

03062



**UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO**

**CENTRO DE INVESTIGACION SOBRE FIJACION DE NITROGENO**

2  
Ref

**ANALISIS DE SECUENCIAS DE ADN REITERADO  
Y PREDICCION DE REARREGLOS GENOMICOS  
EN *Rhizobium***

**T E S I S  
PARA OBTENER EL GRADO DE:  
MAESTRO EN INVESTIGACION  
BIOMEDICA BASICA  
QUE PRESENTA:  
DANIEL BALLEZA MEJIA**

258062

Cuernavaca, Mor.

1998

**TESIS CON  
FALLA DE ORIGEN**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



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

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

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

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

*Dedico esta Tesis a Mama, Alis y Humberto  
A Papá...  
A la Memoria de mi Abuelo...  
An die ferne Geliebte...*

## AGRADECIMIENTOS

Ante todo, agradezco a la Universidad Nacional Autónoma de México y al CIFN-UNAM por permitirme formar parte de su comunidad.

El presente trabajo fue dirigido por el Dr. Rafael Palacios de la Lama, a quien agradezco esta valiosa oportunidad. Asimismo, agradezco a los demás miembros del Comité Tutorial que semestre a semestre fueron testigos de lo aquí expuesto: Dr. Guillermo Dávila Ramos, Dra. Georgina Hernández Delgado y Dr. Patrick Mavingui Malonda.

Gracias al Dr. Patrick Mavingui Malonda, al Dr. Luis Eguiarte Fruns, al Dr. Lorenzo Segovia Forcella y al Dr. Alejandro Garcarrubio Granados por los valiosos comentarios al manuscrito.

Una parte muy importante de este trabajo fue realizada mediante el empleo del paquete de análisis de secuencia Molpac en la Universidad de Texas en San Antonio, EUA. Extiendo mis más sinceras gracias al grupo del Dr. Ming-Ying Leung por las facilidades para el uso de este software.

El apoyo recibido por el CONACyT fue de gran ayuda... Gracias.

Y cómo olvidar a los compañeros que hacen que todo esto sea más divertido: Charlie "Bad-Boy" Hernández, Michael "Free-Spirit" Ramírez, Vero "Queen" Quintero, J. A. "Radiactive" Gama, Hilda "Little-Friend" Vázquez, Javier "Blue-Demon" Izquierdo, George "Jungle-Boy" Pineda y la Marcelita.

*Soy feliz todas las veces  
que venzo alguna dificultad...*

Beethoven

*An die Hoffnung...*

---

## Índice de Contenido

---

<b>Resumen</b>	1
<b>Consideraciones Previas</b>	2
<b>INTRODUCCIÓN</b>	3
El Genoma simbiótico de <i>Rhizobium meliloti</i> y de <i>Rhizobium</i> sp. NGR234	5
Dinámica del Genoma en <i>Rhizobium</i>	6
Un Modelo Dinámico del Genoma de <i>Rhizobium</i>	8
Significado Biológico de la Amplificación Génica	10
El Análisis de Genomas Completos y su Importancia	10
<b>OBJETIVOS</b>	12
<b>MÉTODOS</b>	13
Obtención de Secuencias Depositadas en el GenBank	13
Programas para Análisis de Secuencias Reiteradas	13
Búsqueda Primaria de Secuencias Directas Repetidas	13
Repeticiones Grandes en Secuencias Largas	13
Extensión de Motivos Centrales de Secuencia	14
Generación de Secuencias Azarosas	15
Identificación de Patrones de Secuencia	15
Búsqueda de Motivos Protéicos	16
Predicción de Estructuras Secundarias a Nivel Protéico	16
Secuenciación Automática de ADN	16

<b>RESULTADOS</b>	1 7
Integración del Archivo de Secuencia de la Región de Nodulación de <i>R. meliloti</i> : Rmel_Regulon_47.1_kb.txt	1 7
Búsqueda Primaria de Secuencias Reiteradas en la Región de Nodulación de <i>R. meliloti</i>	2 0
Extensión de los Motivos Centrales de Secuencia	2 5
Definición de Familias de Elementos Reiterados (FERs)	2 5
<b><i>R. meliloti</i>: Familias de Elementos Reiterados (FERs)</b>	2 5
FER A: Genes <i>nodD</i> y <i>syrM</i>	2 5
FER B: <i>nod</i> box	2 7
FER C: Regiones Promotoras Reps I: P1-P2-P3	3 0
FER D: Regiones Promotoras Reps II	3 1
FER E: Regiones Promotoras Reps III: Región -5' <i>fixA</i>	3 4
FER F: Regiones Intergénicas Reps IV: <i>nifH-fixA</i> y <i>nodD2-nolQa</i>	3 4
FER G: Regiones Intragénicas Reps V: <i>fixB</i> y <i>nifA</i>	3 7
Búsqueda Primaria de Reiteraciones Grandes en el pNGR234a de <i>Rhizobium</i> sp. NGR234	3 8
Extensión de las Reiteraciones Grandes Encontradas en el pNGR234a de <i>Rhizobium</i> sp. NGR234a	3 8
<b><i>Rhizobium</i> sp. NGR234: Familias de Elementos Reiterados (FERs)</b>	4 1
FER I: Operones <i>nifHDK1</i> y <i>nifHDK2</i>	4 1
FER II: Secuencias de inserción NGRIS4a y NGRIS4b	4 1
FER III: Secuencias de inserción NGRIS3b y NGRIS3c	4 1
FER IV: Secuencias de inserción NGRIS5a, NGRIS5b y NGRIS5c	4 2
FER V: Secuencias de inserción NGRIS2a y NGRIS2b	4 2
FER VI: Genes <i>y4sH</i> y <i>y4uA</i>	4 2
Amplicones Propuestos en la Región de Nodulación de <i>R. meliloti</i> y en el pNGR234a de <i>Rhizobium</i> sp. NGR234	4 2
Regiones Repetidas en Orientación Inversa. Búsqueda de Secuencias Invertidas	4 3

<b>DISCUSIÓN Y CONCLUSIONES</b>		4 6
Familias Multigénicas y Bancos de Información de Secuencia		4 6
Fenómenos Recombinatorios en <i>Rhizobium</i>		4 7
Reiteraciones en la Región de Nodulación de <i>R. meliloti</i>		5 0
Reiteraciones en el pNGR234a de <i>Rhizobium</i> sp. NGR234		5 4
Estrategias Experimentales para obtener Evidencia Funcional de Amplicones		5 5
Perspectivas a Futuro		5 5
<b>Apéndices</b>		
A-1 Número Total de Repeticiones encontradas en el pNGR234a		5 8
A-2 PCR cuantitativo: Selección de Amplificaciones y Delecciones Genómicas		5 9
A-3 PCR cuantitativo: Selección de Inversiones Genómicas		6 0
<b>LITERATURA CITADA</b>		6 1

## Resumen

El mapeo físico y los proyectos de secuencia de megaplásmidos en bacterias del género *Rhizobium* han permitido plantear nuevos enfoques para la comprensión de la organización y la función de los genomas bacterianos. Con estos antecedentes se planteó una búsqueda de aquellas secuencias de ADN reiterado que, por sus características pudieran ser relevantes en términos de recombinación homóloga RecA-dependiente para generar rearrreglos genómicos. Como punto de partida se buscaron regiones repetidas en dos especies diferentes: (i) *R. meliloti* y (ii) *Rhizobium* sp. NGR234. Mediante el uso de métodos computacionales se determinaron reiteraciones directas e invertidas en dos archivos generados a partir de los datos de secuencia de estas especies. Para *R. meliloti* se integró un archivo de secuencia de 471 kb que comprende la región de nodulación (Regulón-*nod*); por otra parte se empleó la secuencia total (536 165 pb) del pNGR234a de *Rhizobium* sp. NGR234 (Freiberg *et al.*, 1997; Acc. U00090). De esta manera se describen Familias de Elementos Reiterados (FERs) y en algunas de estas familias se proponen unidades genéticas capaces de amplificar y deletar fragmentos específicos de ADN, conocidas como **amplicones**. Como principal conclusión, de los resultados de esta búsqueda se describen 6 posibles amplicones dentro de la región de nodulación de *R. meliloti*, de los cuales el Amplicón I (sistema *nodD1-nodD2*) es el más interesante de todos. Los cinco amplicones restantes para este caso son funcionalmente menos probables, debido al nivel de similaridad y al tamaño de la región reiterada entre las secuencias repetidas encontradas en cada uno de ellos. De la misma forma se proponen al menos 8 amplicones para el pNGR234a de *Rhizobium* sp. NGR234; en este caso, cada uno de los 8 amplicones es de gran interés dados los valores de identidad encontrados en los elementos reiterados que los constituyen. El estudio de los amplicones propuestos en cada caso será de gran interés para demostrar un modelo dinámico del genoma de *Rhizobium* y definir su significado biológico. Este modelo pudiera ser aplicable al genoma bacteriano en general.

## Consideraciones Previas

Originalmente, el presente trabajo incluía solamente el análisis de la región de nodulación de *R. meliloti*, sin embargo, durante el transcurso de este análisis se reportó la secuencia total del plásmido simbiótico de *Rhizobium* sp. NGR234. Dada la disponibilidad de la secuencia plasmídica de esta especie decidimos incluirla en este trabajo, no sin antes hacer las algunas consideraciones.

Dadas las dimensiones de cada secuencia (47.1 kb y 536.1 kb respectivamente), la profundidad del análisis en ambos casos, fue diferente. Para *R. meliloti*, los elementos de cada familia descrita tienen bajos niveles de similitud, mientras que, en *Rhizobium* sp. NGR234 se describen familias con niveles de similitud más elevados en cada uno de sus elementos. Como resultado del análisis de ambas secuencias se describen 17 FERs para la región de nodulación de *R. meliloti*. En una de estas familias (familia B) se describen 2 copias tipo *nod* box adicionales a las 6 copias que se conocían para esta región. Por otra parte, en el pNGR234a se describen 6 FERs. Estos dos grupos de familias de ADN reiterado son de diferente naturaleza. Los elementos repetidos de las familias descritas para la región de nodulación de *R. meliloti* son en general de baja similitud entre sí (> 50 %) y pequeños (hasta 1.1 kb); mientras que para el caso de los elementos reiterados descritos en el pNGR234a, los niveles de identidad entre las secuencias reiteradas son muy altos (> 98.5 %) y son más grandes (hasta 4.3 kb).

## INTRODUCCIÓN

Las especies bacterianas que han sido clasificadas dentro de los géneros *Rhizobium*, *Azorhizobium*, *Bradyrhizobium* y *Sinorhizobium* (Rhizobiaceae) son capaces de establecer relaciones especie-específicas con ciertos géneros de plantas leguminosas y formar estructuras bien diferenciadas dentro del cuerpo vegetal, donde se lleva a cabo un complejo proceso bioquímico de reducción y asimilación del nitrógeno atmosférico. Este proceso de fijación simbiótica de nitrógeno tiene lugar en los nódulos radiculares de estas plantas (Fisher y Long, 1992). La formación de los nódulos es el resultado de procesos de diferenciación celular y es controlada por intercambio de señales moleculares entre los dos organismos: planta-bacteria. Este intercambio de señales químicas está controlado a su vez por la expresión de ciertos genes específicos en cada organismo. El proceso simbiótico se puede ver como una interacción eventual entre dos organismos que han adaptado sus genomas para la sobrevivencia y el crecimiento independientes de cada uno de ellos. Se sabe que la planta libera a la rizósfera compuestos de naturaleza flavonoide y que la bacteria - en respuesta- produce toda una gama de lipoquitooligosacáridos, llamados factores de nodulación (factores Nod). Los factores Nod son las principales moléculas responsables de provocar en la planta las respuestas necesarias que conducen al desarrollo del nódulo (Fisher y Long, 1992). En *Rhizobium*, los genes que codifican para las enzimas que participan en la biosíntesis de los factores Nod han sido llamados genes *nod*, *nol* y *noe*. Una molécula tipo factor Nod, típicamente tiene una cadena oligomérica de quitina que está substituída por diversos radicales en cada extremo de ella (Figura 1). Estas substituciones específicas en la estructura de los factores Nod son codificadas por ciertos genes de nodulación, conocidos anteriormente como genes *hsn* (host specific genes) (Schultze *et al.*, 1994).

El estudio que se presenta en este trabajo es un análisis de la secuencia nucleotídica de regiones simbióticas en dos especies dentro del género *Rhizobium*. De las especies de bacterias rhizobiáceas descritas hasta la fecha, el estudio de *R. meliloti* y *Rhizobium* sp. NGR234 es de gran interés debido a que la primera de ellas es la especie mejor conocida hasta la fecha y *Rhizobium* sp. NGR234 es la primera rhizobiácea, cuyo plásmido simbiótico (pNGR234a) ha sido recientemente secuenciado por completo. Asimismo, el espectro de hospederos de ésta especie es muy amplio (Pueppke y Broughton, datos no publicados). Aquí se presentan 17 familias de elementos reiterados (FER) para la región de nodulación de *R. meliloti* y 6 FERs para *Rhizobium* sp. NGR234; de la misma forma se describen amplicones potencialmente funcionales para cada secuencia. En *Rhizobium* sp. NGR234, las posiciones de cada elemento reiterado de las FERs I, II, III, IV, V y VI se esquematizan más adelante



**El Genoma simbiótico de *Rhizobium meliloti* y de *Rhizobium* sp. NGR234.** En las bacterias del género *Rhizobium*, los genes de nodulación se han localizado principalmente en grandes plásmidos, llamados plásmidos simbióticos o pSym (Martínez *et al.*, 1990). En *R. meliloti* estos genes se encuentran en el megaplásmido pSym-a (1 340 kb), el resto del genoma se encuentra formando parte de un segundo megaplásmido, el pSym-b (1 700 kb) y otra parte en el cromosoma (3 540 kb). En esta misma especie se han identificado hasta 18 genes *nod* flavonoide-inducibles, que están organizados en cinco unidades transcripcionales localizadas hacia el extremo -3' de las secuencias regulatorias *nod box* (Schultze *et al.*, 1994). Las secuencias tipo *nod box* son de 47 pb y contienen 3 motivos conservados no sobrelapados de secuencia que miden 7, 5 y 25 pb respectivamente. En *R. meliloti* se han descrito 6 copias de *nod box* que se encuentran en una zona de 100 kb dentro de la región *nod/nif* (*nif*: genes para la fijación de nitrógeno), del plásmido simbiótico pSym-a (Rostas *et al.*, 1986). En *R. meliloti*, al menos las primeras 5 copias de este consenso están asociadas con unidades de transcripción inducibles. Este modelo de regulación coordinada de los genes de nodulación es conocido como **Regulón *nod***. Las unidades de transcripción que integran el Regulón *nod* comprenden los genes *nodABCIIJ*, *nodFEG*, *nodH* y *nodLnoeAB* que son controlados por las *nod box* **n1**, **n2**, **n3** y **n5** respectivamente, además del operón *nodMN/nolFGHI* que está precedido por la *nod box* **n4** (Baev *et al.* 1991). Recientemente se han descubierto tres nuevos genes. *nolQaQbS*. precedidos por la **n6** pero al parecer no están controlados por ésta (Plazanet *et al.*, 1995). El conjunto de estas unidades de transcripción se encuentran en una zona de unas 100 kb aproximadamente, dentro del megaplásmido pSym-a. Durante el proceso simbiótico, los genes *nod* de *R. meliloti* determinan la infección, nodulación y especificidad al hospedero vegetal. Dentro de los genes *nod* se puede hacer la distinción de genes comunes (*nod* comunes: *nodABC-IJ*) y los hospedero-específicos (antiguamente conocidos como *hsn*; ahora: *nodFEGHLPQ*).

La expresión de los genes *nod* está controlada por la proteína reguladora de la transcripción, NodD (Schlaman *et al.*, 1992), que es sintetizada de manera constante por la bacteria. En respuesta a las señales moleculares de las plantas, las proteínas NodD adquieren una conformación distinta y se unen de manera específica a las secuencias consenso *nod box* (Fisher y Long, 1993) en las regiones reguladoras de los operones *nod*. En *R. meliloti* se han descrito tres copias de genes *nodD*, cuyos productos protéicos (NodD) tienen diferentes afinidades por compuestos inductores vegetales (Dénarié *et al.*, 1992). De esta manera, los diferentes genes y operones *nod* están regulados de manera coordinada y compleja en el sistema Regulón *nod* (Kondorosi, 1992).

Por otra parte, *Rhizobium* sp. NGR234 es de gran interés debido a que puede establecer simbiosis con más de 110 géneros de plantas leguminosas e incluso con la no-leguminosa *Parasponia andersonii* (Ulmaceae) (**Pueppke y Broughton**, datos no publicados). Esta especie bacteriana posee un plásmido simbiótico llamado pNGR234a, donde se encuentran los genes determinantes de la simbiosis. En este plásmido se encuentran codificados -entre otros- genes *nod* y genes *nif*fix. El pNGR234a mide 536 165 pb su secuencia completa ha sido establecida (**Freiberg et al., 1997**).

El plásmido simbiótico pNGR234a tiene una gran cantidad de secuencias de inserción y reductos de secuencias presumiblemente de otros orígenes (entre otros: *Agrobacterium*, *Bacillus escherichia*, *Pseudomonas* y *Rhizobium*), los elementos mosaico (en conjunto 18%) así como un loci de transferencia conjugativa similar a la descrita en *Agrobacterium* (**Freiberg et al., 1997**). Con base en el análisis de la secuencia de este plásmido se predice la existencia de 416 marcos de lectura abierta (*orfs*); la función del 56% de estas probables proteínas ya ha sido descrita en alguna otra especie. El análisis de la estructura y organización génica del plásmido sugiere que en algún momento durante la evolución de esta especie hubo transferencia de información genética a partir de diferentes bacterias del suelo (**Freiberg et al., 1997**). Gracias a la disponibilidad de los datos de secuencia de esta especie es posible ahora pensar en estudios genómicos comparativos. Este tipo de análisis seguramente permitirá establecer bases sólidas acerca del origen y evolución de los organismos fijadores de nitrógeno.

**Dinámica del Genoma en *Rhizobium*.** Se ha demostrado que una característica del genoma de *Rhizobium* es la presencia de secuencias de ADN reiterado (para una revisión ver: **Martínez et al., 1990**). En *R. etli* se han descrito tres regiones similares a los genes estructurales de la nitrogenasa (**Quinto et al., 1982**). Dos de estas regiones contienen copias funcionales del operón estructural *nifHDK* (5 kb) en orientación directa (**Quinto et al., 1985**). De esta manera, los genes estructurales de la nitrogenasa en *R. etli* constituyen uno de los pocos ejemplos de familias multigénicas funcionales en bacterias. Recientemente, también se ha reportado en *Rhizobium* sp. NGR234 una reiteración del operón *nifHDK* (**Freiberg et al., 1997**).

En *R. etli* se ha inferido la presencia de unas 200 familias de ADN reiterado y un análisis de la localización de algunas de ellas indica que ciertas familias forman parte del cromosoma o de plásmidos específicos; mientras que otras pueden encontrarse en más de un plásmido o incluso en el cromosoma (**Flores et al., 1987**). El nivel de reiteración genómica que se describe para *R. etli* es poco usual para bacterias; sólo en *Halobacterium* (**Sapienza y**

Doolittle, 1982) y *Streptomyces* (Schrempf, 1985) se tienen datos de elevados niveles de reiteración.

En *R. meliloti* se ha encontrado que las regiones promotoras del operón *nifHDK* (Better *et al.*, 1983) y los genes *fixKN* (Renalier *et al.*, 1987) también están reiterados. Por otra parte, además de las reiteraciones en *R. meliloti*, el gen *nodD* se encuentra de igual manera repetido en varias especies de *Rhizobium* (Schlaman *et al.*, 1992). Otro tipo de secuencias repetidas en el genoma de *Rhizobium* son los elementos transponibles, como es el caso de las secuencias de inserción (IS) (Ruvkun *et al.*, 1982; Dusha, *et al.*, 1987; Ogawa *et al.*, 1992). El número de copias de los elementos IS varía según el tipo de secuencia de inserción y la cepa analizada; sin embargo pueden estar presentes desde 2 hasta 12 copias por genoma. Esta clase de secuencias tiene gran importancia en la generación de la diversidad genética.

Un factor que parece de particular interés es la ubicación relativa de las secuencias de ADN repetido. En el pSym de *R. etli*, además de la familia de los operones *nifHDK* se han descubierto otros elementos reiterados, distribuidos en diferentes familias. Estos elementos se encuentran agrupados en el pSym, cerca de las reiteraciones *nifHDK* (Dávila *et al.*, 1988). El mismo fenómeno de distribución diferencial para secuencias reiteradas se ha observado en *Bradyrhizobium japonicum* (Hahn y Hennecke, 1987) y *R. fredii* (Masterson y Atherly, 1986). En el caso de ciertas *Archaeobacteria*, las cuales también poseen grandes cantidades de ADN repetido, estas secuencias parecen estar limitadas hacia ciertas zonas o "islas" de elevado contenido en [A + T] (Pfeifer y Betlach, 1985).

Las secuencias de ADN reiterado pueden participar como sitios de recombinación homóloga, generando rearrreglos genómicos. Esta clase de rearrreglos han sido observados en diferentes especies de *Rhizobium* a frecuencias diversas (dentro del intervalo entre  $10^{-2}$  a  $10^{-5}$ ). Los principales rearrreglos genómicos que pueden surgir como resultado de este tipo de fenómenos son la cointegración, delección y amplificación de secuencias específicas. De esta manera se ha propuesto el término **amplición** para referirse a regiones específicas de ADN que tienen la potencialidad de ser amplificadas o deletadas como unidad (Romero *et al.*, 1995). Por otra parte esta clase de fenómenos son la base para establecer el concepto de un genoma dinámico, que puede ser de gran importancia para el entendimiento de la estructuración, organización y evolución del genoma bacteriano. Ciertas evidencias indican que esta dinámica puede manifestarse a diferentes niveles del genoma. Sin embargo, los mecanismos moleculares que puedan intervenir en ello no son del todo comprendidos todavía.

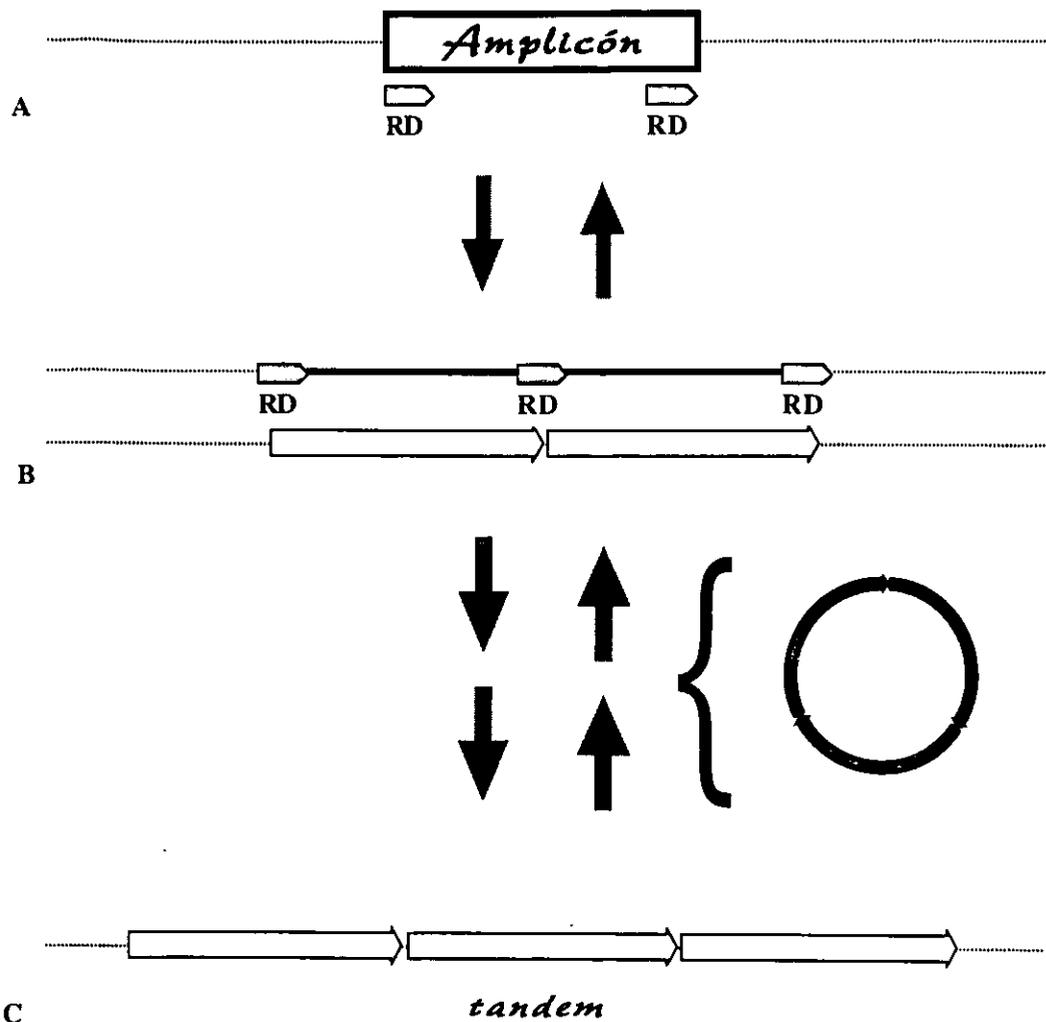
**Un Modelo Dinámico del Genoma de *Rhizobium*.** Las unidades básicas de este modelo (Palacios *et al.*, 1995) son los amplicones. Cada amplicón se encuentra limitado por secuencias de ADN repetido que se orientan de manera directa entre sí. Los fenómenos de recombinación homóloga entre estas secuencias repetidas tienen como resultado la duplicación o delección del amplicón completo (Anderson y Roth, 1977). Las frecuencias de duplicación varían entre diferentes amplicones y sus valores dependerán principalmente de la longitud y el grado de similitud de las regiones repetidas que limitan cada amplicón.

En este modelo, el paso limitante para la amplificación es el evento de la duplicación inicial. Una vez que se ha conseguido ésta, la estructura resultante es una secuencia repetida en *tandem*, donde cada secuencia repetida está constituida por el amplicón completo original (Figura 2). Esta estructura puede recombinar hacia la amplificación de la zona o bien la delección de una de sus unidades. Debido a que la organización de este tipo de estructuras está en *tandem*, la estructura es altamente reactiva en términos recombinatorios. De esta manera, tales estructuras incrementan o disminuyen el nivel de amplificación y eventualmente pueden regresar al estado no amplificado inicial. Teóricamente, durante estos procesos dinámicos, la estructura genómica no se desorganiza y la secuencia de ADN en el amplicón se conserva. Es decir, el orden de los genes incluidos en un amplicón permanece sin alteración, sin embargo es obvio que en un rearrreglo tipo delección estas secuencias se pierden.

Como consecuencia de una actividad recombinatoria alta, las regiones amplificadas liberan continuamente estructuras circulares que contienen monómeros o multímeros de la secuencia del amplicón original. Generalmente estas estructuras circulares no contienen origen de duplicación, por lo que su permanencia dentro de una población bacteriana es cada vez menor. Sin embargo, estas unidades circulares pueden participar también en fenómenos recombinatorios posteriores y generar nuevos rearrreglos genómicos.

Las estructuras tipo amplicón se han descrito desde hace algún tiempo en *Streptomyces* y *Bacillus*. En estas bacterias se han observado fenómenos de amplificación y delección en condiciones naturales, sin presiones de selección (para una revisión ver: Young y Hranueli, 1988). También en *Haemophilus influenzae* se han reportado ya fenómenos de esta naturaleza (Spies *et al.* 1983).

El primer amplicón que ha sido descrito para *Rhizobium* corresponde a una región *nod-nif* del pSym de *R. etli*, cepa CFN42 (Romero *et al.*, 1991).



**Figura 2.** Modelo dinámico del genoma de *Rhizobium*. La Unidad Amplificón (A) consiste de cualquier región de ADN que se encuentre limitada por dos secuencias repetidas en orientación directa (RD). Estas secuencias repetidas directas pueden ser puntos para recombinación homóloga, dependientes de RecA. Como resultado del primer evento recombinatorio, se produce la primera duplicación del sistema y la eventual circularización durante la deamplificación de la estructura (B). El resultado de varios fenómenos recombinatorios entre estas secuencias repetidas, puede conducir a amplificaciones en *tandem* (C) y deleciones del material genético, propio de cada unidad amplificón. Estos fenómenos suceden a diferentes niveles y frecuencias.

Este amplicón mide 120 kb y sus secuencias repetidas directas de 5 kb corresponden a los operones funcionales de la nitrogenasa, *nifHDK* (Quinto *et al.*, 1982). Este amplicón se duplica a frecuencias de  $10^{-3}$  (Romero *et al.*, 1991) y el análisis de las subclonas de la cepa CFN42 indica que 1/300 células contiene la región duplicada (Brom *et al.*, 1991). Actualmente, además de este amplicón se han descrito otros tres en la misma cepa, cuya frecuencia de duplicación es aproximadamente de  $10^{-4}$ - $10^{-5}$ . Estos cuatro amplicones se encuentran sobrelapados y se localizan en una región del pSym que es rica en genes de nodulación y de fijación de nitrógeno (Romero *et al.*, 1995). Además se han detectado más amplicones en otros plásmidos e incluso en el cromosoma. Estos amplicones son muy variados y sus dimensiones fluctúan entre 10 a más de 100 kb (Flores *et al.*, 1993). Por otra parte se ha demostrado la presencia de amplicones naturales en *Rhizobium tropici* CFN299 que incluyen varios genes *nod/nif* (Mavingui *et al.* 1996).

**Significado Biológico de la Amplificación Génica.** Históricamente, la amplificación génica se ha correlacionado con fenómenos de adaptación en las bacterias (Petes y Hill, 1988; Roth *et al.* 1996). La capacidad de responder rápidamente a cambios ambientales es el elemento clave de la sobrevivencia de los microorganismos. Para sobrevivir, las células deben evaluar los cambios en el medio ambiente y responder a ellos de manera adecuada, usualmente vía sistemas de transducción de señales moleculares, que finalmente conducen a cambios en la expresión de determinados genes requeridos o bien, vía rearrreglos del material genómico (Dybving, 1993).

En *Rhizobium* este concepto puede ser de particular relevancia con respecto a su capacidad adaptativa ante condiciones ambientales cambiantes o hacia su interacción con diferentes plantas hospederas. En *R. etli* se ha comprobado que la amplificación génica puede incrementar la fijación de nitrógeno (D. Romero, datos no publicados). Es probable que la implicación más importante de la amplificación génica sea que incrementa la potencialidad del genoma para reorganizar su material genético, sin alterar la estructura global del mismo. Teóricamente, cada genoma tiene su estructura dinámica particular. Tal estructura es el resultado del número de secuencias repetidas y su posición para ser utilizadas como puntos de recombinación homóloga.

**El Análisis de Genomas Completos y su Importancia.** Analizar genomas completos como unidad puede ser la clave para evaluar la complejidad y plasticidad de los genomas bacterianos (R. Palacios, en prensa). Para ello es de gran importancia presentar un panorama global de la estructuración de cualquier genoma en un contexto recombinatorio. Con esto será posible establecer la naturaleza de aquellos elementos de secuencia reiterada, que

resulten de interés para considerar la posibilidad de la generación de rearrreglos genómicos predecibles. Afortunadamente, ahora se cuenta con una gran cantidad de herramientas computacionales para el acceso a los bancos de información genética y la manipulación de la misma. Esto permite que el análisis de genomas completos sea menos laborioso y más manejable.

Durante la realización de este trabajo se reportó la secuencia total del pNGR234a (Freiberg *et al.*, 1997). El presente estudio contemplaba entonces un análisis fino sobre la región de nodulación de *R. meliloti*. Sin embargo, la información recientemente depositada en el GenBank, para el pNGR234a constituyó una interesante materia de análisis, de tal manera que resultó de gran interés incluir esta nueva secuencia (536 165 pb) en el presente estudio.

Debido a que la secuencia que constituye el archivo aquí llamado Regulón *nod* es pequeña -47.1 kb- se decidió realizar un análisis más profundo de esta región. En el análisis de la zona de nodulación en *R. meliloti* se consideraron secuencias más pequeñas y con menor grado de similaridad. En general se describen familias de dos elementos repetidos y en las secuencias intragénicas se proponen posibles motivos protéicos conservados. Por otra parte, para el caso de *Rhizobium* sp. NGR234, dado que la secuencia del pNGR234a es de 536.1 kb se decidió considerar solamente las secuencias más grandes y con un elevado grado de simiaridad (> 98 %). Sin embargo, si el nivel de análisis hubiese sido el mismo en ambos casos se encontrarían muchas familias más dentro del pNGR234a, con niveles de similitud comparables a los descritos aquí para *R. meliloti*. Estas consideraciones deben permanecer en mente para el análisis final de los resultados y su discusión.

De esta forma, aquí presentamos los resultados y el análisis de las secuencias repetidas agrupadas en 17 FERs, encontradas en la compilación de la región de nodulación del pSym-a de *R. meliloti*. De acuerdo a una comparación entre la secuencia biológica real de esta zona y una secuencia generada azarosamente (pero de la misma composición) se determinan las secuencias repetidas de mayor interés. Se proponen 6 posibles amplicones que corresponden a aquellas reiteraciones con más altos grados de identidad a nivel de secuencia. Asimismo, con base en un análisis global de la secuencia del pNGR234a de *Rhizobium* sp. NGR234 se encontraron aquellas reiteraciones grandes (> 1 500 pb) que probablemente puedan participar como puntos para recombinación homóloga. Cada una de ellas se agrupó en 6 FERs diferentes. En este caso, nuestro análisis nos permite proponer la existencia de al menos 8 posibles amplicones. En ambas secuencias, los potenciales amplicones se encuentran sobrelapados.

## OBJETIVOS.

Como ha quedado de manifiesto en la Introducción, existen suficientes antecedentes que nos permiten afirmar la naturaleza dinámica del genoma rizobial (y tal vez del genoma bacteriano en general). Sin embargo, el análisis global de estos genomas simbióticos no ha sido suficientemente integrativo, de manera que sea posible establecer los parámetros básicos de búsqueda de algunos de los elementos genéticos responsables de esta plasticidad genómica.

Estamos convencidos de que una parte muy importante en el estudio de la dinámica del genoma simbiótico es precisamente determinar aquellos elementos genéticos, responsables de fenómenos como los de amplificación y delección, con el objetivo de categorizarlos y agruparlos. Con esta información será posible contar con una visión general de las implicaciones dinámicas que tenga la organización de cualquier genoma dado, con base en la presencia de zonas con potencialidad de generar este tipo de rearrreglos. Con tal razonamiento se planteó el presente estudio, que pretende cubrir los siguientes objetivos:

**Objetivo General:** Búsqueda de secuencias reiteradas dentro de la región de nodulación de *Rhizobium meliloti* y dentro del plásmido simbiótico pNGR234a de *Rhizobium* sp. NGR234, con énfasis en la propuesta de posibles amplicones.

### Objetivos Específicos:

- I. Construir un archivo de la secuencia conocida para la región de nodulación de *R. meliloti*, definiendo y localizando los intervalos de secuencia no conocida.
  - II. Obtener de manera manipulable, la secuencia total reportada para el pNGR234a.
  - III. Diseñar y realizar una búsqueda global de secuencias repetidas en orientación directa para ambas regiones.
  - IV. Definir las principales familias de elementos reiterados para cada región analizada.
  - V. Determinar los posibles Amplicones dentro de la región de nodulación de *R. meliloti* y en el plásmido simbiótico pNGR234a.
  - VI. Construir un mapa dinámico para el plásmido simbiótico de *Rhizobium* sp. NGR234.
-

## MÉTODOS.

**Obtención de Secuencias Depositadas en el GenBank.** Las secuencias de interés se obtuvieron haciendo una búsqueda tipo ENTREZ en el NCBI (National Center for Biotechnology Information). El NCBI cuenta con la base de datos **GenBank** de secuencias de ADN. Para la búsqueda de las secuencias depositadas en esta base de datos es necesario introducir el nombre del gen o el número de acceso que tiene éste. La búsqueda de la secuencia de genomas completos también es posible con esta base de datos. En este caso el formato de la secuencia puede ser en un solo documento o en varios, dependiendo de su tamaño. Las secuencias de las zonas que integran la región de modulación de *R. meliloti* fueron obtenidas en la base de datos GenBank. Para el acceso de la secuencia total del pNGR234a se ejecutó un FTP (*File Transfer Protocol*) anónimo. La base de datos viene comprimida en forma *gzip*. La secuencia puede ser consultada en la dirección siguiente vía NetScape versión 3.0: <http://genome.imb-jena.de/~cfreiber/pNGR234a2.html>.

**Programas para Análisis de Secuencias Reiteradas.** Los programas para análisis de secuencia citados en este trabajo se incluyen en el paquete de análisis de secuencia GCG (Genetics Computer Group Inc.), versión 8.0 y en el paquete Molpac (telnet 129.115.10.30). La construcción de los árboles se realizó usando el programa UPGMA (*Unweighted Pair Group Method with Arithmetic mean*) (Nei, 1987) del paquete GeneWorks, versión 2.5 (IntelliGenetics Inc.). Los primeros dos programas pueden ser ejecutados vía una conexión InterNet mediante el programa NCSA-Telnet 2.6. GeneWorks requiere un software particular para el uso de sus programas.

**Búsqueda Primaria de Secuencias Directas Repetidas.** Para esta clase de búsqueda se utiliza el programa Repeat en el paquete GCG. Repeat encuentra repeticiones directas en las secuencias. Se pueden establecer el tamaño, grado de severidad e intervalo en el cual se desean encontrar las repeticiones; todas las repeticiones encontradas de ese tamaño o mayores son desplegadas como alineamientos cortos en la pantalla. Repeat informa el número de repetidas encontradas según los parámetros establecidos antes de desplegar los resultados en la pantalla. Repeat puede mostrar varias secuencias repetidas que son parte de una misma estructura. Como restricciones, el programa Repeat no puede encontrar más de 1 000 repeticiones en secuencias no mayores a 350 000 pb y tampoco busca repeticiones en orientación inversa.

**Repeticiones Grandes en Secuencias Largas.** Las secuencias repetidas grandes con pocos errores de sustitución o *indeles* (inserciones-deleciones) se obtienen a partir de una búsqueda global en cualquier secuencia nucleotídica

utilizando el algoritmo descrito por **Leung et al. (1991)**. Para la búsqueda de reiteraciones en secuencias genómicas grandes (de hasta millones de bases) se utiliza este algoritmo. En el presente estudio, la secuencia total del pNGR234a fue analizada con este método. Cada reiteración recibe el nombre de *match*. El *match* es un agregado de bloques de identidad extendidos entre dos o más secuencias. El algoritmo utilizado por este programa encuentra cada *match* localizando primero los *core block*, que son aquellas secuencias que se presentan al menos como parejas y que son idénticas entre ellas. Cada *core block* es entonces extendido en ambos extremos mediante bloques cortos de secuencia llamados *extended blocks*, los cuales están separados por errores o *error blocks* de longitud variable. Estos parámetros pueden ser especificados por el usuario o por los parámetros "default" del mismo programa, de manera tal que se asegura que las reiteraciones localizadas sean poco probables de encontrarse de manera azarosa.

Los parámetros del programa que pueden ser manipulados para la búsqueda de las reiteraciones (*matches*) son la longitud de los bloques de secuencia básica (*core block length*), el error de concatenación de cada uno de estos bloques (*error block*) y el error de alineamiento entre una repetición y cualquier otra (*alignment error*). Asimismo, se pueden manipular los niveles de impresión (*printing levels*) para observar en pantalla las reiteraciones, según el tamaño que se especifique por el usuario. Los programas dentro del paquete de análisis **Molpac**, donde se pueden ejecutar esta clase de búsquedas son los programas **rep** e **icrep**. El primero de ellos es utilizado para búsqueda de reiteraciones en orientación directa y el segundo para secuencias en orientación inversa.

**Extensión de Motivos Centrales de Secuencia.** Cada secuencia repetida puede ser extendida en cada extremo, si se reduce la severidad de su alineamiento con alguna otra secuencia. Con la finalidad de extender estos motivos de secuencia centrales -encontrados con el programa Repeat- el siguiente objetivo es extenderlos lo más posible sin incluir errores muy grandes. La extensión de los motivos centrales de secuencia se obtiene mediante un alineamiento tipo **FastA** (Fast Alignment) entre las dos secuencias repetidas encontradas en cada caso. FastA es una búsqueda de similitudes tipo **Pearson-Lipman (1988)**, entre una secuencia dada y cualquier grupo de secuencias en el GenBank o secuencias especificadas por el usuario.

Este método compara una secuencia de interés con otra especificada por el usuario y reporta la mejor similitud en el alineamiento. El programa FASTA calcula un puntaje de similitud "local" donde se encuentra la mejor

región de similaridad entre ambas secuencias. Si existen porciones de alineamiento pobres fuera de la esta región de similaridad, estos no influyen en el puntaje final. FASTA reporta solamente los valores de similaridad para el mejor alineamiento encontrado entre las dos secuencias.

El Algoritmo FASTA (Pearson, 1990) utiliza cuatro pasos para calcular los tres valores que caracterizan la similaridad entre secuencias; estos valores son el porcentaje de identidad entre las dos secuencias sobrelapadas, la cantidad de símbolos sobrelapados y su ubicación. En el primer paso identifica las regiones de mayor densidad de identidades, aquí se especifica un parámetro llamado *ktup* que determina cuantas identidades consecutivas se requieren para generar un grupo de identidades (*match*). Para comparaciones de secuencias de ADN, este parámetro puede variar entre 1 a 6; se recomiendan valores entre 4 y 6. Cuando la secuencia base es corta o se trata de un oligonucleótido se prefiere un *ktup* = 1. En esta etapa el programa genera una matriz de "diagonales" donde se buscan todas las regiones de similaridad entre las dos secuencias, contando los mejores grupos de *ktup* (tamaño de la palabra) y penalizando los malos alineamientos (*mismatches*). Con este método se identifican las regiones en la diagonal con mayor densidad de grupos de identidad. En el paso siguiente cada extremo de la región se delimita para lograr el mejor alineamiento posible, esto resulta en un grupo de alineamientos parciales sin brechas (*gaps*). Las brechas de secuencia no conocida en cada alineamiento no pueden ser muy grandes; asimismo, cualquier alineamiento entre dos secuencias no puede ser alterado por más de 32 *indeles*. Durante el tercer paso, las regiones cercanas se unen en una sola región incluyendo errores tipo *mismatches* y *gaps* necesarios para unir las regiones iniciales. Finalmente, en el cuarto paso se genera un histograma donde se calculan los valores de similaridad promedio de la secuencia dada y el resto de las comparadas (cuando se trata de comparaciones con muchas de estas); calcula también la desviación estándar de los valores iniciales.

**Generación de Secuencias Azarosas.** Para esta clase de manipulación, se utiliza el programa **Shuffle** en el paquete GCG. Shuffle altera el orden de los símbolos (A, C, G, T) en una secuencia sin cambiar su composición. Shuffle utiliza un generador de números azarosos para mezclar las posiciones de los símbolos en una secuencia. El generador numérico se inicializa según la hora, de tal manera que al ejecutar el programa a diferentes tiempos se obtienen resultados diferentes. Se ejecutaron 10 repeticiones aleatorias en la secuencia estudiada.

**Identificación de Patrones de Secuencia.** FindPatterns identifica secuencias que contienen pequeños patrones de secuencia conservada. Los

patrones se pueden definir de manera arbitraria y se permiten ciertos errores pero sin brechas. Cada patrón puede ser parte de una secuencia determinada o puede tipificarse en la misma terminal, en la computadora personal. La búsqueda se realiza en las dos hebras. FindPatterns toma la secuencia blanco y escanea el resto de la secuencia generando un listado de alineamientos con los resultados de la búsqueda de esa secuencia.

**Búsqueda de Motivos Protéicos.** El programa **ProfileScan** utiliza una base de datos de perfiles protéicos para encontrar motivos estructurales y de secuencia en proteínas. Este programa utiliza el método de **Gribskov et al. (1988)** para realizar una búsqueda en un listado de los motivos protéicos en una base de datos. Cada posible alineamiento permite un margen de error. El resultado de la comparación es una lista de alineamientos donde se observan ambas secuencias y su grado de identidad.

**Predicción de Estructuras Secundarias a Nivel Protéico.** Aquellas secuencias que se encuentran formando parte de regiones codificantes, pueden ser visualizadas gráficamente a nivel peptídico y establecer la estructura secundaria de ese péptido en particular. Con el fin de evaluar la posibilidad de encontrar ciertas similitudes a nivel de estructura protéica, las secuencias reiteradas que son intragénicas, deben ser traducidas a secuencias peptídicas mediante **Translate**. Una vez traducidas las secuencias reiteradas, la predicción de su estructura secundaria se basa en el método de **Chou-Fasman (1978)** para establecer la probabilidad de que el polipéptido forme hélices-alfa, láminas-beta, giros y enrollamientos. Este algoritmo es la base del programa **PeptideStructure**. Para visualizar estas estructuras, el programa **PlotStructure** despliega gráficamente las predicciones de PeptideStructure.

**Secuenciación Automática de ADN.** La secuenciación de una parte de la secuencia no conocida entre los genes *nifH* y *fixA*, dentro de la región de nodulación analizada en *R. meliloti*, se realizó con el método de terminación de cadena por dideoxinucleótidos fluorados usando el kit *Taq DyeDeoxy Terminator Cycle Sequencing* (Applied Biosystems Inc., Foster City, CA). Los productos de la reacción de secuenciación fueron separados por electroforesis en geles de poliacrilamida, bajo condiciones desnaturizantes y su análisis se realizó en un secuenciador automático de ADN, modelo 373 A (Applied Biosystems Inc.).

## RESULTADOS

**Integración del Archivo de Secuencia de la Región de Nodulación de *R. meliloti*: Rmel\_Regulon\_47.1\_kb.txt.** La compilación de un Archivo de Secuencia constituye la parte inicial básica de cualquier proyecto de análisis genómico. Con base en una búsqueda en el GenBank se obtuvieron todas aquellas secuencias que han sido descritas para la región de nodulación de *R. meliloti*. Se encontraron 30 entradas de secuencia que corresponden a los genes de nodulación y fijación de nitrógeno descritos en la **Tabla 1**. Esta región contiene al menos 5 operones que integran el Regulón *nod*, un modelo regulatorio que comprende la regulación coordinada entre estos operones, que participan en la interacción *Rhizobium*-planta (**Kondorosi, 1992**). Cada secuencia encontrada tiene diferente longitud y algunas se han secuenciado en más de una cepa.

Cada secuencia encontrada en el GenBank corresponde a genes u operones completos pero no fue posible estructurar un archivo completo de secuencia a partir de una sola cepa. Sin embargo, en este estudio en particular, trabajar con un archivo de secuencia compuesto permite contar con la mayor información posible -a nivel de secuencia- para esta región. Trabajar con la secuencia conocida de una sola cepa resultaría limitante debido a que muchas reiteraciones serían ignoradas. Por otra parte, la secuencia utilizada aquí, tampoco es ideal debido a que la secuencia puede variar entre una cepa y otra e incluso la organización génica también.

Sin embargo, en el presente caso la variabilidad entre las secuencias de una y otra cepa, podría ser mínima. Un análisis de alineamiento entre los genes *nodD* de las cepas Rm 1021 y Rm 41 (alineamientos no mostrados) revelaron sólo pequeñas diferencias entre ambas secuencias. Los valores de divergencia entre las secuencias de cada gen *nodD* son: *nodD1-nodD1* (0.18% en 1.10 kb), *nodD2-nodD2* (1.46% en 1.09 kb) y *nodD3-nodD3* (0.79% en 1.61 kb).

En general son dos las cepas que se encontraron como fuente de las secuencias: Rm 1021 y Rm 41. Con base en los mapas genéticos conocidos reportados en la literatura (Para una revisión ver: **San Juan et al., 1992**), una vez ordenada esta información y localizados los solapamientos se organizaron seis grupos continuos de secuencia que en total cubren 45745 pb e incluyen 5 brechas de secuencia no conocida (*gaps*). Estos seis grupos de secuencia se integraron en un solo archivo de secuencia que denominamos "Rmel\_Regulon\_45.7\_kb.txt". Las longitudes de estas seis secuencias, que integran el archivo total se muestran en la **Tabla 2**.

**Tabla 1.***Secuencias reportadas para la región de nodulación de R. meliloti.*

Loci	Números de Acceso	Longitud <sup>a</sup>	Cepa	Referencias
1) n4nodM-NnollFGHI	X58632	7080 pb	41	MGG 228: 113-124 (1991)
2) fixF	M18272	3073 pb	2011	J. Bacteriol. 169: 5393-5400 (1987)
3) nodD1	X04473	1106 pb	41	J. Mol. Biol. 191: 411-420 (1986)  Nucleic Acids Res. 12: 9509-9524 (1984)
4) nodD1-n1ABC	M10923 M11268 M11269	4230 pb	1021	DNA 4: 241-248 (1985) J. Bacteriol. 162: 469-476 (1985)
5) n1nodABC	X01649 M13287	3373 pb	41	Nucleic Acids Res. 12: 9509-9524 (1984)
6) nodP1Q1	X14809	4037 pb	SU47	Mol. Microbiol. 3: 745-755 (1989)
7) nodP1Q1	M68858 J03676	3476 pb	1021	MPMI 2: 181-194 (1989)
8) ORF downstream nodQ1	L08667	1235 pb	1021	MPMI 7: 151-153 (1994)
9) n2nodFEG	Y00604 M37416	4110 pb	1021	Genetics 117: 191-201 (1987)
10) n2nodFEG	X04379	3932 pb	2011	Nucleic Acids Res. 14: 7453-7472 (1986)
11) n3nodH	X04380	2078 pb	2011	Nucleic Acids Res. 14: 7453-7472 (1986)
12) n3nodH	M37417	1277 pb	1021	Genetics 117: 191-201 (1987)
13) syrM	X61396	1550 pb	41	Mol. Microbiol. 5: 3035-3048 (1991)
14) syrM	M33495	2636 pb	1021	J. Bacteriol. 172: 3695-3700 (1990)
15) Intergénica syrM-nodD3	L76274	912 pb	1021	J. Bacteriol. 178: 1782-1787 (1996)
16) nodD3	X53820	2053 pb	1021	Nucleic Acids Res. 19: 921-927 (1991)

*Continúa en la página siguiente*

**Tabla 1.**  
*cont.*

17) <i>nodD3</i>	X61457	2182 pb	41	Mol. Microbiol. 5: 3035-3048 (1991)
18) <i>nodD3</i>	S63823	2053 pb	NC <sup>a</sup>	Science in China. Series B, Chemistry, Life Sciences and Earth Sciences 36: 225-236 (1993)
19) <i>nifH</i>	V01215	1329 pb	41	Nucleic Acids Res. 9: 5711-5722 (1981)
20) <i>fixABCX</i>	M15546	4785 pb	1021	J. Bacteriol. 169: 1127-1136 (1987)
21) " <i>fixD</i> " ( <i>fixX-nifA</i> )	X03065	2111 pb	2011	EMBO J. 4: 2793-2800 (1985)
22) <i>nifA</i>	X02615	1875 pb	1021	Nucleic Acids Res. 13: 4539-4555 (1985)
23) <i>nifB</i>	M15544	1722 pb	1021	J. Bacteriol. 169: 1120-1126 (1987)
24) <i>fdxN</i>	X52662	944 pb	2011	MGG 216: 293-302 (1989)
25) <i>fdxN</i>	M21841	255 pb	2011	J. Bacteriol. 170: 4406-4410 (1988)
26) <i>n5nodL</i>	X61083	900 pb	41	Plant Mol. Biol. 18: 843-846 (1992)
27) <i>n5nodLnoeAB</i>	U26430	4376 pb	2011	Mol. Microbiol. 17: 687-699 (1995)
28) <i>nodD2</i>	M29367	1317 pb	1021	J. Bacteriol. 172: 901-911 (1990)
29) <i>nodD2</i>	X04474	1096 pb	41	J. Mol. Biol. 191: 411-420 (1986)
30) <i>n6nolQaQbS</i>	X91350	2577 pb	2011	FEMS Microbiol. Lett. 133: 285-291 (1995)

<sup>a</sup> Para la compilación del archivo de Secuencia Total, se tomaron las secuencias equivalentes más largas.

<sup>b</sup> No clasificada.

**Tabla 2**

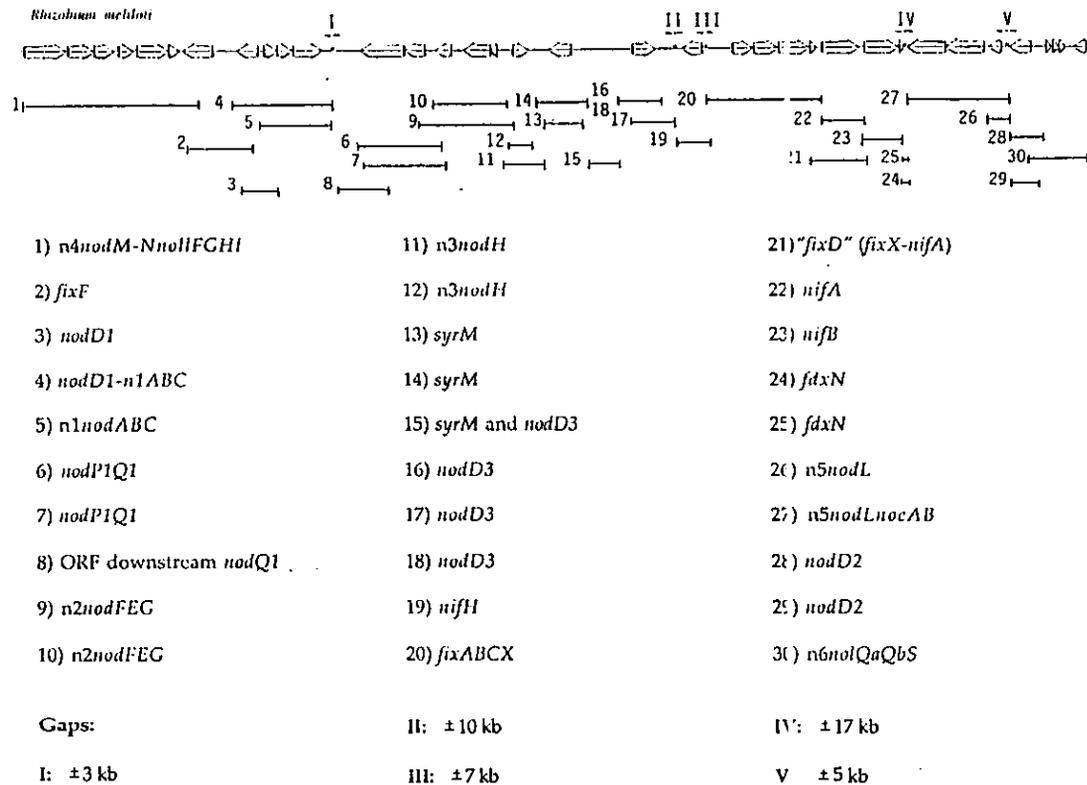
*Continuos de secuencia encontrados para la región de nodulación de R. meliloti.*

Continuo	Genes incluidos	Longitud
Secuencia I:	n4nodMnolFGHInodNnifNnodD1n1ABC	13 440 pb
Secuencia II:	nodQ1P1GEFn2n3HsyrMnodD3	14 850 pb
Secuencia III:	nifH	1 275 pb
Secuencia IV:	fixABCXnifABfdxN	8 470 pb
Secuencia V:	noeBANodLn5	4 380 pb
Secuencia VI:	nodD2n6no1QaQbS op. stach.	3 330 pb
Secuencia Total:	'Regulón nod'	45 745 pb

Los intervalos de secuencia no conocida entre los seis fragmentos encontrados miden aproximadamente 3 kb, 10 kb, 7 kb, 17 kb y 5 kb respectivamente. Finalmente, después del diseño de oligonucleótidos específicos para secuenciar la zona entre los genes *nifH* y *fixA* se secuenciaron 1 400 pb. Con esta última "entrada" de secuencia se obtuvo finalmente una secuencia de 47 145 pb que recibió el nombre de "Rmel\_Regulon\_47.1\_kb.txt". De esta manera se introdujeron los datos de toda la búsqueda en el GenBank, en forma de 6 segmentos de secuencia de ADN, separados por 5 intervalos y la secuencia determinada en este laboratorio correspondiente a una parte del intervalo entre los genes *nifH* y *fixA* (1 400 pb, **Figura 3**).

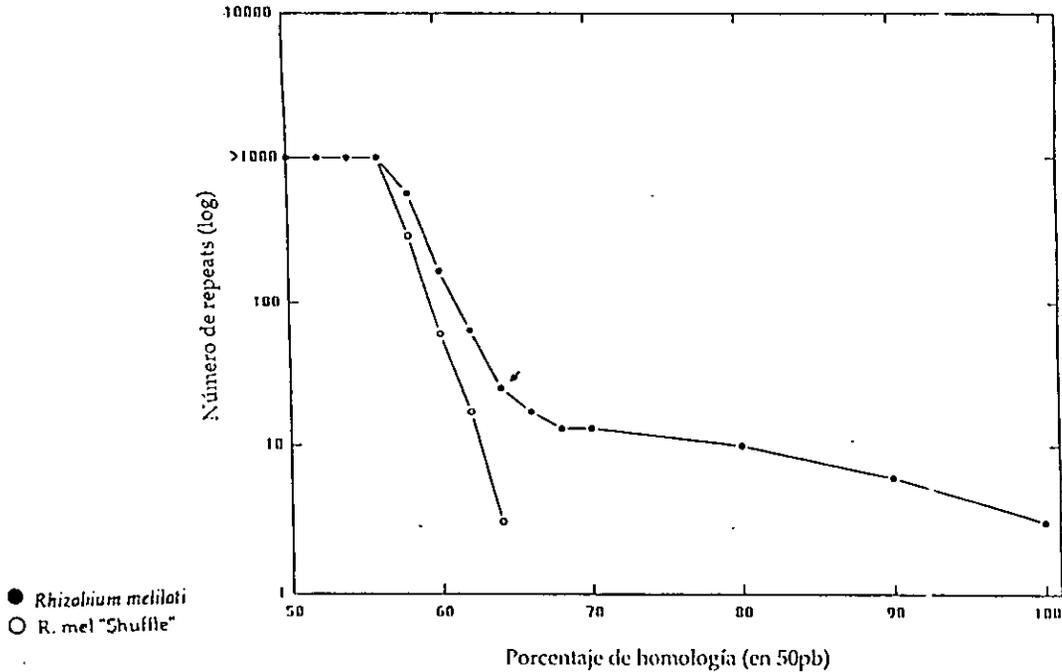
**Búsqueda Primaria de Secuencias Reiteradas en la Región de Nodulación de *R. meliloti*.** Una vez integrado el Archivo de Secuencia se buscaron los motivos centrales de los elementos repetidos utilizando el programa Repeat (GCG). Como principal objetivo se determinaron aquellas secuencias reiteradas que pudieran ser funcionales en el contexto de la recombinación homóloga. Para el cumplimiento de este objetivo se diseñaron dos tipos de búsqueda mediante el empleo de este programa, con las siguientes características: La primera búsqueda se hizo en cualquier intervalo de 50 pb (**Figura 4a**) y la segunda en cualquiera de 100 pb (**Figura 4b**) en la misma secuencia y su contraparte generada azarosamente. El programa Repeat se ejecutó pidiendo a la computadora encontrar secuencias con un 50% de identidad y hasta un 100%, en intervalos ascendentes de 2%. Con estos resultados obtenidos se graficaron el número de secuencias repetidas encontradas en cada caso.

Para el análisis de la secuencia de *R. meliloti*, las secuencias repetidas de interés se consideraron a partir del razonamiento siguiente: En cualquier secuencia dada resultan interesantes todas aquellas secuencias que se encuentren repetidas al menos una vez y cuya probabilidad de ser encontradas



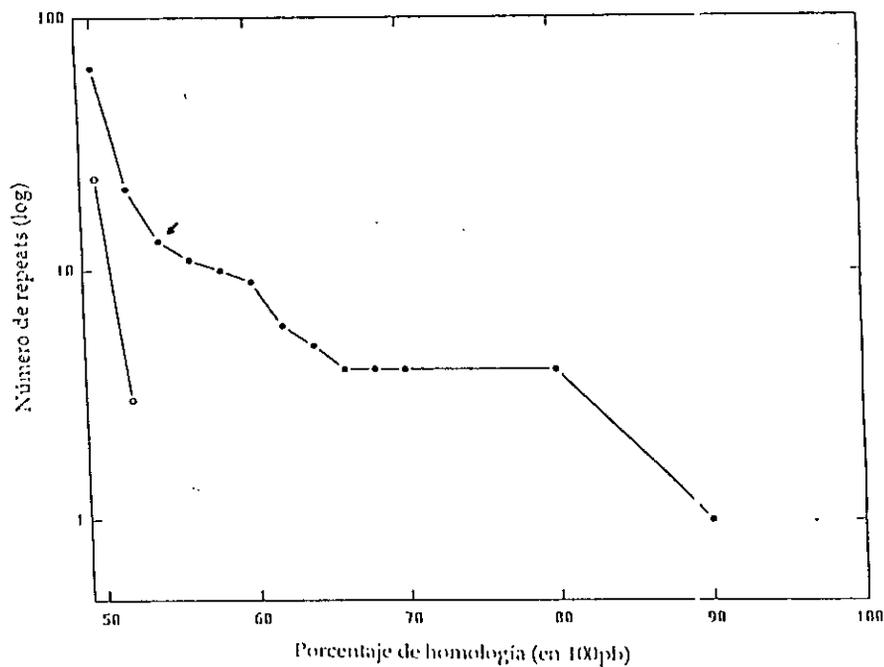
**Figura 3.** Compilación de las entradas de secuencia encontradas en el GenBank. Cada entrada de secuencia se representa como una línea horizontal que se alinea según los genes que porta cada una. Algunas entradas de secuencia sobrelapan con otras en ambos extremos. Los sobrelapamientos de estas secuencias permiten construir un mapa global de la región (arriba), llamado R. meliloti Regulón *nod*. Aquellos sitios donde no se encontró sobrelapamiento fueron considerados como intervalos de secuencia no conocida y en base a una revisión bibliográfica se incluyen las longitudes aproximadas de cada uno de estos intervalos no secuenciados (abajo).

Número de secuencias repetidas para el Regulón-*nod*



**Figura 4a.** Comparación entre el número de secuencias repetidas en la secuencia Regulón *nod* de *R. meliloti* y las secuencias repetidas encontradas en una secuencia generada azarosamente en base a la primera. Los puntos negros corresponden al número de repeticiones en la secuencia original y los puntos blancos a las repeticiones en la secuencia azarosa. La tendencia general de ambas gráficas es tener un gran número de secuencias repetidas de baja similitud, mientras que son escasas aquellas donde el grado de similaridad es superior. En la comparación entre ambas gráficas, el punto de inflexión en la gráfica (flecha) es considerado aquí como el punto a partir del cual las secuencias repetidas son de interés y pueden tener importancia en fenómenos recombinatorios dependientes de RecA. Las repeticiones graficadas se encontraron con el programa Repeat (GCG) a partir de una búsqueda en ventanas de 50 pb.

Número de secuencias repetidas para el Regulon-*nod*



- *Rhizobium meliloti*
- R. mel "Shuffle"

Figura 4b. Comparación entre el número de secuencias repetidas en la secuencia Regulon *nod* de *R. meliloti* y las secuencias repetidas encontradas en una secuencia generada azarosamente en base a ésta. Las repeticiones graficadas se encontraron con el programa Repeat (GCG) a partir de una búsqueda en ventanas de 100 pb. Para detalles ver Figura 4a.

sea superior a la de ser encontradas en una secuencia azarosa de la misma composición. De este modo se empleó el programa Shuffle (GCG) y se generó una secuencia azarosa alterando el orden de los nucleótidos dentro del archivo de secuencia de *R. meliloti*; para obtener esta secuencia se hicieron 10 repeticiones con este programa, de tal manera que la secuencia azarosa final fue resultado de 10 ensayos consecutivos de reordenamiento aleatorio de la secuencia original.

Como siguiente etapa se ejecutó el programa Repeat en la secuencia generada al azar de la misma composición a la de *R. meliloti*. En las figuras 4a y 4b ambas gráficas se muestran comparadas. El hecho de haber graficado sendos datos en una misma figura, nos permite comparar el número de reiteraciones que se pueden encontrar en una secuencia azarosa y establecer las diferencias que se observen con una secuencia biológica real (secuencia original, *R. meliloti*). Por otra parte es interesante hacer notar que en esta figura se observa una cierta tendencia de encontrar secuencias repetidas cortas en gran cantidad. La primera fase de nuestro análisis consistió en encontrar las posibles diferencias entre esta secuencia real (región de nodulación) y su contraparte azarosa. También se puede observar que las gráficas de la Figura 4 tienen dos fases; durante la primera de ellas, ambas secuencias se comportan de manera similar y muestran la misma tendencia con pendientes muy parecidas. Sin embargo, hay un incremento en el número de las repeticiones encontradas en la secuencia real; la curva que representa a la región de nodulación se separa de la curva que representa a la secuencia azarosa. Este fenómeno se discutirá más adelante. La primera fase de cada gráfica puede ser considerada de poco interés en términos de repeticiones grandes para recombinación homóloga dependiente de *recA* (ver *Discusión*).

La segunda fase de las gráficas (después del punto de inflexión en la secuencia real) está representada por repeticiones que son más grandes y que por su tamaño son probabilísticamente difíciles de encontrar repetidas en cualquier secuencia azarosa de ADN. Estas repeticiones son sucesivamente más escasas. Conforme estas secuencias son más grandes tienen mayores posibilidades de intervenir como puntos de recombinación, dependientes de *RecA*, para generar rearrreglos genómicos. De esta manera -en teoría- aquellas secuencias que mostrasen un comportamiento diferente al esperado por el azar resultan de interés para su análisis y podrían tener significado biológico a nivel molecular. En efecto, para *R. meliloti* ambas gráficas muestran una tendencia de la secuencia real a separarse de la secuencia azarosa.

En este punto se cumplió con el objetivo de determinar los motivos centrales mediante una búsqueda primaria. Aquellas secuencias que se

apartaron de las secuencias esperadas por el azar, fueron consideradas importantes como posibles puntos para recombinación homóloga. De esta manera, los puntos de inflexión son el punto de partida para analizar cada secuencia por separado. Cada una de estas secuencias fueron extendidas mediante el uso del algoritmo FastA (GCG).

**Extensión de los Motivos Centrales de Secuencia.** Para ambas secuencias estudiadas aquí (*R. meliloti* y *Rhizobium* sp. NGR234), con los resultados obtenidos a partir de la búsqueda primaria de reiteraciones, la siguiente etapa corresponde a la extensión de cada reiteración genómica con posible significado biológico. Para tal objetivo se empleó el programa FastA (Pearson y Lipman, 1988). El alineamiento tipo FastA es capaz de extender a ambos extremos un alineamiento dado entre dos secuencias, permitiendo un cierto margen de errores como pequeñas deleciones, inserciones y/o corrimiento de bases.

**Definición de Familias de Elementos Reiterados (FERs).** Por definición, una FER es un grupo de secuencias nucleotídicas similares entre sí con algún margen de error. Cada familia puede estar formada al menos de dos elementos reiterados, sin embargo cuando son más de dos elementos se pueden formar subfamilias. Los principales atributos que tiene una FER son la longitud de cada secuencia repetida y el grado de identidad entre estas. Cada secuencia puede encontrarse en orientación directa o inversa y también pueden estar sobrelapadas. Con los resultados obtenidos en la primera parte del análisis -las primeras secuencias repetidas ya extendidas- y los resultados del alineamiento múltiple entre ellas se integraron finalmente las familias que definiremos a continuación para ambas secuencias bacterianas.

#### ***R. meliloti*: Familias de Elementos Reiterados (FERs).**

En la secuencia compuesta de *R. meliloti* se encontraron 17 familias diferentes de elementos repetidos de 2 o más elementos. Las familias encontradas dentro de la región de nodulación en *R. meliloti* se muestran en la **Figura 5**. A continuación se incluyen las características de cada una de las diferentes subfamilias posibles para esta zona.

**FER A: Genes *nodD* y *syrM*.** Estos elementos reiterados corresponden a las regiones codificantes de los genes reguladores *nodD* y *syrM*. En *R. meliloti* ya habían sido reportados tres genes *nodD*; estos se denominaron *nodD1*, *nodD2* y *nodD3* además de *syrM* (Honma y Ausubel, 1987; Barnett y Long, 1990). Se sabe que el producto protéico de estos genes corresponde a los activadores transcripcionales de los otros genes *nod*, aunque también pueden actuar como



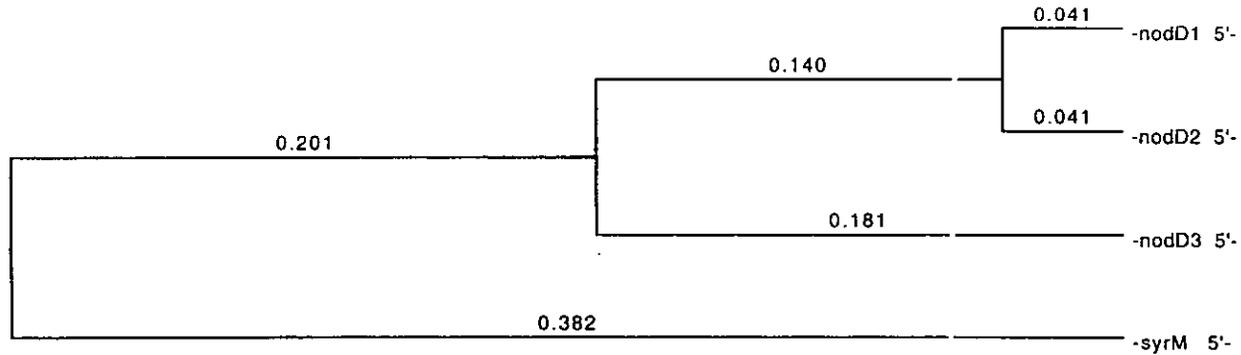
represores (Schlaman *et al.*, 1992).

Mediante el empleo de los análisis por hibridización ADN-ADN se ha comprobado un elevado grado de similaridad entre estas tres copias de los genes *nodD* (Honma *et al.*, 1990). Un análisis de secuencia reveló también un alto grado de similitud entre los genes *nodD1* y *nodD2* (de hasta el 85%, **Figura 6**). Un estudio detallado de esta familia con base en su homología ha clasificado a las proteínas NodD dentro de la familia de reguladores transcripcionales tipo LysR (Schlaman *et al.*, 1992). Estas proteínas necesitan un compuesto inductor de naturaleza flavonoide para su activación.

El resultado presentado aquí del alineamiento entre los genes *nodD1*, *D2*, *D3* y *syrm* nos muestra un grado de identidad del 45.7% en 543 pb sobrelapadas. Este valor corresponde a la identidad promedio entre todos los miembros de la familia. El alineamiento coincide con el extremo-N de las proteínas codificadas por cada uno de estos genes. Además es precisamente en este lugar donde se ha reportado ya un motivo de unión al ADN, tipo H-T-H (Schlaman *et al.*, 1992). Este motivo H-T-H es el que reconoce la *nod* box para controlar la transcripción de los operones *nod* en un circuito regulatorio complejo (Kondorosi, 1992). Las relaciones de similaridad entre cada elemento en esta familia se ha esquematizado en la **Figura 6**.

A partir de aquí, la descripción de cada familia se presenta por grupos grandes de elementos reiterados. Sin embargo, un análisis más profundo de cada una de estas familias puede evidenciar subfamilias de elementos repetidos para cada familia propuesta. En general, una familia está integrada por un cierto número de elementos reiterados, con cierto nivel de similitud. Una subfamilia es considerada como un subgrupo dentro de cada familia, la cual tiene menos elementos repetidos pero cuyo grado de similaridad y/o extensión siempre son superiores a los valores de similaridad de las reiteraciones en cada familia. En la **Tabla 3** se muestran las principales subfamilias que se encontraron dentro de las FERs A-G.

**FER B: *nod* box.** En un estudio previo de ciertas secuencias de ADN en la cepa Rm 41 de *R. meliloti* se había descubierto ya una secuencia conservada de 47 pb en las regiones -5' de los operones *nodABC*, *nodFEG* y *nodH*. En el mismo estudio también se identificaron otras tres copias de lo que se llamó *nod* box (Rostas *et al.*, 1986). Estudios posteriores identificaron las regiones codificantes del resto de los operones que no habían sido aún identificados y de esta manera se detectaron el operón *nodMno1FGHInodN*, controlado por la *nod* box n4 (Baev *et al.*, 1991); el operón *nodL-noeAB*, controlado por la n 5 (Baev y Kondorosi, 1992) y el recientemente descrito operón *nolQaQbS*, al



**Figura 6.** FER A: Relación de distancia por similitud entre las secuencias nucleotídicas de los genes *nodD* y *syrM*. Los números indican el porcentaje de diferencia entre una secuencia y la(s) otra(s).

**Tabla 3**

*Subfamilias encontradas para las FER A-G dentro de la región de nodulación de R. meliloti.*

FAMILIA	Número de elementos	Orientación	Extensión (pb)	Identidad (%)	Distancia de separación (pb)
<b>FER A</b> ( <i>nodD1-2-3-syrM</i> )	4	Dir. e Inv.	543	45.7	- <sup>a</sup>
<b>Subfam. A.a.</b> ( <i>nodD1-2-3</i> )	3	Dir. e Inv.	1087	94.4	-
<b>Subfam. A.b.</b> ( <i>nodD1-D2</i> )	2	Directas	1106	82.9	75500
<b>FER B</b> ( <i>n1-n8</i> )	8	Dir. e Inv.	47	27.7	-
<b>Subfam. B.a.</b> ( <i>n3-n4-n7-n8</i> )	4	Dir. e Inv.	47	36.2	-
<b>Subfam. B.b.</b> ( <i>n5-n2-n6-n1</i> )	4	Dir. e Inv.	47	38.3	-
<b>FER C</b> ( <i>P1-P2-P3</i> )	3	Dir. e Inv.	340	47.9	-
<b>Subfam. C.a.</b> ( <i>P1-P3</i> )	2	Directas	319	78.1	7930
<b>FER D</b> (Reps II)	3	Directas	76	73.7	-
<b>Subfam. D.a.</b> (Reps II <i>nifH-fixA</i> )	2	Directas	77	92.2	8420
<b>FER E</b> (Reps III)	2	Directas	160	80.0	360
<b>FER F</b> (Reps IV)	2	Directas	81	77.8	35265
<b>FER G</b> (Intragen. <i>fixB-nifA</i> )	3	Directas	113	45.1	-
<b>Subfam. G.a.</b> ( <i>fixB-nifA</i> )	2	Directas	110	55.5	2365

<sup>a</sup> Solamente se incluyen las distancias entre dos elementos reiterados (Amplicones).

parecer no controlado por *n6* aunque en la misma dirección a ésta (Plazenet *et al.*, 1995).

La presencia de estas 6 copias está bien documentada en la literatura y al menos las primeras 5 son funcionales. El análisis que se presenta aquí propone al menos otras 2 copias tipo *nod* box en la región de nodulación de *R. meliloti*. Nuestros resultados indican que una probable copia de *nod* box (*n7*) se encuentra antes del gen *syrM*. Este dato ya ha sido previamente reportado en un estudio sobre los sitios de inicio transcripcional de los genes *syrM* y *nodD3* (Barnett *et al.*, 1996). En dicho estudio se reporta una secuencia muy parecida al consenso *nod* box que está localizada 26 nucleótidos antes del sitio +1 de inicio de la transcripción para *syrM*. Esta séptima secuencia muestra un buen alineamiento con las seis copias anteriores y respeta el consenso que se muestra en la **Figura 7**.

Finalmente, el resultado de nuestro análisis también revela una posible octava copia tipo *nod* box en el extremo -5' de la región codificante de *fixC*, aunque sobrelapada en el extremo -3' del gen inmediatamente anterior a éste, *fixB*. Estas dos copias adicionales encontradas en nuestro análisis parecen cumplir con ciertas características que podrían considerarse necesarias para ser su inclusión como *nod* box o reductos de estas: *i*) Muestran un consenso elevado en el grado de identidad, organizado en los tres motivos de secuencia ya descritos, *ii*) estos tres motivos de secuencia se encuentran separados entre sí aproximadamente por la misma distancia y *iii*) se localizan en la región -5' de secuencias codificantes de genes *nod* estructurales o reguladores y -al menos para *n8*- de un gen *fix*. La relación encontrada entre las 8 copias para *nod* box se muestra en la **Figura 8**.

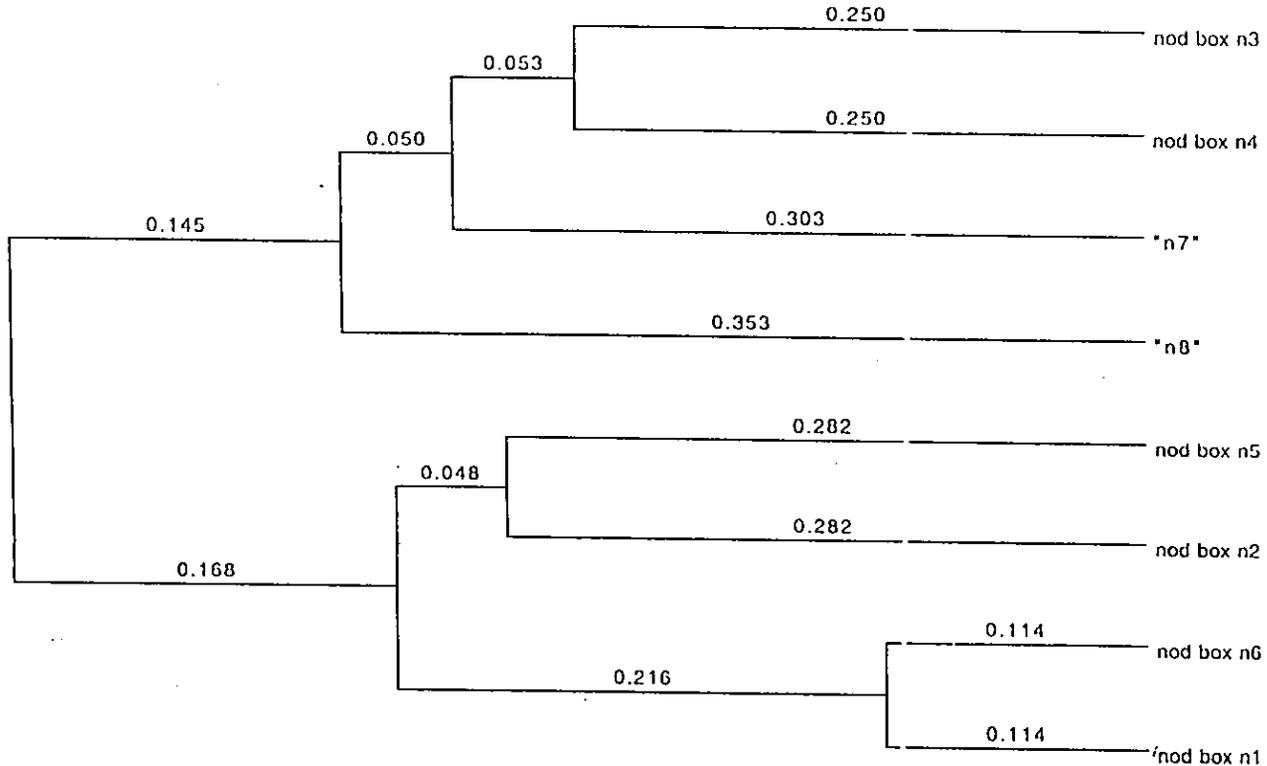
**FER C: Regiones Promotoras Repts I: P1-P2-P3.** La tercera FER corresponde también a secuencias previamente descritas en la literatura y cuyo papel principal está en algún tipo de regulación transcripcional de los genes *nif*. Este grupo de secuencias reiteradas han sido identificadas como regiones promotoras que comparten una buena similitud a nivel de secuencia de ADN, de hasta más de 160 pb (Better *et al.*, 1983). En la **Figura 9** se muestran las relaciones de similaridad entre los elementos P1, P2 y P3.

Estas secuencias fueron descritas para *R. meliloti* 102F34 y localizadas mediante hibridización ADN-ADN. La región *nif* de esta cepa contiene las tres copias denominadas P1, P2 y P3, localizadas antes de unidades transcripcionales específicas. Con base en nuestro análisis podemos inferir que P1 es parte de la región reguladora del gen *nifH* y que P2 podría ser parte de la región reguladora del gen *fixA*.

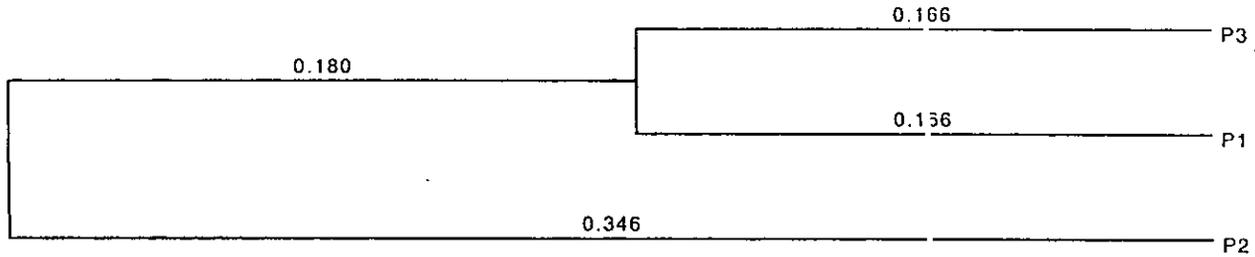
**ALINEAMIENTOS nod box:**  
 nod boxes: 13 id/47 pb (8 elementos) 27.7%

	CATGTGCGGCATCCATATCGCAGATGATCGTTATCCAAACAATCAATTTACCAATCTT-GCAGAGTCCT	nod box n1... nodA
	CGTCACAGACATCCATTTACGGATGGCCGACATCCAAACAATCGATTTTACCAATCCCCTGATATGTA	nod box n2... nodF
	CGTT-CTCATATTCACAGGCTGGATCCCTCTCATAAAAAACAATCGATTTTACCAATCTGGCGACTCCGA	nod box n3... nodH
	GTCTTTCTTATCCATAGGGTGGATGATGCTATCCTCATAATCGATTTTACCAATCTT-GCTGAATGTT	nod box n4... nodM
	AGCTGGTGTCTATCCACCGCGGATAAAGGTGATCCAAACAATCGATTTTACTAATCATCAAGTGTGCTC	nod box n5... nodL
	CATGTGCGGTATCCATAGCCGAGATGATTGTT-CCCAACAATCGATTTTACCAATCCCCTAGAGAACT	nod box n6... noiQa ?
	GAGGCTATGTATCCAAACCTGGATGACAGTCAATGCGAAAACCGGATTTCTCCGTGCTC-GCCGAATGCC	nod box 'n7'... syrM ?
	CATTGTTGCTATCAACA---CCGATC-CGAACGCGCCATCTTCGATTTCCGCCACCTCGCGTGTGGC	nod box 'n8'... fixC ?
	..... <u>ATVMAHH</u> ..... <u>VGATV</u> ..... <u>RYVMNMAHMHKCRATTTYNMVHYVC</u> .....	Consenso
n1 >	ATTAGAGAACCCCTGAAGTTAATGGAATCAAGGTGCGGCGGAGAAAAGTTTACAAAGTACAGGATGGGTC	
	CGAATTTAGAGCCGTCATCTAAGCGCTCGACCAACGGTCCAGCGCTACGGTTGGCGTCCCCGGTGTAACT	
n2 >	GCACAAGCTCCGACAGATAGGGAGGCCAATGATGTTCTTCGTCATCGGAGGCTTCGACACAGACAGCCG	nodA
	GATGTCCGGCTCTGCGGGCATPAGGCTTAGCCAGTCCGCGACGCTGATGATAATTTTCGTAATCGGGCCG	
	CCTCAGGAAATTTAGCCGCGGTCGCAACACAGCTAAAGGGAACAGAAATG	nodF
n3 >	TTAAAACGCTAAGCAAGTCCGGATTGCGAGTGAGCGGGTCCGGGAATCAGCGCTGAGGTGTCGGCCGATA	
	GAAGCAGCAAACTAGCAAACGTCGAGCTCTGACCGAGTCAATGGAAGCTAACCATAGGTGATCAACG	
	CTTAAGAGCAGCGCGCTCCCCAGACTAAGCAAGCAGCGGGGGAACATAATCTCCGCTGAAACCACCAC	
	AACTGCTAAATGGAAAACACGCGCTGAATG	nodH
n4 >	ATTTAGGAGACCCCTGAATGAAGGCCGCGCGGCTCGTCCGTAGTCTACTCGTGCCTAGATTATTTTC	nodM
	GGAGAGAACAATG	
n5 >	GATAGGAGACAAGCCGAAAGAATTCAAAGTGAGTGGGCGCGTGCCTAAAGACGCGCGGCATAGTCAAG	nodL
	GCGATGCAATAGTCAACATCGGCAACCCCTATCAACCATCAAAAGACGTTCTGAGAGGAGACGATG	
n6 >	CTGAAGTTAACGAAAACAAGCTTCAGCGAAGGATCTTTCCCTCATGACCTCGCTGGACCGACATTTCAAT	
	CCCTCATTCATCCATCGATCGCCGACCGCGGATTTTCCTGCCAAGTCAAGCGCGGAGCCAAAGACA	
	GGCGTCAATGCTTTTTTCAGTTACTTGCTCAGGGCTCGCTTGAGCTTACCGCCACGCGGTTGCGAAC	
	CCCCCTTGCCGCAAAACCGGATCTGACGCAGAGCTGCCACCTTGTCGTTTCATATGATCATTTTCCAATCT	
	ACCGCTGAAAGAAAGCTATPAGTATTGTAATTAGTCACTCCTGCCACCTGGCATGACTTTGGCTTGATTG	
	CGGGGTGAGAAATTTCCAGCAATGCCTTTCCCTAGCCGGCTTTGCGACCAGCCAGCATGCTCTGACTC	
	CCTACGTAGTTGCACCATTTGCGTTGCAAGACCTGCGTCGGCTGATGCATTCGCCCAAGGTTGGACCGGG	
	CACTGACAGCAAGAAGATTCCGCTCGATCCACCGTCCACCAGCTTGGCACATAGTTTGATG	noiQa ?
n7 >	ATTACAAGAACGTTAGATATAACCGGAGGGCAGAAGTGGGAGCGAGATCAATCCGACAGACTCCTTTG	
	CAATGAAAGCATGATTTGCGACCAAGGCCCTCAGATGTGTTGAGGCAGTTTCTTCAGTAAGTTTCAGGCG	
	GGGCCACTTTGGCGAGATCATTTTGTGGCTTGCTCGTTGGCACTGAAATATGATGTGAGCCGCAATTC	
	TGAAAGCATTTCTGACAGCGGGCTTTCGACATCCATCAGGCTAGGCTGCGAGGCCGCGGGGGCGTCA	
	CAGCCAAAGTCACTTCGATTAACGGAATAACGAGCGCTTCCCATTTGTGCGCGGGCCAGGAAAGTCTTTG	
	GCACGCACACCCATATTGTGACACCTGGATCTTCGGGATGTGGCGGGGAACATTACCTGCGGAGGAGA	
	ATAGATCCATG	syrM ?
n8 >	CGATGCGATCAGTTTCTGCTGCCCTGACCGAAGTCTTCACTAAACGGTTGGAGCCACGCAATCTTGAG	fixC ?
	AAGTTTCTGCACTGAGACAAGGACCAATG	

Figura 7. Similitud de secuencia nucleotídica entre las 8 copias tipo *nod*-box encontradas. Se muestra el consenso para *nod*-box y la distancia que separa a cada una de ellas del primer gen. Al menos las primeras 5 copias (n1-n5) son funcionales. No existe evidencia de funcionalidad para las copias n6, n7 y n8 (ver texto). Se indican con letras minúsculas los sitios de inicio de transcripción conocidos.



**Figura 8.** FER B: Relación de distancia por similitud entre las secuencias nucleotídicas *nod box n1 - n8*. Los números indican el porcentaje de diferencia entre una secuencia y la(s) otra(s). Es evidente que las secuencias más parecidas son la *n1 (nodA)* y *n6 (nolQa?)*. Asimismo, las posible *n7* y *n8* son las que menos se parecen a las otras.



**Figura 9.** FER A: Relación de distancia por similitud entre las secuencias nucleotídicas de los elementos regulatorios P1, P2 y P3 encontrados en la secuencia de *Rhizobium meliloti*. Los números indican el porcentaje de diferencia entre una secuencia y la(s) otra(s).

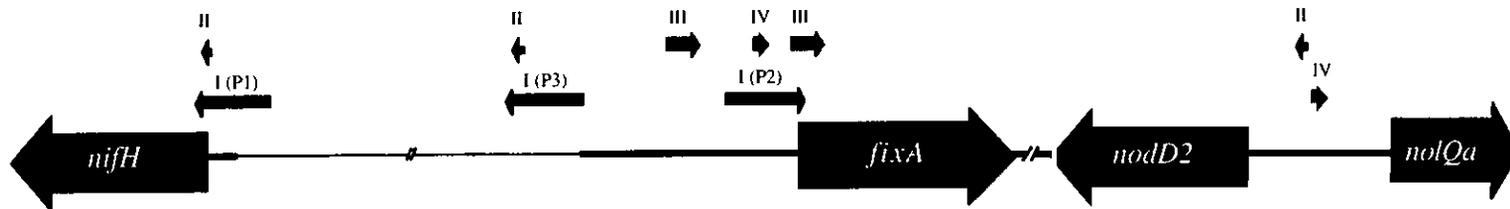
Se sabe que la región entre los promotores P1 y P2 es transcrita específicamente en nódulos radiculares, aunque al parecer no es del todo indispensable para la fijación de nitrógeno (**Better et al., 1983** y refs. en ésta). Se piensa que algún producto protéico codificado por esta región no es esencial pero puede tener cierta importancia en la simbiosis a nivel de algún tipo de ARNm regulador.

**FER D: Regiones Promotoras Reps II.** Las primeras dos secuencias encontradas en esta familia sobrelapan con las regiones promotoras P1 y P3 (**Figura 10**) descritas anteriormente. Sin embargo un tercer elemento se encontró entre las regiones codificantes de los genes *nodD2* y *nolQa*. Las tres secuencias se encuentran en orientación directa y se agrupan como un familia distinta a la anterior debido principalmente a que no se encontró alineamiento con la región promotora P2, además de que su grado de extensión fue menor a cualquier miembro de la familia C. La relación entre estas tres secuencias se puede ver en la **Figura 11**.

**FER E: Regiones Promotoras Reps III: Región -5' *fixA*.** El operón *fixABCX* contiene una región promotora que es homóloga a otros promotores controlados por *nifA*. En esta región promotora se ha encontrado previamente una duplicación hacia el extremo -5' no codificante de *fixA*. La primera copia de esta duplicación se encuentra separada a unas 370 pb de la segunda copia y ésta a su vez sobrelapa con los primeros 32 codones de *fixA*. Asimismo, una pequeña región de este segundo elemento sobrelapa en unas 100 pb con el extremo -3' de la región promotora P2 (ver **Figura 10**).

Sin embargo, un estudio más fino de esta región indica que el promotor *fixA* no está duplicado en sí mismo. Se ha pensado que los promotores de *nifH* y *fixA* -originalmente adyacentes- se separaron mediante una duplicación de la región -5' de ambos operones (**Earl et al., 1987**). Como se puede ver en la **Figura 10**, entre P3 y el primero de estos elementos repetidos, Reps III (extremo -5' de *fixA*) hay unas 500 pb aproximadamente.

**FER F: Regiones Intergénicas Reps IV: *nifH-fixA* y *nodD2-nolQa*.** Esta familia está integrada por dos elementos reiterados en orientación directa que se encuentran separados entre sí por 35.3 kb. El primero de los dos elementos está sobrelapado con la región promotora P2 pero no se alinea con ninguno de los otros dos elementos de la familia C. El segundo elemento se encuentra en la región intergénica *nodD2-nolQa*. De la misma manera como se definió a los elementos de la familia D, ésta familia no se considera como parte de la familia C (P1-P2-P3) porque el consenso no se encuentra sobrelapando más que a P2 y no a P1 o P3 (ver **Figura 10**).

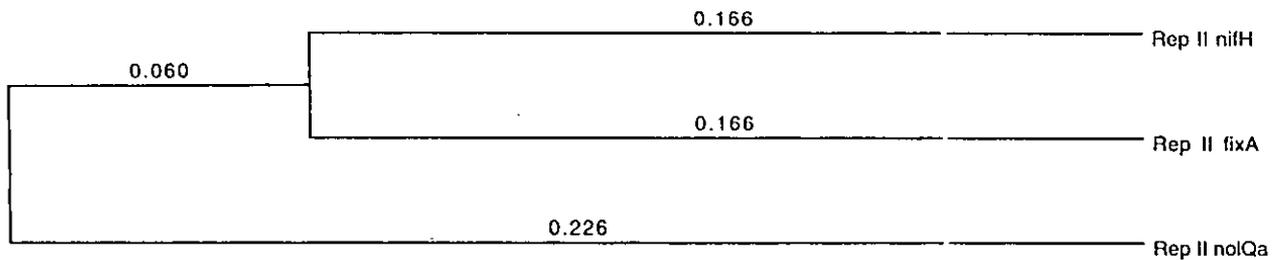


35

- FER C: Repeats I:** 163 id/340 bp (3 elementos) **47.9%**
- FER D: Repeats II:** 56 id/76 bp (3 elementos) **73.7%**
- FER E: Repeats III:** 128 id/160 bp (2 elementos) **80.0%**
- FER F: Repeats IV:** 63 id/81 bp (2 elementos) **77.8%**

1 Kb  
100 pb

**Figura 10.** Organización genética de los elementos repetidos Reps I a Reps IV en la región *nifH-fixA* y *nodD2-nolQa* de la región de nodulación de *R. meliloti*. Los genes se representan como flechas grandes, las reiteraciones se representan con flechas pequeñas. El eje grueso en este mapa corresponde a la secuencia obtenida por GenBank, el eje delgado indica la región secuenciada en laboratorio, cubre 1 400 pb en el intervalo entre los genes *nifH* y *fixA*. Se muestra a escala.



**Figura 11. FER A:** Relación de distancia por similitud entre las secuencias nucleotídicas de los tres elementos Reps II que integran la FER D. Los números indican el porcentaje de diferencia entre una secuencia y la(s) otra(s).

**FER G: Regiones Intragénicas Repts V: *fixB* y *nifA*.** Los tres elementos de esta familia se encuentran en las regiones codificantes de los genes *fixB* y *nifA*. Están en orientación directa y su longitud conservada de máxima extensión es de 113 pb. A partir de los elementos de esta familia, la mayoría de las secuencias reiteradas se encuentran en regiones codificantes y considerando esto se pensó en la posibilidad de que pudiesen ser parte de motivos protéicos una vez que la secuencia de ADN fuera traducida. Por ello, a partir de aquí el resto de las secuencias analizadas dentro de la región de nodulación de *R. meliloti* que se encontraron en regiones codificantes fueron comparadas con un banco de "perfiles protéicos" para encontrar motivos estructurales de secuencia polipeptídica. Con el empleo del programa Profilescan (GCG) se inició una búsqueda de motivos protéicos para cada secuencia intragénica. Sin embargo, sólo en ésta familia (familia G) se encontró un alineamiento (**Figura 12**).

Este motivo protéico se encuentra en la familia de las metaloproteínas, como lo es el caso del zimógeno de la adamalisina II del veneno de *Crotalus adamanteus* (*Viperidae*) (**Grams et al., 1993**). Como se ve en el alineamiento encontrado en Profilescan, el grado de identidad a nivel polipeptídico que muestra NifA es alto con respecto al motivo "Cystein-switch" de las metaloproteínas. Se ha propuesto que este motivo interviene en la activación de las metaloproteínas de matriz celular (**Springman et al., 1990**) y el mecanismo por el que lo hace es conocido como "Cystein-switch". Las metaloproteínas de matriz celular (MMPs) son capaces de metabolizar los componentes principales de la matriz extracelular.

	Perfil: cysteine_switch.prf	alineamiento: 1
NifA	3	TVPKAAIDQVMTAGRLVVPDVCNSELFRKDQIKWRG 37
		: . . . . :   :  .    . : .:   .
Cys-Switch	8	GKLDSDTLEVMRKPRCGVPDVGDFMVPFGKPKWVN 42

**Figura 12.** Alineamiento entre el motivo intragénico encontrado en NifA y un motivo conocido como Cysteine-Switch de ciertas metaloproteínas (*ver Texto*).

Con el fin de encontrar alguna interpretación estructural de los posibles motivos protéicos encontrados a partir de los elementos de la familia G y hasta los de la familia Q, cada secuencia de ADN que fue encontrada formando parte de una región codificante (es decir, intragénica) fue traducida a su versión de secuencia peptídica mediante **Translate** (GCG), respetando el marco de lectura. Cada secuencia peptídica fue analizada por separado. Esto se logró mediante el análisis de cada secuencia con los programas Peptidestructure-Plotstructure

(GCG).

Para el caso de la familia G el primer elemento reiterado (*fixB1*), una vez traducido correctamente, parece adoptar la estructura [H<sub>2</sub>N- $\alpha$ - $\beta$ -COOH]; el segundo elemento reiterado (*fixB2*) adquiere la estructura [H<sub>2</sub>N- $\alpha$ -COOH] y el tercer elemento (*nifA*), cuya secuencia es parecida al motivo de activación "Cystein-switch" (ver arriba) se pliega con mayor facilidad para adoptar la estructura [H<sub>2</sub>N- $\alpha$ - $\beta$ -giro- $\alpha$ -COOH]. Estas predicciones se basan en el algoritmo diseñado por **Chou-Fasman** (1978).

Cada uno de los elementos intragénicos de las siguientes 10 familias (Fam. H-Q) fueron analizadas de la misma forma ya descrita para el caso de los elementos de la familia G. Sin embargo, ninguna de las secuencias intragénicas de estas familias mostraron alineamientos en el banco de Profilescan. El análisis estructural de cada uno de los elementos reiterados de estas familias, una vez traducidos correctamente y mediante el uso de los programas Peptidestructure-Plotstructure se muestran en la **Tabla 4**.

**Búsqueda Primaria de Reiteraciones Grandes en el pNGR234a de *Rhizobium* sp. NGR234.** La secuencia total del pNGR234a mide 536 165 pb y el análisis de esta secuencia se llevó a cabo mediante el programa **rep** (Molpac), diseñado para secuencias genómicas grandes. Para encontrar las reiteraciones de mayor interés en esta secuencia se planteó una búsqueda de aquellas secuencias que fuesen idénticas (100%) y cuya longitud total fuera superior a 1 000 pb. Teóricamente estas características son óptimas para favorecer fenómenos de recombinación homóloga dependientes de RecA. Los parámetros del programa **rep** que fueron utilizados para la búsqueda de estas reiteraciones son buscar secuencias base (*core block sequences*) sin error de concatenación (*error block* 0). Tampoco ningún error de alineamiento entre una repetición y cualquier otra (*alignment error* 0). Los niveles de impresión (*printing levels*) solicitados fueron 1 000 pb. Con base en estos parámetros se encontraron 13 reiteraciones de interés en orientación directa con longitudes de entre 1 600 y 4 300 pb (**Tabla 5**). En el **Apéndice A-1** se incluyen el resto de las repeticiones pequeñas encontradas en este plásmido.

**Extensión de las Reiteraciones Grandes Encontradas en el pNGR234a de *Rhizobium* sp. NGR234a.** Las 13 reiteraciones encontradas fueron agrupadas en 6 familias diferentes de secuencias repetidas idénticas entre sí, superiores a 1 000 pb (**Tabla 5**). Con el objetivo de buscar el máximo de extensión para estas secuencias, cada una de ellas se analizó por separado con FastA (GCG). Los resultados de la extensión de cada secuencia se observan en la **Tabla 6**.

**Tabla 4**  
**Estructuras secundarias de cada elemento intragénico en las FERs H-Q,**  
**una vez traducido.**

FER (Elementos reiterados)	Estructura Secundaria <sup>a</sup>
H. Regiones Intragénicas <i>nodM</i> Elemento 1: <i>nodM1</i> Elemento 2: <i>nodM2</i>	[H <sub>2</sub> N-α-giro-α-COOH] [H <sub>2</sub> N-β-giro-COOH]
I. Regiones Intragénicas <i>nodM</i> y <i>fixC</i> Elemento 1: <i>nodM</i> Elemento 2: <i>fixC</i>	[H <sub>2</sub> N-giro-β-giro-COOH] [H <sub>2</sub> N-β-α-giro-COOH]
J. Regiones Intragénicas <i>nodM</i> y <i>nifN</i> Elemento 1: <i>nodM</i> Elemento 2: <i>nifN</i>	[H <sub>2</sub> N-α-giro-COOH] [H <sub>2</sub> N-giro-β-giro-COOH]
K. Región Intragénica <i>nodH</i> y <i>nifH-5'</i> Elemento 1: <i>nodH</i>	[H <sub>2</sub> N-giro-α-giro-β-COOH]
L. Regiones Intragénicas <i>nifN</i> y <i>syrM</i> Elemento 1: <i>nifN</i> Elemento 2: <i>syrM</i>	[H <sub>2</sub> N-α-β-giro-COOH] [H <sub>2</sub> N-α-COOH]
M. Regiones Intragénicas <i>nodD1</i> y <i>noeA</i> Elemento 1: <i>nodD1</i> Elemento 2: <i>noeA</i>	[H <sub>2</sub> N-α-β-COOH] [H <sub>2</sub> N-α-giro-COOH]
N. Región Intergénica <i>nodQ1-3'</i> e intragénica <i>nodD3</i> Elemento 2: <i>nodD3</i>	[H <sub>2</sub> N-α-β-COOH]
O. Regiones Intragénicas <i>nodQ1</i> y op. <i>stach.</i> Elemento 1: <i>nodQ1</i> Elemento 2: op. <i>stach.</i>	[H <sub>2</sub> N-giro-β-COOH] [H <sub>2</sub> N-α-β-COOH]
P. Regiones Intragénicas <i>nodQ1</i> y op. <i>stach.</i> <sup>b</sup> Elemento 1: <i>nodH</i> Elemento 2: op. <i>stach.</i>	[H <sub>2</sub> N-α-COOH] [H <sub>2</sub> N-α-COOH]
Q. Región Intergénica <i>nifH-5'</i> e intragénica <i>fixC</i> Elemento 2: <i>fixC</i>	[H <sub>2</sub> N-giro-COOH]

<sup>a</sup> Según el algoritmo de Chou-Fasman (1978).

<sup>b</sup> Esta familia es la única que presenta una estructura secundaria bastante bien conservada entre ambos elementos. Las otras secuencias no mantienen dicha identidad hasta el nivel de aminoácidos.

**Tabla 5***Extensión máxima de identidad para las reiteraciones del pNGR234a.*

Repetición <sup>a</sup>	Extensión máxima (100 %)
I. <i>nifHDK1/nifHDK2</i>	4 246 pb
II. <i>NGRIS-4a/4b</i>	3 325 pb
III. <i>NGRIS-3b/3c</i>	2 624 pb
IV. <i>NGRIS-5a/5b/5c</i>	2 642 pb (5a/5b) 2 643 pb (5a/5c) 2 644 pb (5b/5c)
V. <i>NGRIS-2a/2b</i>	2 561 pb
VI. <i>y4sH/y4uA</i>	1 558 pb

<sup>a</sup> Se incluyen los nombres de cada secuencia, correspondiente a la nomenclatura RIS/NGRIS o genes conocidos (Freiberg et al. 1997).

**Tabla 6***Extensión mediante el algoritmo FastA de cada reiteración encontrada en el pNGR234a.*

Repetición	Extensión original (100 %)	Máxima extensión con FastA
I. <i>nifHDK1/nifHDK2</i>	4 246 pb	4 409 pb
II. <i>NGRIS-4a/4b</i>	3 325 pb	3 456 pb
III. <i>NGRIS-3b/3c</i>	2 624 pb	2 685 pb
IV. <i>NGRIS-5a/5b/5c</i>	2 642 pb 2 643 pb 2 644 pb	2 642 pb (5a/5b) 2 679 pb (5a/5c) 2 705 pb (5b/5c)
V. <i>NGRIS-2a/2b</i>	2 561 pb	2 564 pb
VI. <i>y4sH/y4uA</i>	1 558 pb	1 604 pb

## ***Rhizobium* sp. NGR234: Familias de Elementos Reiterados (FERs).**

Con base en la búsqueda primaria de secuencias reiteradas grandes y de alta identidad dentro del pNGR234a de *Rhizobium* sp. NGR234 se proponen 6 FERs. Estas familias son las que presentaron los más altos puntajes en toda la secuencia plasmídica. A continuación se describen las características generales de cada una de ellas, empezando por las que son más extensas.

**FER I: Operones *nifHDK1* y *nifHDK2*.** Esta familia está representada por los operones funcionales *nifHDK1* y *nifHDK2*. Los genes *nifH1* y *nifH2* codifican para el componente II de la nitrogenasa (proteína-Fe). Los genes *nifD1* y *nifD2* codifican para la cadena alfa del componente I de la nitrogenasa (complejo protéico Mo-Fe) y los genes *nifK1* y *nifK2* codifican para la cadena beta del mismo componente I de la nitrogenasa (Freiberg *et al.* 1997). El operón *nifHDK1* también comprende los genes *nifENX* para la síntesis del cofactor Mo-Fe pero estos no se encuentran repetidos en *nifHDK2*.

**FER II: Secuencias de inserción NGRIS4a y NGRIS4b.** Esta familia está representada por dos elementos reiterados que corresponden a las secuencias de inserción NGRIS4a y NGRIS4b. La primera de ellas, NGRIS4a contiene los genes *y4bABCD*; la segunda, NGRIS4b incluye los genes *y4pHIJK*. Cada uno de estos genes son idénticos. Los genes *y4bD* y *y4pK* codifican hipotéticamente para una proteína de 10.2 kD. Esta posible proteína es homóloga a los reguladores transcripcionales Ros y MucR. Los genes *y4bCBA* y *y4pJIH* codifican hipotéticamente para una proteína de 16.8 kD. Para estas posibles proteínas no se ha reportado homología en las bases de datos y por tanto su función es desconocida (Freiberg *et al.* 1997). Sin embargo se ha calculado que las posibles proteínas Y4bB y Y4pI tendrían una masa molecular de 9.7 kD. Las proteínas Y4bA y Y4pH tendrían una masa molecular de 78.7 kD. Se sabe que estas secuencias están repetidas también en el cromosoma.

**FER III: Secuencias de inserción NGRIS3b y NGRIS3c.** Esta familia está representada por dos elementos reiterados que corresponden a las secuencias de inserción NGRIS3b y NGRIS3c. La primera de ellas, NGRIS3b contiene los genes *y4kI* y *y4kJ*. El posible producto protéico Y4kI, tiene homología con proteínas de unión a ATP y es idéntica a las proteínas Y4bM y Y4tA. Y4kJ tiene homología con transposasas y es idéntica a Y4bL y Y4tB. Estas homologías se encuentran en las proteínas de *Pseudomonas fluorescens*. La segunda reiteración corresponde a la secuencia de inserción NGRIS3c que incluye los genes *y4tAB* (Freiberg *et al.* 1997). Además de estas dos copias existe una tercera (NGRIS3a) en orientación inversa y cerca del sitio *oriV*. En el cromosoma se encuentran repetidas estas secuencias.

**FER IV: Secuencias de inserción NGRIS5a, NGRIS5b y NGRIS5c.** Esta familia está representada por tres elementos reiterados que corresponden a las secuencias de inserción NGRIS5a, NGRIS5b y NGRIS4c. La primera de ellas, NGRIS5a contiene los genes *y4iQ* y *y4jA*. La posible proteína Y4iQ tiene homología con proteínas de unión a ATP de *Escherichia coli* y es idéntica a Y4nD y Y4sD. Y4jA es una posible transposasa (homología con *E. coli*) idéntica a Y4nE y Y4sE. La segunda reiteración corresponde a la secuencia de inserción NGRIS5b, que incluye los genes *y4nDE* y la tercera, NGRIS5c contiene los genes *y4sDE* (Freiberg *et al.* 1997). Estas secuencias también se encuentran repetidas en el cromosoma.

**FER V: Secuencias de inserción NGRIS2a y NGRIS2b.** Esta familia está compuesta por dos elementos reiterados que corresponden a las secuencias de inserción NGRIS2a y NGRIS2b. La secuencia de inserción NGRIS2a contiene los genes *y4pEFG*. El gen *y4pE* codifica para una posible proteína de 15.5 kD y es similar al extremo-N de la proteína RFRS9 (25 kD) de *R. fredii*. Esta proteína es idéntica a Y4sA (NGRIS2b). El segundo gen, *y4pF* codifica para una posible transposasa y es idéntica a Y4sB (NGRIS2b). Finalmente, el posible producto del gen *y4pG* es una proteína, Y4pG de 21.1 kD e idéntica a Y4sC (NGRIS2b); es similar al ORFA de la IS 1238 de *Azotobacter xylinum* (Freiberg *et al.* 1997). La segunda reiteración corresponde a la secuencia de inserción NGRIS2b que incluye los genes *y4sABC*. Las proteínas codificadas por estos genes posiblemente participen en funciones de integración y/o recombinación.

**FER VI: Genes *y4sH* y *y4uA*.** Esta familia está representada por las regiones codificantes de los genes *y4sH* y *y4uA*. Hipotéticamente, las proteínas Y4sH y Y4uA tienen una masa de 50.7 kD. Ambas son idénticas y muestran una elevada similitud con la proteína de encapsulación CapA de *Bacillus anthracis* (Freiberg *et al.* 1997).

**Amplicones Propuestos en la Región de Nodulación de *R. meliloti* y en el pNGR234a de *Rhizobium* sp. NGR234.** Los resultados presentados hasta aquí corresponden a dos diferentes tipos de análisis (ver arriba). Sin embargo, con las 17 FERs en *R. meliloti* y las 6 FERs para *Rhizobium* sp. NGR234 se proponen 6 y 8 amplicones respectivamente. Para *R. meliloti* el amplicón más probable es el sistema *nodD1-nodD2*: dos regiones repetidas directas de poco más de 1100 pb, separadas por unas 75.5 kb de distancia y 82.9% de similaridad entre ambas. Cada uno de los 6 amplicones propuestos se incluyen en la **Tabla 7**. Para cada posible amplicón, las secuencias reiteradas en orientación directa podrían ser puntos para recombinación homóloga dependiente de RecA. Cada amplicón propuesto corresponde a las secuencias reiteradas que encontramos con los valores más altos en relación a su grado de

identidad, longitud total y distancia de separación entre ambas.

Por otra parte, los resultados presentados en este trabajo en relación a la búsqueda de reiteraciones en el pSym de *Rhizobium* sp. NGR234 permiten establecer un panorama general de la posible estructura dinámica de este plásmido. Las secuencias repetidas encontradas aquí se obtuvieron con base en una búsqueda más estricta de identidad. Las características generales de cada amplicón propuesto están descritas en la **Tabla 8**.

Los datos anteriores sugieren la existencia de al menos 8 amplicones. Estas regiones tienen la potencialidad de participar en la generación de amplificaciones y deleciones en el pNGR234a. En la **Figura 13** presentamos un panorama global de la ubicación de las secuencias reiteradas en este plásmido y los 8 amplicones propuestos. Como se puede notar, cada amplicón propuesto no incluye el *oriV* (ver Figura 13).

**Regiones Repetidas en Orientación Inversa. Búsqueda de Secuencias Invertidas.** El programa Repeat (GCG) tiene el inconveniente de encontrar solamente secuencias repetidas en orientación directa. Para el análisis de las secuencias invertidas en particular, la estrategia seguida fue el empleo del programa FindPatterns (GCG), el cual encuentra patrones cortos de secuencia y los estandariza como consenso para ser buscados en el resto de la secuencia, en cualquiera de las dos cadenas de ADN. Este programa permite encontrar secuencias repetidas en orientación inversa siempre que haya al menos dos de ellas en orientación directa previamente identificadas. Utiliza tales secuencias como consenso, permitiendo pequeños errores.

De este modo, no es posible encontrar secuencias invertidas como parejas con el software GCG. No obstante, gracias a FindPatterns fueron encontradas las principales secuencias reiteradas en orientación inversa conocidas dentro de la región de nodulación en *R. meliloti*: los 3 genes *nodD* y *syRM*, así como las *nod* boxes y los elementos P1-P2-P3. Además, el empleo del programa **icrep** en el paquete Molpac (ver *Métodos*) permite encontrar aquellas secuencias invertidas que puedan presentarse en parejas o grupos.

En *R. meliloti*, de las secuencias encontradas mediante el uso del programa **icrep**, todas -salvo una- ya habían sido localizadas con el paquete GCG y corresponden a los genes *nodD1-2-3*, los elementos P1-2-3 y algunas *nod* boxes. La única secuencia que no fue encontrada más que con este programa (**icrep**) es un palíndromo perfecto intragénico al gen estructural *nodQ1* con la secuencia siguiente: 5'-AGCTGGTCTTCGAAGACCAGCT-3'. No se profundizó en el análisis de esta secuencia. Asimismo, tampoco se buscaron las reiteraciones en orientación invertida dentro del pNGR234a de *Rhizobium* sp.

**Tabla 7**

*Amplicones propuestos dentro la región de nodulación de R. meliloti. <sup>a</sup>*

	<b>Amplicón (Subfamilia)</b>	<b>Extensión (pb) <sup>b</sup></b>	<b>Identidad (%)</b>	<b>Distancia de separación (pb)</b>	<b>Genes incluidos</b>
I.	<i>nodD1/D2</i>	1 106	82.9	75 500	<i>nodABCQ1P1GEFHsyrM nodD3nifHfixABCX nifABfdxNnoeBANodL</i>
II.	P1-P3	319	78.1	7 930	¿?
III.	<i>nodD1/syrM</i>	226	57.5	16 700	<i>nodABCQ1P1GEFH</i>
IV.	<i>syrM/nodD2</i>	202	57.9	59 850	<i>nodD3nifHfixABCX nifABfdxNnoeBANodL</i>
v.	familia B	160	80.0	300	Ninguno
VI.	G-2 ( <i>fixB/nifA</i> )	110	55.5	2 365	<i>fixB*CX-nifA*</i>

<sup>a</sup> Extensión, grado de identidad y distancias de separación para las secuencias repetidas en orientación directa que limitan cada amplicón.

<sup>b</sup> Máxima extensión con FastA.

**Tabla 8**

*Amplicones propuestos dentro del pNGR234a de Rhizobium sp. NGR234.*

	<b>Amplicón (DRs) <sup>a</sup></b>	<b>Extensión (pb)</b>	<b>Identidad <sup>b</sup> (%)</b>	<b>Distancia de separación (pb)</b>	<b>Genes incluidos de función conocida <sup>c</sup></b>
I.	<i>nifHDK1/nifHDK2</i>	4 409	98.8	27 610	<i>nifENXfdxBnifWS</i>
II.	NGRIS-4a/4b	3 456	98.2	228 535	<i>syrM1gabDfixUnifZfdxN nifABfixABCXnifQdctA1 (Amp. I) nodD2nolWX nolBTUVhrcNQRSTUnolK noeHnodZnoeIUnodD1 nodSUotsABSyrM1gabD</i>
III.	NGRIS-3b/3c	2 685	98.8	180 440	<i>nodSU</i>
IV.	NGRIS-5a/5b	2 642	100.0	90 650	<i>nodSUotsABSyrM1</i>
	NGRIS-5a/5c	2 679	99.4	191 130	<i>otsABSyrM1</i>
	NGRIS-5b/5c	2 705	98.9	97 840	<i>syrM</i>
V.	NGRIS-2a/2b	2 564	99.9	54 680	<i>gabD</i>
VI.	<i>y4sH/y4uA</i> (CapA)	1 604	98.8	7 930	

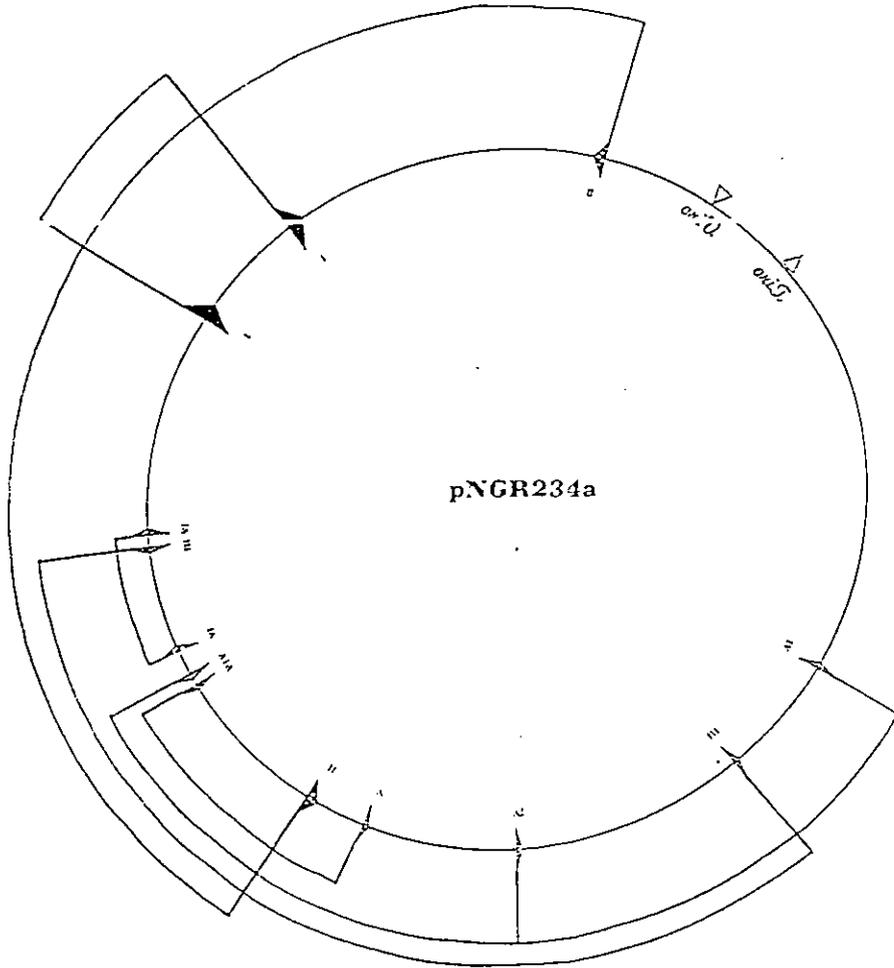
<sup>a</sup> Secuencias Directas Repetidas.

<sup>b</sup> Alineamiento tipo FastA (GCG).

<sup>c</sup> Funciones asignadas a los genes identificados: *dctA1*, transportador de C4-dicarboxilatos; *gabD*, succinato-semialdehído deshidrogenasa; *hrc*, sistema de secreción; *otsAB*, sistema trehalosa- fosfatasa sintasa. Información adicional:

<http://genome.imb-jena.de/~cfreiber/orfclassif.html>.

NGR234a. Esta clase de secuencias no participan en rearrreglos tipo amplificación y delección. Las secuencias repetidas en orientación inversa pueden participar en fenómenos de inversión de secuencias de ADN. Esta clase de fenómenos, aunque pueden ser un elemento muy importante de la dinámica de un genoma no se analizaron con profundidad en este estudio.



**Figura 13.** Estructura esquemática del pNGR234a de *Rhizobium* sp. NGR234. Se muestra la localización de las diferentes reiteraciones grandes, su extensión relativa y orientación. Estas reiteraciones fueron encontradas a partir del análisis global de la secuencia total reportada en el GenBank, recientemente depositada (<http://genome.imb-jena.de/~cfreiber/pNGR234a2.html>). Para cada reiteración se proponen los amplicones más probables con base en su tamaño (líneas externas al mapa). Cada uno de los 8 amplicones propuestos es el más pequeño posible (ver *Texto*) y todos ellos se ubican al lado opuesto de los sitios *oriV* y *oriT* (orígenes de duplicación y transferencia).

## DISCUSIÓN Y CONCLUSIONES

**Familias Multigénicas y Bancos de Información de Secuencia.** Las familias multigénicas son comunes en organismos eucariontes. Sin embargo, hay pocos ejemplos de reiteraciones genéticas bien caracterizadas en procariontes. Por ejemplo, en *R. etli* se ha caracterizado la familia multigénica de los genes *nifH* para la Nitrogenasa Reductasa (Quinto *et al.* 1985). En *E. coli* y *Salmonella typhimurium* (*S. enterica* serovar. Typhimurium) otros ejemplos de familias multigénicas son los operones *rrn* para ARNr y las secuencias *rhs*, de función desconocida (Anderson y Roth, 1977; Petes y Hill, 1988). En este estudio hemos presentado los resultados del análisis de secuencia de las regiones simbióticas en dos especies de *Rhizobium*, con el objetivo de encontrar y describir aquellas secuencias reiteradas que probablemente puedan participar en fenómenos recombinatorios y generación de rearrreglos genómicos tipo amplificación-delección. Las secuencias analizadas aquí corresponden a la región de nodulación de *R. meliloti* y a la secuencia completa del pSym de *Rhizobium* sp. NGR234 (pNGR234a). Las secuencias analizadas son de gran interés debido al antecedente de que en general los miembros de la familia *Rhizobiaceae* poseen una gran cantidad de secuencias reiteradas (para una revisión: Martínez *et al.* 1990). Por una parte, *R. meliloti* es la especie rhizobial más ampliamente estudiada hasta ahora y de la que se tiene mayor conocimiento de la organización y función de los genes de nodulación, los genes *nod*. Por otra parte, *Rhizobium* sp. NGR234 es la primera rhizobiácea cuyo pSym ha sido completamente secuenciado. Estas características, además de los antecedentes mencionados anteriormente hacen de estas dos especies, una interesante materia de análisis. Con los datos de secuencia disponibles, el principal objetivo de este estudio fue encontrar secuencias repetidas que puedan ser sustrato de recombinación para la generación de rearrreglos genómicos del tipo amplificaciones y delecciones.

Según el GenBank, en *R. meliloti* se han secuenciado unas 220 000 pb (220 kb) de manera separada, respondiendo a los intereses particulares de diferentes grupos de investigación en el mundo. Recientemente se ha manifestado un interés por la secuenciación total del genoma de esta especie (Romero y Palacios, *Ann. Rev. Genet.*: en prensa). El genoma completo de *R. meliloti* mide aproximadamente unas 6580 kb (6.58 Mb), de tal manera que se ha secuenciado por separado sólo un 3.35% del genoma total de esta especie pero seguramente esta cifra sigue creciendo día con día.

El Archivo de Secuencia generado aquí, "Rmel\_Regulon\_47.1\_kb.txt" corresponde a una secuencia compuesta de 30 entradas reportadas en el GenBank y es representativa de un 0.71% del genoma total de *R. meliloti*. De la

misma forma, considerando sólo la comparación con el pSym-a (1340 kb), el porcentaje de secuencia analizado aquí es de un 3.5%. Por otra parte, como se mencionó anteriormente, la secuencia completa del plásmido simbiótico de *Rhizobium* sp. NGR234 (pNGR234a-536165 pb) ya ha sido reportada y puede ser consultada en los bancos de información en la dirección InterNet siguiente: (<http://genome.imb-jena.de/~cfreiber/pNGR234a2.html>). De la misma forma, se pueden consultar las secuencias de los genomas completos de otros microorganismos (<http://www.tigr.org/tdb/mdb/mdb.html>). Con estos antecedentes se ha abierto una nueva etapa en la investigación genómica que permitirá analizar genomas completos como un todo y hacer comparaciones entre especies diferentes.

**Fenómenos Recombinatorios en *Rhizobium*.** Las bacterias usualmente portan una sola copia de cada gen en su genoma, por ello se les ha considerado como organismos haploides ( $n$ ). Sin embargo, se ha reportado en diferentes especies bacterianas la presencia de secuencias de ADN duplicadas (**Anderson y Roth 1977**). Estas duplicaciones pueden ser una situación inestable a nivel genómico. Si se encuentran en orientación directa, tales regiones repetidas pueden recombinar entre sí dando como resultado la delección o la amplificación del fragmento de ADN en cuestión. Estos fenómenos de reorganización genómica requieren de mecanismos de recombinación legítima (ADN-ADN) entre las dos secuencias homólogas; por esta razón, tales fenómenos pueden ser muy frecuentes. En los genomas de *E. coli* y *S. typhimurium* por ejemplo, algunos rearrreglos cromosomales ocurren frecuentemente en el laboratorio. Las duplicaciones se pueden formar y segregarse frecuentemente y su valor selectivo puede permitirle a la célula adaptarse a condiciones de estrés (**Roth et al. 1996**). Algunos rearrreglos cromosomales ocurren frecuentemente y es posible que algunos de estos se mantengan de manera estable en la organización del genoma.

Las regiones repetidas grandes son puntos para recombinación homóloga y generación de rearrreglos genómicos. Por ejemplo, en *E. coli* y *S. typhimurium* los loci *rrn* son secuencias repetidas de 5 kb de homología en diversas orientaciones a lo largo del cromosoma y que participan en la generación de rearrreglos tipo delección, duplicación e inversión (**Petes y Hill 1988**). Por otra parte, en *S. typhimurium* las secuencias de inserción (IS's) y elementos REP pueden ser usados también como sitios recombinantes que generen duplicaciones (**Shyamala et al. 1990**). En *E. coli*, las secuencias *rhs* también son utilizadas en la generación de rearrreglos (**Lin et al. 1984**). Las frecuencias de recombinación entre las diferentes repeticiones en un genoma bacteriano dependen de la longitud de la repetición y el grado de similaridad de las secuencias (**Roth et al. 1996**). Se sabe que la formación de

duplicaciones como resultado de la recombinación entre secuencias repetidas grandes requieren de la proteína RecA (Petit *et al.* 1991). En cuanto a la generación de duplicaciones por recombinación entre secuencias repetidas pequeñas, los mecanismos recombinatorios que participan son diferentes a los dependientes de RecA y generalmente son referidos como fenómenos de recombinación RecA-independientes (Roth *et al.* 1996). Esta clase de fenómenos ocurren cuando se tienen secuencias directas repetidas cercanas entre sí (Bi y Liu 1994). Si la distancia que separa a las repeticiones se incrementa, se reduce drásticamente la frecuencia de recombinación. Este sistema independiente del producto del gen *recA* no necesita de una repetición muy extensa, siempre que las repetidas se encuentren cerca, preferentemente en *tandem*. De hecho, longitudes mayores a 100 pb en sistemas *tandem* tienen efectos contraproducentes en la recombinación independiente de *recA* (Bi y Liu 1996). Las frecuencias para este tipo de duplicaciones son entonces más bajas.

Se sabe que el genoma de *Rhizobium* contiene una gran cantidad de secuencias de ADN reiterado. Estas secuencias repetidas se presentan como operones completos, genes específicos, secuencias reguladoras y elementos IS, así como regiones aún no caracterizadas de ADN (Martínez *et al.*, 1990). En el genoma de *R. etli* CFN42 se ha inferido la presencia de unas 200 familias de ADN reiterado, de manera dispersa (Flores *et al.*, 1987). Como ya se dijo, una de estas familias está integrada por dos operones funcionales para la nitrogenasa (*nifHDK*) en orientación directa y separados por 120 kb. Cada elemento reiterado de esta familia mide 5 kb y un tercer elemento reiterado (*nifH*) está entre las dos copias de *nifHDK* y en orientación inversa a estos (Girard *et al.*, 1991). Se ha demostrado la recombinación homóloga entre estos dos elementos y se han descrito también fenómenos de amplificación y delección en esta cepa (Romero *et al.*, 1991). De la misma forma, se han descrito otros amplicones dentro del cromosoma, el pSym o plásmidos crípticos de *R. etli* CFN42 (Flores *et al.*, 1993). Los tamaños de estos amplicones varían entre 10 y más de 100 kb.

La amplificación genética probablemente es un fenómeno muy frecuente en los genomas de organismos procariotas. El número, ubicación y orientación de las secuencias de ADN reiterado determinan la "estructura tipo amplicón" de un genoma determinado (Romero y Palacios, *Ann. Rev. Genet.*: en prensa). Estas características están relacionadas con la capacidad de un organismo para adaptarse e interactuar con otros. El amplicón descrito en *R. etli* es funcional y se ha demostrado que esta estructura es capaz de recombinar mediante mecanismos dependientes de *recA* y duplicar la región de 120 kb a una frecuencia de  $10^{-3}$  (Romero *et al.*, 1991). En esta misma cepa (CFN42) se han

descubierto al menos otros tres amplicones que sobrelapan con el primero descrito (120 kb). Los otros amplicones son grandes también (90-175 kb) y para ser funcionales requieren la actividad del gen *recA* (Romero *et al.*, 1995). Curiosamente estos amplicones están concentrados en una zona donde se encuentran varios genes simbióticos. En el pSym de *R. tropici* se ha encontrado recientemente otro amplicón simbiótico que cubre una región de 60 kb y donde se localizan diferentes genes simbióticos (P. Mavingui, datos no publicados). En este caso se ha demostrado también que los fenómenos amplificatorios que lleva a cabo esta estructura incrementa la producción de Factores Nod y altera la nodulación en diferentes hospederos.

La complejidad y flexibilidad del genoma bacteriano nos permite cuestionar la naturaleza de los rearrreglos genómicos y las consecuencias que puede tener una reorganización del material genético. De acuerdo con un modelo recientemente propuesto (Kolsto 1997), el orden genómico es mantenido mediante la adaptación de la expresión genética al contenido superhelicoidal local del cromosoma. El superenrollamiento está distribuido diferencialmente en dominios dentro del cromosoma y los niveles locales de superenrollamiento dentro de los dominios están influenciados por la expresión genética. De esta manera muchos promotores responden a los niveles de superenrollamiento dentro de cada dominio. Por tanto, la localización de los genes dentro del cromosoma debe tener gran importancia para la capacidad que tiene la célula de adaptarse a determinadas condiciones de crecimiento. La alteración del orden genómico puede tener efectos drásticos.

Las secuencias repetidas en orientación directa pueden recombinar para generar duplicaciones y deleciones. El caso de las inversiones es complejo e interesante. Cuando recombinan dos secuencias invertidas se espera invertir toda la secuencia que separaba ambas reiteraciones. Sin embargo, este tipo de alteraciones son las que probablemente ocurran a menor frecuencia. Se ha propuesto que en una molécula de ADN circular -cromosoma- ciertos lugares son "permisivos" para que se invierta la secuencia en cuestión pero existen otros lugares "prohibidos", donde una inversión puede resultar letal para la bacteria y donde preferencialmente no ocurren. Las principales limitaciones para que estos fenómenos de inversión se lleven a cabo son de naturaleza mecánica o funcional y se sabe que los intervalos permisivos y no permisivos para invertir no están distribuidos de manera azarosa, sino que muestran una distribución mas bien regional. En el cromosoma de *E. coli* los lugares no permisivos se encuentran principalmente cerca del *oriV* (Segall *et al.*, 1988).

**Reiteraciones en la Región de Nodulación de *R. meliloti*.** En la primera parte de nuestro estudio se describen aquellas secuencias reiteradas que pudieran ser biológicamente importantes en el contexto de su participación en fenómenos recombinatorios. Para el caso de *R. meliloti* las secuencias consideradas de interés fueron aquellas que se encontraron luego de una comparación de las secuencias repetidas entre la secuencia original, compuesta por 6 contínuos de secuencia y otra generada azarosamente pero de la misma composición nucleotídica. Para ambas secuencias se ejecutó el programa Repeat (GCG) y se buscaron secuencias repetidas de 50 y 100 pb como mínimo permitiendo márgenes de error de 2% sucesivamente. Los resultados obtenidos luego de esta búsqueda (**Figura 4**) permiten establecer las diferencias que se observan en cada caso. Para ambas búsquedas las secuencias repetidas encontradas en la secuencia *real* son más grandes y más numerosas que aquellas encontradas en la secuencia *azarosa*. Ambas curvas son diferentes. La curva representativa de la secuencia *real* muestra la misma tendencia que la curva representativa de la secuencia generada azarosamente para repeticiones con baja similaridad (50-60%) pero se separa ligeramente en esta región. Como se puede observar para cada caso (**Figuras 4a y 4b**) aún para estas secuencias de baja similitud existe una diferencia en el número de repeticiones encontradas, de tal manera que la curva de la secuencia *real* se separa de la curva *azarosa* en esta zona. Por otra parte, ambas curvas se separaran cuando las repeticiones son cada vez más similares entre sí (60-100%). Los puntos de inflexión donde las curvas *reales* se separan de las curvas *azarosas* son los puntos a partir de los cuales se consideraron importantes las secuencias repetidas y éstas se analizaron.

Las diferencias observadas en ambas curvas indican una clara tendencia de una secuencia biológica a tener repeticiones en mayor cantidad (incluso repeticiones pequeñas) en comparación a las repeticiones esperadas en cualquier secuencia generada azarosamente. Esta característica es de gran interés y una posible explicación a este fenómeno es que toda secuencia biológica está compuesta de cientos de señales para interactuar con otras moléculas (v. gr.: proteínas). La mayor parte de las repeticiones pequeñas que encontramos en la zona previa al punto de inflexión en la secuencia biológica podrían pertenecer a este grupo de secuencias-símbolo. Por citar sólo algunos ejemplos mencionaré los sitios para enzimas de restricción, sitios chi ( $\chi$ ), promotores  $\sigma^{54}$ , terminadores dependientes de rho ( $\rho$ ), etc. A nivel de ADN todos estos sitios señal tienen consensos específicos y en general esta clase de secuencias siempre son pequeñas. En las regiones codificantes la frecuencia en el uso de codones también es un factor que contribuye de manera importante a la cantidad de secuencias repetidas pequeñas en esta zona.

Los resultados presentados aquí para *R. meliloti* nos muestran 17 grupos de secuencias repetidas que hemos considerado como Familias de Secuencias Reiteradas y que según nuestra definición y parámetros de búsqueda resultan de interés. Estas secuencias pueden tener importancia en la generación de rearrreglos genómicos en dependencia de *recA*. El orden con que han sido presentadas en este estudio no responde al azar. En general los determinantes que hemos seguido aquí para su presentación son la extensión máxima que se obtuvo, el grado de similaridad entre cada una de las reiteraciones y el número de elementos que integran cada familia.

Es interesante hacer notar que en general las secuencias reiteradas encontradas en *R. meliloti* se hallan dispersas a lo largo de toda la secuencia analizada. En *R. meliloti* la estructura tipo amplicón más probable, está constituida por una región de unas 75.5 kb, limitada en ambos extremos por los genes funcionales *nodD1* y *nodD2*. Estos genes comparten una identidad de más del 82% en 1.1 kb y se encuentran en orientación directa (ver **Tabla 7**). Este posible amplicón, como los encontrados en *R. etli*, también contiene varios genes simbióticos pero aún falta demostrar que pueda ser funcional.

Dentro de la región de nodulación de *R. meliloti*, en segundo término y bajo esta misma perspectiva, al parecer los elementos de la familia C (P1-P2-P3) son los siguientes candidatos para recombinación dependiente de *RecA*. No podemos hacer predicciones reales de las frecuencias recombinatorias que se obtendrían en un sistema de esta naturaleza. Sin embargo, los argumentos presentados parecen suficientes para apoyar este razonamiento. Como tercer candidato, los elementos de la familia E parecen satisfacer también los requerimientos mencionados aquí para recombinación homóloga dependiente de *RecA*. Estos dos elementos están separados por unas 360 pb y entre ellos se encontró una similaridad del 80.0% en una longitud de 160 pb.

Los 6 amplicones propuestos para la región de nodulación de *R. meliloti* en la **Tabla 7**, corresponden al 20% del total de secuencias reiteradas encontradas aquí. El resto de las secuencias reiteradas que se encontraron tienen menos posibilidades de participar como puntos recombinatorios dependientes de *RecA* pero otros sistemas independientes de *RecA* pueden ser de gran importancia para generar rearrreglos genómicos en esta clase de secuencias. Estas secuencias podrían tener algún otro significado a nivel biológico. Es difícil especular sobre esta posibilidad pero en este estudio se trató de encontrar alguna explicación al respecto. La mayoría de las secuencias cortas encontradas corresponden a secuencias codificantes y con estos resultados se pensó en la posibilidad de que formasen parte de motivos estructurales a nivel protéico. De esta forma se efectuó una búsqueda que

pretendía encontrar algún alineamiento con motivos protéicos conocidos. Esta búsqueda sólo nos permitió encontrar un alineamiento para el caso del tercer elemento de la familia G (NifA) el cual tiene buena similitud con el motivo "Cystein-switch" de las metaloproteínas (Grams *et al.*, 1993). Se ha propuesto que este motivo participa en la activación de las metaloproteínas y en este caso en particular podría participar en la activación de NifA. Con el objetivo de encontrar una posible función para el resto de las secuencias cortas intragénicas se pensó en que pudieran adquirir estructuras secundarias compartidas que se conservaran desde la secuencia de ADN y hasta el nivel de secuencias peptídicas. Cada elemento se analizó con esta idea en mente aunque los resultados encontrados mediante el empleo del algoritmo de Chou-Fasman (1978) no parecen indicar tal conservación entre cada elemento a lo largo de las últimas 11 familias. Tal vez la única posibilidad sean los dos elementos de la familia P, la cual si se conservó bastante bien hasta el nivel peptídico (ver Tabla 4). Sin embargo, es necesario aclarar que la estructura secundaria de un polipéptido no depende exclusivamente de la secuencia de este. Características tales como el plegamiento de los residuos adyacentes puede ser de gran importancia, así como la interacción de cada proteína con otras. Las predicciones en base a este tipo de algoritmos deben tomarse con cierta precaución y tener siempre en mente que la estructura secundaria real depende de tales otros factores.

En otro orden de ideas, dentro de los elementos reiterados descritos para *R. meliloti*, el caso de las 8 copias de *nod* box es interesante. Previamente han sido identificadas en *R. meliloti* seis copias tipo *nod* box. Cada una de estas secuencias se encuentra en precediendo a un operón *nod*; la sexta copia n6, aparentemente inactiva, precede al operón *nodQaQbS* pero la expresión de este locus no responde a los inductores flavónicos que sí activan al resto de los genes *nod* (Plazanet *et al.*, 1995). En este estudio se encontraron al menos otros dos elementos tipo *nod* box en la secuencia analizada. La primera de ellas -n7- ya había sido reportada anteriormente (Barnett *et al.*, 1995) y se piensa que no fue localizada originalmente por Rostas y cols. (1986) debido a las condiciones de hibridización empleadas. La sonda empleada originalmente sólo es 64% similar a la posible n7. El papel funcional que ha sido propuesto para esta *nod* box es favorecer la unión de NodD3 al promotor de *syrM* e incrementar la expresión de este gen (Barnett *et al.*, 1995). Se sabe que el gen *syrM* está regulado por NodD3 y se sabe también que NodD3 reconoce secuencias tipo *nod* box mediante interacciones proteína-ADN. Algunos modelos pretenden explicar que NodD es una proteína de membrana que se une a los compuestos flavonoides y se activa *in situ* induciendo un cambio conformacional que le permite translocar el complejo hasta las *nod* box, donde se une como tetrámero (Kondorosi, 1992).

El caso de la posible **n8** es intrigante; aunque esta secuencia no es una *nod* box perfecta, se encuentra precediendo al gen *fixC* a una distancia que es comparable con la distancia entre **n4** y *nodM* (Figura 7). Asimismo, el análisis presentado en la Figura 8 nos muestra a las 8 copias de *nod* box agrupadas estrechamente entre sí. A decir del árbol presentado aquí (Figura 8), la **n7** guarda una relación más estrecha con las *nod* box **n3** y **n4** mientras que **n8** se asocia a su vez con estas tres últimas aunque es la más distante. La posible funcionalidad de esta octava copia está en duda debido principalmente a que se encuentra en una región codificante y anterior a un gen *fix*. Estos argumentos no apoyan su posible funcionalidad pero también hay que mencionar que en *Rhizobium* sp. NGR234, una copia *nod* box se encuentra precediendo al gen *fixF* (Freiberg *et al.*, 1997). Con este antecedente y con base en los razonamientos expuestos aquí resulta interesante manifestar el argumento de ampliar el concepto *nod* box: ¿Es este motivo exclusivo de genes *nod*?

Se sabe que las *nod* boxes controlan la transcripción de los genes *nod* (Kondorosi, 1992; Fisher y Long, 1993) pero sin duda es interesante pensar en alguna clase de regulación de los genes *nif/fix* (genes para fijación de N<sub>2</sub>) vía *nod* box. Se sabe que los genes *nif/fix* están controlados de manera compleja en circuitos regulatorios dependientes principalmente de NifA y promotores -24/-12 tipo  $\sigma^{54}$ . Esta clase de control se caracteriza por cascadas regulatorias de genes y operones organizados jerárquicamente (Para una revisión: Fischer, 1994). La acción concertada de los elementos regulatorios de estos circuitos integra la captación de señales ambientales óptimas para la fijación del N<sub>2</sub> y la respuesta molecular a éstas a nivel de la expresión genética. En este orden de ideas la *nod* box **n8** podría ser interpretada de dos maneras: i) se trata de una copia reminiscente de una *nod* box que antes era funcional y controlaba genes *nod* que ahora ya no existen por algún tipo de reorganización genómica o bien ii) es un elemento regulatorio adicional al sistema NifA- $\sigma^{54}$  (*nif* box) que responde a señales diferentes o incluso al mismo tipo de señales *nod* box-dependientes (NodD-flavonoide) para el control de los genes para fijación del N<sub>2</sub> (genes *nif/fix*).

Con base en nuestros parámetros de búsqueda iniciales los valores que tienen las *nod* box no cumplen con los valores mínimos de nuestra selección. Estas ocho copias representan las repeticiones más pequeñas encontradas (aún por debajo de 50 pb) y su inclusión como repeticiones importantes responde a que las pocas *nod* box encontradas en la búsqueda primaria (particularmente n1-n6) fueron la base para encontrar el resto de ellas, además del antecedente de su importante papel a nivel regulatorio (Schultze *et al.*, 1994).

Por otra parte, en *R. meliloti* Rm 2011 secuenciamos solamente 1400 pb correspondientes al intervalo de secuencia entre *nifH* y *fixA* (originalmente un *gap* de 7 kb). En esta secuencia encontramos el elemento reiterado P3 y el segundo de los elementos de la familia D (**Figura 10**). Como ya se mencionó, estas regiones reguladoras fueron ya previamente descritas (**Better et al., 1983**); sin embargo este grupo de investigación secuenció tales fragmentos de ADN por separado. En el presente estudio se localizó dicha secuencia de manera más precisa. Ya se tenía el mapa físico de las regiones P1-P2-P3 y ahora se integran en el mapa completo analizado aquí.

**Reiteraciones en el pNGR234a de *Rhizobium* sp. NGR234.** Para *Rhizobium* sp. NGR234, la secuencia analizada fue el plásmido simbiótico completo, pNGR234a. En la región de nodulación de *R. meliloti* se describen secuencias con baja similaridad entre sí y se proponen familias con base en estas secuencias; mientras que en el pSym de *Rhizobium* sp. NGR234 el nivel de análisis es menos profundo y sólo se consideraron aquellas reiteraciones grandes con identidad casi perfecta de secuencia (> 98.5%). Así se fundamenta aún más la potencialidad funcional de los amplicones propuestos en este caso. De esta manera, se proponen entonces dos diferentes grupos de amplicones con base en el grado de similaridad que presentan las secuencias reiteradas que se encontraron. Estas diferencias también deben considerarse en el contexto de la participación de cada secuencia reiterada en fenómenos recombinatorios. Evidentemente, las secuencias reiteradas grandes y de alta similaridad son las que más posibilidades tienen para ser puntos de recombinación homóloga. Por tanto los amplicones propuestos para el pNGR234a de *Rhizobium* sp. NGR234 tienen mayores posibilidades de ser funcionales. Dentro de la región de nodulación de *R. meliloti* sólo aquellas secuencias que cumplan tales parámetros de elevado grado de similitud y tamaño constituyen los amplicones funcionales más probables.

El análisis para el pNGR234a de *Rhizobium* sp. NGR234 indica la existencia de varios amplicones simbióticos potenciales. Los 8 amplicones propuestos en este caso son grandes (en un intervalo de 54.7 kb a 228.5 kb), están sobrelapados y cubren unas dos terceras partes del pNGR234a (**Figura 13**). Los amplicones propuestos son aquellos donde no se incluye el *oriV*. Cada uno de estos amplicones es el más pequeño posible; ya que en una molécula circular como el pSym, cada pareja de secuencias repetidas podría recombinar en ambos sentidos. Teóricamente la amplificación de un sector que contenga el *oriV* traería consecuencias serias para la estabilidad del plásmido y podría provocar un colapso durante su duplicación.

**Estrategias Experimentales para obtener Evidencia Funcional de Amplicones.** Una vez identificados los posibles amplicones resulta de gran interés diseñar estrategias experimentales que permitan seleccionar cepas portadoras de amplificaciones y/o deleciones. En general se propone seguir dos caminos: i) construcciones tradicionales tipo cassettes selectivos y/ o ii) PCR cuantitativo. El primero de ellos se refiere al uso de las técnicas clásicas de Biología Molecular que permitan insertar construcciones selectivas dentro de la estructura tipo amplicón y poder aislar así cepas que hayan amplificado o deletado ese material genético. Este tipo de selección puede lograrse insertando elementos genéticos tipo GDYN1 (Romero *et al.*, 1991) en regiones específicas. El elemento genético GDYN1 es un sistema capaz de seleccionar positivamente diferentes tipos de rearrreglos genómicos. Este cassette contiene un segmento del plásmido pSA, el cual porta los marcadores de resistencia para Km/Gm y Sm/Sp. Además porta los genes *sacR* y *sacB* de *B. subtilis*. El gen *sacR* confiere sensibilidad a sacarosa en bacterias gram-negativas y posibilita aislar cepas con deleciones específicas (Gay *et al.*, 1985). Introduciendo esta clase de sistemas genéticos en posibles amplicones se pueden seleccionar cepas amplificadas resistentes a elevados niveles de kanamicina en el medio; por otra parte, en medios ricos en sacarosa, se pueden seleccionar cepas que porten deleciones de este elemento y por tanto del posible amplicón en el que estaba clonado. Con este tipo de construcciones es posible seleccionar de manera positiva deleciones, amplificaciones y deamplificaciones variando las condiciones de selección. Esta estrategia experimental ya ha sido empleada para describir los amplicones descritos por Romero y colaboradores (1995).

El PCR cuantitativo que proponemos aquí podría ser otra estrategia experimental que puede permitir seleccionar posibles rearrreglos genómicos derivados de las estructuras tipo amplicón descritas en este estudio. El enfoque se basa en el conocimiento que se tiene acerca de la naturaleza de los rearrreglos genómicos. En rearrreglos del tipo amplificación y deleción las secuencias repetidas que recombinan generan una reorganización de la secuencia en los puntos de recombinación. Es posible diseñar oligonucleótidos específicos capaces de bordear los límites de cada amplicón propuesto y luego probar su funcionalidad obteniendo productos de PCR específicos que resulten del rearrreglo genómico correspondiente. Tales productos de PCR reflejan de manera indirecta este tipo de rearrreglos. La abundancia del producto de PCR es cuantificable y una extrapolación de las frecuencias a las que pueden ocurrir dichos rearrreglos.

**Perspectivas a Futuro.** En conclusión, este estudio propone la existencia de 8 amplicones en el pSym de *Rhizobium* sp. NGR234 (pNGR234a) y al menos otros 6 en la región de nodulación de *R. meliloti*. Sin embargo, estos amplicones

son de diferente naturaleza. Los primeros son potenciales amplicones con secuencias reiteradas más grandes y casi idénticas entre sí; los segundos poseen en general secuencias reiteradas más pequeñas y de menor grado de similitud. Es necesario mencionar que si se hubiera analizado la secuencia nucleotídica del pNGR234a con los mismos parámetros de búsqueda que se utilizaron para la secuencia analizada de *R. meliloti* seguramente se habrían encontrado otros posibles amplicones.

Estos resultados son alentadores en el sentido de buscar evidencia de la funcionalidad de cada uno de ellos -principalmente dentro del pNGR234a-. Para este objetivo se cuenta con la poderosa herramienta del análisis de secuencia, la tecnología del PCR y se conoce la naturaleza de los rearrreglos tipo amplificación-delección. Como ya se dijo, será posible diseñar oligonucleótidos específicos que limiten los puntos de recombinación homóloga para cada caso. Se podrán obtener productos de PCR precisos que correspondan al punto de unión de las reiteraciones (*join point*), pensando en el resultado que tiene una amplificación o delección a nivel de la generación secuencia nueva. Estos productos de PCR serán evidencia de la generación de tales rearrreglos. La cantidad de producto de PCR hará de la tecnología del PCR cuantitativo una herramienta que permitirá también establecer indirectamente las frecuencias de amplificación-delección de cada caso. En los **Apéndices A-2** y **A-3** se pueden observar las generalidades teóricas de estos métodos experimentales de PCR cuantitativo propuestos en este trabajo.

Los resultados presentados en este estudio son preliminares. Hasta aquí hemos podido definir de manera general qué tipo de análisis es de interés para cualquier secuencia nucleotídica procarionte en una concepción dinámica del genoma bacteriano. Estableciendo los parámetros de búsqueda de las principales FERs descritas aquí se cuenta con una poderosa herramienta de análisis que permitirá hacer comparaciones globales a nivel genómico y comparaciones entre especies diferentes. Los proyectos de secuenciación genómica completados recientemente y los que están en marcha, así como el mapeo físico de los genomas han abierto una nueva área de investigación donde este tipo de análisis permitirá comprender más profunda y globalmente estos genomas.

Dentro del CIFN el objetivo general de ésta línea de investigación (**Estructura, dinámica y manipulación del genoma de microorganismos fijadores de nitrógeno**) es demostrar un modelo dinámico del genoma de *Rhizobium* y definir su significado biológico. Este modelo pudiera ser aplicable al genoma bacteriano en general. Como ya se mencionó, el modelo establece que el genoma de *Rhizobium* está sujeto

contínuamente a eventos reversibles de amplificación génica que ocurren a frecuencias altas. Se propone que la amplificación génica aumenta la potencialidad del genoma sin alterar su estructura y permite a las cepas bacterianas adaptarse a condiciones ambientales cambiantes y a la interacción con diferentes plantas. Esto abre nuevos horizontes a la manipulación del genoma bacteriano y a la obtención de cepas con propiedades óptimas para fines específicos como el incremento en la competitividad bacteriana (Mavingui *et al.* 1997).

El tipo de estudio presentado aquí es integrativo. Propone un enfoque nuevo de análisis y es una herramienta adicional para el estudio de la dinámica del genoma bacteriano. A la par del empleo de esta herramienta se encuentran otras que son tradicionales como la construcción de cassettes genómicos para selección experimental de amplicones. Sin embargo, la información precisa de la secuencia nucleotídica y la predicción de amplicones puede contribuir al diseño de estrategias experimentales más dirigidas y precisas. Como perspectivas a futuro es interesante proponer el diseño y optimización de algoritmos específicos para su implementación en programas de simulación que permitan obtener en cortos tiempos las predicciones de rearrreglos genómicos de diferente naturaleza para un genoma determinado.



## Apéndice

### A-1 Número Total de Repeticiones encontradas en el pNGR234a.

Niveles de Impresión: 1000

Máximo error en la longitud del bloque: 0

Longitud Total de la Secuencia = 536165

---

\*\*\* Sequence # 1 ( RSU00090 ) Length = 536165 \*\*\*

---

#### No. de Segmentos de Identidad

Longitud del bloque básico	2	3	4	5	6 o más
12:	7070	1104	245	41	5
13:	2518	262	66	7	0
14:	700	81	17	0	1
15:	233	22	4	0	0
16:	65	4	2	0	0
17:	39	7	0	0	0
18:	12	2	0	0	0
19:	12	0	0	0	0
20:	11	2	0	0	0
21:	5	0	0	0	0
22:	2	2	0	0	0
23:	5	1	0	0	0
24:	3	1	0	0	0
25:	2	1	0	0	0
26:	4	1	0	0	0
28:	2	0	0	0	0
29:	3	2	0	0	0
30:	2	0	0	0	0
32:	0	1	0	0	0
33:	1	0	0	0	0
34:	1	0	0	0	0
35:	2	0	0	0	0
37:	1	0	0	0	0
38:	1	0	0	0	0
40:	0	1	0	0	0
42:	1	0	0	0	0
43:	0	1	0	0	0
44:	1	0	0	0	0
49:	3	0	0	0	0
53:	1	0	0	0	0
60:	1	0	0	0	0
69:	1	0	0	0	0
81:	1	0	0	0	0
>= 112:	7	1	0	0	0

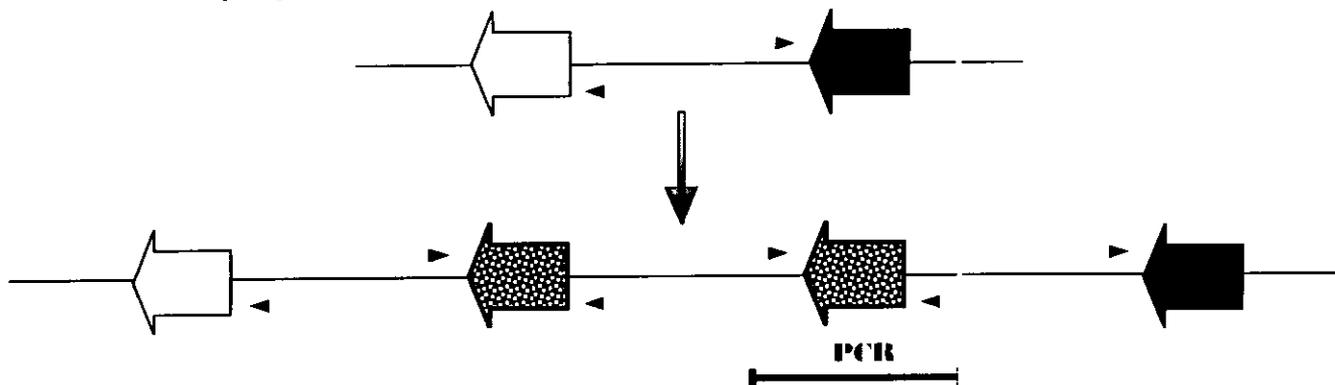
Número Total de bloques básicos = 13597

Número Total de Repeticiones Significativas = 6

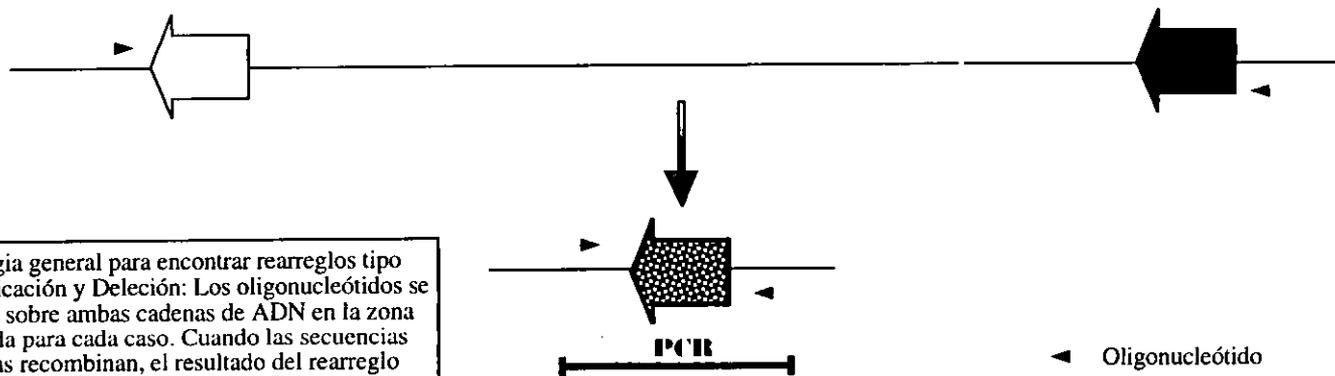
# Apéndice

## A-2 PCR cuantitativo: Selección de Amplificaciones y Delecciones Genómicas.

### Amplificación



### Delección



◄ Oligonucleótido  
◄ Reiteraciones Directas

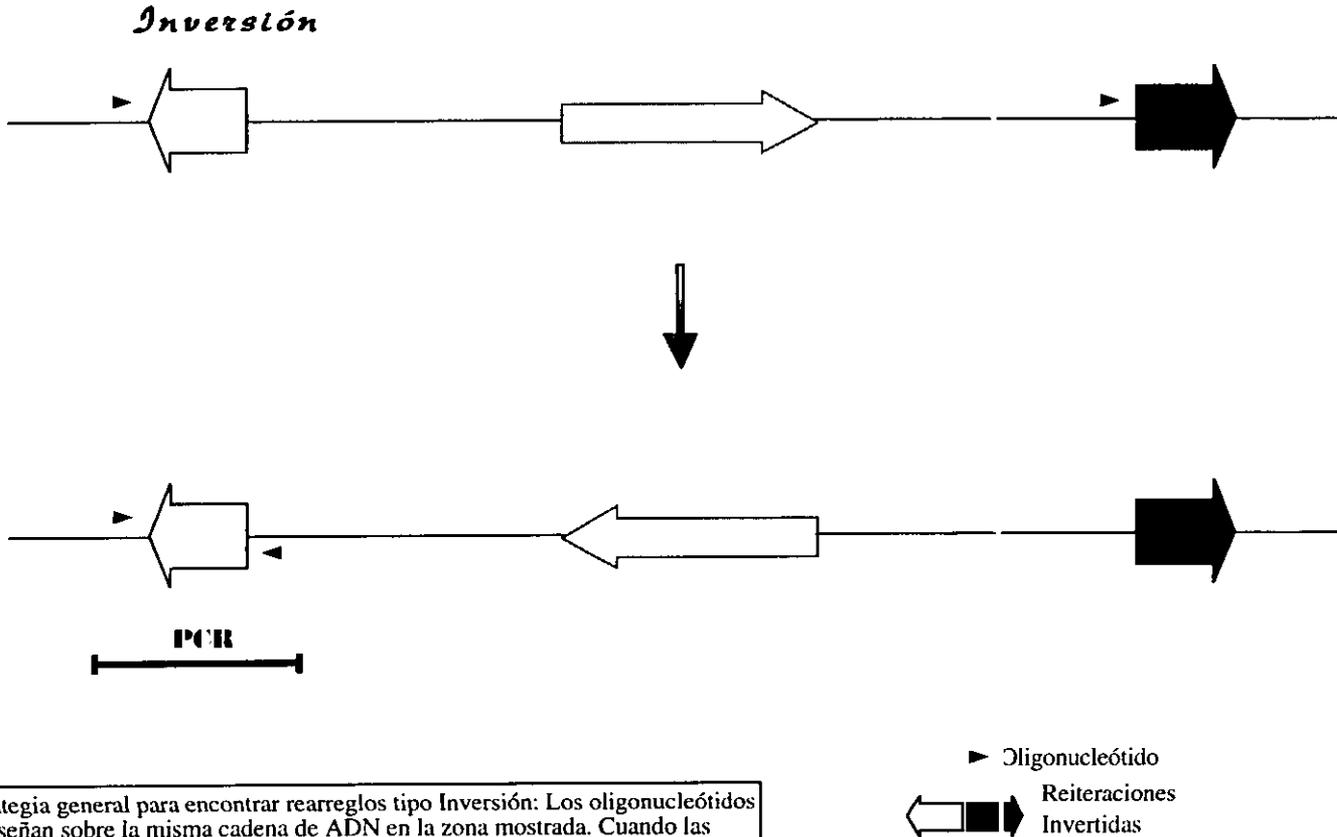
Estrategia general para encontrar rearruglos tipo Amplificación y Delección: Los oligonucleótidos se diseñan sobre ambas cadenas de ADN en la zona mostrada para cada caso. Cuando las secuencias repetidas recombinan, el resultado del rearruglo genómico genera fragmentos de ADN nuevos, haciendo posible la síntesis de productos de PCR específicos.

ESTR  
SALIR  
NO  
PERO  
ESTR  
SALIR

## Apéndice

### A-3 PCR cuantitativo: Selección de Inversiones Genómicas.

69



Estrategia general para encontrar rearrreglos tipo Inversión: Los oligonucleótidos se diseñan sobre la misma cadena de ADN en la zona mostrada. Cuando las secuencias repetidas recombinan se contraponen la región completa y ahora el segundo oligonucleótido se enfrenta al primero, haciendo posible la síntesis de un producto de PCR específico.

## LITERATURA CITADA

- Anderson, R. P. y Roth, J. R. 1977. Tandem Genetic Duplications in Phage and Bacteria. *Ann. Rev. Microbiol.* **31**: 473-505.
- Baev, N., Endre, G., Petrovics, G., Banfalvi, Z. y Kondorosi, A. 1991. Six Nodulation Genes of *nod* box Locus 4 in *Rhizobium meliloti* are Involved in Nodulation Signal Production: *nodM* Codes for D-glucosamine Synthetase. *Mol. Gen. Genet.* **228**: 113-124.
- Baev, N. y Kondorosi, A. 1992. Nucleotide Sequence of the *Rhizobium meliloti nodL* Gene Located in Locus n5 of the *nod* Regulon. *Plant Mol. Biol.* **18**: 843-846.
- Barnett, M. J. y Long, S. R. 1990. DNA Sequence and Translational Product of a New Nodulation-Regulatory Locus: *SyrM* Has Sequence Similarity to NodD Proteins. *J. Bacteriol.* **172**: 3695-3700.
- Barnett, M. J., Rushing, B. G., Fisher, R. F. y Long, S. R. 1995. Transcriptional Start Sites for *syrM* and *nodD3* Flank an Insertion Sequence Relic in *Rhizobium meliloti*. *J. Bacteriol.* **178**: 1782-1787.
- Better, M., Lewis, B., Corbin, D., Ditta, G. y Helinski, D. R. 1983. Structural Relationships among *Rhizobium meliloti* Symbiotic Promoters. *Cell.* **35**: 479-485.
- Bi, X. y Liu, L.F. 1994. *recA*-independent and *recA*-dependent Intramolecular Plasmid Recombination. *J. Mol. Biol.* **235**: 414-424.
- Bi, X. y Liu, L.F. 1996. A Replicational Model for DNA Recombination between Direct Repeats. *J. Mol. Biol.* **256**: 849-858.
- Brom, S., García de los S., A., Girard, Ma. de L., Dávila, G. y Romero, D. 1991. High-Frequency Rearrangements in *Rhizobium leguminosarum* bv. *phaseoli*. *J. Bacteriol.* **173**: 1344-1346.
- Chou, P. Y. y Fasman, G. D. 1978. Prediction of the Secondary Structure of Proteins from their Amino Acid Sequence. *Adv. in Enzymol.* **47**: 45-148.
- Dávila, G., Brom, S., Flores, M., Girard M. de L., Gonzalez, V., Louzada, M., Martínez, J., Martínez, E., Palacios, R., Piñero, D., Romero, D. y Valdéz, A. M. 1988. The Symbiotic Genome of *Rhizobium phaseoli*. In: *Molecular Plant-Microbe Interactions*, Verma, D.P.S. y Palacios, R., Eds., APS Press, St. Paul MN. 187-191.
- Dénarié, J., Debellé, F. y Rosenberg, C. 1992. Signaling and Host range Variation in Nodulation. *Annu. Rev. Microbiol.* **46**: 497-531.
- Dusha, I., Kovalenko, S., Banfalvi, Z. y Kondorosi, A. 1987. *Rhizobium meliloti*

- Insertion Element *ISRm2* and its use for Identification of the *fixX* gene.  
*J. Bacteriol.* **169**: 1403-1409.
- Dybvig, K.** 1993. DNA Rearrangements and Phenotypic Switching in Prokariotes.  
*Mol. Microbiol.* **10**: 465-471.
- Earl, C. D., Ronson, C. W. y Ausubel, F. M.** 1987. Genetic and Structural Analysis of the  
*Rhizobium meliloti fixA, fixB, fixC, and fixX* Genes. *J. Bacteriol.* **169**: 1127-1136.
- Fischer, H.M.** 1994. Genetic Regulation of Nitrogen Fixation in Rhizobia. *Microbiol. Rev.* **58**:  
 352-386.
- Fisher, R.F. y Long. S.R.** 1992. *Rhizobium*-plant Signal Exchange. *Nature.* **357**: 655-660.
- Fisher, R.F. y Long. S.R.** 1993. Interactions of NodD at the *nod* Box: NodD Binds to Two  
 Distinct Sites on the Same Face of the Helix and Induces a Bend in the DNA.  
*J. Mol. Biol.* **233**: 336-348.
- Flores, M., Brom, S., Stepkowski, T., Girard, M., Dávila, G., Romero, D. y Palacios, R.** 1993. Gene Amplification in *Rhizobium*: Identification and *in vivo* Cloning of Discrete Amplifiable DNA Regions (Amplicons) from *Rhizobium leguminosarum biovar phaseoli*. *Proc. Natl. Acad. Sci.* **90**: 4932-4936.
- Flores, M., González, V., Brom, S., Martínez, E. Piñero, D., Romero, D., Dávila, G. y Palacios, R.** 1987. Reiterated DNA Sequences in *Rhizobium* and *Agrobacterium* spp.  
*J. Bacteriol.* **169**: 5782-5788.
- Freiberg, C., Fellay, R., Bairoch, A., Broughton, W. J., Rosenthal, A. y Perret, X.**  
 1997. Molecular Basis of Symbiosis between *Rhizobium* and Legumes.  
*Nature.* **387**: 394-401.
- Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T. y Kado, C. I.** 1985. Positive Selection  
 Procedure for Entrapment of Insertion Sequence Elements in Gram-Negative Bacteria.  
*J. Bacteriol.* **164**: 918-921.
- Girard, M. de L., Flores, M., Brom, S., Romero, D., Palacios, R. y Dávila, G.** 1991.  
 Structural Complexity of the Symbiotic Plasmid of *Rhizobium leguminosarum* bv.  
*phaseoli*. *J. Bacteriol.* **173**: 2411-2419.
- Grams, F., Huber, R., Kress, L. F., Moroder, L. y Bode, W.** 1993. Activation of Snake  
 Venom Metalloproteinases by a Cysteine Switch-like Mechanism. *FEBS.* **335**: 76-80.
- Gribskov, M., Homyak, M., Edenfield, J. y Eisenberg, D.** 1988. Profile Scanning for  
 Three-Dimensional Structural Patterns in Protein Sequences. *CABIOS.* **4**: 61-66.
- Hahn, M. y Hennecke, H.** 1987. Mapping of a *Bradyrhizobium japonicum* DNA Region  
 Carrying Genes for Symbiosis and an Asymmetric Accumulation of Reiterated  
 Sequences. *Appl. Environ. Microbiol.* **53**: 2247-2252.

- Honma, M. A. y Ausubel, F. M. 1987. *Rhizobium meliloti* has Three Functional Copies of the *nodD* Symbiotic Regulatory Gene. *Proc. Natl. Acad. Sci.* **84**: 8558-8562.
- Honma, M.A., Asomaning, M. y Ausubel, F.M. 1990. *Rhizobium meliloti nodD* Genes Mediate Host-Specific Activation of *nodABC*. *J. Bacteriol.* **172**: 901-911.
- Kashanov, F. K., Zvingila, D. J., Zainullin, A. A., Prozorov, A. A. y Bashkirov, V. I. 1992. Homologous Recombination between Plasmid and Chromosomal DNA in *Bacillus subtilis* Requires Approximately 70 bp of Homology. *Mol. Gen. Genet.* **234**: 494-497.
- Kolstø, A-B. 1997. Dynamic Bacterial Genome Organization. *Mol. Microbiol.* **24**: 241-248.
- Kondorosi, A. 1992. Regulation of Nodulation Genes in Rhizobia. *In: Molecular Signals in Plant-Microbe Communications*. D.P.S. Verma (ed.). CRC Press. London.
- Leung, M.-Y., Blaisdell, B. E., Burge, C. y Karlin, S. 1991. An Efficient Algorithm for Identifying Matches with Errors in Multiple Long Molecular Sequences. *J. Mol. Biol.* **221**: 1367-1378.
- Lin, R.J., Capage, M. y Hill, C. W. 1984. A Repetitive DNA Sequence, Rhs. Responsible for Duplication within the *Escherichia coli* K12 Chromosome. *J. Mol. Biol.* **177**: 1-18.
- Martínez, E., Romero, D. y Palacios, R. 1990. The *Rhizobium* Genome. *Crit. Rev. Plant Sci.* **9**: 59-93.
- Masterson, R. V. y Atherly, A. G. 1986. The Presence of Repetead DNA Sequences and a Partial Restriction Map of the pSym of *Rhizobium fredii* USDA 193. *Plasmid.* **16**: 37-44.
- Mavingui, P., Flores, M., Romero, D., Martínez-R., E. y Palacios, R. 1997. Generation of *Rhizobium* Strains with Improved Symbiotic Properties by Random DNA Amplification (RDA). *Nature Biotech.* **15**: 564-569.
- Mavingui, P., Flores, M., Laeremans, T., Martínez-R., E. y Palacios, R. 1996. Identification of a *nod-nif* Amplicon in the Symbiotic Plasmid of *Rhizobium tropici* CFN299. 8th International Congress: Molecular Plant-Microbe Interactions. Knoxville, TN. July 14-19, 1996.
- Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press. USA, pp.: 293-298.
- Ogawa, J., Brierley, H.I. and Long, S.R. 1991. Analysis of *Rhizobium meliloti* Nodulation Mutant WL131: Novel Insertion Sequence IS $Rm3$  in *nodG* and Altered *nodH* Protein Product. *J. Bacteriol.* **173**: 3060-3065.
- Palacios, R., Flores, M., Hernández, G., Mavingui, P. y Romero, D. 1995. Dynamics of the *Rhizobium* Genome. *In: Nitrogen Fixation: Fundamentals and Applications*. I. A. Tikhonovich *et al.* (eds.). 353-357.

- Pearson, W. R. y Lipman, D. J. 1988. Improved Tools for Biological Sequence Comparison. *Proc. Natl. Acad. Sci.* **85**: 2444-2448.
- Pearson, W. R. 1990. Rapid and Sensitive Sequence Comparison with FASTP and FASTA. *Meth. Enzymol.* **183**: 63-98.
- Petes, T. D. y Hill, C. W. 1988. Recombination Between Repetead Genes in Microorganisms. *Annu. Rev. Genet.* **22**: 147-168.
- Petit, M. A., Dimpfl, J., Radman, M. y Echols, H. 1991. Control of Large Chromosomal Duplications in *Escherichia coli* by the Mismatch Repair System. *Genetics.* **129**: 327-332.
- Pfeifer, F. y Betlach, M. 1985. Genome Organization in *Halobacterium halobium*: a 70 kb Island of More (AT) Rich DNA in the Chromosome. *Mol. Gen. Genet.* **198**: 440-455
- Plazanet, C., Réfrégier, G., Demont, N., Truchet, G. y Rosenberg, C. 1995. The *Rhizobium meliloti* Region Located Downstream of the *nod* box n6 is Involved in the Specific Nodulation of *Medicago lupulina*. *FEMS Microbiol. Letters.* **133**: 285-291.
- Quinto, C., de la Vega, H., Flores, M., Fernández, L., Ballado, T., Soberón, G. y Palacios, R. 1982. Reiteration of Nitrogen Fixation Gene Sequences in *Rhizobium phaseoli*. *Nature.* **299**: 724-726.
- Quinto, C., de la Vega, H., Flores, M., Leemans, J., Cevallos, M. A., Pardo, M. A., Azpiroz, R., Girard, M. L., Calva, E. y Palacios, R. 1985. Nitrogenase Reductase: A functional Multigene Family in *Rhizobium phaseoli*. *Proc. Natl. Acad. Sci.* **82**: 1170-1174.
- Renalier, M. H., Batut, J., Ghai, J., Terzaghi, B., Gherardi, M., David, M., Garnerone, A. M., Vasse, J., Truchet, G., Huguet, T. y Boistard, P. 1987. A New Symbiotic Cluster on the pSym Megaplasmid of *Rhizobium meliloti* 2011 Carries a Functional *fix* Gene Repeat and a *nod* Locus. *J. Bacteriol.* **169**: 2231-2238.
- Romero, D., Brom, S., Martínez-Salazar, J., Girard, M., Palacios, R. y Dávila, G. 1991. Amplification and Deletion of a *nod-nif* Region in the Symbiotic Plasmid of *Rhizobium phaseoli*. *J. Bacteriol.* **173**: 2435-2441.
- Romero, D., Martínez-Salazar, J., Girard, M., Brom, S., Dávila, G., Palacios, R., Flores, M. y Rodríguez, C. 1995. Discrete Amplifiable Regions (Amplicons) in the Symbiotic Plasmid of *Rhizobium etli* CFN42. *J. Bacteriol.* **177**: 973-980.
- Rostas, K., Kondorosi, E., Horvarth, B., Simoncsits, A. y Kondorosi, A. 1986. Conservation of Extended Promoter Regions of Nodulation Genes in *Rhizobium*. *Proc. Natl. Acad. Sci.* **83**: 1757-1761.
- Roth, J. R., Benson, N., Galitski, T., Haack, K., Lawrence, J. G. y Miesel, L. 1996.

- Rearrangements of the Bacterial Chromosome: Formation and Applications. *En: Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2a. ed. Vol. 2. F. C. Neidhardt (Ed.). ASM Press. Washington, D. C. 2256-2276.
- Ruvkun, G.B., Long, S.R., Meade, H.M. van den Bos, R.C. and Ausubel, F.M. 1982. ISRm1: a *Rhizobium meliloti* Insertion Sequence that Transposes Preferentially into Nitrogen Fixation Genes. *J. Mol. Appl. Genet.* 1: 405-418.
- Sapienza, C. y Doolittle, W.F. 1982. Unusual Physical Organization of the *Halobacterium* Genome. *Nature.* 295: 384-389.
- SanJuan, J., Luka, S. y Stacey, G. 1992. Genetic maps of *Rhizobium* and *Bradyrhizobium* Species. *Genetic Maps.* 2:136-2:144.
- Schlaman, H.R.M., Okker, R.I.H. y Lugtenberg, R.I.I. 1992. Regulation of Nodulation Gene Expression by NodD in Rhizobia. *J. Bacteriol.* 174: 5177-5182.
- Schrempf, H. 1985. Genetic Instability: Amplification, Deletion and Rearrangement within *Streptomyces* DNA. In: *Microbiology*. Leive, L. (ed.). Washington, D.C.: American Society for Microbiology, pp.: 436-444.
- Schultze, M., Kondorosi, E., Ratet, P., Buiré, M. y Kondorosi, A. 1994. Cell and Molecular Biology of *Rhizobium*-Plant Interactions. *Internat. Rev. Cyt.* 156: 1-75.
- Segall, A., Mahan, M. J. y Roth, J. R. 1988. Rearrangement of the Bacterial Chromosome: Forbidden Inversions. *Science.* 241: 1314-1318.
- Shyamala, V., Schneider, E. y Ames, G. F. 1990. Tandem Chromosomal Duplications: Role of REP Sequences in the Recombination Event at the Joint-Point. *EMBO J.* 9: 939-946.
- Spies, T., Laufs, R. y Riess, F-C. 1983. Amplification of Resistance Genes in *Haemophilus influenzae* Plasmids. *J. Bacteriol.* 155: 839-846.
- Springman, E. B., Angleton, E. L., Birkedal-Hansen, H. y Van Wart, H. E. 1990. Multiple Modes of Activation of Latent Human Fibroblast Collagenase: Evidence for the Role of a Cys<sup>73</sup> Active-Site Zinc Complex in Latency and a "Cysteine Switch" Mechanism for Activation. *Proc. Natl. Acad. Sci.* 87: 364-368.
- Young, M. y Hranueli, D. 1988. Chromosomal Gene Amplification in Gram-Positive Bacteria. In: *Recombinant DNA and Bacterial Fermentation*. J. A. Thomson (ed.) CRC Press. Florida, U.S.A. 157-200 pp.
-