*, 1995), donde se han estudiado principalmente los genes ARNr y se ha observado un alto número de copias (>100 repeticiones), la mayoría arregladas en tándem; mientras que en procariontes se ha observado un número bajo de copias (<10 repeticiones). Si bien el número de repeticiones en bacterias es más bajo que en eucariontes, existen suficientes para producir extensos rearrreglos genómicos (Rocha & Blanchard, 2002).

Asimismo, se ha observado que las secuencias reiteradas son importantes para la respuesta al estrés (Rocha *et al.*, 2002), la plasticidad (Aras *et al.*, 2003), en la variación de fase para su adaptación (Moxon *et al.*, 2006), como en algunas especies de *Mycoplasma* (Rocha & Blanchard, 2002) y en la dinámica del genoma (Mavingui *et al.*, 2002; Guo *et al.*, 2003). La plasticidad genómica provocada por estos rearrreglos suele tener un impacto fenotípico importante, como en la competencia de nodulación (Castillo *et al.*, 1999). Sin embargo,

algunas anormalidades son causadas por la recombinación homóloga no-alélica entre repeticiones de bajo número de copias (Lupski, 1998); así, varias enfermedades han sido atribuidas a los rearrreglos genómicos (Ji *et al.*, 2000), también llamadas desórdenes genómicos (Stankiewicz & Lupski, 2002).

Los elementos reiterados, principalmente las familias multigénicas, han sido bien estudiados en diferentes organismos; sin embargo, una visión más amplia de estas características a nivel genómico falta aún por hacer. Se han desarrollado algunos estudios experimentales (Aras *et al.*, 2003), pero su alcance es aún limitado a nivel de genomas. Debido a ello, se han desarrollado técnicas computacionales que hacen posible el estudio de las reiteraciones a nivel genómico y en varios organismos a la vez (Treangen *et al.*, 2009). Con estas herramientas, algunos estudios computacionales ya han mostrado numerosas repeticiones directas en diferentes grupos de organismos bacterianos y arqueas (Rocha *et al.*, 1999; Achaz *et al.*, 2002).

### **1.6. Rearreglos Genómicos**

El estudio de las secuencias reiteradas integra un amplio campo de estudio y, como ya se ha mencionado, mediante la recombinación homóloga, se generan distintos tipos de rearrreglos genómicos, tales como amplificaciones, deleciones o cointegraciones. Existe una gran evidencia de que las amplificaciones existen de manera natural en los organismos procariontes (Petes y Hill, 1988) y ocurren gracias a la recombinación de secuencias orientadas en la misma dirección, comúnmente denominadas amplicones (Flores *et al.*, 1993). Se ha visto que tales amplificaciones se encuentran asociadas con respuestas adaptativas como el crecimiento de organismos en fuentes de carbono limitantes, incrementar su virulencia o patogenicidad e incrementando la resistencia a antibióticos, entre otros (Romero y Palacios, 1997).



El conocimiento de las secuencias de los genomas permite la localización y orientación precisa de las secuencias reiteradas, lo cual a su vez da la posibilidad de predecir los diferentes tipos de rearrreglos genómicos que se puedan generar. Tal predicción fue probada por Flores *et al.* (2000), en el plásmido simbiótico de la bacteria *Rhizobium* sp. NGR234, donde además de predecir los distintos rearrreglos genómicos, éstos fueron comprobados mediante una estrategia de PCR. Un aspecto importante es que además de poder predecir, es que se pueden seleccionar aquellos rearrreglos que impliquen una ventaja adaptativa al organismo. Mavingui *et al.* (1997), realizaron una selección de cepas derivadas de *Rhizobium tropici* con amplificaciones de genes al azar que les daban una mayor competencia y mejores capacidades de nodular distintas plantas leguminosas.

Mediante este diseño genómico natural, se han obtenido distintas arquitecturas genómicas novedosas en *S. meliloti*, gracias a la presencia de las secuencias reiteradas, tanto en cromosoma como en plásmidos, se han podido identificar distintos rearrreglos en sus tres replicones, tales como la cointegración de ambos plásmidos, cointegración de un plásmido con el cromosoma, e incluso la unión de los tres replicones en una sola molécula, los cuales además, fueron capaces de regresar a su estado original (Guo *et al.*, 2003). Cabe destacar que dichos rearrreglos no afectaron la capacidad de nodulación de esta bacteria, lo cual puede indicar que ocurren de manera natural sin afectar al organismo. Aún se desconoce si estos rearrreglos tienen algún otro fenotipo, aparte del observado por Mavingui *et al.* (1997), tal como alguno relacionado con la interacción ecológica con otros microorganismos de la rizósfera, el cual se podría estudiar en proyectos próximos.

## 2. JUSTIFICACIÓN

---

Las rhizobias son un grupo de bacterias poco estudiadas en el área de la recombinación homóloga y en la detección de eventos de conversión génica en familias multigénicas, aún cuando hay indicios de que sus genomas contienen una gran cantidad de secuencias repetidas. Estudiar los eventos de conversión génica experimentalmente a nivel genómico es una tarea casi imposible de realizar, por ello la importancia de utilizar programas bioinformáticos y filogenéticos que ayuden a detectar eventos de conversión génica en familias multigénicas. Asimismo, se desconoce la función de las secuencias reiteradas en los genomas de rhizobias.

### 3. HIPÓTESIS

---

En las bacterias del Orden Rhizobiales existe una gran cantidad de secuencias reiteradas y familias multigénicas que pueden llevar a cabo procesos de recombinación homóloga y conversión génica.

### 4. OBJETIVOS

---

#### 4.1 Objetivo General

Localizar y caracterizar las secuencias reiteradas y detectar eventos de conversión génica en bacterias del Orden Rhizobiales.

#### 4.2 Objetivos específicos

- Detectar eventos de conversión génica en familias multigénicas de rhizobias.
- Localizar *in silico* las secuencias repetidas en genomas de rhizobias.
- Caracterizar las secuencias reiteradas en cuanto a su número de copias, localización genómica y densidad.
- Determinar la función de las reiteraciones.

## 5. MATERIALES Y MÉTODOS

---

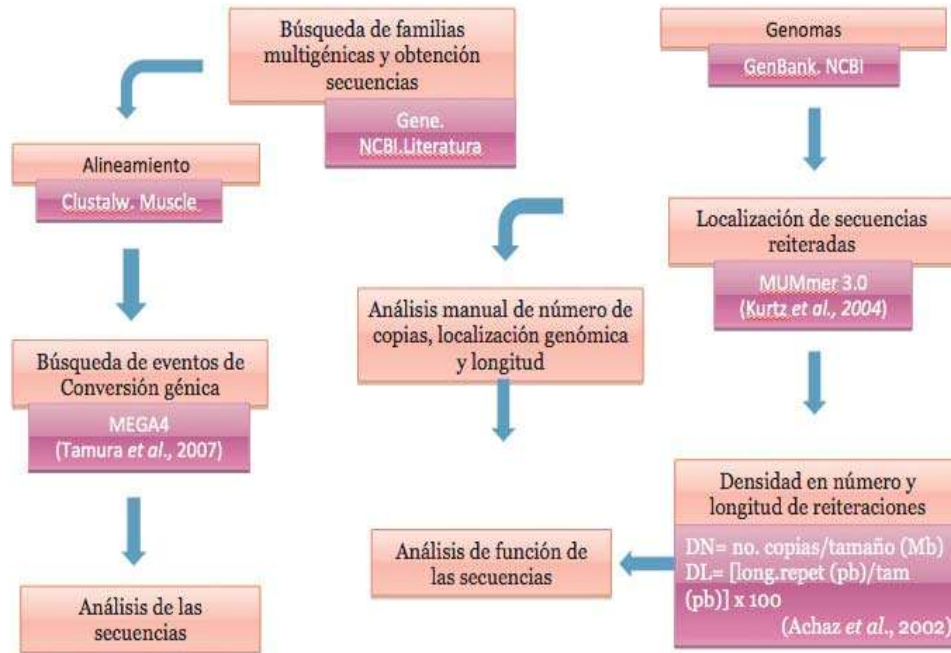
### 5.1 Búsqueda de familias multigénicas en rhizobias

Se realizó una búsqueda exhaustiva de familias multigénicas en el genoma de *Rhizobium etli* CFN42, como cepa de referencia, en la base de datos PUBMED del NCBI (<http://www.ncbi.nlm.nih.gov/pubmed>); a partir de la cual se seleccionaron y analizaron sus genes ortólogos en una búsqueda por similitud tipo BLAST en los siguientes genomas: *Rhizobium leguminosarum*, *Rhizobium* sp. NGR234, *Mesorhizobium loti*, *Sinorhizobium meliloti* y *Bradyrhizobium japonicum*. La secuencia de nucleótidos de cada copia fue obtenida del GenBank (NCBI) con los siguientes números de acceso: *R. etli* CFN42, *nifH* (1005068, 1005033, 1004995), *nifD* (1005069, 1004996), *nifK* (1005079, 1004997), *nodD* (6402815, 6402862, 6402864), *ccm* (3896046, 3896045, 3896047), *adhC* (3895921, 3892269), *tuf* (3892582, 3892567), *fix* (1005327, 1005055, 1005054), *groEL* (3894519, 3891062), *dnaJ* (3890921, 3890921, 3894442) y *purU* (6400149, 6398142); *R. etli* CIAT652, *nifH* (6402850, 6402804, 6402760), *nifD* (6402759, 6402805, 6402849), *nifK* (6402758, 6402848), *nodD* (6402815, 6402862, 6402864), *fix* (6402772, 6402773, 6402774), y *groEL* (6403417, 6399413); *Rhizobium* sp NGR234, *nifH* (962478, 962500), *nifD* (962502, 962482), *nifK* (962485, 962471), *fix* (962188, 962189), *groES* (7788251, 7791635), *groEL* (7792170, 7788321, 7788250, 7791634), y *dnaJ* (7792773, 7791032, 7791742); *R. leguminosarum*, *fix* (4403906, 4403966), *groEL* (4402951, 4398962, 4401984, 4398208), *etfA* (4401546, 4398246), *etfB* (4401547, 4398247), y *hemA* (4398963, 4402319); *M. loti*, *nifA* (1229107, 1229090, 1229044), *repA* (1224477, 1231331), *repB* (1224476, 1231330), *repC*

(1224475, 1231329), *groES* (1231535, 1230998, 1226411, 1226287, 1229067), *groEL* (1231471, 1230997, 1226412, 1226286, 1229066) y *dnaJ* (1228214, 1230424); *B. japonicum*, *groES* (1053428, 1055491, 1048455, 1054913, 1051464), *groEL* (1053427, 1052432, 1055489, 1048307, 1051467, 1051462, 1047245), *rpoN* (1054956, 1054195), *dnaJ* (1052529, 1052410, 1049300, 1048922) y *fix* (1055510, 1054820, 1054853), y *S. meliloti nod* (1235512, 1235224, 1235494), *fix* (1235483, 1235482), *fixO3* (1235141, 1235447, 1235700), *groES* (1235290, 1235216, 1232434) y *groEL* (1235289, 1235215, 1237337, 1232840, 1232433).

## **5.2 Análisis filogenéticos para inferir eventos de conversión génica**

Una vez construida la base de datos, se realizó un alineamiento múltiple de las secuencias génicas utilizando los programas Muscle y ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). Se seleccionaron las familias multigénicas compartidas por más de tres genomas para realizar el análisis de conversión génica mediante evidencia filogenética. Los arboles filogenéticos se construyeron con el programa MEGA 4.0 (Tamura *et al.*, 2007), utilizando el algoritmo Neighbor joining, basado en el método de Kimura 2-parámetros (Kimura, 1980); asimismo, se utilizaron otros métodos como el de mínima evolución y máxima parsimonia, sin embargo, la topología fue similar en todos los casos. Para obtener un valor de confianza, se estableció un valor Bootstrap de 1000 repeticiones (Fig. 5).



**Figura 5.** Metodología general. Por un lado se realizó la búsqueda de familias multigénicas y conversión génica y, por el otro, la localización y caracterización de las secuencias reiteradas a nivel genómico, así como su función.

### 5.3 Análisis de secuencias reiteradas en genomas

Se obtuvieron las secuencias completas de los genomas de *R. etli* CFN42 (NC\_007761.1), *R. leguminosarum* bv. *viciae* 3841 (NC\_008380.1), *S. meliloti* 1021 (NC\_003047), *M. loti* MAFF303099 (NC\_002678.2) y *B. japonicum* USDA110 (NC\_004463.1) del sitio FTP del NCBI (ftp.ncbi.nih.gov) (Fig. 5).

Con la ayuda del programa MUMmer 3.0 (Kurtz *et al.*, 2004) se realizó la búsqueda de secuencias reiteradas mayores a 300 pares de bases (pb) de cada uno de los genomas, incluyendo sus plásmidos. Posteriormente, se realizaron los siguientes análisis por conteo manual: número total de reiteraciones en el

genoma, cantidad de reiteraciones dependiendo del número de copias, longitud de las reiteraciones (pb), localización genómica (cromosoma o plásmidos), así como el porcentaje de las mismas de acuerdo a su localización (Fig. 5).

Con el objetivo de establecer una medida comparativa de las reiteraciones con el tamaño del genoma, se realizó un análisis de densidad, anteriormente propuesto por Achaz *et al.* (2002), el cual relaciona los datos de número de reiteraciones y la longitud respecto al tamaño del genoma:

**Densidad en número (DN)**= no. de copias/tamaño de cromosoma (Mb)

**Densidad en longitud (DL)**= [tamaño de la secuencia repetida (pb)/tamaño del cromosoma (pb)] x 100

Este análisis se realizó, por un lado, tomando en cuenta el genoma completo incluyendo los plásmidos; posteriormente, se llevó a cabo por separado: por un lado aquellas reiteraciones encontradas en el cromosoma y, por el otro, las contenidas en los plásmidos.

Para la búsqueda de la función de las secuencias reiteradas se buscaron inicialmente los marcos de lectura abiertos en el programa ORF Finder del NCBI y se seleccionaron aquellos de mayor cobertura, posteriormente se corroboraron en la base de datos Genome, del NCBI (Fig. 5).

## 6. RESULTADOS

---

### 6.1. Familias multigénicas y conversión génica en rizobias

Los genomas de rizobias son modelos interesantes de estudiar en el área de la recombinación y dinámica de genomas ya que contienen una gran cantidad de familias multigénicas. En este estudio encontramos reportadas en las bases de datos y la literatura 18 familias multigénicas compartidas, en algunos casos, por los cinco genomas analizados (Tabla 1). Por ejemplo, la familia génica de la chaperonina *groEL* aparece reiterada en los cinco organismos, sin embargo, el número de copias es variable en cada genoma, desde cuatro copias en *R. etli*, hasta siete en *B. japonicum*. Algunos otros genes, como los *nifHDK*, que codifican la nitrogenasa, aparecen reiterados solo en *R. etli* y en *R. sp.* NGR234 (Tabla 1).

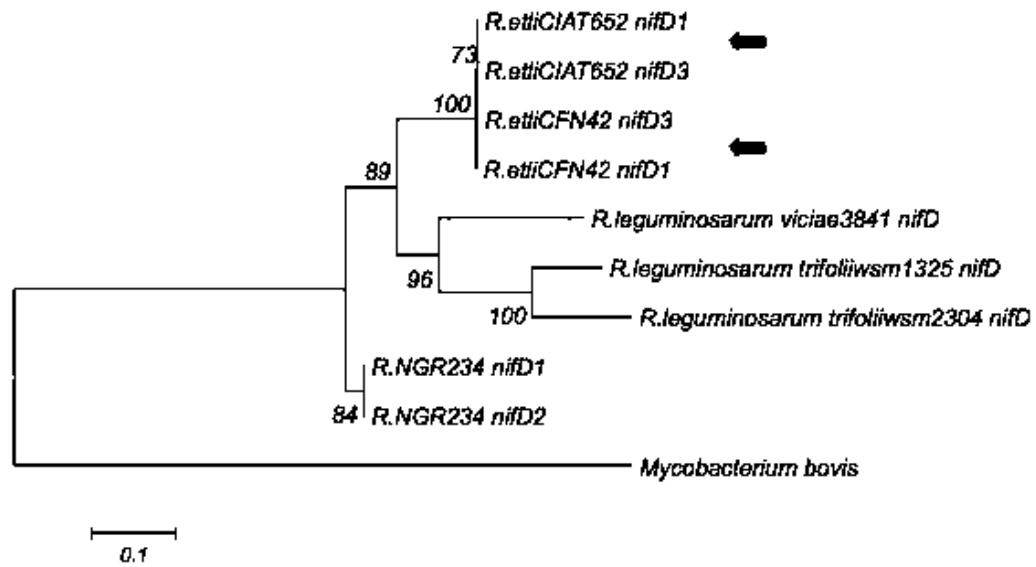
Es interesante resaltar que las copias génicas se localizan a su vez en distintos replicones, algunas se localizan únicamente en cromosoma o en plásmidos, mientras que algunas se distribuyen en ambos replicones (Tabla 1). Esto representa un interesante tema de estudio debido a la cantidad de rearrreglos genómicos que pueden llevarse a cabo, tales como amplificaciones, deleciones o bien, cointegraciones.



**Tabla 1.** Familias multigénicas encontradas en algunas especies de rizobias representativas.

Genoma	Familia Génica	Función	Número de copias	Localización genómica	Referencia	
<i>Rhizobium etli</i> CFN42	<i>nifH</i>	Nitrogenasa reductasa	3	Plásmido	(Quinto <i>et al.</i> , 1982)	
	<i>nifD</i>	Nitrogenasa subunidad $\alpha$	3	Plásmido	GenBank	
	<i>nifK</i>	Nitrogenasa subunidad $\beta$	2	Plásmido	GenBank	
	<i>nodD</i>	Regulador transcripcional	3	Plásmido	GenBank	
	<i>ccm</i>	Citocromo <i>c</i>	4	Ci/P(3)	(González <i>et al.</i> , 2006)	
	<i>adhC</i>	Formaldehído deshidrogenasa	2	Ci/P	(González <i>et al.</i> , 2006)	
	<i>tuf</i>	Factor de elongación	2	Cromosoma	GenBank	
	<i>fix</i>	Transferencia de electrones	4	Plásmido	GenBank	
	<i>groEL</i>	Chaperonina	4	C(3)/P	GenBank	
	<i>dnaJ</i>	Chaperona	3	Cromosoma	GenBank	
	<i>purU</i>	Biosíntesis de purinas	2	Cromosoma	(González <i>et al.</i> , 2006)	
	<i>Rhizobium</i> sp. NGR234	<i>nifH</i>	Nitrogenasa reductasa	2	Plásmido	(Freiberg <i>et al.</i> , 1997)
		<i>nifD</i>	Nitrogenasa subunidad $\alpha$	2	Plásmido	(Freiberg <i>et al.</i> , 1997)
<i>nifK</i>		Nitrogenasa subunidad $\beta$	2	Plásmido	(Freiberg <i>et al.</i> , 1997)	
<i>fix</i>		Transferencia de electrones	3	Plásmido	GenBank	
<i>groES</i>		Chaperonina	2	Ci/P	GenBank	
<i>groEL</i>		Chaperonina	4	C(2)/P(2)	GenBank	
<i>dnaJ</i>		Chaperona	3	Cromosoma	GenBank	
<i>Rhizobium leguminosarum</i>		<i>fix</i>	Transferencia de electrones	2	Plásmido	GenBank
	<i>groEL</i>	Chaperonina	4	C(2)/P(2)	GenBank	
	<i>etfA</i>	Aceptor de electrones	2	Ci/P	GenBank	
	<i>etfB</i>	Aceptor de electrones	2	Ci/P	GenBank	
	<i>hemA</i>	Coenzima	2	Ci/P	GenBank	
<i>Mesorhizobium loti</i>	<i>nifA</i>	Proteína regulatoria	3	Cromosoma	GenBank	
	<i>ccm</i>	Citocromo <i>c</i>	6	Plásmido	(Kaneko <i>et al.</i> , 2000)	
	<i>groES</i>	Co-chaperonina	5	C(4)/P	(Kaneko <i>et al.</i> , 2000)	
	<i>groEL</i>	Chaperonina	5	C(4)/P	(Kaneko <i>et al.</i> , 2000)	
<i>Bradyrhizobium japonicum</i>	<i>groES</i>	Co-chaperonina	5	Cromosoma	(Kaneko <i>et al.</i> , 2002)	
	<i>groEL</i>	Chaperonina	7	Cromosoma	(Kaneko <i>et al.</i> , 2002)	
	<i>rpoN</i>	Factor sigma	2	Cromosoma	(Kaneko <i>et al.</i> , 2002)	
	<i>dnaJ</i>	Chaperona	4	Cromosoma	GenBank	
	<i>fix</i>	Transferencia de electrones	3	Cromosoma	GenBank	
<i>Sinorhizobium meliloti</i>	<i>nodD</i>	Regulación de la nodulación	3	Plásmido	(Honma <i>et al.</i> , 1987)	
	<i>fix</i>	Transferencia de electrones	2	Plásmido	(David <i>et al.</i> , 1987)	
	<i>fixN,O</i>	Citocromo <i>c</i> oxidasa	2	Plásmido	GenBank	
	<i>groES</i>	Chaperonina	3	Plásmido	GenBank	
	<i>groEL</i>	Chaperonina	5	C(2)/P(3)	GenBank	

Asimismo, la Tabla 1 incluye las familias multigénicas que se analizaron para detectar eventos de conversión génica de las cuales fueron seleccionadas de aquellas presentes en más de tres genomas. La figura 6 muestra el árbol filogenético de la familia de genes *nifD*, en la cual fue posible detectar conversión génica en las cepas de *R. etli* CFN42 y CIAT652. Esto se puede determinar ya que los genes parálogos están más cercanamente relacionados filogenéticamente que los ortólogos, lo cual sugiere un mecanismo de homogenización en esta familia multigénica. La conversión génica en estos genes fue convincentemente comprobada al analizar las secuencias de los genes *nifD* (Fig. 7). Al comparar los parálogos de la cepa *R. etli* CFN42, se encontró que su secuencia es idéntica; lo mismo para el caso de la cepa *R. etli* CIAT652, mientras que al comparar los genes de ambas cepas se detectaron dos polimorfismos (Fig. 7), los cuales indicaron los probables eventos de conversión de un tamaño aproximado de 170pb.



**Fig. 6.** Árbol filogenético de la familia multigénica *nifD*. Los posibles eventos de conversión génica se indican con flechas, donde los genes parálogos se encuentran mas cercanamente relacionados que los ortólogos. Se utilizó el algoritmo Neighbor joining con un Bootstrap de 1000.

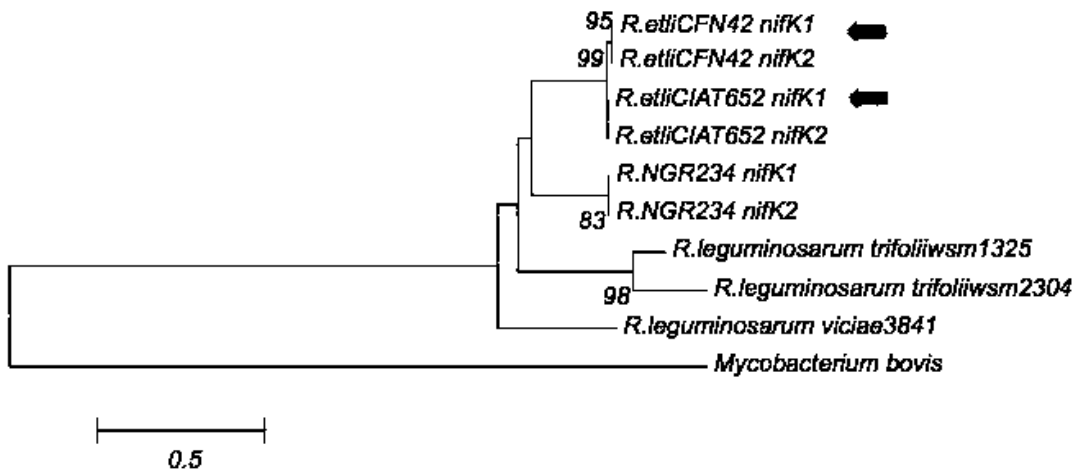
```

R.etliCIAT652_nifD1      ATCGAAGGCATCCGCCCGACCTTGTGGGTCCGGCATCAAGAGAAGTATCCGGTGCAG 1320
R.etliCIAT652_nifD3      ATCGAAGGCATCCGCCCGACCTTGTGGGTCCGGCATCAAGAGAAGTATCCGGTGCAG 1320
R.etliCFN42_nifD1        ATCGAAGGCATCCGCCCGACCTTGTGGGTCCGGCATCAAGAGAAGTATCCGGTGCAG 1320
R.etliCFN42_nifD2        ATCGAAGGCATCCGCCCGACCTTGTGGGTCCGGCATCAAGAGAAGTATCCGGTGCAG 1308
*****
R.etliCIAT652_nifD1      AAGATGGGCATCCCCTTCCGCCAGATGCACTCCTGGGATTATCCGGCCCGTATCACGGC 1380
R.etliCIAT652_nifD3      AAGATGGGCATCCCCTTCCGCCAGATGCACTCCTGGGATTATCCGGCCCGTATCACGGC 1380
R.etliCFN42_nifD1        AAGATGGGCATCCCCTTCCGCCAGATGCACTCCTGGGATTATCCGGCCCGTATCACGGC 1380
R.etliCFN42_nifD2        AAGATGGGCATCCCCTTCCGCCAGATGCACTCCTGGGATTATCCGGCCCGTATCACGGC 1368
*****
R.etliCIAT652_nifD1      TATGACGGCTTGGCCATATTGCCCCGCGACATGGATCTGGCCATCAATAATCCGGTGTGG 1440
R.etliCIAT652_nifD3      TATGACGGCTTGGCCATATTGCCCCGCGACATGGATCTGGCCATCAATAATCCGGTGTGG 1440
R.etliCFN42_nifD1        TATGACGGCTTGGCCATATTGCCCCGCGACATGGATCTGGCCATCAATAATCCGGTGTGG 1440
R.etliCFN42_nifD2        TATGACGGCTTGGCCATATTGCCCCGCGACATGGATCTGGCCATCAATAATCCGGTGTGG 1428
*****
R.etliCIAT652_nifD1      GATCTCTACGACGTCCCCTGGAAAAAAGGGCCGCGCCAGCTGAGGCGGTTGCGGCCGAA 1500
R.etliCIAT652_nifD3      GATCTCTACGACGTCCCCTGGAAAAAAGGGCCGCGCCAGCTGAGGCGGTTGCGGCCGAA 1500
R.etliCFN42_nifD1        GATCTCTACGACGTCCCCTGGAAAAAAGGGCCGCGCCAGCTGAGGCGGTTGCGGCCGAA 1500
R.etliCFN42_nifD2        GATCTCTACGACGTCCCCTGGAAAAAAGGGCCGCGCCAGCTGAGGCGGTTGCGGCCGAA 1488
*****

```

**Fig. 7.** Alineamiento parcial de los genes *nifD* en *Rhizobium etli*. Se indican los polimorfismos en los recuadros. Se empleó el programa Clustalw para realizar el alineamiento.

En el caso de la familia génica *nifK*, se observó la misma relación en cuanto a la cercanía filogenética de los genes parálogos (Fig. 8); sin embargo, se presentó un patrón más complejo al encontrar 14 polimorfismos a lo largo de la secuencia génica entre las copias ortólogas (Fig. 9) y se propone que el tamaño del segmento en el cual se llevo a cabo la conversión es de aproximadamente 1254pb.



**Fig. 8.** Árbol filogenético de la familia multigénica *nifK*. Los posibles eventos de conversión génica se indican con flechas, donde los genes parálogos se encuentran más cercanamente relacionados que los ortólogos. Se utilizó el algoritmo Neighbor joining con un Bootstrap de 1000.

```

R.etliCIAT652nifK1      GAGCAAACGATGTCTTCCATCCAGCCAGCCAGGGCTGCGTCCCTATTACCGTTCCGCAT 306
R.etliCIAT652-nifK2    GAGCAAACGATGTCTTCCATCCAGCCAGCCAGGGCTGCGTCCCTATTACCGTTCCGCAT 306
R.etliCFN42-nifK1      GAGCAAACGATGTCTTCCATCCAGCCAGCCAGGGCTGCGTCCCTATTACCGTTCCGCAT 306
R.etliCFN42-nifK2      GAGCAAACGATGTCTTCCATCCAGCCAGCCAGGGCTGCGTCCCTATTACCGTTCCGCAT 360
*****

R.etliCIAT652nifK1      GCGGTGTTCCGGCGGGTTGAAGAACATGGTCCAGCGGGCTCGCCAATACATACAAGCTCTAC 426
R.etliCIAT652-nifK2    GCGGTGTTCCGGCGGGTTGAAGAACATGGTCCAGCGGGCTCGCCAATACATACAAGCTCTAC 426
R.etliCFN42-nifK1      GCAGTGTTCGGCGGGTTGAAGAACATGGTCCAGCGGGCTCGCCAATACATACAAGCTCTAC 426
R.etliCFN42-nifK2      GCAGTGTTCGGCGGGTTGAAGAACATGGTCCAGCGGGCTCGCCAATACATACAAGCTCTAC 480
*****

R.etliCIAT652nifK1      TTGAATGCCGAAGCGACCCCTGTCGCTGCAGCACTATAACACGCGCAAGACGCTGGGAATAT 906
R.etliCIAT652-nifK2    TTGAATGCCGAAGCGACCCCTGTCGCTGCAGCACTATAACACGCGCAAGACGCTGGGAATAT 906
R.etliCFN42-nifK1      TTGAATGCCGAAGCGACCCCTGTCGCTGCAGCACTATAACACGCGCAAGACGCTGGGAATAT 906
R.etliCFN42-nifK2      TTGAATGCCGAAGCGACCCCTGTCGCTGCAGCACTATAACACGCGCAAGACGCTGGGAATAT 960
*****

R.etliCIAT652nifK1      TGCAAGGAGGTCGGTTCAGGCTACGGCTTCGTTCCATTATCCGCTGGGTGTTCCAGGCGACC 966
R.etliCIAT652-nifK2    TGCAAGGAGGTCGGTTCAGGCTACGGCTTCGTTCCATTATCCGCTGGGTGTTCCAGGCGACC 966
R.etliCFN42-nifK1      TGCAAGGAGGTCGGTTCAGGCTACGGCTTCGTTCCATTATCCGCTGGGTGTTCCAGGCGACC 966
R.etliCFN42-nifK2      TGCAAGGAGGTCGGTTCAGGCTACGGCTTCGTTCCATTATCCGCTGGGTGTTCCAGGCGACC 1020
*****

R.etliCIAT652nifK1      CTGGAACGCGGCCGCTCTCGTCGACGCTATGGCAGATAATCCAGGCTGGCTGCACGGGAAG 1086
R.etliCIAT652-nifK2    CTGGAACGCGGCCGCTCTCGTCGACGCTATGGCAGATAATCCAGGCTGGCTGCACGGGAAG 1086
R.etliCFN42-nifK1      CTGGAACGCGGCCGCTCTCGTCGACGCTATGGCAGATAATCCAGGCTGGCTGCACGGGAAG 1086
R.etliCFN42-nifK2      CTGGAACGCGGCCGCTCTCGTCGACGCTATGGCAGATAATCCAGGCTGGCTGCACGGGAAG 1140
*****

R.etliCIAT652nifK1      ACCGGCGGTGAGCCGACCCACTGCCTTGCTACCAACGGCACCCCGGCTGGGAAGCCGAG 1206
R.etliCIAT652-nifK2    ACCGGCGGTGAGCCGACCCACTGCCTTGCTACCAACGGCACCCCGGCTGGGAAGCCGAG 1206
R.etliCFN42-nifK1      ACCGGCGGTGAGCCGACCCACTGCCTTGCTACCAACGGCACCCCGGCTGGGAAGCCGAG 1206
R.etliCFN42-nifK2      ACCGGCGGTGAGCCGACCCACTGCCTTGCTACCAACGGCACCCCGGCTGGGAAGCCGAG 1260
*****

R.etliCIAT652nifK1      ATGAAGGCGTTGCTCGCATCCTCGCCCTTCGGCCAGGATGCCAGGTCTGGGCGGGCAAG 1266
R.etliCIAT652-nifK2    ATGAAGGCGTTGCTCGCATCCTCGCCCTTCGGCCAGGATGCCAGGTCTGGGCGGGCAAG 1266
R.etliCFN42-nifK1      ATGAAGGCGTTGCTCGCATCCTCGCCCTTCGGCCAGGATGCCAGGTCTGGGCGGGCAAG 1266
R.etliCFN42-nifK2      ATGAAGGCGTTGCTCGCATCCTCGCCCTTCGGCCAGGATGCCAGGTCTGGGCGGGCAAG 1320
*****

R.etliCIAT652nifK1      GACCTGTGGGCGTTGCGCTCGCTGCTCTTTACCGAGCCAGTAGATCTTATGATCGGCAAT 1326
R.etliCIAT652-nifK2    GACCTGTGGGCGTTGCGCTCGCTGCTCTTTACCGAGCCAGTAGATCTTATGATCGGCAAT 1326
R.etliCFN42-nifK1      GACCTGTGGGCGTTGCGCTCGCTGCTCTTTACCGAGCCAGTAGATCTTATGATCGGCAAT 1326
R.etliCFN42-nifK2      GACCTGTGGGCGTTGCGCTCGCTGCTCTTTACCGAGCCAGTAGATCTTATGATCGGCAAT 1380
*****

R.etliCIAT652nifK1      GGTGTGACGGACTATTCTTACGACCTGACCCGCTAG 1542
R.etliCIAT652-nifK2    GGTGTGACGGACTATTCTTACGACCTGACCCGCTAG 1542
R.etliCFN42-nifK1      GGTGTGACGGACTATTCTTACGACCTGACCCGCTAG 1542
R.etliCFN42-nifK2      GGTGTGACGGACTATTCTTACGACCTGACCCGCTAG 1596
*****

```

Fig. 9. Alineamiento parcial de los genes *nifK* en *Rhizobium etli*. Se indican los polimorfismos en los recuadros. Se empleó el programa Clustalw para realizar el alineamiento.

Las filogenias realizadas en el resto de las familias multigénicas no mostraron aparentes eventos de conversión, ya que en todos los casos se encontró que los genes ortólogos están más cercanamente relacionados que los parálogos, lo cual indica una evolución divergente en dichas familias de genes. En la figura 10 se muestra el árbol filogenético de la familia multigénica *groEL*, una familia de chaperoninas muy abundante en los cinco genomas analizados, sin embargo, no se pudieron localizar eventos de conversión en ninguno de los organismos.

Estos resultados muestran evidencia de que las familias multigénicas *nifDK* han sido sujetas a procesos de homogenización de sus secuencias, junto con la familia *nifH* (Santoyo *et al.*, 2005), antes reportada y además abre una área de estudio interesante al establecer que las rizobias contienen un gran número de familias multigénicas, las cuales pueden ser responsables de llevar a cabo importantes rearrreglos genómicos en este grupo bacteriano.

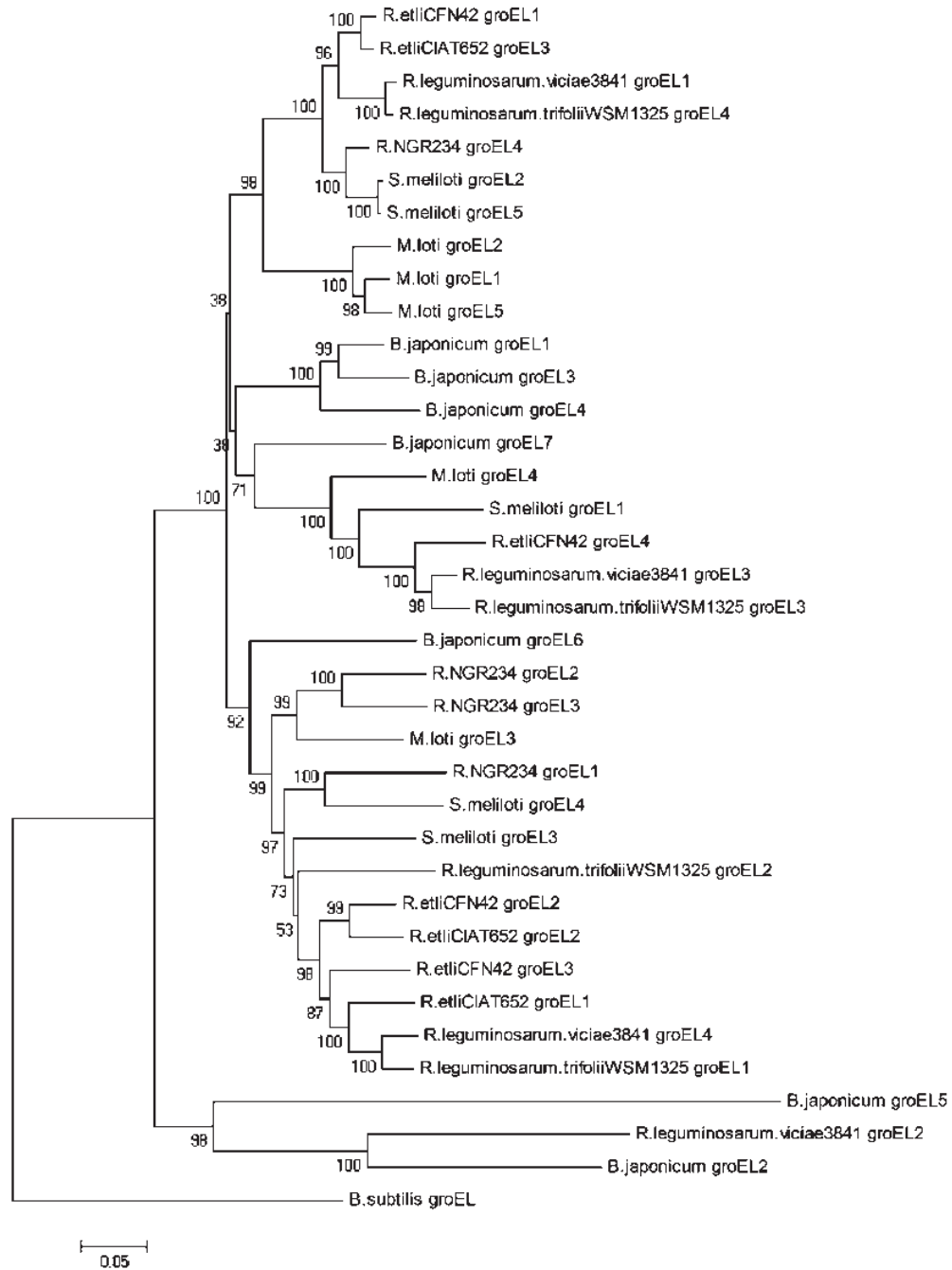


Fig. 10. Árbol filogenético de la familia multigénica *groEL*. El árbol se realizó con el algoritmo neighbor joining. Bootstrap de 1000.



## 6.2. Estructura genómica de las rhizobias analizadas

Las características genómicas que presenta cada uno de los organismos analizados se muestran de forma resumida en la Tabla 2. Si bien se trata de un grupo de organismos que comparten ciertas características como en su estilo de vida, su capacidad de nodular las plantas leguminosas y establecer una simbiosis, a nivel de estructura genómica se pueden notar distintas características. La bacteria *R. etli* contiene el genoma más pequeño, sin embargo, contiene un número elevado de plásmidos, mientras que *B. japonicum*, siendo el genoma más grande, carece de ellos. Asimismo, todos los genomas cuentan con un solo cromosoma cuyo tamaño varía en cada uno de ellos. La suma de los plásmidos también muestra una heterogeneidad en cuanto a los tamaños de los plásmidos, *R. etli* que cuenta con 6 plásmidos, suman aproximadamente 2.1 Mb, mientras que *S. meliloti* que cuenta con solo 2 plásmidos, la suma de ellos equivalen a 3 Mb, casi el tamaño del propio cromosoma. Por lo tanto, se puede observar de manera general que las rhizobias son un grupo muy diverso en cuanto a su tamaño y arquitectura genómica.

**Tabla 2.** Arquitectura genómica de las rhizobias.

Genoma	Cromosoma (pb)	Número de plasmidos (pb)	Tamaño del genoma (bp)	Referencia
<i>Rhizobium etli</i>	4,381,608	6 (2,148,620)	6,530,228	González et al., 2005
<i>Sinorhizobium meliloti</i>	3,654,135	2 (3,037,559)	6,691,694	Galibert et al., 2001
<i>Mesorhizobium loti</i>	7,036,071	2 (560,226)	7,596,297	Kaneko et al., 2000
<i>Rhizobium leguminosarum</i>	5,057,142	6 (2,694,167)	7,751,309	Young et al., 2006
<i>Bradyrhizobium japonicum</i>	9,105,828	0	9,105,828	Kaneko et al., 2002

### 6.3. Secuencias reiteradas en los genomas de rhizobias

Una vez analizadas las familias multigénicas, se inició la búsqueda de las secuencias reiteradas a nivel genómico, las cuales podrían incluir tanto elementos codificantes como no codificantes. Como ya se ha mencionado anteriormente, la recombinación homóloga puede ocurrir entre secuencias idénticas o muy similares a partir de una longitud de 25 pb. Para los fines de este estudio, se realizó la búsqueda de reiteraciones mayores a 300 pb para incrementar la probabilidad de localizar secuencias codificantes y, por lo tanto, familias multigénicas que pueden ser sujetas a eventos de recombinación homóloga y conversión génica con altas frecuencias.

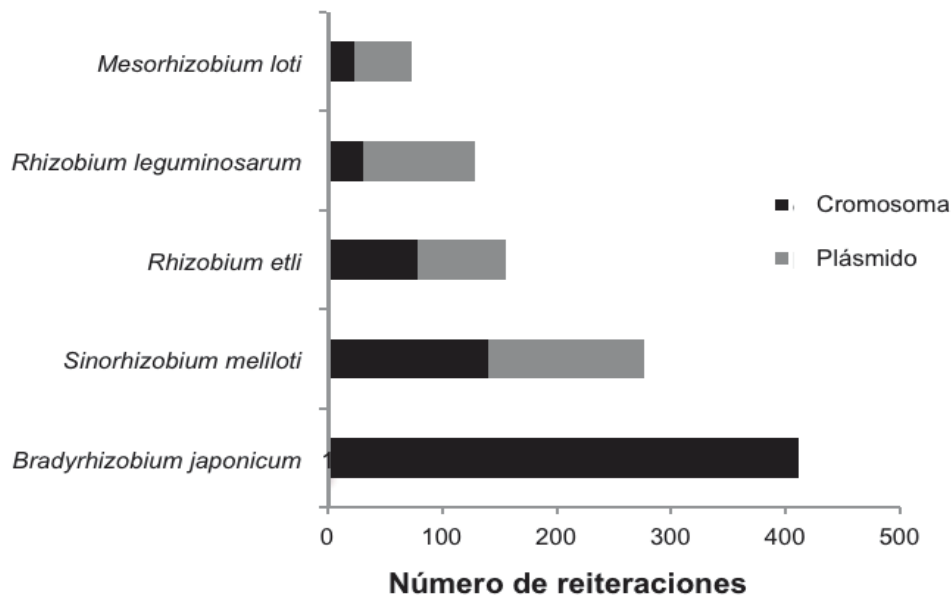
Para realizar dicho análisis, se eligió el programa MUMmer 3.0 (Kurtz *et al.*, 2004). Este programa es capaz de alinear genomas de un tamaño mayor a 5 Mb de manera rápida con la introducción de los paquetes Nucmer y Promer, con la finalidad de encontrar alineaciones exactas entre nucleótidos especificando una longitud mínima. Posteriormente los alineamientos se extienden usando el algoritmo Smith-Waterman, que es el mismo utilizado por BLAST.

El resultado que arroja el programa es una tabla general, donde se indica el número de la posición de inicio de una repetición y la posición de inicio de la segunda repetición, así como la longitud de las mismas. Todas las repeticiones son mostradas por pares, por lo que fue necesario hacer una depuración de toda la matriz de datos, para poder determinar el número real de repeticiones por genoma, así como el número total de copias de cada reiteración.

El análisis del número total de reiteraciones >300pb en cada uno de los genomas se muestra en la figura 11, así como la distribución de las reiteraciones localizadas tanto en el cromosoma como en los plásmidos. Es importante destacar que *B. japonicum* contiene el mayor número de reiteraciones (410), siendo este además el genoma más grande analizado; es

posible pensar que esta relación ocurra en el resto de los genomas, entre mayor tamaño de genoma exista un mayor número de secuencias repetidas; sin embargo, tal correlación no se repite en el resto de los organismos, como en el caso de *S. meliloti* quien es el segundo genoma con mayor número de secuencias reiteradas y contiene uno de los genomas más pequeños.

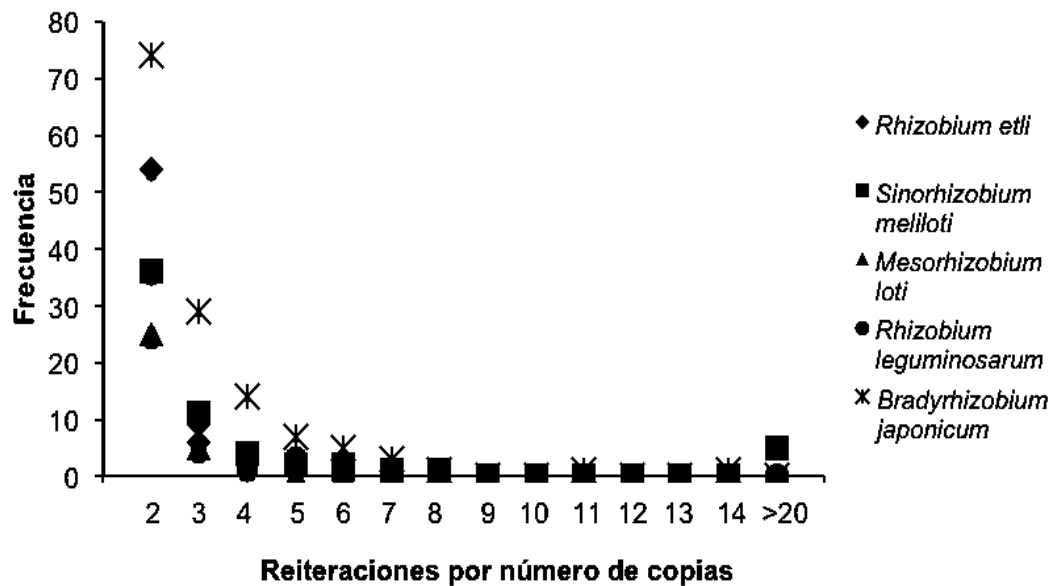
Asimismo, la distribución de las reiteraciones varía entre los diferentes replicones, *S. meliloti* y *R. etli* contienen una distribución similar de reiteraciones tanto en el cromosoma como en los plásmidos y en el caso de *R. leguminosarum* y *M. loti*, la mayoría de sus secuencias reiteradas están presentes en sus plásmidos. Con estos datos se desecha la idea que las rizobias cuentan con un número reducido de secuencias repetidas y, por lo tanto, favorece el interés en el estudio de la recombinación y conversión génica en este grupo bacteriano.



**Fig. 11.** Secuencias reiteradas localizadas en genomas de rizobias. Se muestra el número total de reiteraciones en cada uno de los genomas; en barras color negro se indican las reiteraciones localizadas en el cromosoma y en color gris las localizadas en los plásmidos.

### 6.3.1. Reiteraciones por número de copias en genomas de rizobias

Como parte de la caracterización de las secuencias reiteradas localizadas, se analizaron de acuerdo al número de copias que presentó cada reiteración. En general, se observó que existen una mayor cantidad de secuencias repetidas en aquellas con bajo número de copias, que van desde 2 a 7, observándose la misma tendencia en los cinco genomas analizados (Fig. 12). Además, se localizaron reiteraciones con 11 y 14 copias en *B. japonicum* y algunas reiteraciones con más de 20 copias en *S. meliloti*, lo cual resulta interesante.



**Fig. 12.** Reiteraciones por número de copias en genomas de rizobias. Se indica la frecuencia en que aparecen las reiteraciones por cada número de copias en cada uno de los genomas. Se muestra una alta frecuencia de reiteraciones con bajo número de copias con una tendencia a disminuir conforme aumenta el número de copias.

### 6.3.2 Longitud de las reiteraciones localizadas en genomas de rizobias

Al analizar las reiteraciones de acuerdo a su longitud (Fig. 13), se observó una distribución más o menos homogénea en cuanto a los distintos tamaños de las reiteraciones; si bien, de manera general, se puede observar que la mayoría de las secuencias reiteradas corresponden a tamaños que van desde 300pb hasta 1000pb, no así el caso de *S. meliloti*, que muestra una abundancia similar desde los 400 hasta los 1400pb (Fig. 13). De hecho, los cinco organismos muestran reiteraciones de longitudes mayores a 1000pb, incluso algunos de hasta 6000 pb, lo cual es relevante ya que se trata de tamaños que pueden corresponder a una secuencias codificantes e incluso operones.

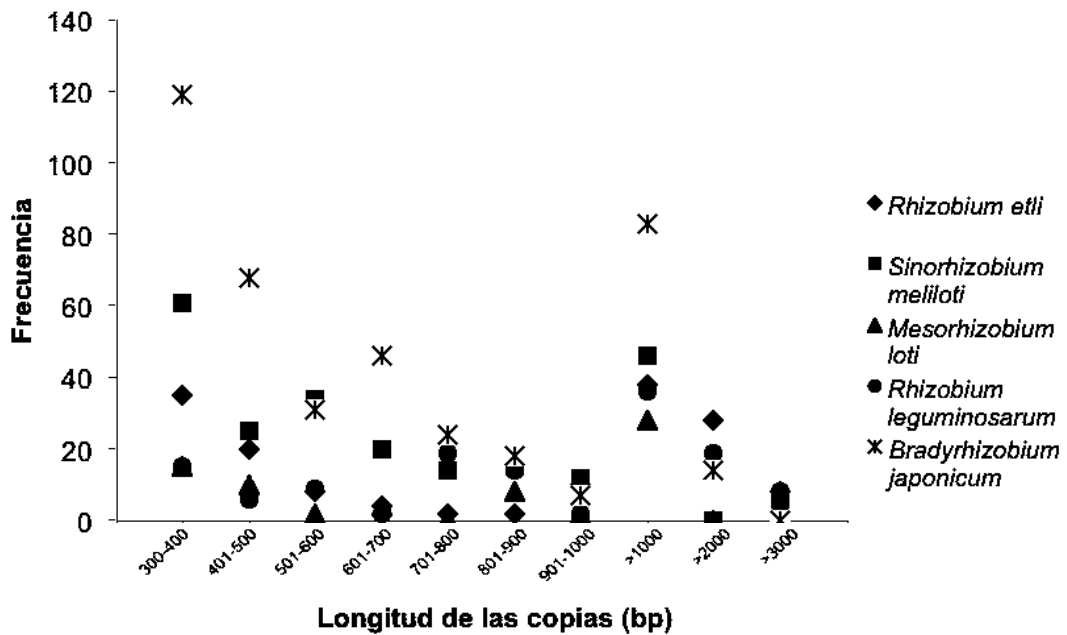


Fig. 13. Longitud de las reiteraciones localizadas en genomas de rizobias. Se muestra la frecuencia de las reiteraciones de distintas longitudes, desde 300 hasta mas de 3000 pb en cada uno de los genomas.

### 6.3.3 Densidades de las reiteraciones por genoma

Con la obtención de los datos de número de reiteraciones y su longitud, se determinó el análisis de densidad propuesto por Achaz *et al.* (2002) para establecer una medida comparativa y homogénea entre los resultados de cada genoma. En cuanto a la densidad en número de las secuencias reiteradas a nivel genómico, se encontró que *B. japonicum* es el genoma con una mayor densidad de reiteraciones (45.1), mientras que *M. loti* contiene la menor densidad (9.6), con lo cual nuevamente se descarta la posibilidad de correlacionar un mayor tamaño de genoma con mayor número de reiteraciones (Tabla 3).

**Tabla 3.** Densidad de las secuencias reiteradas a nivel genómico.

Genoma	Tamaño (Mb)	Total de reiteraciones	Longitud de reiteración	Densidad (N)	Densidad (L)
<i>Rhizobium etli</i>	6.5	155	175,052	23.8	2.7
<i>Sinorhizobium meliloti</i>	6.7	276	239,329	41.2	3.6
<i>Rhizobium leguminosarum</i>	7.8	129	183,907	16.5	2.4
<i>Mesorhizobium loti</i>	7.6	73	74,448	9.6	1.0
<i>Bradyrhizobium japonicum</i>	9.1	410	311,614	45.1	3.4

**N=** número de reiteraciones

**L=** longitud de las reiteraciones

Por otra parte, el análisis de densidad de la longitud de las reiteraciones a nivel genómico indica que *S. meliloti* es el organismo que cuenta con una mayor cobertura de reiteraciones en su genoma, con un 3.6%, mientras que *M. loti* sigue siendo el organismo con menor porcentaje de densidad en longitud (Tabla

3). Estos resultados indican que tampoco existe una relación entre densidad en número y una mayor cobertura en cuanto a la longitud de las reiteraciones, ya que en el caso de *B. japonicum*, el cual mostró la más alta densidad en número, no cuenta con el valor más alto de densidad en sus reiteraciones.

#### 6.3.4 Densidad de las reiteraciones por localización genómica

Para hacer un análisis más profundo sobre la distribución de estas reiteraciones, se llevó a cabo el mismo procedimiento anterior de determinar la densidad de las reiteraciones, en este caso de manera separada, por un lado en el cromosoma y, por el otro, en los plásmidos de cada organismo. Se obtuvieron resultados interesantes; primero, el análisis de densidad en número, mostró que existe una mayor ocurrencia de secuencias reiteradas en los plásmidos de todos de los genomas, siendo los valores muy elevados respecto a los obtenidos en el cromosoma, desde 36.4 en *R. leguminosarum* hasta un 89.2 en *M. loti* (Tabla 4). Esto sugiere una alta dinámica de las reiteraciones en este tipo de replicones, lo cual se refleja a su vez, en la longitud que abarcan, ya que este mismo análisis de densidad en la longitud indica de igual manera, una mayor cobertura de reiteraciones en plásmidos que en cromosomas, solo existe una excepción, en el caso de *S. meliloti*, donde la densidad es ligeramente mayor en su cromosoma.

En general, la densidad de las reiteraciones es más alta en plásmidos en todos los genomas, lo cual sugiere una elevada dinámica en dichas estructuras.

**Tabla 4.** Densidad de las reiteraciones de acuerdo a su localización en el genoma.

Genoma	Número de reiteraciones		Densidad (Número)		Densidad (Longitud)	
	Cromosoma (%)	Plásmido (%)	Densidad (C)	Densidad (P)	Densidad (C)	Densidad (P)
<i>Rhizobium etli</i>	79 (51)	76 (49)	18.0	35.4	2.6	3.8
<i>Mesorhizobium loti</i>	23 (32)	50 (68)	3.3	89.2	0.4	8
<i>Rhizobium leguminosarum</i>	31 (24)	98 (76)	6.1	36.4	1.5	4
<i>Sinorhizobium meliloti</i>	140 (51)	136 (49)	38.3	44.8	4.0	3.1

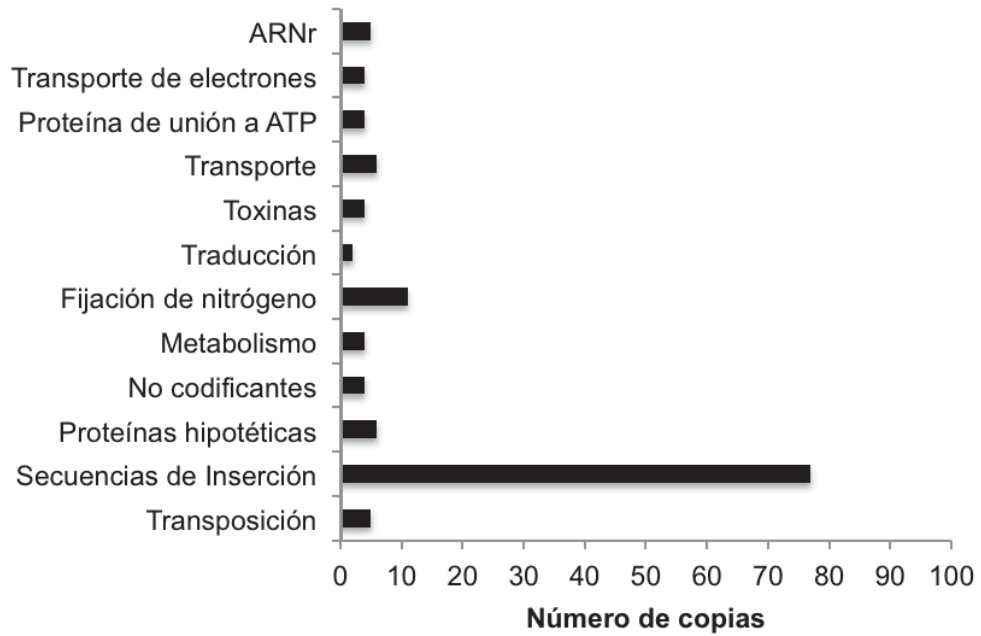
**C=** cromosoma

**P=** plásmido

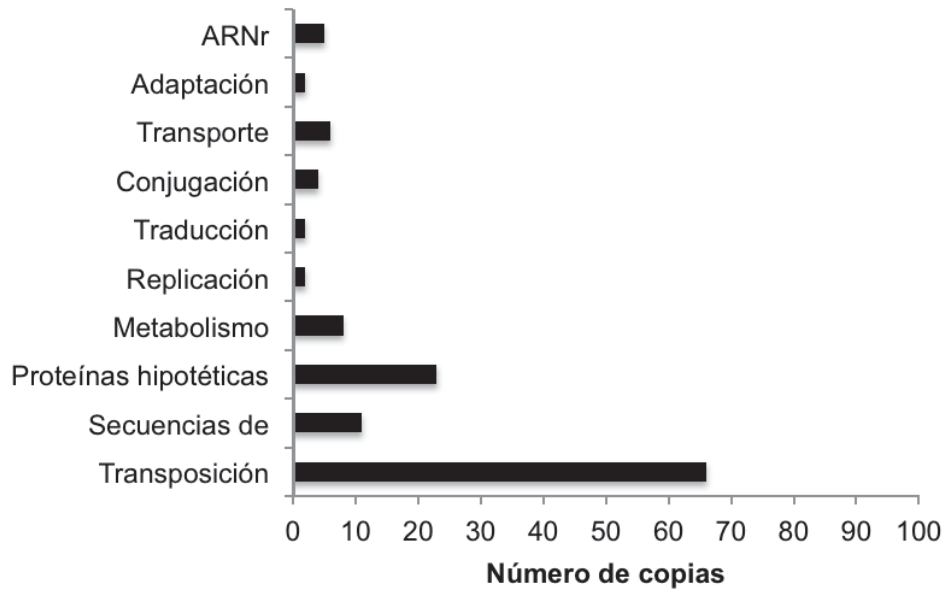
### 6.3.5 Funciones de las reiteraciones en genomas de rizobias

Finalmente, se realizó la búsqueda de la función de las secuencias reiteradas en cada cepa analizada. Los resultados se muestran en la figura 14, en la cual se incluyen todas las reiteraciones agrupadas de acuerdo a la función que desempeñan en el organismo. Tal como se esperaba, se encontró una gran cantidad de secuencias codificantes pertenecientes a familias multigénicas. Sin embargo, las reiteraciones más abundantes en los cinco organismos corresponden a elementos móviles, los cuales se encontraron desde 28 copias en *M. loti* hasta 360 en *B. japonicum*; en el caso de *R. etli* predominaron las secuencias de inserción con 78 copias.





**Fig. 14.** Funciones de las secuencias reiteradas en *R. etli*. Las barras indican el numero de copias localizadas para cada una de las funciones en el lado izquierdo.



**Fig. 15.** Funciones de las secuencias reiteradas de *R. leguminosarum*. Las barras indican el numero de copias localizadas para cada una de las funciones en el lado izquierdo.

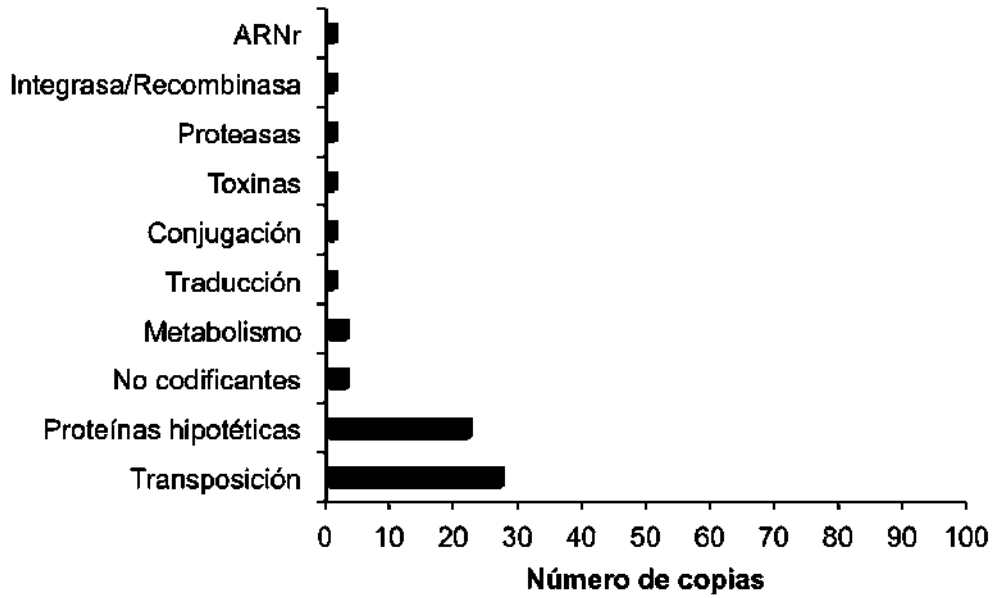


Fig. 16. Funciones de las secuencias reiteradas de *M. loti*. Las barras indican el numero de copias localizadas para cada una de las funciones en el lado izquierdo.

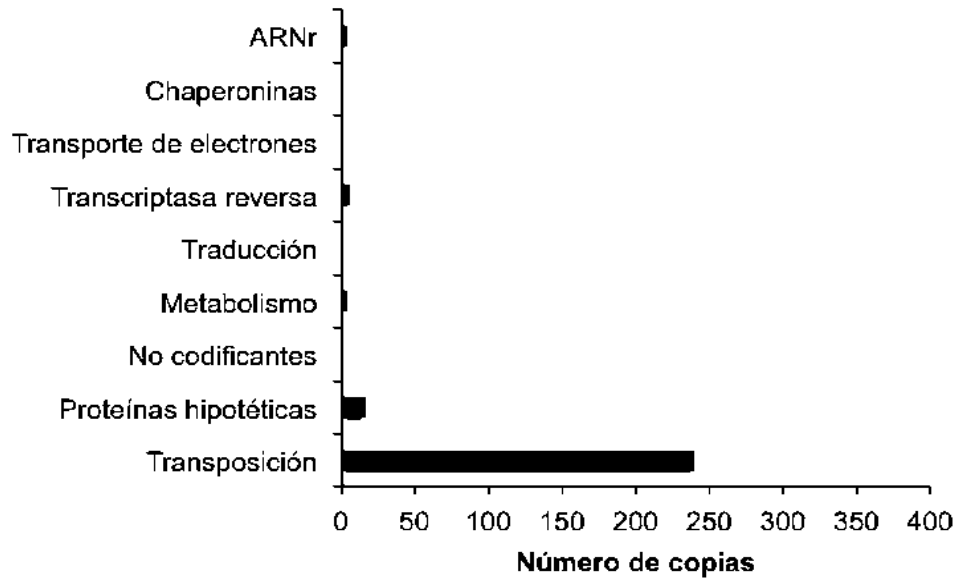
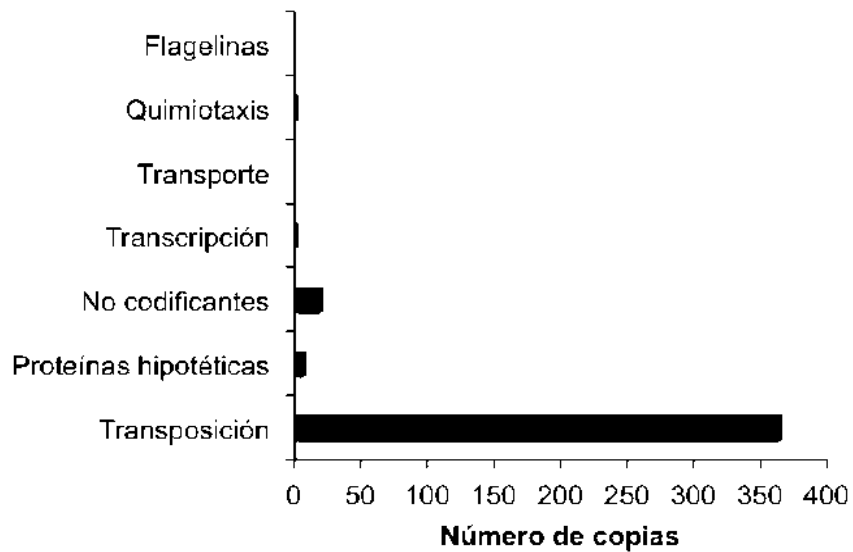


Fig. 17. Funciones de las secuencias reiteradas de *S. meliloti*. Las barras indican el numero de copias localizadas para cada una de las funciones en el lado izquierdo.



**Fig. 18.** Funciones de las secuencias reiteradas de *B. japonicum*. Las barras indican el número de copias localizadas para cada una de las funciones en el lado izquierdo.

Asimismo se encontraron reiteraciones correspondientes a genes que codifican para funciones vitales para los organismos bacterianos, tales como: metabolismo, replicación, transcripción, traducción, conjugación, genes ARNr, etc. Dichas funciones son relevantes para la sobrevivencia de los organismos, lo que sugiere que las reiteraciones juegan un papel importante en el mantenimiento de dichos genes importantes para la sobrevivencia.

## 7. DISCUSIÓN

---

Las bacterias simbiotes de plantas leguminosas capaces de fijar el nitrógeno, conocidas como rizobias, son un grupo de origen polifilético cuya forma de vida se asemeja entre los diferentes organismos, pueden vivir de manera saprofítica habitando la rizósfera, o bien como endosimbiontes en su hospedero específico (Sadowsky y Graham, 1998). Si bien se describen ciertas similitudes en cuanto a su ecología, las diferencias en cuanto a su distribución taxonómica han llevado a discutir ampliamente su historia evolutiva. Las reconstrucciones filogenéticas de los genes *nod*, que codifican para los factores nod, indican un ancestro común entre ellos (Ueda *et al.*, 1995; Laguerre *et al.*, 2001), por lo que la hipótesis de una evolución independiente es poco probable. La hipótesis más apoyada por las evidencias es la de un linaje simbiótico único del que se transfirió lateralmente la información al resto de los organismos (Young y Johnston, 1989).

Una muestra de lo anterior es probablemente la presencia de múltiples familias de genes presentes en diferentes organismos (Tabla 1). En este estudio se encontraron distintas familias multigénicas compartidas por los cinco organismos, lo cual sugiere un alto intercambio de material genético entre organismos, además de que muchos de estos genes se localizan en los plásmidos, que son estructuras altamente dinámicas. Además, las diferentes reiteraciones encontradas se localizaron en diversos replicones, ya sea cromosoma o plásmido, o en ambos. Esta amplia localización de las reiteraciones sugiere que pueden ser blanco para llevar a cabo eventos de recombinación homóloga y así generar distintos arreglos genómicos y/o conversión génica (Orozco-Mosqueda *et al.*, 2009).

Asimismo se comprobó la factibilidad de dichas secuencias a ser sujetas a recombinación homóloga, y por lo tanto, a eventos de conversión génica en dos familias multigénicas de *R. etli*: *nifD* y *nifK* (Fig. 6 y 8), las cuales representan apenas las tres primeras familias (incluyendo *nifH*), donde se han localizado estos eventos en esta bacteria (Santoyo *et al.*, 2005). Se ha sugerido que el principal papel de la conversión es la homogenización de las secuencias, como en el caso de *nifH* (Santoyo *et al.*, 2005); sin embargo, se desconoce aún el significado biológico del mantenimiento homogéneo de estos genes. Una razón puede ser debido a que un mayor número de genes podría mejorar algún evento biológico adaptativo al producir más proteínas. Estudios posteriores pueden revelar si la conversión desempeña algún otro papel, además de la homogenización de las secuencias.

Por otro lado, el tamaño de los segmentos de conversión localizados para el caso de los genes *nifD* y *nifK*, son apropiados para llevar a cabo recombinación y, por lo tanto, conversión génica, ya que se ha reportado una longitud de conversión en los genes *nifH* de 100-800pb (Santoyo y Romero, 2005); en el caso de la familia *nifD*, se encontró una distancia de 170pb entre los dos polimorfismos (Fig. 7). En la familia *nifK* se encontraron múltiples polimorfismos con una distancia de 1254pb desde la posición del primer polimorfismo hasta el último (Fig. 9). Por lo tanto, probablemente se lleven a cabo eventos de conversión entre estos segmentos de ADN.

Por otra parte, se ha sugerido que la presencia de dos o más genes relacionados con la fijación del nitrógeno podría tener ventajas desde el punto de vista biotecnológico, ya que se podrían seleccionar cepas con amplificaciones de dichos genes que presenten mayores capacidades de nodulación y fijación de nitrógeno y así mejorar de manera indirecta el crecimiento de la planta (Mavingui *et al.*, 1997). Esto además puede tener una

mayor alcance a distintos genes que puedan ser amplificados al azar, conociendo su localización y seleccionarse de acuerdo a sus ventajas adaptativas (Flores *et al.*, 2000).

Con los avances en el área bioinformática y genómica, la secuenciación de gran cantidad de organismos de diversos taxa ha ampliado el conocimiento de los organismos a nivel genómico; sin embargo, el conocimiento que se tiene sobre las rhizobias es aún limitado y fragmentado. De acuerdo a las características generales de los organismos analizados, es posible establecer una diferencia en cuanto a su arquitectura genómica, que probablemente refleje características evolutivas distintas. Es importante resaltar la diferencia en cuanto al número de plásmidos, ya que pueden ser producto de una gran dinámica entre replicones e incluso entre organismos (Tabla 2). Se sugiere que la adquisición de plásmidos puede tener diferentes orígenes, por ejemplo en *S. meliloti* se propone que el plásmido pSymA se ha adquirido por transferencia horizontal (Galibert *et al.*, 2001), mientras que el plásmido pSymB pudo haber sido parte del cromosoma. Asimismo, se han analizado distintas cepas de *R. leguminosarum* donde se ha observado un número variable de plásmidos, de 2 a 6 (Young *et al.*, 2006), lo cual sugiere que los organismos no son estables sino que tienen una alta dinámica en sus replicones, ya sea intra o inter organismos.

Lo anterior tiene relación con la hipótesis inicial del origen de las rhizobias propuesto por Young y Johnston (1989) sobre la transferencia lateral de información genética a partir de un linaje simbiótico, donde también se cuenta con evidencia de tal transferencia en la cepa *M. loti* R7A, la cual adquirió su capacidad de nodular gracias a la transferencia lateral de la isla simbiótica de cepas nodulantes (Sullivan y Ronson, 1998).

Es interesante notar la relación que se muestra entre el genoma de mayor tamaño y el mayor número de reiteraciones en el caso de *B. japonicum* (Fig. 11), lo cual se esperaría para el resto de las especies; sin embargo, esto no ocurre así. No existe una correlación entre el tamaño y mayor número de reiteraciones. Esto se ha observado en un estudio de Rocha *et al.* (1999), donde se concluye que la densidad del número de repeticiones no cambia respecto al tamaño en los genomas que analizó; sin embargo, en organismos patógenos sí se observa una tendencia a tener mayor densidad de reiteraciones. En un trabajo de Romero *et al.*, 1999, se observó que algunos organismos causantes de enfermedades como *Helicobacter pylori*, *Escherichia coli* y *Mycobacterium tuberculosis*, contienen un elevado número de copias por cada reiteración, lo cual probablemente pueda atribuirse a su estilo de vida, ya que la gran cantidad de secuencias reiteradas en patógenos tiene implicaciones importantes para la variación antigénica; si embargo, cabe mencionar que su estudio solo se enfocó a las reiteraciones localizadas en el cromosoma y no a nivel genómico.

A diferencia de los organismos causantes de enfermedades, en las rizobias, es notable la gran cantidad de secuencias con bajo número de copias encontradas (la mayoría de 2 y 7 copias) (Fig. 12), las cuales pueden indicar algún papel importante, propio de este grupo de organismos. Por lo tanto, es interesante indagar sobre cuales son las secuencias reiteradas encontradas y qué posible papel estén desempeñando en las rizobias. Por otro lado, se encontraron reiteraciones con un alto número de copias (11 y 14) en *B. japonicum*, e incluso cinco reiteraciones con más de 20 copias en *S. meliloti* (Fig. 12), todas ellas corresponden a ISs. En *B. japonicum* se desconoce este tipo de datos, sin embargo, en *S. meliloti* se ha observado que la presencia de gran cantidad de reiteraciones puede dirigir a distintas arquitecturas genómicas (Guo *et al.*, 2003). De igual manera, existe evidencia de que las ISs están

involucradas en la evolución de los genomas, al ser responsables de que distintos rearrreglos genómicos ocurran, incluso se ha observado su participación en la reducción del tamaño del genoma de algunas especies de *Bordetella* (Parkhill et al., 2003) y de amplificaciones en distintos patovares de *Yersinia pestis* (Parkhill et al., 2001; Chain et al., 2004).

En lo que respecta a la longitud de las reiteraciones, cabe destacar que se encontraron longitudes variables que van de 300 a 1000 pb, lo cual puede asegurar que probablemente estén sujetas a recombinación (Fig. 13). Este resultado coincide con lo reportado por Treangen *et al.*, 2009, donde se analizaron las secuencias reiteradas de los cromosomas de 720 organismos y se encontró que el 80% de las reiteraciones contienen tamaños menores a 2000pb; sin embargo, en este estudio se encontraron gran cantidad de reiteraciones entre los rangos de 1001-2000, 2000-3000, y >3000 pb, encontrando incluso algunas de hasta 6000 pb (Fig. 13), lo cual indica que probablemente la mayoría se trate de secuencias codificantes e incluso operones. Además se confirma la idea de que los procariontes, si bien cuentan con genomas pequeños comparados con los eucariontes, contienen una gran cantidad de reiteraciones y además de gran tamaño (>20 pb) (Rocha *et al.*, 1999).

Así pues, es importante resaltar que la presencia de reiteraciones largas pueden otorgar al organismo distintos mecanismos evolutivos para adaptarse a su medio, como la transferencia de genes, o en el caso de patógenos, de llevar a cabo variación antigénica; aunque se desconoce cual sea su papel principal en las rizobias o si existen mecanismos evolutivos similares. Por otro lado, se sugiere que las reiteraciones largas son muy difíciles de persistir, a menos de que ocurran diversos rearrreglos genómicos que las coloquen a una mayor



distancia en el genoma (Achaz *et al.*, 2002), dadas las altas frecuencias de recombinación entre ellas que involucran deleciones.

El análisis comparativo a nivel genómico requiere una medida con la cual se representen los datos de manera homogénea, por lo que se realizó el análisis de densidad, por un lado a nivel de genoma (incluyendo cromosoma y plásmidos) (Tabla 3) y, por otro lado, de acuerdo a la localización en los distintos repicones (Tabla 4). A nivel genómico no se observó una correlación entre una mayor densidad en el número de reiteraciones y su densidad o cobertura en el genoma; es decir, de acuerdo a Treangen *et al.* (2009), un genoma muy grande podría implicar un mayor número de familias reiteradas, pero ello no implica que a su vez incremente el tamaño y la cobertura de tales reiteraciones. Achaz *et al.* (2002) sugieren que la densidad de la reiteración es importante ya que puede reflejar la historia evolutiva del cromosoma y la tolerancia de éste a ser reiterado.

A la fecha existen pocos estudios donde se analiza la densidad de las reiteraciones; en el trabajo ya mencionado de Treangen *et al.* (2009), reportan una correlación muy baja entre tamaño de genoma y la cobertura de las repeticiones (densidad), esto solamente en cromosomas, lo cual resulta difícil de comparar ya que en el presente estudio se realizó tal análisis incluyendo los plásmidos, en los cuales se observó una mayor densidad en cuanto a número de reiteraciones respecto al cromosoma (Tabla 4). Esto se puede explicar debido a que, como ya hemos visto, existe una alta dinámica en los genomas, especialmente en los plásmidos, ya que éstos pueden no ser esenciales para la sobrevivencia de la bacteria; en el caso del cromosoma, al contener los genes esenciales, es importante la conservación de dichos genes importantes para la sobrevivencia.

Por otra parte, si se comparan los valores de densidad en cuanto a la longitud o cobertura de las reiteraciones obtenidos en este grupo de bacterias (en un rango de 0.4 a 4%) con aquellos obtenidos por Treangen *et al.* (2009), es muy notoria la diferencia, siendo *Orientia tsutsugamushi* el organismo con el genoma más reiterado, con 42.2%; de hecho, los 10 primeros genomas con mayor cobertura rebasan el 20%. Por lo anterior se puede asumir, que en general las rhizobias cuentan con una baja cobertura de reiteraciones, aunque cabe mencionar que los organismos de Treangen *et al.* (2009) son todos simbioses obligados y patógenos. Por lo tanto, se puede establecer que las rhizobias poseen un alto número de secuencias reiteradas en sus genomas, principalmente en sus plásmidos; sin embargo, no cuentan con una alta cobertura de reiteraciones, tanto en su cromosoma como en los plásmidos, lo cual probablemente se deba a alguna característica propia de este grupo bacteriano que aún falta por establecer.

Para ahondar en este trabajo, se determinó la función de las secuencias reiteradas encontradas, así como con sus múltiples copias. La presencia de un alto número de elementos de transposición era de esperarse (Fig. 14-18), al ser estos los elementos más abundantes en los organismos (Aziz *et al.*, 2010); si bien se desconoce su función y la razón de ser tan abundantes, por un lado pueden favorecer la ocurrencia de eventos de recombinación y, por lo tanto, generar diferentes rearrreglos. Por otro lado, se ha observado que la presencia de un transposón en el genoma de *Pseudomonas fluorescens* juega un papel relevante en la colonización de espacios en la rizósfera (Dekkers *et al.*, 1997) lo cual puede asumirse para otras rhizobias como una posible razón de la alta reiteración de estos elementos; por lo tanto, sería interesante llevar a cabo este tipo de estudios en bacterias fijadoras de nitrógeno.

Solo en el caso de *R. etli*, las secuencias de inserción fueron las de mayor frecuencia, siendo éstos los ejemplos más abundantes de los elementos móviles (Fig. 14). Si bien se sabe que las IS están involucradas en distintos rearrreglos genómicos, recientemente se ha observado que en diversas cepas de *R. etli* provenientes de distintas regiones geográficas, las ISs mantienen un alto grado de conservación en el contexto genómico y menores eventos de recombinación que los genes “housekeeping” cromosomales (Lozano *et al.*, 2010). Por lo que es difícil determinar el papel que desempeñan estas secuencias en los diferentes organismos.

Además de los elementos móviles, se encontraron las proteínas hipotéticas como el segundo grupo más abundante entre los genomas, seguido de genes involucrados en el metabolismo, factores de traducción, genes ARNr y de transporte. Dichas funciones son relevantes para la sobrevivencia de los organismos, lo que sugiere que las reiteraciones juegan un papel importante en el mantenimiento de dichos genes para la sobrevivencia, hipotetizando que duplicaciones de genes importantes podrían ser una reserva de funciones esenciales para la sobrevivencia.

Este trabajo representa además, el primero en mostrar la función para la cual codifican las secuencias reiteradas encontradas en este grupo de organismos y cabe mencionar, que se encontraron muy pocas secuencias no codificantes, la mayoría pertenecen a familias multigénicas con funciones importantes para el organismo. Finalmente, este estudio a nivel genómico abre una ventana que permite entender más la organización y distribución en cuanto a la presencia de secuencias reiteradas y, probablemente ayude a entender más aspectos relacionados con la ecología de este grupo bacteriano, así como la evolución y dinámica de sus replicones.

## 8. RESUMEN DE RESULTADOS

---

- Se encontraron 18 familias multigénicas compartidas entre los cinco genomas.
- Se detectaron eventos de conversión génica mediante análisis filogenéticos en dos familias génicas de *R. etli*: *nifD* y *nifK*.
- El tamaño probable de los segmentos de conversión en las dos familias génicas es de 170 y 1254 pb respectivamente.
- En el análisis genómico se encontraron un total de 410 reiteraciones en *B. japonicum*, 276 en *S. meliloti*, 155 en *R. etli*, 129 en *R. leguminosarum* y 73 en *M. loti*.
- Los cinco genomas contienen principalmente una alta cantidad de reiteraciones de bajo número de copias (2-7).
- La longitud de las repeticiones generalmente contienen tamaños entre 300 a 1000 pb.
- El genoma de *B. japonicum* contiene la mayor densidad de número de repeticiones, mientras que el genoma de *M. loti* tiene la menor densidad.
- En cuanto a la densidad del tamaño de las repeticiones, *S. meliloti* es el de mayor densidad y nuevamente *M. loti* el más bajo.
- De acuerdo a la localización genómica de las reiteraciones, se encontró que hay una mayor densidad de número de repeticiones en los plásmidos que en el cromosoma. Lo mismo ocurre en la densidad del tamaño de las repeticiones, excepto en el caso de *S. meliloti*, que es mayor en el cromosoma.
- La mayoría de las repeticiones pertenecen a elementos móviles en todos los genomas analizados. Otras funciones importantes: metabolismo, transcripción, traducción, conjugación y transporte.

- Un pequeño porcentaje de las repeticiones no tienen funciones codificantes o tienen funciones hipotéticas.

## 9. CONCLUSIONES

---

Los genomas de las rizobias analizados contienen un gran número de familias multigénicas y secuencias reiteradas, la mayoría de las cuales pertenecen a secuencias codificantes, que pueden llevar a cabo eventos de recombinación homóloga y conversión génica.

## 10. RESULTADOS EXPERIMENTALES ADICIONALES

---

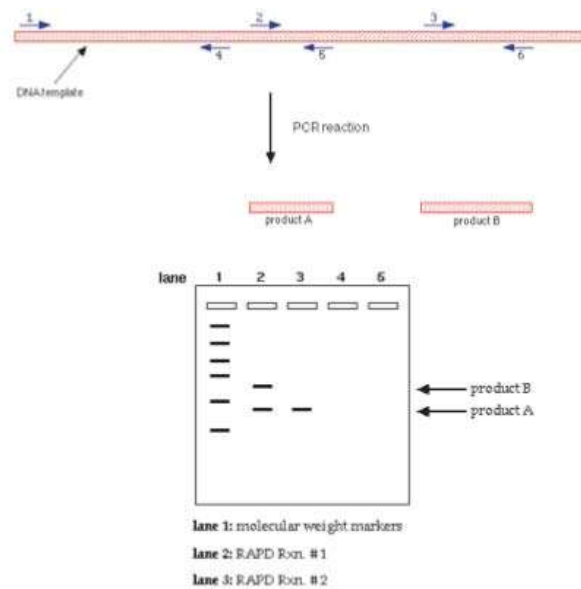
Debido a que se desconoce la relación entre la densidad de reiteraciones en los genomas de rizobias y su capacidad de adaptación al ambiente rizosférico, se planea probar dicha hipótesis en futuros trabajos de investigación. En especial, sería interesante realizar un estudio poblacional de rizobias en ambientes rizosféricos. Para ello, es necesario establecer la diversidad de organismos que habitan la rizósfera en un inicio, para después determinar cuáles serían los genomas de nuestro objeto de estudio. Cabe destacar que si se logra determinar que las poblaciones de rizobias más abundantes en la rizósfera tienen una mayor densidad de reiteraciones, se podrían correlacionar o asociar ambas características y probar nuestra hipótesis. Por lo tanto, un primer paso para probar dicha hipótesis consiste en conocer la diversidad genética de un sistema como la rizósfera, por lo que se requiere de técnicas eficaces que faciliten el análisis de gran cantidad de organismos de manera rápida y a bajo costo. Por tal motivo, se plantea buscar un método que sea capaz de facilitar dicho análisis en la rizósfera.

La rizósfera, se define como la zona del suelo que rodea a la raíz, en donde ocurre una intensa actividad biológica y química influida por los exudados de la raíz y los microorganismos del suelo (Barea y Azcón-Aguilar, 1982). Se trata de un ambiente muy diverso en cuanto a microorganismos. Para el estudio de esta diversidad se han empleado técnicas como la metagenómica, el análisis de los genes ARNr 16S, la tipificación como RFLP (Restriction Fragment Length Polymorphism), ARDRA (Amplified rDNA Restriction Analysis) y RAPD (Random Amplified Polymorphic DNA).

La técnica RAPD presenta las características deseadas para este análisis, ya que utiliza oligonucleótidos de 10 nt, para lo cual no se requiere conocer la

secuencia del genoma de manera previa; cuenta con ciertas ventajas ya que requiere pequeñas cantidades de ADN y no se necesita clonar ni secuenciar, por lo que representa una técnica de bajo costo.

La forma en cómo funciona la técnica se ilustra en la figura 19; los oligonucleótidos se alinean al azar en la molécula de ADN tanto en dirección directa como reversa y solo aquellos oligonucleótidos que se orienten de manera adecuada a una distancia apropiada, serán los que generen un producto de PCR. Por lo tanto, una ligera variación en la secuencia, como una inserción o deleción, o bien diferentes rearrreglos genómicos, ocasionarán polimorfismos que se podrán detectar en un gel de agarosa.



**Figura 19.** Técnica RAPD-PCR. Las barras rojas indican la molécula de ADN, las flechas con los números indican los oligonucleótidos y su orientación, una vez que se lleva a cabo la reacción de PCR se amplifican dos productos: A y B. En la parte de abajo se muestra el gel de agarosa donde se observan los productos de PCR.



Esta técnica ya ha sido empleada para analizar la diversidad de rizobias que nodulan la soya (*Glycine max*) (Saldaña *et al.*, 2003), con la cual se estableció una alta diversidad de organismos en dicha leguminosa. Asimismo, en un estudio de Yang *et al.* (2008), se analizaron 42 aislados de rizobias que nodulan el chícharo (*Pisum sativum*) de distintas regiones de China, con lo cual demostraron el efecto geográfico sobre la diversidad genética de las rizobias.

Por lo tanto se decidió el presente análisis probando esta técnica para determinar la diversidad de bacterias presentes en la rizósfera con el siguiente objetivo: **Implementar un método para la determinación de la diversidad genética de bacterias que habitan la rizósfera de plantas leguminosas.**

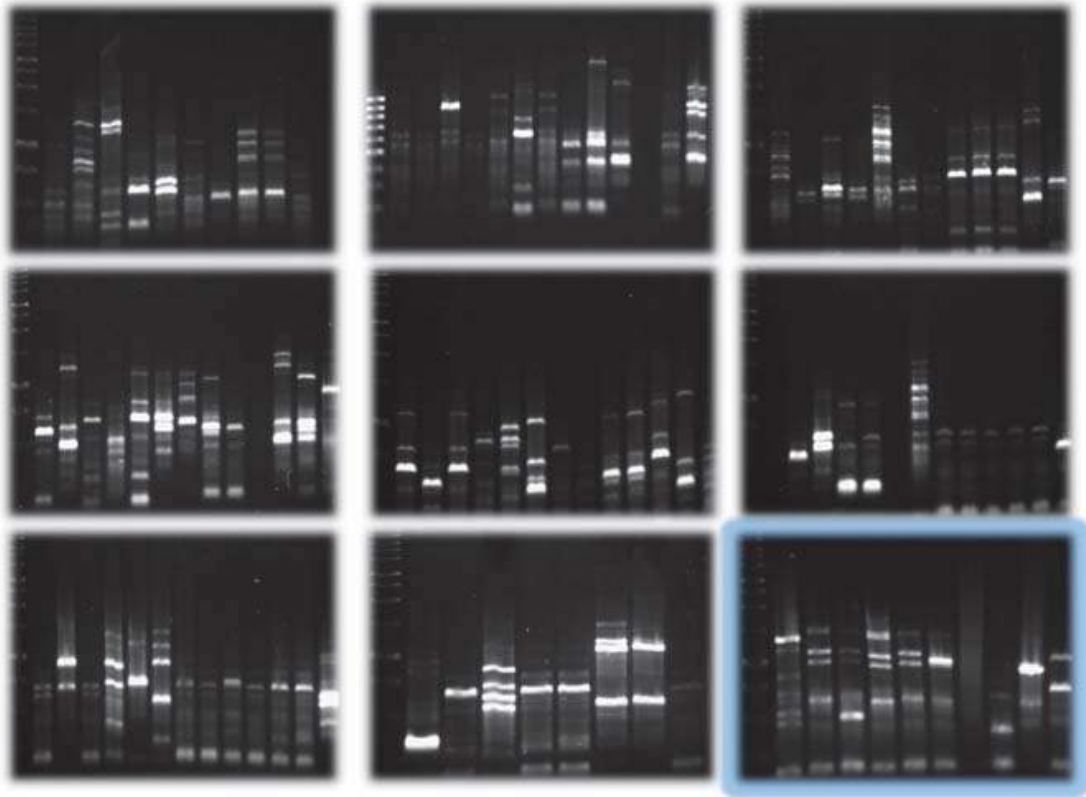
A continuación se describen los materiales y métodos usados:

Se colectaron 20 plantas de *Medicago truncatula*, de las cuales se tomó una muestra de 1 gramo de cada una y se resuspendió en buffer de sales y fosfatos (PBS). Posteriormente se realizaron diluciones hasta  $10^{-8}$  y se plaquearon en medio Py y agar nutritivo y se incubaron 24 horas a 31° C. A partir de entonces, se obtuvieron aproximadamente 1000 aislados, de los cuales se seleccionaron 101 que presentaron una morfología variada. Una vez obtenidas las colonias, se extrajo el ADN total, basado en la técnica establecida por Mahuku, (2004). Se utilizaron 13 cepas de referencia, de las cuales se extrajo el ADN siguiendo el mismo procedimiento. Las cepas utilizadas fueron amablemente proporcionadas por el laboratorio de Ecología Microbiana a cargo del D.C. Eduardo Valencia Cantero, las cuales fueron: cuatro cepas de *Sinorhizobium meliloti* (1021 nat, 1021 seq, GR4, 2011), *Rhizobium etli* C3, *Rhizobium* sp. NGR234, *Pseudomonas aeruginosa* PA01, *Stenotrophomonas maltophilia*, *Bacillus megaterium*, *Arthrobacter luteus*, *A. viscosus*, *A. nicotiana* y *A. agilis*.

Para la técnica de RAPD-PCR se emplearon tres juegos de oligos, siendo el oligo Opa02 (GTTTCCGCCC), el que dió mejores resultados. Las condiciones de PCR empleadas fueron las siguientes: 35 ciclos, 94° 1 min, 38° 1 min, 72° 1 min. Una vez obtenidos los productos de PCR de los 101 aislados y las 13 cepas de referencia se construyó la matriz de datos de presencia/ausencia de banda de acuerdo al tamaño molecular, con la cual se realizó un dendrograma apoyado por el coeficiente de similitud propuesto por Nei y Li, (1979). Para la construcción del dendrograma se utilizó el programa MVSP (Multivariate Statistical Package) (Kovach, 1999).

A continuación se presentan los resultados obtenidos:

El análisis RAPD-PCR mostró una gran heterogeneidad entre los aislados, así como entre las cepas de referencia, lo cual indica una gran diversidad de organismos tipificados (Fig. 20).



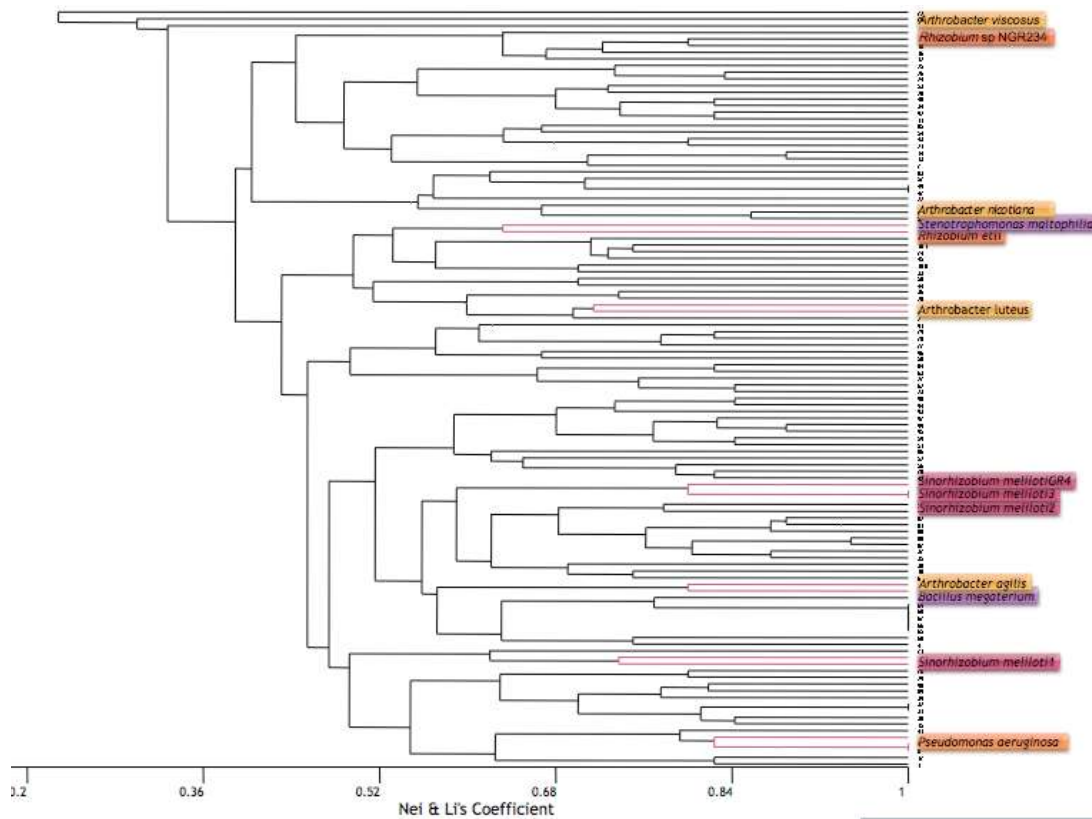
**Figura 20.** Polimorfismos por RAPD-PCR en los aislados de la rizósfera. Se muestra el patrón de bandas de los aislados rizosféricos; en cuadro azul se presentan las bandas obtenidas de las cepas de referencia.

El dendrograma indica de manera general cinco grupos de organismo; sin embargo, no se observó una relación clara con las cepas de referencia encontradas en cada grupo. En el caso de *P. aeruginosa*, se encontraron dos aislados con un 80% de similitud, lo mismo en el caso de una cepa de *S. meliloti* y *A. luteus*, con un aislado cada uno (Fig. 21).

De manera general, se puede observar que el método cuenta con una alta sensibilidad para detectar polimorfismos. Cabe mencionar que en este tipo de análisis no se buscaron relaciones filogenéticas, sino solamente el grado de similitud entre los aislados. Por lo tanto se puede establecer que se encontró

una alta diversidad de microorganismos en la rizósfera de *M. truncatula*, por lo que es necesario introducir una mayor cantidad de organismos de referencia para poder identificar la mayoría de los organismos aislados.

Por lo tanto, se concluye que se implementó un método efectivo basado en RAPDs para determinar la diversidad genética en la rizósfera de plantas de *Medicago truncatula*.



**Figura 21.** Dendrograma de la diversidad genética de la rizósfera de *M. truncatula*. En colores se observan las 13 cepas de referencia, el resto pertenecen a los 101 aislados de la rizósfera. Las ramas del árbol en color rojo indican una alta similitud entre una cepa de referencia y los aislados rizosféricos. Se utilizó el programa MVSP para construir el dendrograma, empleando el coeficiente de similitud de Nei & Li.

## Análisis *in silico* de familias multigénicas en *Rhizobium etli* CFN42 y comparación con otras rizobias

Julie E. Hernández-Salmerón, Gustavo Santoyo

Laboratorio de recombinación y diversidad genómica. Instituto de Investigaciones Químico Biológicas, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Mich. Ciudad Universitaria, Edif. A. C.P. 58030

### Resumen

En el presente trabajo se analizaron *in silico* el número de copias y la función de las familias multigénicas de *Rhizobium etli* CFN42. Además, se compararon con aquellas descritas en *Rhizobium leguminosarum*, *Rhizobium sp.* NGR234, *Bradyrhizobium japonicum*, *Sinorhizobium meliloti* y *Mesorhizobium loti*. Los resultados sugieren que *R. etli* presenta la mayor cantidad de genes reiterados y que no hay una correlación directa entre el número de repeticiones y el tamaño del genoma. Así mismo, se determinó que el mayor porcentaje de las repeticiones encontradas en *R. etli* se localizaron en plásmidos, cuyas funciones están fuertemente asociadas a la simbiosis con plantas.

**Palabras clave:** secuencias reiteradas, rizobias, conversión génica.

### Abstract

In the present study, the number of copies and the function of the multigenic families of *Rhizobium etli* CFN42, were *in silico* analyzed. In addition, they were compared with those described in *Rhizobium leguminosarum*, *Rhizobium sp.* NGR234, *Bradyrhizobium japonicum*, *Sinorhizobium meliloti* and *Mesorhizobium loti*. The results suggest that *R. etli* displays the greater amount of repeated genes and that there is not a direct correlation between the number of repetitions with the size of the genome. Also, we determined that a great percentage of this repetitions found in *R. etli* is located in plasmids, whose functions are strongly associated to the symbiosis with plants.

**Key words:** reiterated sequences, rhizobia, gene conversion.

### Introducción

La recombinación homóloga se define como el proceso por el cual dos secuencias de ADN idénticas o altamente similares intercambian información genética. Este es un mecanismo de intercambio genético que juega importantes papeles durante la reparación del ADN, la segregación de cromosomas, la generación de variación genética, por mencionar algunos (West, 2003; Marians, 2004). Así, los eventos de recombinación homóloga requieren la presencia de secuencias repetidas.

Se ha sugerido que dicha característica es común en diversos genomas bacterianos (Romero *et al.*, 1999) y que tales reiteraciones pueden estar localizadas en cromosomas o plásmidos, o en ambos, lo que podría generar diversos rearrreglos genómicos. Un resultado de un evento de recombinación entre secuencias repetidas del tipo entrecruzamiento o crossover, generaría amplificaciones o deleciones (Orozco-Mosqueda, *et al.* 2009). En una deleción, se podrían perder genes con funciones importantes que pondrían en riesgo la viabilidad celular, mientras que en las amplificaciones se incrementaría el número de copias de un segmento de ADN. Por otra parte, un segundo resultado sería un evento de conversión génica, la cual se define como la donación unidireccional de información genética entre las dos moléculas que interactúan (Szostak *et al.*, 1983). Cabe destacar que la conversión génica está asociada a la evolución concertadas de familias multigénicas, así como la variación antigénica en bacterias patógenas que ayudan a evadir el sistema inmune del huésped (Santoyo y Romero, 2005).

Las reiteraciones multigénicas y los procesos de recombinación homóloga han sido ampliamente estudiados en sistemas eucariotes y procariotes, tales como: *Saccharomyces cerevisiae* (Kuzminov, 2001), *Escherichia coli* (Takahashi y Kobayashi, 1990), *Salmonella typhimurium* (Alani *et al.*, 1994) y *Bacillus subtilis* (Belle *et al.*, 2004). Sin embargo, otros grupos bacterianos como las bacterias fijadoras de nitrógeno (conocidas como rizobias) han recibido poca atención. Una probable razón es por su importancia agronómica, aunque sin duda representan un sistema interesante dentro del área de la recombinación homóloga por la gran cantidad de secuencias repetidas que presentan sus genomas (Romero *et al.*, 1999; González *et al.*, 2006).

En este trabajo se analizó *in silico* la presencia, número de copias y función de familias multigénicas del genoma de *R. etli*. Además, hacemos un análisis comparativo con las reiteraciones encontradas en otros genomas de rizobias representativas (*R. leguminosarum*, *Rhizobium sp.* NGR234, *Bradyrhizobium japonicum*, *Sinorhizobium meliloti* y *Mesorhizobium loti*). Es de resaltar que se analizaron únicamente familias multigénicas por su importancia biológica; aunque existen otro tipo de reiteraciones, tales como transposones, secuencias de inserción o aquellas no codificantes.

### Métodos

Se realizó la búsqueda de genes repetidos reportados en la literatura, empleando la base de datos PUBMED (<http://www.ncbi.nlm.nih.gov/pubmed>) del National Center for Biotechnology

Autor de correspondencia: Dr. Gustavo Santoyo. Laboratorio Biológicas, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Mich. Ciudad Universitaria, Edif. A. C.P. 58030. Email: gsantoyo@yahoo.com, gustavo\_santoyo@yahoo.com

## Phylogenetic analysis reveals gene conversions in multigene families of rhizobia

J.E. Hernandez-Salmeron and G. Santoyo

Laboratorio de Recombinacion y Diversidad Genomica,  
Instituto de Investigaciones Quimico Biologicas,  
Universidad Michoacana de San Nicolas de Hidalgo, Morelia,  
Michoacan, Mexico

Corresponding author: G. Santoyo  
E-mail: gsantoyo@umich.mx / gustavo\_santoyo@yahoo.com

Genet. Mol. Res. 10 (3): 1383-1392 (2011)  
Received October 21, 2010  
Accepted January 16, 2011  
Published July 12, 2011  
DOI 10.4238/vol10-3gmr1118

**ABSTRACT.** Gene families are an important and intrinsic trait of rhizobial species. These gene copies can participate in non-reciprocal recombination events, also called gene conversions. Gene conversion has diverse roles, but it is usually implicated in the evolution of multigene families. Here, we searched for gene conversions in multigene families of six representative rhizobial genomes. We identified 11 gene families with different numbers of copies, genome location and function in CFN42 and CIAT652 strains of *Rhizobium etli*, *Rhizobium* sp NGR234, *Mesorhizobium loti* MAFF303099, *Sinorhizobium meliloti* 1021, and *Bradyrhizobium japonicum* USDA110. Gene conversions were detected by phylogenetic inference in the *nifD* and *nifK* gene families in *R. etli*. Sequence analysis confirmed multiple gene conversions in these two gene families. We suggest that gene conversion events have an important role in homogenizing multigene families in rhizobia.

**Key words:** Rhizobia; Gene conversion; Concerted evolution; Multigene families

## 12. LITERATURA CITADA

---

- Achaz, G., Rocha E.P.C., Netter P. and Coissac E. 2002. Origin and fate of repeats in bacteria. *Nucleic Acids Research* 30:2987-2994.
- Alani, E., Reenan R.A. and Kolodner R.D. 1994. Interaction between mismatch repair and genetic recombination in *Saccharomyces cerevisiae*. *Genetics* 137:19-39.
- Aras, R.A., Kang J., Tschumi A.I., Harasaki Y. and Blaser M.J. 2003. Extensive repetitive DNA facilitates prokaryotic genome plasticity. *Proceedings of the National Academy of Sciences USA* 100:13579-13584.
- Awramik, S.M., Schopf J.W. and Walter M.R. 1983. Filamentous fossil bacteria 3.5 x 10<sup>9</sup> years old from the Archaean Western Australia. *Precambrian Research* 20:357-374.
- Aziz, R.K., Breitbart M. and Edwards R.A. 2010. Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Research* 1-11.
- Barea, J.M. y Azcón-Aguilar C. 1982. La Rizosfera: Interacciones microbio-planta. *Anales de Edafología y Agrobiología* XII:517-1532.
- Belle, A., Landthaler M. and Shub, D.A. 2002. Intronless homing: site-specific endonuclease Seg F of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. *Genes & Development* 16:351-362.
- Castillo, M., Flores M., Mavingui P. *et al.* 1999. Increase in Alfalfa Nodulation, Nitrogen Fixation, and Plant Growth by Specific DNA Amplification in *Sinorhizobium meliloti*. *Applied and Environmental Microbiology* 65:2716-2722.
- Chain, P.S., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O., Regala W.M., Georgescu A.M., Vergez L.M., Motin V.L. *et al.* 2004. Insights into the evolution of *Yersinia pestis* through whole genome comparison with *Yersinia pseudotuberculosis*. *Proceedings of the National Academy of Sciences USA* 101:13826-13831.
- David, M., Domergue O., Pognonec P. and Kahn D. 1987. Transcription patterns of *Rhizobium meliloti* symbiotic plasmid pSym: identification of nifA-independent fix genes. *Journal of Bacteriology* 169:2239-2244.

- Deitsch, K.W., Moxon E.R. and Wellems T.E. 1997. Shared themes of antigenic variation and virulence in bacterial, protozoal and fungal infections. *Microbiology and Molecular Biology Reviews* 61:281-293.
- Dekkers, L.C., Phoelich C.C., Van der Fits L. and Lugtenberg B.J.J. 1998. A site-specific recombinase is required for competitive root colonization by *Pseudomonas fluorescens* WCS365. *Proceedings of the National Academy of Sciences USA* 95:7051-7056.
- de Lajudie, P., Willems A., Pot B., *et al.* 1994. Polyphasic taxonomy of rhizobia: emendation of the genus *Sinorhizobium* and description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov., and *Sinorhizobium teranga* sp. nov. *International Journal of Systematic Bacteriology* 44:715-733.
- Downie, A. 1994. Signalling strategies for nodulation of legumes by rhizobia. *Trends in Microbiology* 2:318-324.
- Flores, M., Brom S., Stepkowski T., Girard M.L., Dávila G. Romero D., *et al.* 1993. Gene amplification in *Rhizobium*: identification and *in vivo* cloning of discrete amplifiable DNA regions (amplicons) from *Rhizobium leguminosarum* biovar *phaseoli*. *Proceedings of the National Academy of Sciences USA* 90:4932-4936.
- Flores, M., Mavingui P., Perret X., Broughton W.J., Romero D., Hernandez G. *et al.* 2000. Prediction, identification, and artificial selection of DNA rearrangements in *Rhizobium*: toward a natural genomic design. *Proceedings of the National Academy of Sciences USA* 97:9138-9143.
- Finan, T.M. 2002. Evolving insights: symbiosis islands and horizontal gene transfer. *Journal of Bacteriology* 184:2855-2856.
- Freiberg, C., Felllay R., Bairoch A., Broughton W.J., Rosenthal A. and Perret X. 1997. Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* 387:394-401.
- Galibert, F., Finan T.M., Long S.R., Puhler A., *et al.* 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293:668-672.
- Garrity, G.M. and Holt, J.G. 2001. The road map to the manual. pp. 119-166. In Bergy's Manual of Systematic Bacteriology. 2nd edition. Volume I. EditBoone, D.R., Castenholz, R.W. and Garrity, G.M. (eds.).
- González, V., Bustos P., Ramírez-Romero M.A., Medrano-Soto A., Salgado H., Hernández-González I., Romero D., Dávila G. *et al.* 2003. The mosaic structure



of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. *Genome Biology* 4:R36

Gonzalez, V., Santamaria R.I., Bustos P., Hernández-González I., Medrano-Soto A., Moreno-Hagelsieb G., Janga S.C., Ramírez M.A., Jiménez-Jacinto V., Collado-Vides J. and Dávila G. 2006. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. *Proceedings of the National Academy of Sciences USA* 103:3834-3839.

Guo, X., Flores M., Mavingui P., Fuentes S.I., Hernández G., Dávila G., Palacios R. 2003. Natural genomic design in *Sinorhizobium meliloti*: novel genomic architectures. *Genome Research* 13:1810-1817.

Hernandez-Salmeron, J.E. and Santoyo G. 2011. Phylogenetic analysis reveals gene conversions in multigene families of rhizobia. *Genetics and Molecular Research* 10:1383-1392.

Holliday, R. 1964. A mechanism for gene conversion. *Genetics Research* 5:282-304.

Honma, M.A. and Ausubel F.M. 1987. *Rhizobium meliloti* has three functional copies of the *nodD* symbiotic regulatory gene. *Proceedings of the National Academy of Sciences USA* 84:8558-8562.

Howell-Adams, B. and Seifert, H.S. 2000. Molecular models accounting for the gene conversion reactions mediating gonococcal pilin antigenic variation. *Molecular Microbiology* 37:1146-1158.

Ji, Y. *et al.* 2000. Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Research* 10:597-610.

Jordan, D.C. 1982. Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a genus of slow-growing, root nodule bacteria from leguminous plants. *International Journal of Systematic Bacteriology* 32:136-139.

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 in Genetics* 17:7-10.

Katti, M.V., Ranjekar P.K. and Gupta V.S. 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular Biology and Evolution* 18:1161-1167.

Kaneko, T., Nakamura Y., Sato S., Asamizu E., *et al.* 2000. Complete genome structure

- of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Research* 7:331-338.
- Kaneko, T., Nakamura Y., Sato S., Minamisawa K., *et al.* 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Research* 9:189-197.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111-120.
- Kovach, W.L. 1999. MVSP - A Multivariate Statistical Package for Windows, ver. 3.1. Kovach Computing Services, Pentraeth, Wales, UK, 133p.
- Kurtz, S., Phillippy A., Delcher A.L., Smoot M., Shumway M., Antonescu C. and Salzberg S.L. 2004. Versatile and open software for comparing large genomes. *Genome Biology* 5:R12.
- Kuzminov, A. 2001. DNA replication meets genetic exchange: chromosomal damage and its repair by homologous recombination. *Proceedings of the National Academy of Sciences USA* 98:8461-8468.
- Laguerre, G., Nour S.M., Macheret V., Sanjuan J., Drouin P. and Amarger N. 2001. Classification of rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts. *Microbiology* 147:981-993.
- Lawson, M.J., Jiao J., Fan W. and Zhang L. 2009. A pattern analysis of gene conversion literature. *Comparative and Functional Genomics* 2009:1-11.
- Lloret, L. y Martínez-Romero, E. 2005. Evolución y filogenia de Rhizobium. *Revista Latinoamericana de Microbiología* 47:43-60.
- Lozano, L., Hernández-González I., Bustos P., Satamaría R.I., Souza V., Young J.P.W., Davila G. and González V. 2010. Evolutionary dynamics of insertion sequences in relation to the evolutionary histories of the chromosome and symbiotic plasmid genes of *Rhizobium etli* populations. *Applied and Environmental Microbiology* 76:6504-6513.
- Lupski, J.R. 1998. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends in Genetics* 14:417-422.
- Mahuku, G.S. 2004. A simple extraction method suitable for PCR-based analysis of plant, fungal and bacterial DNA. *Plant Molecular Biology* 22:71-81.

- Margulis, L. 1993. Symbiosis in cell evolution. H.W. Freeman and Company. 2° ed. 118-130.
- Marians, K.J. 2004. Mechanisms of replication fork restart in *Escherichia coli*. *Philosophical Transactions of the Royal Society of London B* 29:71-77.
- Mavingui, P., Flores M., Romero D., Martinez-Romero E., Palacios R. 1997. Generation of Rhizobium strains with improved symbiotic properties by random DNA amplification (RDA). *Nature Biotechnology* 15:564-569.
- Mavingui, P., Flores M., Guo, X., Davila, G., Perret, X., Broughton, W.J., *et al.* 2002. Dynamics of genome architecture in *Rhizobium* sp. Strain NGR234. *Journal of Bacteriology* 184:171-176.
- Meselson, M.S. and Radding C.M. 1975. A general model for genetic recombination. *Proceedings of the National Academy of Sciences USA* 72:358-361.
- Moulin, L., Munive A., Dreyfus B. and Boivin-Masson C. 2001. Nodulation legumes by members of the  $\beta$ -subclass of proteobacteria. *Nature* 411:948-950.
- Moxon, R., Bayliss C. and Hood D. 2006. Bacterial contingency Loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annual Review of Genetics* 40:307-333.
- Nei, M. and Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences USA* 76:5269-5273.
- Oldroyd, G.E.D. 2007. Nodules and hormones. *Science* 315:52-53.
- Orozco-Mosqueda, M.C., Altamirano-Hernandez J., Farias-Rodriguez R., Valencia-Cantero E. and Santoyo G. 2009. Homologous recombination and dynamics of rhizobial genomes. *Research in Microbiology* 160:733-741.
- Parkhill, J., Sebahia M., Preston A., Murphy L.D., Thomson N., Harris D.E., Holden M.T., Churcher C.M., Bentley S.D., Mungall K.L. *et al.* 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nature Genetics* 35:32-40.
- Parkhill, J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T., Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L. *et al.* 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* 413:523-527.

- Petes, T.D. and Hill C.W. 1988. Recombination between repeated genes in microorganisms. *Annual Review of Microbiology* 31:473-505.
- Quinto, C., H. de la Vega M., Flores L., Fernandez T., Ballado G., Soberon *et al.* 1982. Reiteration of nitrogen fixation gene sequences in *Rhizobium phaseoli*. *Nature* 299:724-726.
- Raymond, J., Siefert J.L., Staples C.R. and Blankenship R.E. 2004. The natural history of nitrogen fixation. *Molecular Biology and Evolution* 21:541-554.
- Rivas, R., García-Fraile P., Velázquez E. 2009. Taxonomy of bacteria nodulating legumes. *Microbiology Insights* 2:51-69.
- Rocha, E.P.C. and Blanchard A. 2002. Genomic repeats, genome plasticity and the dynamics of *Mycoplasma* evolution. *Nucleic Acids Research* 30:2031-2042.
- Rocha, E.P.C., Cornet E. and Michel B. 2005. Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genetics* 1:e15.
- Rocha, E.P.C. Danchin A. and Viari A. 1999. Functional and evolutionary roles of long repeats in prokaryotes. *Research in Microbiology* 150:725-733.
- Rocha, E.P.C., Matic I. and Taddei F. 2002. Over-representation of close repeats in stress response genes: a strategy to increase versatility under stressful conditions? *Nucleic Acids Research* 30:1886-1894.
- 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. *Research in Microbiology* 150:735-743.
- Romero, D. and Palacios R. 1997. Gene amplifications and genome plasticity in prokaryotes. *Annual Review of Genetics* 31:91-111.
- Russelle, M.P. and Birr A.S. 2004. Biological nitrogen fixation. Large-scale assessment of symbiotic dinitrogen fixation by crops: soybean and alfalfa in the Mississippi river basin. *Journal of Agronomy* 96:1754.
- Sadowsky, M.J. and Graham P.H. 1998. Soil biology of the Rhizobiaceae. In: The Rhizobiaceae. H. P. Spaink, A. Kondorosi, and P. J. J. Hooykaas (eds.). Kluwer Academic Publishers, Dordrecht. Pg. 155-172.
- Sakai, K., Ohta T., Minoshima S., Kudoh J., Wang Y., de Jong P.J. and Shimizu N. 1995. Human ribosomal RNA gene cluster: identification of the proximal end containing a novel tandem repeat sequence. *Genomics* 26:521-526.

- Saldaña, G., Martínez-Alcántara V., Vinardell J.M., Bellogín R., Ruíz-Sainz J.E. and Balatti P.A. 2003. Genetic diversity of fast-growing rhizobia that nodulate soybean (*Glycine max* L. Merr). *Archives of Microbiology* 180:45-52.
- Santoyo, G. and Romero D. 2005. Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiology Reviews* 29:169-183.
- Santoyo, G., Martínez-Salazar J.M., Rodríguez C. and Romero D. 2005. Gene conversion tracts associated with crossovers in *Rhizobium etli*. *Journal of Bacteriology* 187:4116-4126.
- Segovia, L., Young J.P.W., Martínez-Romero E. 1993. Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* type I strains as *Rhizobium etli* sp. nov. *International Journal of Systematic Bacteriology* 43:374-377.
- Stankiewicz, P. and Lupski J.R. 2002. Genome architecture, rearrangements and genomic disorders. *TRENDS in Genetics* 18:74-82.
- Szostak, J., Orr-Weaver T., Rothstein R., Stahl F. 1983. The doublestrand-break repair model for recombination. *Cell* 33:25-35.
- Sullivan, J.T. and Ronson C.W. 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis isly that integrates into a phe-tARN gene. *Proceedings of the National Academy of Sciences USA* 95:5145-5149.
- Takahashi, N. and Kobayashi I. 1990. Evidence for the doublestrand break repair model of bacteriophage lambda recombination. *Proceedings of the National Academy of Sciences USA* 87:2790-2794.
- Tamura, K., Dudley J., Nei M. and Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24:1596-1599.
- Treangen, T.J., Abraham A.L., Touchon M. and Rocha E.P.C. 2009. Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiology Reviews* 33:539-571.
- Ueda, T., Suga Y., Yahiro N. and Matsuguchi T. 1995. Phylogeny of Sym plasmids of rhizobia by PCR-based sequencing of a *nodC* segment. *Journal of Bacteriology* 177:468-472.
- Warner, J.R. 1989. Synthesis of ribosomes in *Saccharomyces cerevisiae*. *Microbiological Reviews* 53:256-271.

- Werner. 1992. Symbiosis of plants and microbes. Chapman & Hall. Berlin. pp. 89.
- West, S.C. 2003. Molecular views of recombination proteins and their control. *Nature Reviews Molecular Cell Biology* 4:435-445.
- Woese, C.R., Stackebrandt E., Weisburg W.G., *et al.* 1984. The phylogeny of purple bacteria: The alpha subdivision. *Systematic and Applied Microbiology* 5:315-26.
- Yang, C., Yang J., Li Y. and Zhou J. 2008. Genetic diversity of root-nodulating bacteria isolated from pea (*Pisum sativum*) in subtropical regions of China. *Science in China Series C: Life Sciences* 51:854-862.
- Young, J.P.W. 1992. Phylogenetic classification of Nitrogen-fixing organisms. In: Biological Fixation. Stacey, G., Burris, H.R. & Evans, H.J. (eds.). Chapman and Hall, New York. pp 43-79.
- Young, J.P.W. 1996. Phylogeny and taxonomy of rhizobia. *Plant and Soil* 186:45-52.
- Young, J.P.W. and Johnston A.W.B. 1989. The evolution of specificity in the legume-*Rhizobium* symbiosis. *Trends in Ecology and Evolution* 4:341-349.
- Young, J.P.W., Crossman L.C., Johnston A.W.B., Thomson N.R., Ghazoui Z.F., *et al.* 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology* 7:R34.