



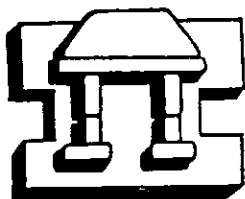
UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES IZTACALA

SECUENCIACION Y ANALISIS DE LAS FAMILIAS DE DNA REITERADO DEL PLASMIDO SIMBIOTICO DE *Rhizobium etli* CFN42

T E S I S

QUE PARA OBTENER EL TITULO DE BIOLOGO PRESENTA JUAN CARLOS HERNANDEZ CELIS



IZTACALA LOS REYES IZTACALA

1998

TESIS CON FALLA DE ORIGEN

264729



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.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES IZTACALA
JEFATURA DE LA CARRERA DE BIOLOGIA

Los Reyes Iztacala, a 9 de junio de 1998

APROBACION DE TESIS

LIC. AMERICA LANDA ROMERO
JEFE DE LA UNIDAD
DE ADMINISTRACION ESCOLAR.
P R E S E N T E .

Por medio de la presente manifestamos a Ud. que como Miembros de la Comisión Dictaminadora del trabajo de Tesis del Pasante de Biología:

JUAN CARLOS HERNANDEZ CELIS

títulado: "Secuenciación y análisis de las familias de DNA reiterado del plásmido simbiótico de Rhizobium etli CFN42".

para obtener el grado de Licenciatura, después de haber sido cuidadosamente revisado y realizadas las correcciones que se consideraron pertinentes, declaramos nuestra aprobación del trabajo escrito, ya que reúne las características, calidad y decoro académico del título al que aspira.

A t e n t a m e n t e
"POR MI RAZA HABLARA EL ESPIRITU"

DR. DIEGO ARENAS ARANDA

DR. JOSE GUILLERMO DAVILA RAMOS

M. EN C. SERGIO VACA PACHECO

M. EN C. RAMON V. MORENO TORRES

BIOL. ELIAS PIEDRA IBARRA

(Nombre completo)

(Firma)

A Mamá,

A Papá,

a quienes debo todo, ¡hasta mi información genética!

***A mis hermanos Fabiola, Alicia, Diana y Alfredo, por
compartir buenos y malos momentos.***

A Patty, por ser mi razón de luchar.

A Fabiancin y Uri, ¡por ser mis primeros sobrinos!

A la familia Ramírez Hernández por su apoyo incondicional.

AGRADECIMIENTOS

Agradezco de forma muy especial al Dr. Guillermo Dávila por darme la oportunidad de pertenecer a su grupo de investigación, por su confianza y su incondicional apoyo.

Al Maestro y Vidente Profesor Majadma Sidartha Gahothama Miguel Angel Ramírez, por iniciarme en las doctrinas herméticas de los "fast protocols" y la supervivencia en medio mínimo, así como por compartir el entusiasmo por la investigación.

A los Doctores José Luis Vargas Sánchez, Bulmaro Cisneros Vega y Diego Arenas Aranda por brindarme sus conocimientos para adentrarme en esta apasionante área de la Biología y por aceptar ser los revisores de este trabajo.

Por su puesto, agradezco a todos mis amigos de la ENEP-Icóptero por hacer del estudio de la Biología toda una aventura (¡Primero las damas!): Rosario, Roxana, Rocio, Sandra, Laura, Luisa, Euse, Angeles, Silvia, (¡luego los cuates!) los Exuvios Antonieto y Angel Face, el tremendo Gilito, Marcox, Emilio Fernández, Rafail Gallego y Carnlos de la Ronsa.

A mis amigos del CIFN, por compartir el largo y sinuoso camino de la investigación y hacerlo una experiencia increíble y fascinante (Por su puesto, ¡primero las damas!): Elizabeth, Vero, Hildiux, Noriux, Rosy, Paty, (¡luego los feos!) Miguelón, Don José, Dennis Boy, Jorjo, Gameto, Javier Soberón y Oscar.

Agradezco muy especialmente a los compas, por conservar la amistad que hemos compartido por mas de 13 años y ¡que seguirá por mucho tiempo más!: Alejandro Olmos (Arquiteito), Daniel Franco (Giógrafo) y Ricardo Pérez Valadez (Ingeniebrío).

A mis amigas Lourdes Morales, Lourdes Luke y Tania, por mantener nuestra amistad a pesar de la distancia y seguir intentando reunirnos.

¡¡¡A TODOS MIL GRACIAS!!!

El presente trabajo se realizó en el laboratorio No. 1 del Departamento de Genética Molecular del Centro de Investigación sobre Fijación de Nitrógeno de la UNAM, bajo la dirección del Dr. José Guillermo Dávila Ramos

ÍNDICE

RESUMEN	1
1. INTRODUCCION	2
1.1 Secuencias de DNA reiterado en el genoma bacteriano	2
1.2 La simbiosis de <i>Rhizobium</i> y su genoma	7
1.3 Secuencias reiteradas en <i>Rhizobium</i>	10
2. OBJETIVOS	14
2.1 Objetivo general	14
2.2 Objetivos particulares	14
4. MATERIAL Y MÉTODOS	15
4.1 Cepas utilizadas	15
4.2 Plásmidos utilizados y construidos	15
4.3 Condiciones de cultivo de microorganismos	17
4.4 Clonación de los fragmentos de restricción	17
4.4.1 Digestión de DNA con endonucleasas de restricción	17
4.4.2 Ligación de DNA	17
4.4.3 Preparación de células competentes con CaCl ₂	17
4.4.4 Transformación de células competentes	18
4.5 Extracción y purificación de DNA plasmídico	18
4.6 Electroforesis verticales de DNA en geles de agarosa	19
4.7. Electroforesis horizontal en minigeles de agarosa	19
4.8 Cuantificación de DNA	20
4.9 Transferencia de DNA a filtros de nylon	20
4.10 Marcaje de la sonda de DNA con [α - ³² P] dCTP	20
4.11 Hibridización tipo Southern	21
4.11.1 Prehibridización de filtros	21
4.11.2 Hibridización de filtros	21
4.11.3 Lavado de los filtros	21
4.10.4 Secado de los filtros	22
4.10.5 Exposición y revelado de la placa fotográfica	22
4.12 Diseño de oligonucleótidos para secuenciación	22
4.13 Secuenciación automática de DNA	22
4.14 Ensamblaje y análisis de las secuencias de DNA	23
5. RESULTADOS	24
5.1 Elementos reiterados de la familia B	24
5.1.1 Análisis de la secuencia del pBG02	25
5.1.2 Conformación de la familia de secuencias reiteradas BF	25

5.1.3	Análisis de las secuencias de los elementos reiterados de la familia BF	26
5.2	Elementos reiterados de la familia C	27
5.2.1	Localización del fragmento de restricción que contiene el elemento reiterado de la familia C de la banda BG83	28
5.2.2	Subclonación del fragmento de restricción que contiene el elemento reiterado de la familia C en el pBG83	28
5.2.3	Secuenciación de las construcciones que contienen el elemento reiterado de la familia C en el pBG83	28
5.2.4	Análisis de la secuencia 83S4ex que contienen el elemento reiterado de la familia C de la clona pBG83	28
5.3	Elementos reiterados de la familia D	29
5.3.1	Localización del fragmento de restricción que contiene el elemento reiterado de la familia D en el pBG85	29
5.3.2	Subclonación del fragmento de restricción que contiene el elemento reiterado de la familia D en el pBG85	30
5.3.3	Secuenciación de las construcciones que contienen el elemento reiterado de la familia D en el pBG85	30
5.3.4	Análisis de la secuencia del p85P4 que contiene el elemento reiterado de la familia D de la clona pBG85	30
5.4	Elementos reiterados de la familia E	32
5.4.1	Localización del fragmento de restricción que contiene el elemento reiterado de la familia E en el pBG79	32
5.4.2	Subclonación del fragmento de restricción que contiene el elemento reiterado de la familia E en el pBG79	32
5.4.3	Secuenciación de las construcciones que contienen el elemento reiterado de la familia E en el pBG79	33
5.4.4	Análisis de la secuencia p79H4 que contienen el elemento reiterado de la familia E en el pBG79	33
5.4.5	Secuencia del elemento reiterado de la familia E del pBG85	34

5.5 Análisis estadístico del porcentaje de GC de las secuencias obtenidas	34
6. DISCUSIÓN	36
7. CONCLUSIONES	46
8. FIGURAS	
Figura 1	47
Cuadro 1	48
Figura 2	49
Figura 3	50
Figura 4	53
Figura 5	56
Figura 6	60
Figura 7	63
Figura 8	64
Figura 9	65
Tabla 1	66
Figura 10	67
Cuadro 2	68
8. BIBLIOGRAFÍA	69

RESUMEN

El análisis de la secuencia nucleotídica de los genomas bacterianos ha demostrado la existencia de un gran número de secuencias reiteradas en procariotes. En los géneros *Halobacterium*, *Streptomyces* y en la familia Rhizobiaceae se ha encontrado un elevado número de estas secuencias que favorecen los eventos de recombinación homóloga. En especies del género *Rhizobium*, las cuales se caracterizan por establecer simbiosis con leguminosas, se ha calculado que existen 200 familias de secuencias reiteradas conteniendo de 2 a 3 elementos, tanto en plásmidos como en cromosoma. Las bacterias de éste género poseen la mayor parte de los genes involucrados en los procesos simbióticos en un plásmido de gran tamaño denominado plásmido simbiótico (pSim). Varias secuencias reiteradas han sido caracterizadas, revelando operones completos, genes, secuencias regulatorias, secuencias de inserción y secuencias con función desconocida. En el pNGR234a de la cepa *Rhizobium* sp. NGR234, el único pSim secuenciado, estas secuencias representan el 18% del plásmido. En el pSim de *Rhizobium etli* CFN42 existen 11 familias de secuencias reiteradas: de tres elementos (A, B, C, D, y E), y seis de dos elementos (F, G, H J y K). Solo la familia A se ha caracterizado como los genes *nifH*.

En este trabajo se presentan las secuencias de los 3 elementos de la familia B (ahora BF), un elemento de la familia C (de la BG83), un elemento de la familia D (de la BG85) y dos elementos de la familia E (de las BG79 y BG85). La secuencia de uno de los elementos BF revela un fragmento que contiene un marco de lectura abierto (MLA) semejante al Y4wP de función desconocida del pNGR234a. La secuencia del elemento de la familia C muestra similitud con la secuencia de la IS407 de *Pseudomonas cepacia*. El elemento de la familia D es similar con la IS427 de *Agrobacterium tumefaciens* y con la IS6501 de *Brucella ovis*, pero ésta secuencia no se encuentra en la BG01 del mapa físico del pSim, donde inicialmente se había adscrito, por lo que se propone la división de la familia D en tres familias D1, D2 y D3, con dos elementos cada una. Finalmente, los dos elementos de la familia E muestran alta identidad con la ISRm6 de *Rhizobium meliloti*.

Los resultados fueron analizados respecto a la definición de las familias reiteradas en el pSim, su posible origen como IS's funcionales, su participación en fenómenos de recombinación, su posible integración como un transposón complejo, su adquisición por medio de un evento de transferencia horizontal y las implicaciones evolutivas dentro del genoma de *Rhizobium*.

1. INTRODUCCIÓN

1.1 Secuencias de DNA reiterado en el genoma bacteriano.

Las secuencias de DNA reiterado se definen como secuencias que se encuentran más de una vez en el genoma y no son producto del azar, de acuerdo al tamaño del genoma, es decir, si se tiene un genoma de 4.6 megapares de bases (Mpb.), como el de *Escherichia coli* (Blattner, 1997), la secuencia reiterada más grande posible generada por azar sería de 11 pares de bases (pb.), ya que solo existen 4 nucleótidos y cualquiera puede ocupar cualquier posición, esto es $4^{11} = 4\ 194\ 304\ \text{pb.} = 4.1\ \text{Mpb.}$ Si se buscara una secuencia de 12 pb., se tendría $4^{12} = 16\ 777\ 216\ \text{pb.} = 16.7\ \text{Mpb.}$ la cual se encuentra fuera del azar. Aunado a lo anterior, estudios en secuencias generadas al azar poseen un número menor de secuencias reiteradas de diversos tamaños que secuencias codificantes (Balleza, 1998).

Los primeros estudios realizados en torno a la presencia de secuencias de DNA reiterado a nivel de genoma mostraron su existencia en un número elevado, hasta un millón de veces, en genomas de organismos eucariotes, desde protozoarios hasta plantas y animales. Sin embargo, dados los experimentos de renaturalización de DNA y de medición de efecto hipercrómico, no se detectaron secuencias repetidas en *E. coli*, por lo que se consideró la ausencia de este tipo de secuencias en organismos procariotes (Waring, 1966; Britten, 1968).

A partir de la primera mitad de los 70's se descubrieron varias copias de los genes de RNA ribosomal en *E. coli* (Jørgensen, 1976; Kenerley, 1977), concluyendo que existen siete copias de los operones de RNAr (Kiss, 1977), lo cual se comprobó al obtenerse la secuencia total de su genoma (Blattner, 1997). Simultáneamente a la descripción de las copias de RNAr, se describieron fenómenos de duplicación génica en tándem entre estos operones (Hill, 1977); poco después se evidenciaron inversiones en la misma zona (Hill, 1981). Rearreglos similares se descubrieron en *Salmonella typhimurium* también en operones de RNAr (Lehner, 1980; Anderson, 1981). Rearreglos a frecuencias de 10^{-4} se observaron en *Halobacterium halobium* por recombinación entre familias reiteradas (Sapienza, 1982b), cuyo número se estimó en más de 50 conteniendo de 2 hasta 20 elementos, con lo cual se estima que existen más de 500 elementos reiterados (Sapienza, 1982a);

estas reiteraciones las constituyen principalmente secuencias de inserción (ISH) (Charleobis, 1989).

La amplificación génica se ha observado como respuesta adaptativa ante el ambiente, tal es el caso del gen de la cloranfenicol-acetiltransferasa de *Staphylococcus aureus* integrado al cromosoma de *Bacillus subtilis*, el cual se amplificó hasta 15 veces en una concentración de 20µg/ml. de cloranfenicol (Young, 1984); de forma similar, se describió la amplificación del gen de resistencia a kanamicina y neomicina bordeado por IS1 del plásmido pRR330 de *E. coli* en altas concentraciones de neomicina (Peterson, 1985). En *S. typhimurium* se comprobó la duplicación de una zona de su cromosoma, que involucra genes de sistemas de transporte de fuentes de carbono, cuando se cultivó en un medio con limitación de sustancias carbonadas (Sonti, 1989).

Un caso especial de reiteración lo constituyen las unidades amplificables de DNA (AUD's, "Amplifiable Units of DNA") en especies del género *Streptomyces* (Bar-Nir, 1994), ya que se observó que después de extensas deleciones del cromosoma, de hasta 800 kb., ocurrían amplificaciones de éstas unidades hasta alcanzar las 3 000 kb. (Birch, 1990); aunque no se conocen los mecanismos por los cuales se llevan a cabo estos fenómenos de dinámica genómica, se proponen eventos de recombinación ilegítima, aunque no se descarta la posible participación de fagos atemperados, elementos transponibles o plásmidos integrativos, los cuales ya han sido descritos (Chater, 1988).

La secuenciación de genomas bacterianos completos permite realizar análisis precisos sobre secuencias de DNA reiterado. Hasta el momento se han secuenciado los genomas de *Haemophilus influenzae Rd*, *Mycoplasma genitalium*, *Synechocystis* PCC6803, *Methanococcus janaschii*, *Mycoplasma pneumoniae*, *Helicobacter pylori*, *E. coli* K-12, *Methanobacterium thermoautotrophicum* ΔH, *B. subtilis*, *Archaeoglobus fulgidus* y *Borelia burgdorferi*. A continuación se hace referencia a sus reiteraciones más importantes, por longitud y nivel de conservación, descubiertas en estos organismos.

El primer genoma bacteriano secuenciado fue el de *H. influenzae Rd*, de 1.8 Mpb. de extensión, en él se encontraron seis operones de RNAr y una

reiteración de dos elementos de 5 340 pb. Las secuencias de los genes de las subunidades de RNAr fueron idénticas en los seis operones, así como dos de las secuencias intergénicas entre tres de estos (Fleschmann, 1995).

En *M. genitalium* se ha estimado que un 4% de su genoma está reiterado a pesar de poseer el genoma mas pequeño hasta el momento descrito (580 kb.) (Peterson, 1993); éste porcentaje (4.7% exactamente) lo representan 9 secuencias que poseen identidad con partes del operón MgPa, descritas dentro de la secuencia total de su genoma (Fraser, 1997a). Se ha evidenciado que probablemente la recombinación entre estas secuencias provee la variabilidad antigénica observada en éste organismo (Peterson, 1995).

El genoma de la arqueobacteria *M. jannaschii* (1.66 Mpb.) posee dos operones de RNAr, así como tres familias de elementos repetidos: ISAMJ1, semejante a IS240 de *Bacillus thuringiensis*, se encuentra repetido 10 veces en el cromosoma y una en el plásmido mayor ECE; dos MLA's (928 pb.) parecidos a la transposasa de la IS982 de *Lactococcus lactis* y 18 copias de secuencias que probablemente participan en la partición del cromosoma, constituidas por repeticiones en tándem, de 1 a 25, de fragmentos pequeños con regiones únicas entre ellas (Bult, 1996).

Alrededor de un 8 % del genoma de *M. pneumoniae* (816.3 Kb.) se encuentra constituido por las secuencias repetidas RepMP1, RepMP2/3, RepMP4 y RepMP5, las cuales poseen similitud con la proteína P1 o con el MLA6 del operón P1, la cual es una adhesina. Se han observado fenómenos de amplificación de segmentos y secuencias completas de genes de lipoproteínas y de los elementos RepMP (Himmelreich, 1996).

La secuencia del genoma del patógeno gástrico *H. pilory* (1.66 Mb) posee dos operones de RNAr y ocho familias de elementos repetidos con 97% de identidad con tamaños que varían entre 0.47 y 3.8 kb., los cuales están relacionados en duplicaciones génicas, secuencias intergénicas ("repeat" 7) y con proteínas de membrana externa ("repeats" 1, 2, 3, y 6), éstas últimas posiblemente participen en eventos de recombinación similares a los observados en *M. genitalium* que le confieren variabilidad antigénica. Las secuencias repetidas no codificantes representan un 2.3% del genoma. Así mismo, existen 13 copias de la IS605 (5 completas y 8 parciales) y 4 copias de IS606 (2 completas y 2 parciales) (Tomb, 1997).

Además de los 7 operones de RNAr mencionados, *E. coli* K-12 posee 5 elementos repetidos *Rhs* de 5.7 a 9.6 kb. de longitud, los que constituyen 0.8% del genoma y de los cuales no se conoce su función, sin embargo se sugiere que probablemente sean elementos móviles. Pequeñas secuencias repetidas en tándem (referidas como REP, BIME "Bacterial interspersed mosaic elements" o PU) componen los denominados complejos REP, que se encuentran en un número de 314 y se desconocen su origen y su función. Otras reiteraciones descritas son los elementos IRU o ERIC (19 copias), las secuencias de término de la duplicación (*Ter*, 12 copias), las cajas C (Box C, 33 copias), las secuencias LRD (4 copias), las repeticiones *iap* (23 copias en grupos de 14, 7 y 2) y los RSA (6 copias). Existen varias copias de las secuencias de inserción IS1 (7 copias), IS2 (7 copias), IS3 (5 copias), IS5 (11 copias), IS186 (3 copias), IS30 (3 copias) e IS911 (2 copias) (Blattner, 1997).

La secuencia del genoma de *M. thermoautotrophicum* (1.75 Mb.) posee dos secuencias repetidas en orientación directa de aproximadamente 8 kb., las cuales contienen más de 30 genes que codifican polipéptidos relacionados con factores de iniciación de la transcripción y con proteínas de división celular. Existen dos operones de RNAr (5.3 Kb.) y dos regiones con un gran número de secuencias reiteradas, R1 y R2, constituidos por repeticiones, 47 y 124 copias respectivamente, de una secuencia de 30 pb., que en conjunto tienen una extensión de 3.6 y 8.6 kb. Así mismo, se encuentran dos genes de la enzima citratosintetasa y *nifH* (una lista completa de todas las secuencias reiteradas encontradas en éste genoma se encuentra disponible en Internet, cuya dirección es: http://www.cric.com/htdocs/sequences/methanobacter/ repeat_summary.prn) (Smith, 1997).

En el genoma de la bacteria Gram-positiva *B. subtilis* (4.2 Mpb.) existen 10 operones de RNAr en regiones aledañas al origen de replicación, las secuencias flanqueantes a estos operones también se encuentran reiteradas. Otras regiones que se encuentran repetidas las conforman los genes *pkS* y *srf*, una región de 182 pb. dentro de los genes *yyaL* y *yyaO*, una secuencia de 410 pb. entre los genes *yxAK* y *yxBL*, una duplicación interna de 174 pb. en el gen *ydcl*, duplicaciones involucradas en el control transcripcional de varios genes como las 118 pb. repetidas 3 veces entre los genes *yxB* y *yxBC*, una secuencia de 190 pb. con 10 copias divididas en dos familias de 6 y 3 elementos cada una, y por último, un número considerable de repeticiones en

los borde de secuencias de probables regiones de integración de fagos (Kunst, 1997).

El genoma de la arqueobacteria *A. fulgidus* posee un gran número de duplicaciones de genes, lo cual contribuye de forma importante en el tamaño de su genoma, que, comparado con el de otras arqueobacterias, resulta ser bastante grande (2.1 Mb.). En contraste con el resto de los genomas publicados, *A. fulgidus* sólo contiene un operón para RNAr, mientras que otras secuencias codificantes poseen varias copias como las de los genes para la NADH-flavin oxidoreductasa (1 886 pb., 2 copias), el MLA NifS-NifU (1 549 pb., 2 copias), la transposasa potencial IS1214 con el ISORF2 (1 214 pb., 6 copias), la ISA1083 con el ISORF2 (1 083 pb., 3 copias), el gen para una proteína del sistema de secreción tipo II (1 014 pb., 4 copias), una posible transposasa ISA0963 (963 pb., 7 copias), un MLA homólogo a MJ0794 (836 pb., 3 copias), dos MLA's para proteínas conservadas hipotéticas (696 y 628 pb., dos copias cada una). Existen otras secuencias repetidas idénticas pequeñas (30 pb.) en orientación directa no codificantes, las cuales se encuentran en conjuntos de 48 (SR-1A) y de 60 (SR-B). También se encuentra otro grupo de secuencias de 37 pb. repetidas con 42 copias similares a SR-1 (Klenk, 1997).

La espiroqueta *B. burgdorferi* posee un cromosoma lineal de 910 Kb., así como varios plásmidos circulares y lineales que en conjunto llegan a 17 (533 kb.). Entre estos últimos se localizan segmentos con 99% de identidad con longitudes entre 3 y 5 kb. Existen varias copias de la transposasa potencial IS891 en plásmidos, lo cual se ha asociado con los rearreglos frecuentes entre ellos, dado que no posee secuencias invertidas en los extremos, probablemente se trate de una IS no funcional. En el cromosoma se encuentran dos copias del gen *fliG*, componente del motor flagelar. Los 900 pb. de la secuencia telomérica derecha posee una alta identidad con el extremo izquierdo de los plásmidos lineales lp17y lp28-3, al igual que la región comprendida entre las bases 3 600 y 8 000 del extremo derecho, la cual muestra alta identidad en aproximadamente 4500 pb. con varios genes plasmídicos que incluyen un MLA de una posible transposasa (Fraser, 1997b).

Es claro que la presencia de secuencias reiteradas en los genomas procariotes es común, inclusive en organismos como *M. genitalium* que posee el genoma más pequeño hasta ahora descrito. Se ha reportado la existencia

de un gran número de secuencias reiteradas en miembros de la familia Rizobiaceae, los cuales se caracterizan por establecer relaciones con plantas, ya sean simbióticas o patológicas. Dichas secuencias se han estudiado por su importancia en fenómenos de expresión génica, de recombinación inter e intra replicón, de amplificación y delección, así como en fenómenos de transferencia horizontal e integración.

1.2 La simbiosis de *Rhizobium* y su genoma.

La importancia de la fijación biológica de nitrógeno radica en que éste elemento se encuentra presente en todos los organismos, como parte de las unidades fundamentales de proteínas y ácidos nucleicos entre otros (Lehninger, 1978). Sin embargo, a pesar de que la atmósfera lo contiene en un 79%, éste se encuentra como molécula diatómica químicamente inerte y sumamente estable, por lo que se requieren de $+940 \text{ kJ mol}^{-1}$ y poder reductor para convertirlo a amonio (Mathews, 1990), por lo cual no es asequible a la mayoría de organismos. Los procesos mediante los cuales se reduce el N_2 a amonio comprenden la fijación biológica de nitrógeno (95%), procesos químicos industriales y atmosféricos por efecto de descargas eléctricas y radiación ultravioleta (5%) (Sprenst, 1987).

La capacidad para llevar a cabo la fijación biológica de nitrógeno está confinada únicamente a organismos procariotes, los cuales comprenden cianobacterias como *Anabaena*, *Nostoc*, *Calotrix* (Wolk, 1996), *Synechococcus* (Mitsui, 1986), *Cyanotece* (Colón-López, 1997); actinobacterias del género *Frankia* (Valdés, 1996); arqueobacterias tales como *Methanococcus thermolithotrophicus* (Sibold, 1988), *Methanococcus maripaludis* (Cohen-Kupiec, 1997) *Methanosarcina barkeri* 227 (Chien, 1996)); bacterias de los géneros *Azotobacter*, *Clostridium* y *Klebsiella*, los cuales reducen nitrógeno en vida libre (Merrick, 1992); bacterias que se relacionan con plantas no leguminosas como los géneros *Azospirillum* y *Azoarcus* (Baldani, 1997) y el endosimbionte *Acetobacter* (Jiménez-Salgado, 1997) así como bacterias pertenecientes a la familia Rizobiaceae, las que establecen asociaciones simbióticas con miembros de la familia Leguminoseae (Long, 1989a).

La familia Rhizobiaceae comprende los géneros *Agrobacterium*, *Rhizobium*, *Azorhizobium* y *Bradirhizobium*, de acuerdo a grupos de inoculación cruzada ("cross inoculation groups") (Long, 1989a), sin embargo, estudios a nivel molecular han revelado el agrupamiento de especies genéticamente distantes, por lo que la clasificación actual se encuentra en revisión (Schultze, 1994).

Los miembros de la familia Rhizobiaceae son capaces de establecer simbiosis generalmente con las raíces de plantas leguminosas por medio del desarrollo de estructuras especializadas denominadas nódulos, en donde, tras un complejo proceso de reconocimiento e infección, las bacterias se establecen como bacteroides intracelulares fijadores de nitrógeno. El establecimiento de la simbiosis es considerado como un proceso específico, una especie de bacteria solo puede infectar y nodular a un hospedero particular. De acuerdo a lo anterior, *Rhizobium etli* nodula frijol (*Phaseolus vulgaris*), la soya (*Glycine max*) es nodulada por *Bradirhizobium japonicum*, la alfalfa (*Medicago sativa* y *Melilotus albus*) por *Rhizobium meliloti*, *Sesbania* por *Azorhizobium caulinolans*, etc. (Fisher, 1992). Sin embargo, se han descrito casos donde la especificidad no es absoluta (Martínez, 1990) y especies de amplio espectro de hospederos como *Rhizobium* sp NGR234, la cual nodula más de 110 diferentes géneros de leguminosas y la no leguminosa *Parasponia andersonii* (Pueppke, S. J. y Broughton, W. J.; datos no publicados).

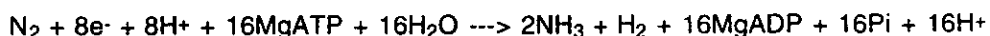
La interacción inicial de la bacteria con la planta se lleva a cabo en la raíz, donde productos excretados son quimioattractores bacterianos. Los exudados radiculares incluyen aminoácidos, carbohidratos y flavonoides. Se ha demostrado que puede afectarse la producción de flavonoides por la presencia de nitrógeno combinado en el medio. Inicialmente, la síntesis y liberación de flavonoides se encuentra controlada por la planta, pero existe evidencia de un posible control ejercido por la bacteria (Schultze, 1994). Se han identificado también lectinas, las cuales posiblemente se encuentran involucradas en el proceso de adhesión a la rizósfera (Halverson, 1986; Díaz, 1995).

La primera respuesta macroscópica es la deformación de los pelos radiculares (Nap, 1990), donde las bacterias quedan atrapadas en el sitio de rizado. La liberación de los factores Nod específicos

(lipoquitooligosacáridos) (Carlson, 1994) inducen la expresión de nodulinas (proteínas específicas sintetizadas por la planta durante la formación y funcionamiento del nódulo). Las nodulinas tempranas (ENOD) se encuentran relacionadas con la formación de hilos de infección que conducen a las bacterias hasta el primordio nodular que se establece simultáneamente en la corteza radicular. Finalmente, los hilos de infección penetran las células centrales del primordio nodular liberando a las bacterias en el citoplasma. El éxito de la infección involucra la expresión de los genes *exo* (exopolisacáridos), *lps* (lipopolisacáridos) y *ndv* (cicloglucanos), partícipes en la nodulación específica del hospedero (Halverson, 1986; Rolfe, 1988; Nap, 1990; Schultze; 1994).

El proceso para liberación de las bacterias en las células del primordio nodular aún no es totalmente entendido. Estudios de microscopía revelaron que las bacterias son envueltas en una matriz vegetal y éstas aparecen discontinuamente en el hilo de infección (Robertson, 1978; Mellor, 1987), posteriormente la bacteria disminuye su nivel de replicación y división celular hasta cesarlo, aunado a ello cambian su expresión genética y morfología, denominándose bacteroides (Sutton, 1981), estos son envueltos en una membrana compuesta por material vegetal denominada membrana peribacteroidal, la cual posee especificidad de transporte y permeabilidad (Long, 1989). Las nodulinas tardías componen la maquinaria de asimilación y transporte de amonio, metabolismo de compuestos de carbono así como la manutención de un ambiente microaeróbico por medio de la leghemoglobina (Dilworth, 1984; Long 1989b) para llevar a cabo el proceso de fijación de nitrógeno. La leghemoglobina (Lb) es la nodulina tardía más abundante, constituye el 30% de las proteínas totales del nódulo (Nap, 1990).

La capacidad de fijar nitrógeno está dada por la nitrogenasa, la cual lleva a cabo la reducción de N_2 a amonio. Los genes *nifD* y *nifK* codifican las subunidades α y β de la proteína FeMo, también llamada componente I o dinitrogenasa (220 kDa), mientras que el gen *nifH* codifica la proteína homodimérica Fe, denominada también dinitrogenasa reductasa o componente II (60 kDa) (Fischer, 1994). El sistema Fe-MoFe lleva a cabo la siguiente reacción :



Se han reportado nitrogenasas con núcleos de vanadio y de hierro, sin contenido de Mo (Pau, 1989; Newton 1993). La expresión de la nitrogenasa se presenta después de la expresión de las nodulinas tardías (Nap,1990).

En la mayoría de las especies del género *Rhizobium* el genoma se encuentra compartimentalizado en un cromosoma y varios plásmidos, a diferencia de los géneros *Bradyrhizobium* (Noti, 1985; Van den Eede, 1987) y *Azorhizobium* (Boivin, 1997), los cuales posee todos los genes involucrados en fijación de nitrógeno y nodulación en el cromosoma; la cantidad y tipo de plásmidos varía entre especies y cepas. Los principales genes involucrados en nodulación y fijación de nitrógeno (*nod*, *nif* y *fix*) se localizan en un plásmido de alto peso molecular denominado plásmido simbiótico (pSim). Usualmente el número de plásmidos varía de uno a seis, representando entre el 25 y el 50% del genoma total (Martínez, 1990). A pesar de que los pSim contienen los principales genes involucrados en la simbiosis, se ha demostrado que otros replicones son requeridos para llevar a cabo el proceso de manera óptima o para sobrevivir en diferentes medios (Brom, 1992). Genes involucrados en producción de polisacáridos de superficie, en aprovechamiento de metabolitos vegetales, compuestos aromáticos y azúcares diversos, así como la competitividad han sido localizados en otros plásmidos diferentes del pSim (García, 1996).

1.3 Secuencias reiteradas en *Rhizobium*.

La presencia de secuencias reiteradas a nivel de DNA es otra característica importante en el genoma de *Rhizobium*. (Martínez, 1990). Se ha estimado el número aproximado de secuencias reiteradas en cepas de la familia Rhizobiaceae, encontrándose que existen alrededor de 200 familias reiteradas por genoma, con un promedio de 3.5 elementos cada una, lo cual es considerado inusual en procariotes (Flores, 1987). Niveles de reiteración de esta magnitud solo son característicos de los géneros *Halobacterium* (Sapienza, 1982) y *Streptomyces* (Birch, 1990). Se han encontrado operones completos reiterados, como *nifHDK* en *R. etli* CFN42 (Quinto, 1982; Martínez, 1985; Quinto, 1985); genes completos funcionales tales como *nifA*, *fix* (aun no caracterizado totalmente) (David, 1987) un nuevo *locus nod* y *nodD* (Renalier, 1987; Honma, 1987) en *R. meliloti*; *nodD* en *B. japonicum* (Appelbaum, 1988); *nifDH* en *Rhizobium fredii* (Barbour, 1985); *nifH* en *R.*

ORS571 (Norel, 1987), entre otros (Charles, 1990). Se han identificado secuencias de inserción en distintas cepas de *Rhizobium* que varían en número de repeticiones y son funcionales, por ejemplo ISRM1 (Ruvkun, 1982; Wheatcroft, 1988), ISRM2 (Dusha, 1987), ISRM3 (Ogawa, 1991; Wheatcroft, 1991), ISRM4 (Soto, 1992), ISRM5 (Laberge, 1995) y ISRM6 (Zecrí, 1996) en *R. meliloti* ; ISR1 en *Rhizobium* Clase IV (Priefer, 1989), secuencias repetidas similares a secuencias de inserción en *B. japonicum* (Kaluza, 1985). Se han descrito en *B. japonicum* (Hahn, 1987) y *R. fredii* USDA193 (Masterson, 1986) secuencias reiteradas sin identidad aparente con las secuencias reportadas hasta ahora en bancos de datos.

Dentro de la secuencia completa del plásmido simbiótico pNGR234a (536.16 kb.) de *R. sp.* NGR234 existen reiteraciones de 85 genes posibles de proteínas de la misma familia que aparecen interrumpidas por posibles genes involucrados en transposición, integración y recombinación. Los marcos de lectura abiertos (MLA) probables que poseen identidad con proteínas relacionadas con secuencias de inserción componen un 18% de este plásmido y son semejantes a secuencias de inserción de *Agrobacterium*, *Bacillus*, *Escherichia*, *Pseudomonas* y *Rhizobium*. La mayoría de éstas secuencias se localizan entre las posiciones 300 000 y 390 000 del pNGR234a. Cabe mencionar que estos MLA se encuentran bordeando conjuntos de genes relacionados (*oriV-oriT*, *nif*, *fix*, *hsn*), por lo cual se propone la integración de estos genes como un transposón complejo, llamados islas de secuencias de inserción. El contenido de GC de estas secuencias de inserción es 3% más alto que el del pNGR234a cuya composición posee un 58.5%. Hasta ahora no se ha demostrado que estas secuencias de inserción sean funcionales (Freiberg, 1997). Recientemente se describió la sucesión de eventos de transposición que originaron al elemento NGRRS-1, el cual posee una longitud de 6 kb. Este elemento es parte de una familia de cuatro reiteraciones, una de ellas se encuentra en el pNGR234a y las tres restantes en el cromosoma. Se identificaron 4 secuencias de inserción que componen a los elementos NGRRS, las cuales se propone se integraron en secuencias de inserción previamente establecidas; una posible explicación para éste fenómenopudiera ser que, de ésta forma, se minimiza el efecto mutagénico en genes importantes (Perret, 1997).

En particular, *R. etli* CFN42 posee seis plásmidos de alto peso molecular: p42a, p42b (ambos alrededor de 150 kb), p42c (270 kb), p42d

(pSim, 390 kb), p42e (500 kb) y p42f (600 kb). Al llevar a cabo la conformación del mapa físico del pSim de *R. etli* CFN42 por medio de la construcción de una genoteca de cósmidos, de los cuales 16 cubren la totalidad del plásmido, así como una conteniendo los 85 fragmentos producto de la digestión total del pSim con Bam HI, se descubrieron y localizaron 11 familias de secuencias reiteradas (A-K) dentro del propio plásmido, de las cuales, los genes *nifH* habían sido previamente reportados (Quinto, 1982). Existe alta homología con otros replicones de *R. etli* CFN42, tal es el caso del plásmido p42a, con el que posee 19 fragmentos de Bam HI homólogos, el plásmido p42f es homólogo en 3 fragmentos contiguos de Bam HI, mientras que un sólo fragmento lo es con el plásmido p42b. Con el mapa físico, se localizaron los dos operones *nifHDK* flanqueando una región de 120 kb dentro de la cual se localizó el gen *nodD1*. Los operones mencionados corresponden a dos elementos de la familia reiterada A (*nifHa* y *nifHb*), el tercer elemento de esta familia, *nifHc*, se encuentra también en la región de 120 kb. Por otra parte, la secuencia de la BG02, donde se encuentra un elemento de la familia B, presenta homología con el gen *nifX* (Dávila, comunicación personal). No se conoce la identidad del resto de familias reiteradas del pSim de *R. etli* CFN42 (Girard, 1991) (Fig. 1).

La presencia de un número tan elevado de secuencias reiteradas favorece la frecuencia de recombinación intramolecular (Flores, 1988). Los rearreglos en el pSim de *R. etli* CFN42 han sido profusamente estudiados (Romero, 1995). Se han descubierto fenómenos de cointegración con otros plásmidos (p42b::p42d) (Brom, 1991), amplificación y delección de la región comprendida entre los operones *nifHDK* (Romero, 1991), así como de regiones definidas del pSim denominadas amplicones (Romero, 1995; Flores, 1993). En otra cepas de *R. etli* se han observado fenómenos semejantes (Soberón, 1986; Rodríguez, C. en preparación).

Se ha postulado que la capacidad dinámica del genoma plasmídico de *Rhizobium* le provee de adaptabilidad a posibles cambios microambientales o para la interacción con diferentes plantas por sobreexpresión de los genes amplificados (Romero, 1995), lo cual establece nuevas perspectivas ecológicas y evolutivas.

El establecimiento de la naturaleza y características de las familias de secuencias reiteradas (B-K) resulta esencial para proponer un modelo que

explique el origen evolutivo del pSim, dado que este replicón al parecer es producto de eventos de integración con secciones de DNA foráneo por posibles secuencias de inserción flanqueantes, como ha sido demostrado para algunos de los plásmidos pTi de *Agrobacterium tumefaciens* (Otten, 1992). Existe evidencia que apoya el postulado de que el pSim de *R. etli* CFN42 fue transferido desde otra cepa por conjugación, lo cual concuerda con lo propuesto por Schofield y cols. (1987), dado que no se encuentran zonas de homología con las reiteraciones del plásmido en el cromosoma, siendo factible el que, en caso de existir secuencias de inserción funcionales, deberían aparecer aún en baja frecuencia en otros replicones. Si bien existen zonas de homología con los plásmidos p42a, p42f y p42b no parece haber un patrón consistente entre las reiteraciones interplasmídicas y las intraplasmídicas, lo anterior se deduce con base en base al mapa físico del pSim (Girard, 1991).

En este trabajo se presentan los resultados de los estudios realizados con las 4 familias de 3 elementos reiterados del pSim de *R. etli* CFN42: B, C, D y E (Cuadro 1); así mismo, se muestran los análisis de comparación con secuencias de bases de datos de nucleótidos y aminoácidos, grado de conservación entre los elementos secuenciados de la misma familia, sus implicaciones en los rearrreglos y en la historia evolutiva del pSim.

2. OBJETIVOS

2.1 Objetivo general.

- Secuenciar y caracterizar cada uno de los elementos de DNA reiterado de las familias B, C, D y E del plásmido simbiótico (pSim) de *Rhizobium etli* CFN42.

2.2 Objetivos particulares.

- Localizar el fragmento de restricción susceptible a secuenciación directa que contiene a los elementos reiterados a partir de las clonas portadoras.

- Secuenciar las subclonas que contienen los elementos reiterados.

- Determinar la identidad de los elementos secuenciados de cada familia comparándolos con las secuencias reportadas en las bases de datos.

- Definir el grado de conservación entre las secuencias de los elementos de cada familia reiterada.

- Determinar la presencia de secuencias codificantes, marcos de lectura abiertos (MLA) o motivos (secuencias directas o inversas repetidas, sitios de unión de proteínas a DNA, etc.) en las secuencias de cada familia.

4. MATERIAL Y MÉTODOS

4.1 Cepas utilizadas.

En el presente trabajo se utilizaron cepas de *E. coli* previamente reportadas, así como construcciones plasmídicas producidas durante el desarrollo del mismo. Sus características más importantes se enlistan a continuación.

CEPA	CARACTERÍSTICAS	REFERENCIA
<i>Escherichia coli</i> :		
DH5 α	Cepa utilizada para α -complementación por transformación con plásmidos con <i>lacZ</i> .	Hanahan, 1983
BG79	Cepa con la construcción pBR328/BG79, posee un elemento de la familia E del pSim de CFN42. Inserto de 16.8 kb, Cm ^R .	Girard, 1991
BG83	Cepa con la construcción pBR328/BG83, posee un elemento de la familia D del pSim de CFN42. Inserto de 7.2 kb, Cm ^R .	Girard, 1991
BG85	Cepa con la construcción pBR328/BG85, posee un elemento de la familia D y otro de la familia E del pSim de CFN42. Inserto de 12.9 kb, Cm ^R .	Girard, 1991

4.2 Plásmidos utilizados y construidos.

PLÁSMIDO	CARACTERÍSTICAS	REFERENCIA
VECTORES DE CLONACIÓN		
pSK+	Sitio múltiple de clonación en <i>lac Z</i> , Amp ^R , origen de replicación de ColE1.	Short, 1988
pBC+	Sitio múltiple de clonación en <i>lac Z</i> , Cm ^R , origen de replicación de ColE1.	Stratagene

SUBCLONAS DE LAS CEPAS BG79 Y BG85

p79H4	Subclona de la cepa BG79, fragmento Hind III-Hind III de 1.4 kb. en el pSK+. Contiene el elemento reiterado completo.	Este trabajo
p79H4-Cm	Mismo fragmento del plásmido p79H4 en el vector pBC+.	Este trabajo
p79H4B	Delección del plásmido p79H4 a partir del sitio interno Sal I al sitio Sal I del vector, fragmento de 800 pb.	Este trabajo
p79S4	Subclona de la cepa BG79, banda Sal I-Sal I de 1.6 Kb. en el pSK+. Contiene el elemento reiterado truncado a partir del sitio interno de Sal I.	Este trabajo
p79H4ΔE	Delección del plásmido p79H4 a partir del sitio interno de Eco RI al sitio Eco RI del vector.	Este trabajo
p83S3	Subclona del plásmido pBG83, fragmento Sal I-Sal I de 600 pb en pBC+. Contiene un elemento reiterado de la familia C truncado.	Este trabajo
p85P4	Subclona del plásmido pBG85, fragmento Pst I-PstI de 1.2 Kb. en pSK+. Contiene un elemento reiterado de la familia D completo.	Este trabajo
p85P4N	Mismo fragmento del plásmido p85P4 en orientación inversa en el pSK+	Este trabajo
p85P4A	Delección del plásmido p85P4N a partir del sitio interno de Sal I al sitio Sal I del vector, fragmento de 1.0 Kb.	Este trabajo
p85P4B	Delección del plásmido p85P4 a partir del sitio interno de Sal I al sitio Sal I del vector, fragmento de 200 pb.	Este trabajo
p85L-Sal	Subclona del plásmido pBG85 que contiene parte de un elemento reiterado de la familia E.	Este trabajo
p85P4-Cm	Mismo fragmento del plásmido p85P4 en el vector pBC+.	Este trabajo

4.3 Condiciones de cultivo de microorganismos.

E. coli se cultivó en medio sólido en placa y en medio líquido LB (1.5-3 ml.) a 37°C, este último con agitación de 300 rpm.

Las cepas de *Rhizobium* utilizadas se cultivaron en medio sólido en placa y en medio líquido PY (3-5 ml.) a 30°C, éste último con agitación a 250 rpm. (Noel, 1984).

4.4 Clonación de los fragmentos de restricción.

4.4.1 Digestión de DNA con endonucleasas de restricción.

Se tomó 1 µg. del DNA apropiado, el cual fue digerido con 2-4 U. de la enzima de restricción adecuada, con la solución amortiguadora respectiva en un volumen de 20 µl. a 37°C durante una hora. La digestión total del DNA se comprobó por electroforesis en gel de agarosa al 1% en TBE 1X a 100 V. corriendo 2 µl. de cada reacción de restricción mezclado con 2 µl de colorante (ficol al 15%, azul de bromofenol al 0.25% y xilencianol al 0.25% en agua bidestilada).

4.4.2 Ligación de DNA.

Los vectores y fragmentos a clonar digeridos con la endonucleasa de restricción adecuada fueron mezclados en una proporción 5:1 inserto:vector con 1U. de T4 DNA ligasa (Amersham, Life Science) para un total de 100 ng de DNA. El volumen de reacción en ligaciones intermoleculares fue de 20 µl. y en reacciones intramoleculares fue de 50 µl. La reacción se incubó 16 h. a una temperatura de 14°C.

4.4.3 Preparación de células competentes con CaCl₂.

Se inoculó, por estría cruzada, la cepa DH5α en una placa de LB y se incubó durante 24 h. a 37°C. Se tomó una colonia aislada para inocular 100 ml. de medio LB líquido, el cual se incubó con agitación (300 ciclos/s.) hasta fase exponencial (D.O._{600nm} de 0.5 a 0.7). El cultivo se transfirió a tubos de ensayo estériles y se centrifugaron 10 min. a 4 000 rpm. a 4°C; se decantó el

medio de cultivo y cada pastilla celular se resuspendió en 10 ml. de CaCl_2 100 mM. previamente enfriado en hielo para centrifugarse nuevamente 10 min. a 4 000 rpm. a 4°C. Se decantó el sobrenadante y cada pastilla fue resuspendida en 2 ml. de CaCl_2 100 mM. frío; se alicuotaron 200 μl . en tubos de microcentrífuga estériles y se guardaron a -70°C hasta su utilización.

4.4.4 Transformación de células competentes.

Se tomaron de 2 a 5 μl . de la reacción de ligación y se agregaron a los tubos con células competentes. Los tubos se colocaron en hielo durante 60 min. y al término de este tiempo se incubaron a 42°C durante 90 s. Los tubos fueron colocados en hielo nuevamente durante 10 min. A cada tubo se le agregó 1 ml. de medio LB líquido y se incubaron a 37°C con agitación suave (200 ciclos/s aprox.). Se cultivaron 200-500 μl . de la células transformadas en placas de medio LB con los antibióticos apropiados según el vector de clonación utilizado durante 24 h. a 37°C.

4.5 Extracción y purificación de DNA plasmídico.

La cepa de interés se cultivó por estría cruzada en placas de medio sólido LB conteniendo el antibiótico apropiado y se incubó por 12-16 horas a 37°C, para la obtención de colonias aisladas. Se inoculó la colonia en 3 ml. de medio LB líquido contenido en tubos de ensayo de 15 ml. a 37°C con agitación a 300 rpm. durante 12-16 horas. El cultivo líquido se transfirió a un tubo de microcentrífuga estéril y se centrifugó 1 min. a 14 000 rpm.; se descartó el medio y se secó el remanente colocando el tubo boca abajo sobre papel absorbente. La pastilla obtenida se resuspendió en 100 μl . de solución 1 (Glucosa 50 mM., tris-Cl 25 mM. pH 8.0, EDTA 10 mM) estéril previamente enfriada en hielo. Se agregaron 200 μl . de solución 2 (NaOH 0.2 N, SDS 1%) preparada justo antes de su uso; se agitó el tubo con la mezcla por inversión varias veces hasta que la suspensión celular se mostró translúcida, señal de que se realizó lisis completa, y se enfrió en hielo por 5 min. Se agregaron 150 μl . de solución 3 (60 ml. de acetato de potasio 5M, 11.5 ml. de ácido acético glacial, 28.5 ml de agua bidestilada) enfriada en hielo previamente; se homogenizó agitando por inversión varias veces y se colocó en hielo 5 min. Se centrifugó a 14 000 rpm. durante 10 min. a 4°C. El sobrenadante obtenido se transfirió a un nuevo tubo de microcentrífuga y se le agregó 1

volumen de fenol-cloroformo-alcohol isoamílico (25:24:1); la mezcla se agitó hasta que se obtuvo una mezcla homogénea. Se centrifugó 1 min. a 14 000 rpm. y la fase acuosa se transfirió a un nuevo tubo repitiendo el lavado con fenol-cloroformo-isoamílico. La fase acuosa obtenida se sometió a un lavado con cloroformo (1 volumen) para retirar cualquier posible resto de fenol. La fase acuosa fue separada de la fase orgánica, se le agregaron 2 volúmenes de etanol al 100% enfriado en hielo, se agitó por inversión y posteriormente se centrifugó 15 min. a 14 000 rpm. a 4°C. La pastilla obtenida se lavó con 1 ml. de etanol al 70% y se centrifugó nuevamente a 14 000 rpm. 15 min. a 4°C. La pastilla se secó al vacío con centrifugación en un concentrador Savant (Speedvac® DNA110) y se resuspendió en 20 µl. de TE 10:1 para ensayos de digestiones con enzimas de restricción y en 20 µl. de agua bidestilada estéril para secuenciación con 2 µl de RNAsa (1 mg. ml.⁻¹).

4.6 Electroforesis verticales de DNA en geles de agarosa.

Se prepararon geles de agarosa al 1% en TBE 1X en cámaras para electroforesis vertical (Hoefer Scientific Instruments), se limpiaron los restos de agarosa de los pozos con tiras de papel Whatman. Las muestras fueron cargadas en los pozos con colorante (ficol, azul de bromofenol y xilencianol en agua bidestilada) y se corrió 12 h. a 4°C a 40 V. El gel fue teñido con solución de bromuro de etidio durante 10 min. y se enjuagó con agua destilada 10 min. Se tomó una fotografía al gel en un transiluminador de luz U.V. equipado con una unidad de procesamiento de imagen computarizada (Eagle Eye II™, Stratagene®).

4.7. Electroforesis horizontal en minigeles de agarosa.

EL procedimiento es básicamente el mismo utilizado para la electroforesis de geles verticales, salvo que los minigeles (agarosa 1%) se prepararon en cámaras horizontales (Horizon®11.14 Life Technologies™, QSH International Biotechnologies, Inc.) y se corrieron a 100 V. durante 60 min.

4.8 Cuantificación de DNA.

La cuantificación de DNA se llevó a cabo en un fluorómetro para DNA modelo TKO-100 (Hofer Scientific Instruments) el que se dejó calentando 15 min. antes de llevar a cabo la medición. Se agregaron 2 ml. de solución de fluoreceína y el aparato se calibró a cero; se agregaron 2 μ l. del DNA en solución y se homogenizó. La lectura se interpretó directamente en ng. / μ l.

4.9 Transferencia de DNA a filtros de nylon.

Los geles sometidos a electroforesis, previamente constatado su adecuado corrimiento mediante tinción con bromuro de etidio, fueron lavados dos veces con una solución de HCl 0.25 M durante 15 min. a 4°C. Posteriormente se hicieron dos lavados con una solución de NaOH 0.5 M-NaCl 1.5 M durante 15 min. a 4°C y otros dos lavados de 30 min. con una solución Tris 0.5 M-NaCl 1.5 M. Los geles se colocaron en cristales y los filtros de Nylon Hybond de 0.45 μ (Amersham, Life Sciences), previamente humedecidos en una solución SSC 6X, se pusieron sobre la superficie del cada gel, sobre estos se colocaron dos filtros Whatman, también humedecidos en SSC 6X, 32 capas de papel absorbente doble y un cristal; la posición del dispositivo se invirtió para poner los mismos elementos en la otra cara del gel. Una vez montado, se colocó un peso de 1 kg. sobre el dispositivo y se mantuvo a temperatura ambiente durante 16 h., al término de las cuales se marcaron los carriles del gel sobre los filtros de nylon. Los filtros se humedecieron en SSC 6X y se secaron entre papel absorbente en un horno a 80°C y 15 lb/in² durante 2 h.

4.10 Marcaje de la sonda de DNA con [α -³²P] dCTP.

El DNA que se utilizó como sonda para los ensayos de hibridación fue aislado de digestiones totales con las endonucleasas de restricción correspondientes por electroforesis en gel de agarosa y recuperado mediante el sistema de aislamiento de DNA Wizard™ PCR Preps Purification System (Promega). El DNA aislado, de 100 a 300 ng., fue aforado a 45 μ l. y desnaturalizado a 95°C 10 min., se dejó enfriar a temperatura ambiente. El DNA se agregó a un tubo del sistema de marcaje radioactivo Rediprime

(Amersham, Life Science) y se resuspendió golpeando suavemente el tubo. Se agregaron 3 μ l. de [α - 32 P] dCTP (10 μ Ci/ μ l.) (Amersham, Life Science) y se incubó durante 1 h. a 37°C. Se desnaturalizó el DNA marcado de la reacción por incubación a 95°C durante 10 min. y se enfrió en hielo 10 min. Se tomaron 950 μ l. de la solución amortiguadora de prehibridización caliente (65°C) y se agregaron a la reacción de marcaje; una vez homogenizada, se agregó en el fondo del tubo de hibridización respectivo.

4.11 Hibridización tipo Southern.

4.11.1 Prehibridización de filtros.

Se colocaron los filtros de nylon en tubos para hibridización (Amersham, Life Science), hasta 4 filtros, y se agregaron de 15 a 30 ml. de solución amortiguadora de hibridización rápida Rapid-hyb buffer (Amersham, Life Science) y se prehibridizó 4-12 h. en un horno con rotación a 65°C.

4.11.2 Hibridización de filtros.

Se agregó la sonda marcada y se dejó hibridizando 12 h. bajo las mismas condiciones de prehibridización.

4.11.3 Lavado de los filtros.

Al término del tiempo de hibridización se desechó la solución amortiguadora y se enjuagaron los filtros con 50 ml. de solución SSC 2X-SDS 0.1% y se hizo un lavado con otros 50 ml. de la misma solución durante 15 min. con rotación a temperatura ambiente. Posteriormente se hicieron tres lavados de 30 min. a 55°C con rotación con 50 ml. de una solución de SSC 0.1X-SDS 0.1% y un lavado final durante 15 min con una solución SSC 0.1X a 55°C con rotación.

4.11.4 Secado de los filtros.

Los filtros fueron secados en papel absorbente a 55°C durante 20 min., montados en cartulina, se forraron con plástico autoadherible y se colocaron en un cassette para autoradiografías.

4.11.5 Exposición y revelado de la placa fotográfica.

Se colocó en el cassette con los filtros una película fotográfica (Kodak X-Omat KX K1) en oscuridad total y se mantuvo a -70°C un tiempo variable. La placa fue revelada en oscuridad total en solución de revelado durante 30 s, se lavó en agua 30 s. y se sumergió en solución fijadora 1 min.; se enjuagó nuevamente en agua corriente.

4.12 Diseño de oligonucleótidos para secuenciación.

Los oligonucleótidos utilizados fueron diseñados con el programa Oligo 5.0, considerando una $^{\circ}T_M$ de 60°C, ausencia de complementariedad intra e intermolecular y un contenido de GC cercano al 50%.

4.13 Secuenciación automática de DNA.

La secuenciación se las clonas obtenidas, así como sus extensiones con oligonucleótidos, se llevaron a cabo por el método de terminación de cadena por didesoxinucleótidos fluorados con el equipo Taq DyeDeoxy™ Terminator Cycle Sequencing Kit. Las reacciones de secuenciación se sometieron a electroforesis en gel de poliacrilamida en condiciones desnaturizantes y fueron analizados en un secuenciador 373A DNA; tanto el equipo utilizado como el secuenciador son de Applied Biosystems. El DNA utilizado en las reacciones de secuenciación se obtuvo a partir de cultivos líquidos de la clona de interés mediante lisis alcalina con el equipo Wizard™ Minipreps (Promega).

4.14 Ensamblaje y análisis de las secuencias de DNA.

El ensamblaje de las secuencias de las clonas obtenidas, los análisis de búsqueda de sitios de restricción y composición, así como de sus extensiones con oligonucleótidos, se realizaron con el paquete Gene Works 2.5.1 (IntelliGenetics, Inc.). Los análisis FASTA (Person, W. R. y Lipman, D. J., 1988), BLASTX (Altschul, S. F. et al, 1990) y TESTCODE (Fickett, J. 1982)) contenidos en el paquete GCG versión 8.0 (Genetics Computer Group Inc.), se utilizaron para realizar análisis de identidad contra las bases de datos GenBank y Swisprot.

5. RESULTADOS

Al llevarse a cabo la formación del mapa físico del pSim de *R. etli* CFN42 se realizó la descripción de la complejidad de este plásmido tomando en cuenta su contenido de secuencias reiteradas, localizando los elementos de cada familia y se consideró que el tamaño mínimo de cada elemento, dada la resolución de la técnica utilizada, sería de aproximadamente 200 pb. con identidad $\geq 80\%$ (Girard, 1991) (Fig. 1).

La secuenciación y caracterización de los elementos de DNA reiterado de las familias B, C, D y E del plásmido simbiótico (pSim) de *Rhizobium etli* CFN42 (Cuadro 1) se llevó a cabo mediante la búsqueda de un fragmento de restricción susceptible a subclonación, el cual contuviera de forma total o parcial el elemento reiterado; lo anterior se evidenció por experimentos tipo southern. La clona de la genoteca de fragmentos de Bam HI de menor tamaño que contiene al elemento reiterado se utilizó como sonda en transferencias de digestiones sencillas de la otra clona con el otro elemento de la misma familia. Las endonucleasas de restricción utilizadas fueron Eco RI, Hind III, Pst I, Sal I y Xho I. Para definir si los fragmentos revelados poseían secuencias del vector, se realizó otra hibridización con digestiones dobles, Bam HI con el resto de endonucleasas (Bam HI-Eco RI, Bam HI-Hind III, etc.); en este ensayo se pudieron identificar los fragmentos del vector por su tamaño, ya que también presentaron hibridización, debido a contaminación durante la extracción del inserto de la clona utilizada como sonda (Fig. 2). Se seleccionó el fragmento que pudiera secuenciarse en el menor número de pasos posibles y se subclonó en los vectores pBC y pSK. La secuencia de las subclonas obtenidas se llevaron a cabo a partir de los oligonucleótidos para secuenciación que poseen los vectores mencionados, lesiones, y otras construcciones, así como también con oligonucleótidos diseñados para extender la secuencia. El ensamblaje y el análisis de las secuencias se realizaron con los paquetes GCG y GeneWorks.

5.1 Elementos reiterados de la familia B.

La familia B de secuencias reiteradas se compone de tres elementos, detectados en las bandas BG02 (1.6 kb.), BG28 (6.7 kb.) y BG81 (3.8 kb.) del mapa de bandas de Bam HI del pSim de acuerdo a Girard (1991) (Fig 1, Cuadro

1). Los elementos de las bandas BG81 y BG28 forman los extremos de los amplicones tipo IV, previamente descritos (Romero, 1995), lo cual, evidencia que estas se encuentran en orientación directa dentro del pSim.

5.1.1 Análisis de la secuencia del pBG02.

La secuencia de la clona pBG02 se obtuvo durante el estudio de las familias C, D y E, dentro del proyecto general de secuenciación del pSim. La secuencia tiene un tamaño de 1 614 pb., su composición probable es de 56.13% de GC y 43.87% de AT, dado que se detectaron ambigüedades en la secuencia (Y, C o T, posiciones 1 329 y 1 357; S, C o G, posición 1 321; W, A o T, posición 1 314).

La secuencia obtenida de la clona pBG02 se analizó con el programa FASTA contra el banco de datos GeneBank y se obtuvieron identidades en el extremo 5' del 70.07 % en 548 pb. con la sección 42 (bases 5194-5737) y del 75.84 % en 414 pb con la sección 39 (bases 8254-8661) del pNGR234a de *R. sp.* NGR234. Las regiones del pNGR234a poseen una probable ferredoxina (FDIII, MLA y4vS) y un MLA similar a la región intergénica entre *nifX* y *nifW* en bacterias fijadoras de nitrógeno (MLA y4vQ y y4dX), lo anterior se obtuvo por medio de los programas BLASTX y FASTA. Por BLASTN se obtuvieron las mismas zonas de identidad. El extremo 3', a partir de la base 800 aproximadamente hasta la 1614 no mostró identidad con ninguna secuencia reportada con los programas utilizados.

5.1.2 Conformación de la familia de secuencias reiteradas B-F.

Se realizó un FASTA con la secuencia del pBG02 contra el banco de secuencias que se tiene del pSim de *R. etli* CFN42, con el objeto de encontrar el resto de elementos reiterados. Los resultados mostraron una alta identidad (95.8% en 331 pb) con la secuencia parcial de la banda BG81 y la banda BG28 (83.3% en 329 pb), en estas bandas se encuentran los dos restantes elementos de la familia B. La región de pBG02 en que se observaron los resultados anteriores corresponde al extremo 3' aledaño a la banda BG03. Las regiones de las otras bandas donde se encontró identidad se sitúan en los extremos y dado que la identidad aumenta en estos sitios se supuso que el elemento reiterado continua en la banda aledaña. Para probar la hipótesis anterior se tomó la secuencia que une las bandas BG02 y BG03 y

se hizo un FASTA nuevamente contra el banco de secuencias del pSim. Los resultados revelaron una alta identidad de las bandas BG27 (94.0 % en 436 pb) y BG80 (93.6 % en 578 pb) con el extremo perteneciente a la BG03. Dado que los elementos de la familia F se encuentran en las bandas BG80 y BG27, se asume que en realidad los elementos de las familias B y F conforman una sola familia, la cual fue identificada de forma individual por el sitio de Bam HI que se encuentra hacia la zona central de estos elementos. Dentro de la familia F no se había reportado ningún elemento contenido en la banda BG03, lo anterior se debió a que esta banda contiene el TN5-Mob utilizado para movilizar el pSim a *Agrobacterium* (Girard, 1991), lo cual dificultó la búsqueda de elementos reiterados por hibridización (Dávila, comunicación personal). De acuerdo a lo anterior, se propone la designación de estos elementos dentro de la familia BF como BF02-03, BF27-28 y BF80-81.

5.1.3 Análisis de las secuencias de los elementos reiterados de la familia BF.

Los elementos reiterados de la familia BF poseen un tamaño de aproximadamente 800 pb., y muestran alta identidad entre ellos (BF02-03 con BF27-28, 86.6% en 767 pb; BF02-03 con BF80-81, 94.4% en 891 pb; BF27-28 con BF80-81, 87.4% en 788 pb.) con una región en apariencia idéntica de 714 pb. entre los tres elementos (ver discusión) (Figura3).

La secuencia del elemento BF02-03 posee una composición probable de 53.58% de GC y 46.42% de AT, ya que se detectaron ambigüedades en la secuencia (R, A o G, posiciones 568 y 674; Y, C o T, posiciones 185 y 213; W, A o T, posición 170; S, C o G, posición 177). El elemento BF27-28 tiene una composición 52.29% de GC y 47.71% de AT y el elemento BF80-81 53.81% de GC y 46.19% AT.

Las secuencias se analizaron individualmente por FASTA encontrándose resultados similares para los tres elementos, dada la alta conservación entre ellos. Los elementos BF02-03 y BF80-81 mostraron identidades de 61.48% en 296 pb (bases 9470-9766) y 59.81% en 316 pb. (bases 9449-9757) respectivamente con la sección 41 del pNGR234a que contiene un MLA designado Y4wP, evidenciado con BLASTX, el cual se incluye entre MLA's de función desconocida no relacionados con secuencias de inserción. La secuencia del elemento BF27-28 no mostró identidad

significativa con ninguna secuencia reportada, a nivel de DNA o proteína a pesar de alta identidad con los otros miembros de la familia BF (Fig. 7).

La región conservada entre los tres elementos que posee 714 pb., obtenida por alineamiento de las secuencias de los tres elementos, fue analizada también por FASTA, el cual mostró que esta región no posee una identidad significativa con ninguna secuencia reportada. Estos resultados se comprobaron con los programas BLASTX y BLASTN.

Se buscaron regiones codificantes en las secuencias de los elementos de la familia BF con el programa GeneWorks, encontrándose solo un MLA en fase 2 de 202 aminoácidos con alta probabilidad de codificar una proteína en la secuencia BF80-81, evidenciado por los resultados obtenidos con el programa TESTCODE, el cual evalúa la posibilidad de codificación de una secuencia tomando en cuenta la dispensabilidad de la tercer posición del codón. La secuencia fue analizada con distintos programas que identifican motivos de proteínas conocidas (MOTIFS, BLOCKS y PROFILESCAN) sin encontrar ninguno.

Las secuencias de los elementos de la familia BF fueron analizados con el programa REPEAT para buscar posibles secuencias repetidas en orientación directa o inversa que pudieran presentar significado estructural secundario o ser parecidas a los extremos de alguna secuencia de inserción reportada. No se encontraron secuencias repetidas en orientación directa ni inversa significativas.

5.2 Elementos reiterados de la familia C.

Los elementos reiterados que conforman la familia C se localizan en las bandas BG24 (5.19 Kb.), BG82 (3.4 Kb.) y BG83 (7.2 Kb.) del mapa físico del pSim (Fig. 1). Los elementos reiterados de las bandas BG83 y BG24 franquean los amplicones tipo II (Romero, 1995), por lo que se deduce que se encuentran en posición directa dentro del pSim. Se utilizó el inserto de la clona pBG82 de la genoteca de fragmentos de Bam HI como sonda en las transferencias de digestiones de la clona BG83.

5.2.1 Localización del fragmento de restricción que contiene el elemento reiterado de la familia C de la banda BG83.

Los ensayos de hibridación revelaron un fragmento en cada uno de los carriles de la transferencia. El fragmento de menor tamaño se localizó en el carril de Sal I, con un tamaño aproximado de 600 pb (Fig.2). En los ensayos de hibridación con transferencias de digestiones dobles se constató que el fragmento localizado posee extremos idénticos de la misma enzima y se clonó en el vector pBC cortado con la misma enzima.

5.2.2 Subclonación del fragmento de restricción que contiene el elemento reiterado de la familia C en el pBG83.

El fragmento de 600 pb. de la BG83 digerida con Sal I localizado por hibridación con la BG82 fué clonado en el vector pBC en el sitio único de corte para esta enzima, originando el p83S4.

5.2.3 Secuenciación de las construcciones que contienen el elemento reiterado de la familia C en el pBG83.

Dado el tamaño del fragmento clonado (600 pb. aproximadamente), se secuenció completamente en cadena sencilla desde los extremos del vector con los oligonucleótidos SKE y T7, diseñados para secuenciación. Las secuencias se sobrelaparon en la región central del inserto, ya que la resolución de la secuencia logró un promedio de 400 nucleótidos con alta confiabilidad.

Una vez ensamblada la secuencia del p83S4, se comparó con las secuencias obtenidas de otras regiones de la pBG83, revelando una secuencia sobrelapante en uno de sus extremos, la cual corresponde a la clona pBG83-h2T7. Se integraron las dos secuencias en una sola (83S4ex) y se utilizó para realizar los análisis de composición y comparación de secuencias.

5.2.4 Análisis de la secuencia 83S4ex que contienen el elemento reiterado de la familia C de la clona pBG83.

La secuencia 83S4ex posee una longitud de 894 pb. y tiene una composición probable de 56.21% de GC y 44.79% de AT, ya que se encontraron

ambigüedades en las posiciones 436, 517, 518, 526, 541 (S, C o G), 523 (K, G o T), 536 (Y, C o T), 542 (W, A o T) y 528 (R, A o G).

El análisis tipo FASTA de la secuencia 83S4ex mostró una identidad de 59.8% en 443 pb. en uno de sus extremos con el MLA1 de la IS407 de *Pseudomonas cepacia* (1 236 pb.), la cual posee similitud con la ISR1 de *R. sp.* clase IV y con otras IS's de la familias IS3 (Wood, 1991). Por medio del programa BLASTX se obtuvo una identidad del 50% en 136 aa. con el mismo MLA del *P. cepacia* (Fig. 4 y 8), lo cual se comprobó al traducir la secuencia en sus 6 marcos de lectura y encontrar los mismos aa. en la fase 2 de lectura. Aunado a lo anterior, los resultados obtenidos con el programa TESTCODE muestran una región con alta probabilidad de codificación en la misma zona, denotando un MLA continuo desde la base 53 hasta la 625.

La secuencia se analizó con el programa REPEAT sin encontrarse secuencias directas o inversas repetidas significativas.

5.3 Elementos reiterados de la familia D.

Los elementos reiterados de la familia D se localizan en las bandas BG01 (9.4 Kb.), BG83 (7.2 Kb) y BG85 (12.9 Kb.) del mapa de físico del pSim (Girard, 1991)(Fig. 1, Cuadro 1). Se utilizó el inserto de la clona pBG83 como sondas en transferencias de digestiones de DNA de la clona BG85.

5.3.1 Localización del fragmento de restricción que contiene el elemento reiterado de la familia D en el pBG85.

En todos los carriles se presentó hibridización en mas de una banda, siendo una de la digestión con Pst I, con un tamaño de 1.2 Kb., la inicialmente seleccionada, así como las bandas de la digestión con Sal I de 1.6 y 1.1 Kb. Dado que no fue posible mediante este ensayo discernir los fragmentos que hibridizaron por el vector y no por el elemento reiterado, se realizaron los ensayos de hibridización con transferencias de digestiones dobles. Los fragmentos localizados inicialmente mostraron poseer efectivamente el elemento reiterado, y en el caso de los dos fragmentos en el carril de Sal I, el sitio de corte se encontró dentro del elemento reiterado (Fig 2).

5.3.2 Subclonación del fragmento de restricción que contiene el elemento reiterado de la familia D en el pBG85.

La subclonación de los fragmentos localizados, designados 85P4, debido a que en digestión sencilla con Pst I este fragmento corresponde al cuarto en tamaño descendente, se llevó a cabo en los vectores pBC y pSK en su dos orientaciones posibles, obteniéndose las subclonas p85P4-Cm, p85P4N-Cm, p85P4N y p85P4. Las subclonas p85P4N y p85P4 fueron deletadas a partir del sitio interno de Sal I al sitio del vector, obteniéndose las subclonas p85P4B y p85P4A respectivamente.

5.3.3 Secuenciación de las construcciones que contienen el elemento reiterado de la familia D en el pBG85.

La secuenciación de las subclonas p85P4-Cm, p85P4N-Cm, p85P4N, p85P4, p85P4B y p85P4A se llevó a cabo de forma automática a partir de los oligonucleótidos diseñados para secuencia de los vectores, SKE y T7. Dado que el sitio interno de Sal I se localizó a 200 pb. de uno de los extremos de PstI del inserto, no fue posible que las secuencias obtenidas de los extremos se sobrelaparan, por lo que se diseñaron los oligonucleótidos 85P4.SKE.in y 85P4.T7.in, que parten de las secuencias obtenidas de los extremos hacia el interior del inserto. Se adicionó la secuencia de la clona pBG85L con el oligonucleótido T7, ya que al realizar los ensamblajes para obtener la secuencia del p85P4, se encontró que esta secuencia parte del sitio de Sal I hacia uno de los extremos de Pst I de la subclona. El pBG85L fue obtenido por delección con Sal I, a partir de la clona pBG85-SK, la cual posee el mismo inserto que la pBG85, pero en el vector pSK+.

5.3.4 Análisis de la secuencia del p85P4 que contiene el elemento reiterado de la familia D de la clona pBG85.

La secuencia del p85P4 consta de 1 132 pb. y posee una composición probable de 57.11% de GC y un 42.89% de AT, dado que se detectaron ambigüedades en el armado de la secuencia final (W, A o T, posición 309; N, cualquiera de los 4 desoxinucleótidos, posición 337; S, C o G, posición 347; Y, C o T, posición 780; M, A o C, posición 985).

El análisis por FASTA reveló una identidad del 55.9% en 742 pb. con la secuencia de inserción IS427 de *Agrobacterium tumefaciens*, la cual posee en esta región tres MLA, el MLA1 (bases 514 a la 759), el MLA 3 (inicio en la base 293 sin codón de terminación) y MLA4 (bases 397 a la 812, probable marco de una transposasa), los dos últimos en la cadena complementaria (de Mairsman, 1990) (Fig. 5). El análisis por BLASTN mostró identidades de 63% y 71% con dos fragmentos de 118 (bases 195 a 77) y 36 pb (bases 394 a 356) con una transposasa de un elemento de inserción de *Micobacterium tuberculosis*. Los resultados que se obtuvieron por medio de BLASTX mostraron identidad en varios fragmentos que comprenden desde la base 21 hasta la 233 (58%, bases de la 40 a 80; 40%, bases de la 182 a 233; 45%, bases de la 103 a 139; 58%, bases de la 21 a 32; 66%, bases de la 155 a 166) con una probable transposasa de la secuencia de inserción IS6501 de *Brucella ovis* (Fig. 8).

Dado que los análisis de la secuencia del p85P4 revelaron identidad con diferentes marcos de lectura de transposasas, se buscaron posibles MLA con el programa GeneWorks en los seis marcos posibles. Los resultados revelaron cinco MLA probables, en el marco 1 (bases 514 a 882), marco 2 (bases 572 a 835 y 914 a 1021), marco -2 (bases 813 a 463) y marco -3 (bases 416 a 147). El programa TESTCODE mostró alta probabilidad de codificación en la misma zona (bases 300 a 600), sin embargo, la existencia de ambigüedades en la secuencia evidenció la posibilidad de encontrar cambios de fase y, consecuentemente, posibles terminaciones. Para identificar el posible MLA alterado por fallas de secuencia, se realizó un alineamiento con las secuencias de mayor identidad reveladas por BLASTX (transposasas de la IS6501, de un elemento de transposición de *Micobacterium tuberculosis*, de la IS711 y la proteína hipotética 1 de *Agrobacterium tumefaciens*), con lo que se identificaron los aminoácidos conservados y se comprobó su presencia en alguno de los seis marcos posibles de la secuencia p85P4. Se encontró una región conservada hacia el carboxilo terminal constituida por 39 aa para las secuencias alineadas, de los cuales, en la traducción de la secuencia de p85P4, se encontraron 26 aa discontinuos en los tres marcos negativos, lo cual constituye un 66.6% de los aa. comunes a estas secuencias y da evidencia del posible cambio de fase.

La secuencia del p85P4 fue analizada con el programa REPEAT y no se encontraron secuencias directas ni inversas repetidas significantes, así mismo, se comparó con la secuencia de la banda BG01, en la que previamente se había reportado la existencia de un elemento de la familia D, sin encontrarse identidad con esta.

5.4 Elementos reiterados de la familia E.

Los elementos reiterados de la familia E se encuentran en la bandas BG21 (33.6 Kb.), BG79 (16.8 Kb.) y BG85 (12.9 Kb.) (Fig. 1, Cuadro1). Cabe destacar que en la banda BG79 previamente se localizó el origen de replicación del pSim (Ramírez-Romero, 1997). Se utilizó el inserto de la clona pBG85 como sonda en transferencias de las digestiones del DNA de la clona pBG79.

5.4.1 Localización del fragmento de restricción que contiene el elemento reiterado de la familia E en el pBG79.

Los ensayos de hibridización en las transferencias de digestiones sencillas del DNA de la clona pBG79 mostraron tres fragmentos, excepto en el carril de Xho I, ya que esta clona solo posee dos sitios de corte con esta enzima. Se identificó una banda de 1.36 Kb. en el carril de la endonucleasa Hind III, por lo que se procedió a descartar los fragmentos revelados por el vector mediante hibridizaciones en transferencias de digestiones dobles. Además de la banda de 1.36 Kb. localizada, fueron reveladas dos fragmentos de la digestión con Sal I, con tamaños de 3.2 y 1.83 Kb., con lo que se intuyó la presencia de un sitio interno de corte de Sal I en el fragmento de 1.36 Kb de la digestión con Hind III (Fig. 2).

5.4.2 Subclonación del fragmento de restricción que contiene el elemento reiterado de la familia E en el pBG79.

Los fragmentos seleccionados de las restricciones de las endonucleasas Hind III y Sal I, designados 79H4, 79S2 y 79S4 respectivamente se subclonaron en el vector pSK+. Se obtuvieron las subclonas p79H4, p79S2 y p79S4. El fragmento 79H4 también fue clonado en el pBC, tratando de tener el fragmento en sus dos orientaciones, pero esto no

fue posible, todos los ensayos realizados para subclonar este fragmento en la orientación contraria a la presentada en el p79H4, partiendo de digestiones completas de DNA del pBG79 y aisladas de gel de la misma digestión, dieron resultados negativos. A partir de la subclona p79H4 se obtuvo la p79H4B, la cual posee una delección a partir del sitio interno de Sal I que posee, el cual se encontró a 600 nucleótidos del extremo T7 del inserto. La subclona p79H4ΔE, se obtuvo al deletar un fragmento de la subclona p79H4 a partir de un sitio Eco RI, evidenciado por la secuencia del extremo T7; el sitio Eco RI se encontró a 310 nucleótidos del extremo T7 de p79H4B.

5.4.3 Secuenciación de las construcciones que contienen el elemento reiterado de la familia E en el pBG79.

Se obtuvieron las secuencias de los extremos de las construcciones obtenidas (p79H4, p79S2, p79S4 y p79H4B) con los oligonucleótidos T7 y SKE; también se diseñaron los oligonucleótidos 79H4.T7.in y 79H4B.in para la obtención de la zona central de la secuencia, ya que no se logró la extensión suficiente de los extremos para que se sobrelaparan.

5.4.4 Análisis de la secuencia p79H4 que contienen el elemento reiterado de la familia E en el pBG79.

La secuencia del p79H4 mostró una longitud de 1 358 pb. y una composición probable de 60.54 % de GC y 39.46 % de AT debido a ambigüedades en la secuencia (S, C o G posiciones 625 y 1 138; V, A, C, o G posición 432).

El análisis tipo FASTA reveló una identidad del 76.1 % en 707 pb. con el MLAB de la secuencia de inserción ISRM6 de *R. meliloti* (Fig 6), también mostró identidad con la IS511 de *Caulobacter crescentus* (69.1% en 693 bp.) y con la región intergénica entre *syrM* y *Nod D3* de *Rhizobium meliloti* (72.9% en 520 bp); esta última dos MLA's similares a los encontrados en las secuencias de inserción de la familia de IS3 (Barnett, 1996). Los resultados de los análisis por BLASTN y BLASTX mostraron identidad con la misma región del MLAB, sin embargo, de acuerdo con la organización de los MAL's de la ISRM6, no se encontró identidad en el resto de la secuencia del p79H4 con el MLAA. La secuencia del p79H4 se tradujo a los seis MLA's posibles y se

buscaron las regiones de identidad mostradas por el análisis BLASTX de la proteína hipotética de 33.7 KDa, que probablemente codifica el MLAB; lo anterior evidenció cambios de fase de lectura, dados probablemente por errores de secuencia. Se encontró el motivo DDE que poseen las transposasas de la familia IS3 e integrasas de retrovirus (Zecrí, 1996; Polard, 1995), aunque en distintos marcos de lectura (-1 y -2). La zona de identidad de la secuencia del p79H4 con el MLAB de la ISRM6 no abarca en su totalidad esta última, si bien se extiende 720 pb hacia el extremo 5' (Fig. 9) y no hay señal de término de traducción en el extremo 3', por lo que se asume que la zona de identidad se extiende hacia el extremo 3' fuera del p79H4 aproximadamente 100 pb, lo anterior con base en la secuencia de aa. del MLAB. Para constatar lo anterior, se diseñó un oligonucleótido a partir del extremo 3' del p79H4 para obtener la secuencia aledaña. Actualmente se encuentra en proceso de secuenciación con DNA del pBG79.

5.4.5 Secuencia del elemento reiterado de la familia E del pBG85.

Se realizó un análisis tipo FASTA contra el banco de secuencias del pSim de CFN24 para encontrar los elementos restantes de la familia E, que posiblemente estuvieran secuenciados. Se encontraron dos secuencias de dos subclonas del pBG85, pBG85L-h3SK y pBG85L-s4SK, con identidades de 77.4% en 385 pb. y 74.8% en 321 pb. respectivamente; estas dos secuencias abarcan 690 pb. del p79H4 y son contiguas sin sobrelaparse.

Se realizaron análisis tipo FASTA, BLASTN Y BLASTX de las secuencias pBG85L-h3SK y pBG85L-s4SK, los cuales coincidieron con la identidad mostrada con la secuencia del p79H4 a nivel de proteína y de DNA. Dado lo anterior, el elemento reiterado de la familia E de la BG85 corresponde en 700 pb. a la secuencia del p79H4 (Fig. 9).

5.5 Análisis estadístico del porcentaje de GC de las secuencias obtenidas.

Los porcentajes de GC obtenidos de cada secuencia fueron comparados con el promedio calculado a partir de una muestra de 50 secuencias elegidas al azar de 300 pb. de otras regiones del pSim (Tabla 1), lo anterior se realizó con la prueba de t de Student a un nivel del 95% de confianza (Campbell,

1989). Tomando una media muestral de 56.21% de GC y una desviación estandard de 4.4698 de las secuencias del pSim, se obtuvieron diferencias significativas con las secuencias de los tres elementos de la familia reiterada BF y el elemento de la familia E, mientras que las secuencias de los elementos de las familias D1 y C no fueron significativamente diferentes.

6. DISCUSIÓN

La presencia de secuencias de DNA reiterado se ha descrito ampliamente en bacterias (Kiss, 1977; Sapienza, 1982a; Young, 1984; Sonti, 1989; Bar-Nir, 1994), confirmándose con las secuencias de genomas completos (Peterson, 1995; Bult, 1996; Himmelreich, 1996; Tomb, 1997; Blattner, 1997; Smith, 1997; Kunst, 1997; Klenk, 1997; Fraser, 1997). En bacterias Rizobiaceas se ha considerado como una característica distintiva de ésta familia (Flores, 1987; Martínez, 1990). En la conformación del mapa físico del pSim de *R. etli* CFN42 (Girard, 1991) se describieron 11 familias de secuencias reiteradas de dos y tres elementos cada una, con una extensión mínima de 200 pb. y 80% de identidad en su secuencia, excepto la familia A, lo anterior se dedujo por hibridización con la técnica de marcaje radioactivo utilizada ("nick translation"). La única familia reiterada del pSim secuenciada, hasta antes del presente estudio, era la familia A (Dávila, comunicación personal) la cual se compone de los inicialmente descritos genes *nifH* (Quinto, 1982), así como de los genes *nifD* y *nifK* en los elementos de las bandas en orientación directa 1 y 25 del mapa físico; el elemento de la banda 13 solo contiene el gen *nifH* en orientación inversa respecto a los otros dos elementos (Girard, 1991). La extensión de los dos operones *nifHDK* es de 8.7 kb. (Dávila, comunicación personal), que al igual que los 200 pb. calculados para el resto de las familias, se encuentra dentro del amplio rango de tamaños de secuencias reiteradas reportadas, desde los elementos RHS de *E. coli* (hasta 9 kb.) (Blattner, 1997) hasta las pequeñas secuencias repetidas (30 pb.) en tándem que conforman los RS de *Archaeoglobus fulgidus* (Klenk, 1997). Las secuencias de los tres elementos de la familias B-F, evidenciados por comparación de secuencias de otras del pSim (ver Resultados), y dos elementos de la familia E, demuestran que, al menos estas familias, poseen una extensión aproximada de 700 pb. con identidades de alrededor del 80%. Las clonas donde se localizaron los elementos de las familias C y D ya fueron secuenciadas y poseen longitudes de 874 y 1132 pb., sin embargo, no se tendrán la longitudes exactas de los elementos de estas familias hasta no poseer las secuencias de los otros elementos y efectuar su comparación.

Al obtenerse las secuencias que se presentan en este trabajo, no solo se evidenció que las reiteraciones del pSim pueden poseer tamaños cercanos a una kilobase, sino que también sugieren la reordenación de los elementos

de las familias BF y D en el mapa físico del pSim. Los elementos reiterados BF se encuentran justo en torno a los sitios Bam HI entre las bandas BG02 y BG03, BG27 y BG28, así como entre BG80 y BG81. Al comparar la secuencia de la clona portadora del elemento reiterado D de la banda BG85 (p85P4), evidenciada por hibridización con la banda BG83, se comparó con la secuencia completa de la banda BG01, en la que originalmente se había descrito otro elemento de esta familia (Girard, 1991). No se encontró identidad entre la secuencia de la BG01 y el p85P4, lo cual sugiere que la familia D de tres elementos, en realidad está compuesta por tres familias de dos elementos: D1, con elementos en las bandas BG83 y BG85; D2, con elementos en las bandas BG85 y BG01; por último D3, con elementos en las bandas BG01 y BG83 (Fig. 10). El análisis de la secuencia de la banda BG01 muestra MLA's que comprenden los genes *nodA* (Girard 1991), un fragmento de *nodB*, el operón *nifHDK*. (Quinto, 1982; Quinto 1985) y una región intergénica de 1.72 kb. (entre *nodA* y *nifH*) que tiene regiones de baja identidad con MLA's de *Caenorhabditis elegans* y con elementos transponibles de *Drosophila* (Dávila, comunicación personal), por lo que las familias D2 y D3, las cuales posee un elemento en esta banda, probablemente puedan estar constituidas por los genes, o fragmentos de ellos, *nifD* o *nifK*, dado que no se encontraron otras reiteraciones en estudios previos de hibridización con sondas para *nodA* y *nifH* (Quinto, 1982; Quinto 1985; Girard,1991); también podría encontrarse en la zona intergénica y no poseer identidad elevada con alguna secuencia reportada, como es el caso de la familia BF. Los siguientes estudios deberán encaminarse a la secuenciación y análisis de las familias D2 y D3 (Fig. 10, Cuadro 2).

Las secuencias reiteradas en Rizobiceas se han descrito como operones completos (Quinto, 1982; Martínez, 1985; Quinto, 1985; Freiberg, 1997, Dávila, comunicación personal), genes (David, 1987; Renalier, 1987; Honma, 1987; Appelbaum,1988; Barbour, 1985) secuencias regulatorias (Schofield, 1986; Balleza, 1998), secuencias con función desconocida (Masterson, 1986; Hahn, 1987) y secuencias de inserción, estas últimas franqueando grupos de genes relacionados (Freiberg, 1997; Mergaert, 1997) o como inserciones sencillas del propio elemento (ver introducción). Las secuencia de la familia C de la BG83, el elemento de la BG85 de la familia D1 y los dos elementos de la familia E de las bandas BG79 y BG85 muestran identidades con secuencias de inserción a nivel de nucleótidos y aminoácidos de proteínas hipotéticas relacionadas con transposasas. Considerando las identidades encontradas

entre los miembros de las familias con dos o más elementos secuenciados, los datos de hibridación realizados en el establecimiento de cada familia, así como en este trabajo, para delimitar cada elemento secuenciado, se puede asumir una identidad similar entre los elementos de cada familia con las secuencias encontradas por comparación con bancos de datos. De acuerdo a lo anterior, las familias C, D1, y E muestran identidades con la IS407 de *Pseudomonas cepacia* (1 236 pb.) (Wood, 1991), IS427 de *Agrobacterium tumefaciens* (1 271 pb.) (Meirsmán, 1989; Meirsmán, 1990) e IS6501 de *Brucella ovis* (836 pb.) (Ouahrani, 1993), e ISRm6 de *Rhizobium meliloti* (1 269 pb.) (Zecrí, 1996) respectivamente (Cuadro 2). Los resultados obtenidos por comparación de secuencias no aportan datos suficientes para establecer el origen de estas reiteraciones y su participación en la formación de la estructura del pSim, a pesar de poseer una identidad significativa con secuencias de inserción a nivel de nucleótidos y aminoácidos, además de evidenciar en estas mismas regiones alta probabilidad de codificación, además del motivo DDE encontrado en la traducción de la secuencia del p79H4, propio de las transposasas de la familia IS3 y de integrasas de retrovirus (Zecrí, 1996; Polard, 1995). Las secuencias de las clonas obtenidas no poseen las características estructurales básicas de secuencias de inserción (IS) (Galas, 1989), en ninguna de estas se encontraron secuencias inversas repetidas significativas en los extremos de MLA's, secuencias de duplicación ni MLA's continuos, esto último comparado con la secuencia de aminoácidos de las proteínas con las que se encontró identidad, lo anterior podría deberse a 1) por fallas puntuales en la secuencia obtenida, lo cual daría como resultado el cambio de fase del MLA o su interrupción por la generación de una señal de alto, o 2) por mutaciones que realmente hayan originado estos cambios. Lo anterior solo quedará esclarecido cuando se tenga la secuencia en doble cadena de los elementos de cada familia. Por otra parte, posiblemente las clonas secuenciadas en este trabajo no abarcan completamente la secuencia de los elementos, lo cual explicaría la ausencia de inversas repetidas en la secuencia analizada así como los sitios de duplicación. De acuerdo a lo anterior, es necesario extender las secuencias de los elementos descritos para poseer evidencias claras que apoyen el posible origen de estas por transposición.

Los datos de porcentajes de GC de los elementos secuenciados de las familias reiteradas obtenidos ofrecen también evidencia acerca del posible origen de las familias reiteradas por transposición como IS's. Se obtuvieron

50 secuencias al azar con una longitud de 300 pb. del banco de secuencias del pSim, éstas no incluyen las secuencias aquí descritas, se calcularon la media y desviación estándar muestrales del porcentaje de GC (Tabla 1) y se compararon con los porcentajes de cada una de las secuencias de los elementos secuenciados por medio del método de *t* de Student (Campbell, 1989). Los porcentajes de GC de los elementos de las familias reiteradas BF, y E mostraron ser significativamente diferentes a un nivel de 95% de certeza, mientras que los porcentajes de las secuencias de los elementos de las familias C y D1, no fueron diferentes; lo anterior, aunado a las identidades encontradas y lo análisis de probabilidad de codificación de las secuencias, apoya la propuesta de que estas secuencias reiteradas fueron IS's funcionales y fueron adquiridas por transposición.

Si realmente se trata de secuencias derivadas de IS con cambios provocados por mutaciones espontáneas, la dispersión de estas en el pSim aportan datos acerca de posibles eventos de transposición sencilla, ya que, considerando el concepto de isla simbiótica propuesta por Mergaert (1997) basado en las características de las islas de patogenicidad (Carniel, 1996; Hacker, 1997), y dado que se ha demostrado que la secuencia del pSim de *R. etli* CFN42 está representada en otros pSim's de la misma especie por experimentos de hibridación (Rodríguez, O. 1992; Rodríguez, C., 1992), se puede especular que la región de nodulación y fijación de nitrógeno del pSim de CFN42 (bandas 1 a la 25 del mapa físico, 120 kb.) podría tratarse de un transposón complejo; la razón de no llamar isla simbiótica a esta región es que, a diferencia de las islas de patogenicidad (Carniel, 1996; Hacker, 1997) y las probables islas simbióticas del pNGR234a (Mergaert, 1997), no posee todos los determinantes para nodulación y fijación de nitrógeno, dado que se necesitan otros replicones además del pSim para lograr el fenotipo silvestre (Brom, 1992). La interpretación de que esta zona sea producto de la inserción de un transposón complejo tiene un inconveniente, dado que las únicas reiteraciones que rodean la región de nodulación-fijación de nitrógeno son las de la familia BF (BF80-81 y BF 27-28) y curiosamente estos elementos no poseen la mayor conservación entre los elementos de esta familia (87.4% en 788 pb.); sin embargo, queda la posibilidad de que pudiera haber sido la región comprendida entre los elementos BF02-03 y BF80-81, los cuales poseen una conservación de 94.4% en 891 pb. Los elementos de la familia BF BF02-03 y BF80-81 mostraron identidad con el MLA Y4wP del pNGR234, el cual codifica potencialmente para una proteína de

función desconocida no relacionada con IS (Freiberg, 1997); sin embargo, probablemente se trate de una IS no reportada y no se tenga su secuencia completa, como se explicó anteriormente (Fig. 3 y 7). Las orientaciones de los elementos de la familia BF permiten llevar a cabo amplificaciones y deleciones (amplificación tipo IV, elementos BF80-81 y BF27-28) (Romero, 1995) e inversiones (elemento BF02-03 con el BF27-28 o con el BF80-81).

Otro dato que podría apoyar la propuesta de que las familias BF, C, D y E de elementos reiterados pudieron haber sido IS's funcionales es el análisis de las secuencias aledañas, las cuales debieron ser interrumpidas por el evento de transposición. Hasta el momento, solamente se cuenta con la secuencia 5' del elemento BF02-03, la cual posee un MLA que tiene una identidad del 78% a nivel de aminoácidos (88% positivos) con una ferredoxina III potencial del pNGR234a. Este MLA se encuentra completo respecto al del pNGR234 y se localiza a 340 pb. del extremo 5' del elemento BF02-03, por lo que no se encontró evidencia de interrupción de secuencias codificantes. Se propone, una vez secuenciados completamente todos los elementos de las familias BF, C, D y E, extender la secuencia desde los extremos de los elementos reiterados con oligonucleótidos para obtener evidencia de 1) la posible interrupción de secuencias codificantes por transposición (Galas, 1989), 2) preferencia por sitios específicos para transposición, como es el caso de ISRm3 en *nodG* (Ogawa, 1991), ISRm2 en el gene *fixX* (Dusha, 1987) e ISRm1 en la región *nif* (Ruvkun, 1982) y 3) la orientación de los elementos con respecto al resto de su familia.

Las implicaciones que se tendrían bajo la consideración de que las secuencias reiteradas analizadas tuvieron un origen por transposición, son las mismas que se han estudiado ampliamente en IS de enterobacterias. Los efectos causados por IS's de *E. coli* son principalmente el incremento en la frecuencia de mutación, ya sea creando mutaciones polares, cambios en el producto del gen, cambios en la regulación de genes por efecto directo en el promotor, como probablemente ocurrió en la región que se encuentra entre *syrM* y *nodD3* (región promotora) del pSim de *R. meliloti* donde existe una región con identidad significativa con IS's de la familia IS3 (Barnett, 1996), así como duplicaciones en el sitio de inserción (Galas, 1989). En la secuencia que se tiene hasta el momento del pSim de *R. etli* CFN42 no se ha encontrado, proporcionalmente, una cantidad tan grande de segmentos o IS completas como en el pNGR234a (Freiberg, 1997) y posee una organización

diferente (Dávila, comunicación personal). Cabe destacar que el hecho de que el gran número de IS's del pNGR234a se encuentra en zonas delimitadas del plásmido y en los extremos de regiones que poseen genes relacionados; lo anterior se ha atribuido a un efecto de selección, donde solo los plásmidos que conservan las secuencias de los genes importantes en los procesos de nodulación y fijación de nitrógeno son seleccionados, por lo que los eventos de transposición que afectan dichos genes excluyen este fenotipo de la población (Freiberg, 1997). Aunado a lo anterior se encuentra el concepto de isla simbiótica mencionado (Mergaert, 1997).

El poseer un número importante de secuencias repetidas con alta identidad provee la capacidad para llevar a cabo rearrreglos a distintas frecuencias, estos incluyen inversiones, deleciones y amplificaciones dependiendo de las direcciones relativas de los elementos involucrados. Fenómenos de este tipo han sido observados en *E. coli* (Hill, 1981), *Salmonella typhimurium* (Lehner, 1980; Anderson, 1981) y *Halobacterium halobium* (Sapienza, 1982b). La frecuencia de eventos de rearrreglos genómicos esta determinada por la longitud de las secuencias repetidas flanqueantes, su orientación, el grado de conservación entre ellas y la distancia que las separa (Anderson, 1977). En especies del género *Rhizobium* se han realizado estudios acerca de efecto de los rearrreglos en su genoma en general, no encontrándose diferencias significativas en el tamaño de los fragmentos generados por enzimas con baja frecuencia de corte, visualizados por electroforesis de campos pulsantes, tanto en simbiosis como *in vitro* (Sobral, 1991). Con respecto al componente plasmídico de este género, se han hecho amplios estudios respecto a su inestabilidad (Brom, 1991) y en especial del pSim (Soberón-Chávez, 1986 Flores, 1988; Flores, 1993; Romero, 1993; Romero, 1995). Hasta el momento se han descrito 4 tipos de amplicones en el pSim de CFN42, el tipo I corresponde al comprendido entre los operones *nifHDK* que bordean una región de 120 kb., la cual se considera que posee la mayor parte de genes involucrados en los procesos de nodulación y fijación de nitrógeno (Girard, 1991; Romero, 1993); el tipo II lo constituye una región de 141.2 kb. bordeada por los elementos reiterados de la familia C de las bandas BG83 y BG24; el amplicón tipo III comprende una región de 80 kb. y es el único caso donde no se encuentran elementos reiterados descritos bordeando, por lo que se propone que se trata de reiteraciones de menor tamaño no reveladas con las técnicas utilizadas; por último, el amplicón tipo IV abarca una zona de 172.5 kb. y posee a los

elementos, definidos en ese entonces, de la familia B (Romero, 1995). Balleza (1998) describe, con base en un análisis computacional, posibles amplicones en la región de nodulación de *Rhizobium meliloti*, en un rango que abarca secuencias repetidas de 110 pb. con 55.5% de identidad que comprende una región de 2.35 kb. hasta secuencias de 1 106 pb. con 82% de identidad y con un región de 75.5 kb. entre éstas (amplicones VI G-2 fixB/nifA y nodD1/nodD2 respectivamente); sin embargo, con la definición del tamaño de los elementos de la familia BF, el amplicón con las secuencias repetidas más pequeñas que se ha descrito es el amplicón tipo IV del pSim de CFN42, bordeado por los elementos BF80-81 y BF27-28 (87.4% en 788 pb.); posiblemente el amplicón tipo II también posea secuencias repetidas cercanas a una kilobase, de acuerdo al tamaño de la IS407 de *P. cepacia* con la que posee identidad. La consideración de la existencia de otros amplicones potenciales en el pSim de CFN42, así como posibles inversiones (Van de Putte, 1992), está limitada por el desconocimiento de las longitudes, niveles de identidad, orientaciones y características generales de las secuencias reiteradas de este plásmido, por lo que se deben esclarecer los puntos anteriores para poseer una visión integral sobre los eventos dinámicos de este replicón y sus posibles implicaciones en el proceso de simbiosis e impacto en la generación de nuevas variedades de cepas.

Se ha descrito que existen alrededor de 200 familias de secuencias reiteradas en especies pertenecientes a la familia Rhizobiceae (Flores, 1987); sin embargo, las que se encuentran en el pSim de *R. etli* CFN42 no están representadas en el cromosoma, así mismo, tampoco se comparten los elementos repetidos del cromosoma en este plásmido. Considerando la posibilidad que las familias analizadas fueron IS's funcionales, se esperaría que, en caso de que este plásmido hubiera sido originado en esta especie, se encontrarán también en el cromosoma, como es el caso de los elementos NGRIS de *R. sp.* NGR234 que posee una copia en el pNGR234a y otras dos en el cromosoma (Perret, 1997). Lo anterior sugiere que probablemente el pSim de la cepa CFN42 fue transferido desde otra bacteria nodulante y fijadora por transferencia horizontal. Se han descrito evidencias de este fenómeno en *Haemophilus influenzae* (Martin, 1998); en *E. coli*, *Salmonella typhimurium* y *Shigella dysenteriae* con base en las copias de la IS1 (Ohtsubo, 1984; Galas, 1989) e IS3, (Bisercic, 1993a, Bisercic, 1993b); en *Streptococcus thermophilus* y *Lactococcus lactis* por análisis de las secuencias de las IS1191 e iso-IS981 (Guédon, 1995); en bacterias metilotróficas con los

genes de la vía de utilización de diclorometano (Schmid-Appert, 1997). En particular, los estudios filogenéticos de genes *nif*, *nod* y RNA 16S de Rhizobiaceas revelan relaciones diferentes entre las especies bacterianas estudiadas, dependiendo del gen utilizado, por lo que se sugiere eventos de transferencia horizontal (Schofield, 1987; Mergaert, 1997); de forma similar, estudios hechos con base en la secuencia de los genes *nifK*, incluyendo la secuencia de *Frankia* cepa HFPCcl3, no muestra árboles congruentes con los obtenidos con las secuencias de los genes de RNA16S (Hirsch, 1995); así mismo, se propone que el gen *nodD* de *Azorhizobium caulinodans*, el cual está flanqueado por secuencias con identidad a IS's ($\epsilon 1$, $\delta 1$) y se encuentra en el cromosoma, fue transferido en algún momento de su evolución (Geelen, 1995). Otros datos que apoyan esta propuesta son mostrados por Schofield y cols. (1987), en los que se evalúa la transferencia de pSim's de *R. leguminosarum* bv. *trifolii* en poblaciones aisladas del suelo, con marcadores plasmídicos y un marcador cromosomal, así como por inmunoensayo con las cepas analizadas; estos resultados sugieren la transferencia de plásmidos simbióticos entre cepas del suelo. Las secuencias de inserción IS426, IS1312 e IS1313 del pTiBo542, encontrado en las cepas Bo542 y A281 de *Agrobacterium tumefaciens* poseen alta identidad con las IS's ISRm1 e ISRm2 de *R. meliloti*, lo cual también sugiere un evento de transferencia entre estas dos especies (Deng, 1995).

Es claro que existe una basta información que evidencia posibles fenómenos de transferencia horizontal en bacterias en general y en Rhizobiaceas en particular, pero la interrogante principal es si realmente ocurren estos fenómenos: ¿Qué mecanismos están involucrados y que consecuencias pueden tener en la biología del organismo que adquiere esta nueva información? Se ha encontrado que existe transferencia del p42a de *R. etli* CFN42 a *A. tumefaciens* a una frecuencia de 10^{-2} , por lo que se sugiere que este plásmido posee genes involucrados en conjugación que le permiten su transferencia y la de otros plásmidos, como el pSim (p42d), que se transfiere a frecuencias de 10^{-6} junto con el p42a y 10^{-8} solo el pSim; posiblemente exista transferencia del resto de los plásmidos pero no se han observado, por lo que se propone que ocurren a frecuencias menores (Brom, comunicación personal). Una posible explicación de la transferencia conjunta de los plásmidos pSim y p42a es que, dada la existencia de 19 fragmentos de la genoteca de Bam HI del pSim que hibridizan con el p42a (Girard, 1991), existan regiones conservadas entre los dos plásmidos que les permiten

cointegrarse y ser transferidos, sin embargo, las cointegraciones observadas corresponden a los p42b y pSim (p42d) (Brom, 1991), por lo que se sugiere que el cointegrado se separa inmediatamente después de ser transferido.

La capacidad de transferencia de fragmentos o plásmidos completos ha sido considerada como una posibilidad aplicable para dotar a poblaciones bacterianas nuevas características, tal es el caso del pJP4 que posee los genes necesarios para la degradación del ácido 2,4 diclorofenoxiacético (2,4 D) de *Alcaligenes eutrophus* JPM134, que fue evaluado en su capacidad de transferencia a *Varioborax paradoxus*. Dado lo anterior, es posible pensar en transferencia de plásmidos simbióticos manipulados con características mejoradas en laboratorio a cepas naturales del suelo; otra posibilidad es diseñar construcciones que, una vez transferidos, puedan integrarse al cromosoma o bien a plásmidos, tal es el caso del pSM18, el cual fue integrado al cromosoma de *R. leguminosarum* bv. *viciae* pSim- y evaluado por los genes reporteros del plásmido (Selbitschka, 1995).

Las implicaciones evolutivas respecto a la presencia de un número elevado de secuencias reiteradas en el genoma bacteriano repercute, en primera instancia, en su estructura y su dinámica. Tradicionalmente se ha considerado al genoma una estructura constante no sometida a cambios significativos en cuanto a tamaño y contenido, sin embargo, esta visión ha cambiado en los últimos años tras describirse fenómenos complejos de plasticidad genómica (Romero, 1997) y el papel de elementos transponibles como mediadores de transferencia de nueva información, como entidades mutagénicas y como sustrato para recombinación, cointegración, delección y escisión (Galas, 1989).

Dado que en el pSim de *R. etli* CFN42 los elementos de las familias BF y E poseen una extensión conservada, son susceptibles a generar rearrreglos, además de que el resto de las familias tienen identidad con IS's, es factible el considerar un posible transposón complejo que comprende la región de nodulación y fijación de nitrógeno, situada entre las bandas BG01 y BG25, o una parte de esta, entre los elementos BF02-03 y BF80-81. Aunado a lo anterior, la transferencia del pSim mediada probablemente por el p42a, además de la ausencia de correspondencia entre las reiteraciones del cromosoma y el pSim, sugieren un posible evento de transferencia horizontal

por conjugación. Tanto la capacidad de generación de rearrreglos, los cambios potencialmente provocados por posibles IS's (ahora representadas por los elementos de las familias BF, C, D1 y E) adquiridas por eventos probables de transferencia horizontal, ponen en juego un gran número de posibilidades de arreglos genómicos y de expresión genética que pueden ser seleccionadas bajo determinadas condiciones, permitiendo la evolución de un replicón plásmidico por medio de módulos, como es el caso de pTi de *A. tumefaciens* (Otten, 1992). Si bien es difícil evaluar el impacto de cada uno de estos fenómenos en la proliferación de una población bacteriana, su permanencia en el ambiente y, consecuentemente, en su evolución, la interacción de estos ofrece un escenario aun más complejo, sin embargo, el estudio específico de cada uno de estos, como es el caso de los elementos reiterados del pSim, brindarán los elementos indispensables para proponer explicaciones a cerca de la estructura, dinámica y evolución de los genomas bacterianos y su posible manipulación para su uso práctico, como en particular se ha planteado para los miembros de la familia Rhizobiaceae en la agricultura.

El trabajo presentado sugiere los siguientes estudios para continuar la investigación en torno a la evolución del pSim de *R. etli* CFN42 desde el contexto de las familias de DNA reiterado:

- a) Completar la secuencia en doble cadena de las clonas obtenidas para anular las ambigüedades en las secuencias presentadas.
- b) Secuenciar en doble cadena los elementos restantes de las familias C, D1, D2, y E.
- c) Secuenciar los elementos del resto de familia reiteradas (G, H, I, J y K).
- d) Secuenciar los extremos externos a las secuencias de los elementos de cada familia.
- e) Hibridizar DNA total digerido de otras cepas de *R. etli* y de otras especies de la familia Rhizobiaceae utilizando como sonda la región conservada de cada familia.

7. CONCLUSIONES

- 1.- La familia reiterada originalmente definida como B se redefinió como BF, dado que el sitio interno de Bam HI de sus elementos provocó que se interpretaran como dos familias separadas: B y F. Los elementos de ésta familia se renombraron BF02-03, BF27-28 y BF80-81, de acuerdo con el mapa físico del pSim.
- 2.- Los elementos de la familia reiterada BF comparten una región conservada de 714 pb. con una identidad mayor al 80%.
- 3.- Los elementos de la familia BF poseen de alrededor del 60% con el MLA Y4wP de función desconocida del pNGR234 de *Rhizobium* sp. NGR234.
- 4.- La secuencia del elemento de la familia reiterada C de la BG83 tiene identidad de 59.8% a nivel de DNA con la IS407 de *Pseudomonas cepacia*.
- 5.- La familia reiterada D originalmente descrita (Girard, 1991), está integrada por tres familias de elementos reiterados: la familias D1 (con elementos en las bandas BG83 y BG85), D2 (con elementos en las bandas BG85 y BG01) y D3 (con elementos en las bandas BG01 y BG83).
- 6.- La secuencia del elemento reiterado de la BG85 de la familia D1 tiene una identidad del 55.9% con la secuencia de inserción IS427 de *Agrobacterium tumefaciens* a nivel de DNA y de alrededor del 65% con una probable transposasa de la IS6501 de *Brucella ovis*.
- 7.- Los elementos de la familia reiterada E de las bandas BG79 y BG85 tienen alta identidad, a nivel de proteína y DNA, con el MLAB de la ISRm6 de *Rhizobium meliloti*.

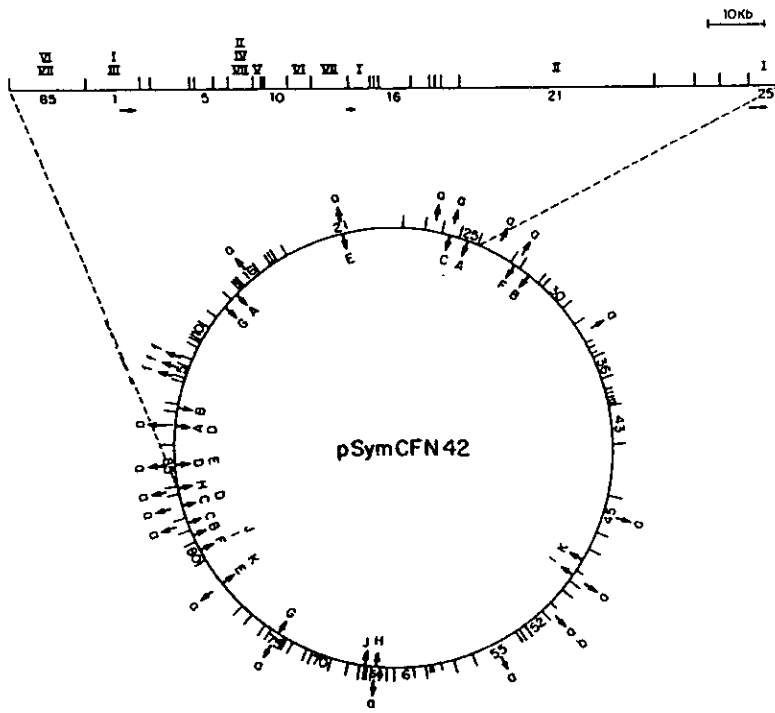


Figura 1. Mapa de restricción con Bam HI del plásmido simbiótico (pSim) de *Rhizobium etli* cepa CFN42 . Las flechas internas denotan la presencia de los elementos reiterados de las 11 familias reportadas (letras mayúsculas). Las flechas externas indican las bandas del plásmido que hibridan con los plásmidos p42A (a), p42b (b) y p42f (f). La línea superior representa una región de 135 kb. (de la banda 85 a la 25) donde se encuentran los genes *nifH* (I), *nifA* (II), *nodA* (III), *nodB* (IV), *nodC* (V), *nodD* (VI) y las cajas *nod* ("nodbox") (VII). Las flechas debajo de bandas 1 y 25 indican la orientación directa de los operones *nifHDK* y la flecha bajo la banda 13 indica la orientación inversa del gen *nifH* (tomado de Girard, 1991)

REITERACION	BANDA	PESO (Kb)
B	2	1.6
	28	6.7
	81	3.8
C	24	5.19
	82	3.4
	83	7.2
D	1	9.4
	83	7.2
	85	12.9
E	21	33.6
	79	16.8
	85	12.9

Cuadro 1.- Elementos de las familias de secuencias reiteradas B, C, D y E del plásmido simbiótico (pSim) de *R. etli* CFN42. Los números de bandas corresponden a la designación de las clonas de la genoteca de Bam HI del pSim que contienen los elementos de cada familia. Aunado, se muestra el peso molecular de los insertos de cada clona, de acuerdo con Girard (1991).

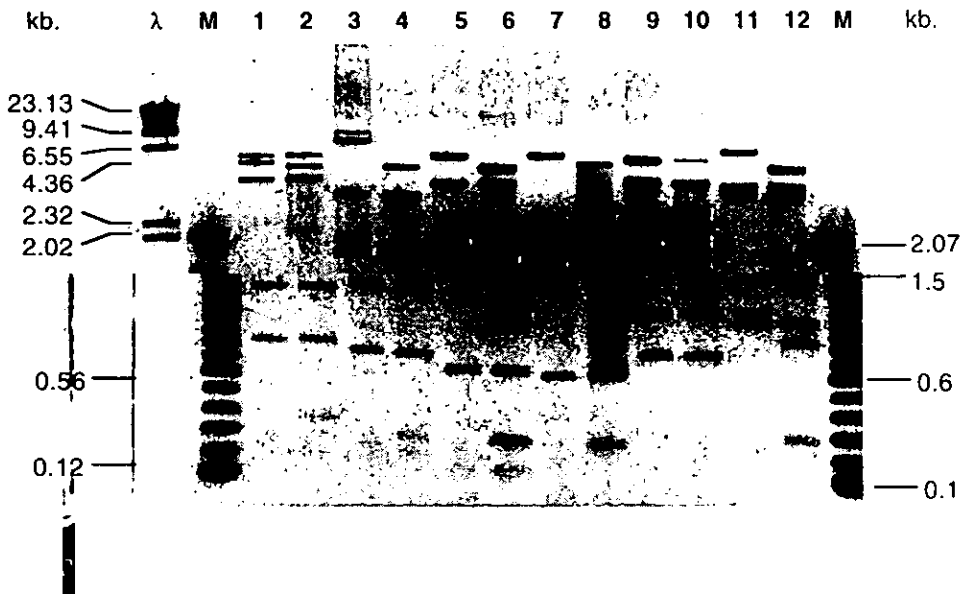


Figura 2. Electroforesis en gel de agarosa de los productos de digestión con enzimas de restricción de las clonas de las bandas de Bam HI utilizadas para aislar los elementos reiterados de las familias C, D y E. Las bandas marcadas con asteriscos mostraron hibridación con las sondas utilizadas (familia reiterada C, BG82; familia D, BG83; familia E, BG85). λ, DNA del fago λ digerido con Hind III; M, marcador de peso molecular 100; carriles 1, 3, 5, 7, 9 y 11 digeridos con Hind III, Sal I, Xho I, Pst I y Sal I respectivamente; carriles 2, 4, 6, 8, 10 y 12 digeridos con las mismas enzimas más Bam HI. Carriles 1, 2, 3 y 4, DNA de la BG79; 5, 6, 7 y 8 DNA de la BG83; 9, 10, 11 y 12 DNA de la BG85.

BF.02-03.seq	GGTGCAAGA AGGCACTGCA ATTGCACGCA ATGTCTCGCG CAGCTGACC	50
BF.27-28.seq	-----	
BF.80-81.seq	----- ---TCCTTTA TTGGCCCCCA AAG---CGGG AAGGTTGTCC	34
Consensus	GGTGCAAGA AGGYMCTKYA WTKGCMSCA AWGTCTCGSG MARGYTGWCC	50
BF.02-03.seq	--GTCBAGGT AA-GCA---C CAGTGGGCGG CTTAC-GGA GGCACCTTGGG	92
BF.27-28.seq	--GCCCGGGA AAAATATTTT TCGGGGGGGG AATATGGGA AAAANAG-G	47
BF.80-81.seq	AAAFCCAGG RRCCCGGAC CAAGGGGTTG -ATTA--GA CCCGCCAG-G	79
Consensus	AAAFCSGGD BAMUYKKLYV YMAKGGVGG AATFVGGGA JUMAMNIGD	100
BF.02-03.seq	GCGAGGACCG CACGGLACT TTCACCCCG CAGG-----C TCAATACGAT	137
BF.27-28.seq	GAGAGCTC- CNCCGAGATT TGATTTCCG GGGGNACCC CCTTARTAN	96
BF.80-81.seq	--GAGGTTT C-ACGGA-- --AGT--C- -GAGTACCC CCTTARCCC	116
Consensus	MGAGGMYCG CNCCGSAVT TKMBWYVCG SRFGINACCC YCMUTAMBNN	150
BF.02-03.seq	CCGACACAGG ---ADGATC CTGAGCGCGG T-TGAAWAA CAGGTAATT	182
BF.27-28.seq	ACGACACAGG GAACATGAGG CCTGGCGCGA TATGTA-GAC CAGGATGATT	145
BF.80-81.seq	ACGA-CA-G GCACATGATC CTTGGCGCGA TTTGAA-GTC CAGGTAATT	163
Consensus	MCGAACAGG GMACBGAKS CVKAGCGCGA TTTGAAIDWM CAGGKLAAT	200
BF.02-03.seq	CCMTTGTCCC CAGGCAACC-GTCGADATT TCGCAATG TGGGATCC	232
BF.27-28.seq	CCMTTGTCCC CAGGCAACC-GTCGADATT TCGCAATA TGGGATCC	195
BF.80-81.seq	CCMTTGTCCC CAGGCAACC-GTCGADATT TCGCAATG TGGGATCC	213
Consensus	CCMTTGTCCC CAGGCAACC-GTCGADATT TCGCAATA TGGGATCC	250
BF.02-03.seq	GCCACTTTG TCAGGTTGT COTGATGGA A-CGTTTGT TCCCGGTCT	281
BF.27-28.seq	GCCACTTTG TCAGGTTGT COTGATGGA A-CGTTTGT TCCCGGTCT	245
BF.80-81.seq	GCCACTTTG TCAGGTTGT COTGATGGA A-CGTTTGT TCCCGGTCT	262
Consensus	SCCAVMITG TCAGGTTGT COTGATGGA A-CGTTTGT TCCCGGTCT	300
BF.02-03.seq	CTCCGTCARA TACACATCAA CCTGCCCTCA ACCGCCCAT TGATAAAGA	331
BF.27-28.seq	CTCCGTCARA TACACATCAA CCTGCCCTCA ACCGCCCAT TGATAAAGA	295
BF.80-81.seq	CTCCGTCARA TACACATCAA CCTGCCCTCA ACCGCCCAT TGATAAAGA	312
Consensus	CTCCGTCARA TACACATCAA CCTGCCCTCA ACCGCCCAT TGATAAAGA	350
BF.02-03.seq	AAATATCTT CTGTCGCGT CCGAGAGAT GGCATGGTGT TTCTGATCC	381
BF.27-28.seq	AAATATCTT CTGTCGCGT CCGAGAGAT GGCATGGTGT TTCTGATCC	345
BF.80-81.seq	AAATATCTT CTGTCGCGT CCGAGAGAT GGCATGGTGT TTCTGATCC	362
Consensus	BAATATCTT CTGTCGCGT CCGAGAGAT GGCATGGTGT TTCTGATCC	400

Figura 3. Alineamiento de las secuencias de los elementos de la familia reiterada BF. Las zonas sombreadas muestran las regiones conservadas.

BF.02-03.seq	CTGGTTACAA GCGGCAGAAA GCAGTCTDCA TATATCCGGT TCATGAARACG	431
BF.27-28.seq	CTGGGGACAA GCGGCAGAAA GCAGTCTETA TATATCCGGT TCATGAARACG	395
BF.80-81.seq	CTGGTTACAA GCGGCAGAAA GCAGTCTDCA TATATCCGGT TCATGAARACG	412
Consensus	CTGGKKACAA GCGGCAGAAA GCAGTCTDCA TATATCCGGT TCATGAARACG	450
BF.02-03.seq	AACCGAFCGG AGCCACCTG ACCGTTTCGTC ATCGGATCC CGTCTACCA	481
BF.27-28.seq	AACCGAFCGG AGCCACCTG ACCGTTTCGTC ATCGGATCC ---TCTACCA	441
BF.80-81.seq	AACCGAFCGG AGCCACCTG ACCGTTTCGTC ATCGGATCC CGTCTACCA	461
Consensus	AACCGAFCGG AGCCACCTG ACCGTTTCGTC ATCGGATCC CGTCTACCA	500
BF.02-03.seq	AACCGGCAT GCA---TCCG CCGGGGAA GGGTCATCC CGATGTCACA	526
BF.27-28.seq	AACCGGCAT GCA---TCCG CCGGGGAA GGGTCATCC CGATGTCACA	485
BF.80-81.seq	AACCGGCAT GCATAGTCCG CCGGGGAA GGGTCATCC CGATGTCACA	511
Consensus	AACCGGCAT GCATAGTCCG CCGGGGAA GGGTCATCC CGATGTCACA	550
BF.02-03.seq	GCTTGCAACC AACTCAAG AGAAGGAAA TGCATGACC AAAATTTCCG	575
BF.27-28.seq	GCTTGCAACC AACTCAAG AGAAGGAAA TGCATGACC AAAATTTCCG	535
BF.80-81.seq	GCTTGCAACC AACTCAAG AGAAGGAAA TGCATGACC AAAATTTCCG	560
Consensus	GCTTGCAACC AACTCAAG AGAAGGAAA TGCATGACC AAAATTTCCG	600
BF.02-03.seq	AAAGATCCTT GAACGATC GAAAACCGG GGTATCGTC GATCAAGTG	625
BF.27-28.seq	AAAGATCCTT GAACGATC GAAAACCGG GGTATCGTC GATCAAGTG	585
BF.80-81.seq	AAAGATCCTT GAACGATC GAAAACCGG GGTATCGTC GATCAAGTG	610
Consensus	AAAGATCCTT GAACGATC GAAAACCGG GGTATCGTC GATCAAGTG	650
BF.02-03.seq	CCAGATCAGT TCGGCGATC GTTGAFAAG GGGACAAFA AGGTGACAF	675
BF.27-28.seq	CCAGATCAGT TCGGCGATC GTTGAFAAG GGGACAAFA AGGTGACAF	635
BF.80-81.seq	CCAGATCAGT TCGGCGATC GTTGAFAAG GGGACAAFC AGGTGACAF	660
Consensus	CCAGATCAGT TCGGCGATC GTTGAFAAG GGGACAAFM AGGTGACAF	700
BF.02-03.seq	GGCTTTTTG AAAATTGCGT TCGGCGCTG AAGCGACACA AAAATGCTG	725
BF.27-28.seq	GGCTTTTTG AAAATTGCGT TCGGCGCTG AAGCGACACA AAAATGCTG	684
BF.80-81.seq	GGCTTTTTG AAAATTGCGT TCGGCGCTG AAGCGACACA AAAATGCTG	709
Consensus	GGCTTTTTG AAAATTGCGT TCGGCGCTG AAGCGACACA AAAATGCTG	750
BF.02-03.seq	CCTCCGATCC TCGAACGAC AAATCTATTC GGCACGAAT TGTGGTGGAA	775
BF.27-28.seq	CCTCCGATCC TCGAACGAC AAATCTATTC GGCACGAAT TGTGGTGGAA	734
BF.80-81.seq	CCTCCGATCC TCGAACGAC AAATCTATTC GGCACGAAT TGTGGTGGAA	759
Consensus	CCTCCGATCC TCGAACGAC AAATCTATTC GGCACGAAT TGTGGTGGAA	800

Figura 3. Continuación.

BF.02-03.seq	AAACATCGC TGCAC TGCAD	GCCGACGCCG	AGGCC	AGCT	TCTCAC	TTT	823	
BF.27-28.seq	-GACATCGC TGCAC TGCAD	GCCGACGCCG	AGGCC	AGCT	TCTCAC	ATT	783	
BF.80-81.seq	-GACATCGC TGCAC TGCAR	GCCGACGCCG	AGGCC	AGCT	TCTCAC	ATT	806	
Consensus	AAACATCGC TGCAC TGCAR	GCCGACGCCG	AGGCC	AGCT	TCTCAC	ATT	850	
BF.02-03.seq	TGDAAGCTTT	GTTGGGCG	CGAA	TTTC	GCCGTCGCA	GAT	CCTC	864
BF.27-28.seq	TGDAAGCTTT	TGGTTGGGCG	CGAARTTTTC	GCCGTCGCA	GAT	CCCTCG	833	
BF.80-81.seq	TGDAAGCTTT	GTTGGGCG	CGAA	TTTC	GCCGTCGCA	GAT	CCTC	847
Consensus	TGDAAGCTTT	TGGTTGGGCG	CGAARTTTTC	GCCGTCGCA	GAT	CCCTCG	900	
BF.02-03.seq	GACACAGAGT	CGACCTTTTT	CCGCAAGCCG	GTTGAGACRA	TTTT	GCAGCA	914	
BF.27-28.seq	GACACAGAGT	C-----	CGNA	C-C	-----	CNT	TTTT	857
BF.80-81.seq	GACACAGAGT	CGACCTTTTT	CCGCAAGCCG	GTTGAGACRA	TTTT	GCAGCA	897	
Consensus	GACACAGAGT	CGACCTTTTT	CCGCAAGCCG	GTTGAGACRA	TTTT	GCAGCA	950	
BF.02-03.seq	CGCCAGGAA	TTCCGGGTC	TCTCAGCG	GGCGGTGGAG	GAAATCTCAA		964	
BF.27-28.seq	-----	TTCC	GN	AAAGCCG	GGTT	-----	874	
BF.80-81.seq	CGCCAGGAA	TTCCGGGTC	TCTCAGCG	GGCGGTGGAG	GAAATCTCAA		947	
Consensus	CGCCAGGAA	TTCCGGGTC	TCTCAGCG	GGCGGTGGAG	GAAATCTCAA		1000	
BF.02-03.seq	AGCCGGTCRA	GGATGCTTTT	GACAGGTGC	TGACGGACCT	CARAGCGACG		1014	
BF.27-28.seq	-----	-----	-----	-----	-----		874	
BF.80-81.seq	AGCCGGTCRA	GGATGCTTTT	GACAGGTCC	TGACGRACCC	CARAGCGACG		997	
Consensus	AGCCGGTCRA	GGATGCTTTT	GACAGGTSC	TGACGRACCY	CARAGCGACG		1050	
BF.02-03.seq	TTAAA-----	-----	-----				1019	
BF.27-28.seq	-----	-----	-----				874	
BF.80-81.seq	-TAAAGTAGA	CTCGGCACTA	AGATCAGG				1024	
Consensus	TTAAAGTAGA	CTCGGCACTA	AGATCAGG				1078	

Figura 3. Continuación.

IS407.Pseudcepa	CGTTGAGGT	CGTGGTCAG	GGAAAGCCCT	GAGCCCGAA	GC	CAATCGCG	350
RC.BG83.seq	-----	-----	-----	-----	AA	GCT--TCGG	10
Consensus	CGTTGAGGT	CGTGGTCAG	GGAAAGCCCT	GAGCCCGAA	GC	CAATCGCG	350
IS407.Pseudcepa	AGGAGTGC	GGCATTCCG	GAGAGGTCA	ACAT	CTCCGA	GCGCCGCC	400
RC.BG83.seq	AGG-A--TC	GGC	-----	ACAT	-----	-----	23
Consensus	AGGAGTGC	GGCATTCCG	GAGAGGTCA	ACAT	CTCCGA	GCGCCGCC	400
IS407.Pseudcepa	TGCCGGCTTG	TCGGCTTTC	TCGCAGCGT	CTGCATTACG	ACCG	AGCC	450
RC.BG83.seq	-----	GGC	CGT	-----	CG	-----	31
Consensus	TGCCGGCTTG	TCGGCTTTC	TCGCAGCGT	CTGCATTACG	ACCG	AGCC	450
IS407.Pseudcepa	AGACCAGAG	AATGAGGTGC	TAGCGCCGC	TCTGGTGGAG	TTGCCADACG	500	
RC.BG83.seq	GACC	-----	-----	-----	CA	ACG	40
Consensus	AGACCAGAG	AATGAGGTGC	TAGCGCCGC	TCTGGTGGAG	TTGCCADACG	500	
IS407.Pseudcepa	AACGTC--GT	CGATTGGCT	ACCGCGATT	GCACGCCCTG	GTGCAACCG	548	
RC.BG83.seq	AACGTCGGG	CGAT--GG	ACT-TCG-TT	CACGACC	----	AACCG	77
Consensus	AACGTCGGG	CGATTGGCT	ACCGCGATT	GCACGCCCTG	GTGCAACCG	550	
IS407.Pseudcepa	AGGCACGCA	CGCAATCAC	ARGCGAICT	ATCGCTGTGA	CCGTGAGGCA	598	
RC.BG83.seq	CA--ACG--	GGTAA	----	GAACT	CGCGTG--	C-TGA--C-	105
Consensus	AGGCACGCA	CGCAATCAC	ARGCGAICT	ATCGCTGTGA	CCGTGAGGCA	600	
IS407.Pseudcepa	GGGTGGCTG	TTCGACCCG	TCCACAGCGC	AGCGTGTAA	TGATTAAGC	648	
RC.BG83.seq	GG-TG--G	TTCGACCC	TCTC--GCGC	TA--CGTG--	CG--GTGCT	141	
Consensus	GGGTGGCTG	TTCGACCCG	TCCACAGCGC	AGCGTGTAA	TGATTAAGC	650	
IS407.Pseudcepa	CGACACCTG	GCATGSCCG	GCGCGCCAA	CGAGGTGTGG	TCAATCGATT	698	
RC.BG83.seq	TGATCCACG	TCATG	-----	CAAT	AGAGCG	--AA--GACG	173
Consensus	CGACACCTG	GCATGSCCG	GCGCGCCAA	CGAGGTGTGG	TCAATCGATT	700	
IS407.Pseudcepa	TGTGATGGA	TGCCCTTTC	AACGGCCGGC	GCGTAAAGTG	CCTGACCGTC	748	
RC.BG83.seq	TGTG-CAAA	----	CCCTTG	AACGG	-----	GT--GTG	208
Consensus	TGTGATGGA	TGCCCTTTC	AACGGCCGGC	GCGTAAAGTG	CCTGACCGTC	750	

Figura 4. Alineamiento de las secuencias del elemento reiterado de la familia C de la BG83 (RC.BG83.seq) con la de la IS407 de *Pseudomonas cepacia* (IS427.Pseudcepa). Las regiones sombreadas denotan los nucleótidos comunes.

IS407.Psuedcepa	GTCGATTA	CTCCGAA	AGCTGTCG	ATCGTGTCC	ACCA	GGGAT	798	
RC.BG83.seq	G-GCTA	TC-CGAA	-----	CGAT	-CCGT	GTCG ACCA	GGGA	242
Consensus	GTCGATTA	CTCCGAA	AGCTGTCG	ATCGTGTCC	ACCA	GGGAT	800	
IS407.Psuedcepa	CTCAGGTTTG	TATGTCGC	CGACHTTGA	CCGTGAGCT	CGCTCCG	TG	848	
RC.BG83.seq	CCGA-GTTTG	-TGTCGC	-GAC-TTGA	CC-TGTGG	-GC	-----	276	
Consensus	CTCAGGTTTG	TATGTCGC	CGACHTTGA	CCGTGAGCT	CGCTCCG	TG	850	
IS407.Psuedcepa	GCTATCCCA	GGCGTGCGA	ACAGACCAGG	GACCCGATT	TACGAGCCGC		898	
RC.BG83.seq	-CTATCCCA	GG-GTG	-----	-----	-----		290	
Consensus	GCTATCCCA	GGCGTGCGA	ACAGACCAGG	GACCCGATT	TACGAGCCGC		900	
IS407.Psuedcepa	GCGCTTGACC	AGTGGCCGT	TGGGACCGC	CTCCACTGA	AGTTGATTC		948	
RC.BG83.seq	-----	TGA	-----	CGT	TG-GA	-C	311	
Consensus	GCGCTTGACC	AGTGGCCGT	TGGGACCGC	CTCCACTGA	AGTTGATTC		950	
IS407.Psuedcepa	GGCGGCAG	CCACTGATA	ATGCTTTCAT	CGAATCGTTC	AAAGGC	AAAT	998	
RC.BG83.seq	--CGGCAG	CCACTGATA	ATGCTTTCAT	CGAATCGTTC	AAAGGC	AAAT	359	
Consensus	GGCGGCAG	CCACTGATA	ATGCTTTCAT	CGAATCGTTC	AAAGGC	AAAT	1000	
IS407.Psuedcepa	TCCGDSHGA	ATGCCTTAC	DAGCATTGGT	TCACGACCT	GGCGALGCT		1048	
RC.BG83.seq	TCCGDSHGA	ATGCCTTAC	DAGCATTGGT	TCACGACCT	GGCGALGCT		409	
Consensus	TCCGDSHGA	ATGCCTTAC	DAGCATTGGT	TCACGACCT	GGCGALGCT		1050	
IS407.Psuedcepa	CG	-----	G	GC	-----	-----	1053	
RC.BG83.seq	CGCGAARA	TGGAGGATTG	GCTTASATA	CTACATCCT	GTTTCATAC		459	
Consensus	CGCGAARA	TGGAGGATTG	GCTTASATA	CTACATCCT	GTTTCATAC		1100	
IS407.Psuedcepa	-----	A	ATCATC	G	CGG	-----	C	1070
RC.BG83.seq	TCCCATCTTC	CTAATGCATC	GAAGGATT	AGATTTTCC	GATCTGGGGC		509	
Consensus	TCCCATCTTC	CTAATGCATC	GAAGGATT	AGATTTTCC	GATCTGGGGC		1150	
IS407.Psuedcepa	-----	GTCGGA	-----	C	T	ACRACGA	-----	1086
RC.BG83.seq	AATCCGSSG	GGTKGGGGA	AGGTTTYTC	CSWACRACGA	TTCGAGGAT		559	
Consensus	AATCCGSSG	GGTKGGGGA	AGGTTTYTC	CSWACRACGA	TTCGAGGAT		1200	

Figura 4. Continuación.

IS407. Psuedcepa	-----	-----	GC-----	AAAGGCCTG	-----	CA---	1098	
RC.BG83.seq	CTGACGTGCT	TGCGGGGTCG	GCAGGTTTGG	GAAGGCCTGA	AGATCCATGG		609	
Consensus	CTGACGTGCT	TGCGGGGTCG	GCAGGTTTGG	GAAGGCCTGA	AGATCCATGG		1250	
IS407. Psuedcepa	--CA--GTGC	A--C--TGAA	CTACCTT--	-----	GC--	-----	1125	
RC.BG83.seq	GCCATCGGGT	AGCGGATGAG	CCCCCTTAT	ATCGCCAGCC	TCGGCCGCCA		659	
Consensus	GCCATCGGGT	AGCGGATGAG	CCCCCTTAT	ATCGCCAGCC	TCGGCCGCCA		1300	
IS407. Psuedcepa	GA-----	GT TTCCGGCGAA	ACATCGGGCA	A--CGG--CGG	A--TGCTC--		1163	
RC.BG83.seq	GARGCAGCGT	TTTTCGGCAT	ATCCCGAGTA	ATTTTGCCG	ACCGCCGAGA		709	
Consensus	GARGCAGCGT	TTKYGGCGAA	ATMYCGAGTA	ARTYVGGCG	ATVGLYCAGA		1350	
IS407. Psuedcepa	CT--GCC---	GCTT--TCCA	G--GAGGT--	GGTTTAAAG-	--GG-AC-TTT		1198	
RC.BG83.seq	TTTAGCCATG	GCTCAACCC	GTCGATTTGC	GGACGGAGAG	CCGGATCITT		759	
Consensus	TTTAGCCATG	GCTYAAVCC	GTCGATVTC	GGYKPAAG	CCGGAVVLT		1400	
IS407. Psuedcepa	GC--TA--GA--	-----	AGCC--	--ATTGGC--	-----	CTAT	TGATGG--G	1227
RC.BG83.seq	GCATATGAC	CTCAGCCCGG	TGACCGGCTC	CCGTGTCCAG	CGATTGCCAT		809	
Consensus	GCATATGAC	CTCAGCCCGG	TGAYSGCTC	CCGTGYCAG	YGBTKGCCAK		1450	
IS407. Psuedcepa	GGCAGGTCA-	-----	-----	-----	-----		1236	
RC.BG83.seq	TGCCGGTGT	CAGTCCGTTG	CGATTGAGGA	CACCGGCAAT	CACATCATCA		859	
Consensus	KGCAGTVRT	CAGTCCGTTG	CGATTGAGGA	CACCGGCAAT	CACATCATCA		1500	
IS407. Psuedcepa	-----	-----	-----	-----			1236	
RC.BG83.seq	TTGGCGATCA	AGACGAGCTG	TTTCACAGCG	TCGAC			894	
Consensus	TTGGCGATCA	AGACGAGCTG	TTTCACAGCG	TCGAC			1535	

Figura 4. Continuación.

IS427. Agrotum	GATGATTACA	CACTGAAAA	AGT--GATG	AA--CCGC	GA	TTGCATCTAG	47
REVIS. 85P4	-----	-----C	TGCAGGAGCC	CATCCGC	GGA	AGCC--TCTAG	30
Consensus	GATGATTACA	CACTGAAAA	AG--AGGAGC	AA--CCGC	GA	TTGCATCTAG	50
IS427. Agrotum	GTCGTAGACT	CATAGCCCG	DAKCCAVGG	CGCATTG	CG	CC--AACTG	94
REVIS. 85P4	GTCGTGTGGA	CATTTATCTG	AAKCCATA--	GCATTGT	CG	CCCGCATGT	76
Consensus	GTCGTAKRSW	CAKHWVYCG	DAKCCAVGG	CGCATTG	CG	CCCGAVMGT	100
IS427. Agrotum	GACDATTGGC	AGGAGTTTT	CGCCDAGOTT	GTCCGAGCCG	GTAGCGACCC	144	
REVIS. 85P4	GACDATTGGC	AGGAGTTTT	CGCCDAGOTT	GTCCGAGCCG	GTAGCGACCC	126	
Consensus	GACDATTGGC	AGGAGTTTT	CGCCDAGOTT	GTCCGAGCCG	GTAGCGACCC	150	
IS427. Agrotum	GACDATTGGC	CTTLAGDTT	DAAAGAAC	GCATGATCAG	GTTCGGITCG	194	
REVIS. 85P4	TTCTGAACTG	CTTLAGDTT	DAAAGAAC	ATTCGATCAG	GTTCGGITGA	176	
Consensus	KLCKGAAITG	CTTLAGDTT	DAAAGAAC	ATTCGATCAG	GTTCGGITGA	200	
IS427. Agrotum	CGATAATGCC	ATATGCTGA	--AGTACGGT	TTTGAC--CT	ATGGITTTGCC	239	
REVIS. 85P4	GCATAATGAT	GTATGCTGAG	CGGTAATTTT	CGGGACTTCA	ATGGATTTG--	224	
Consensus	SSATAATGMY	ATATGCTGAG	CGGTAATTTT	CGGGACTTCA	ATGGATTTG--	250	
IS427. Agrotum	TTGGDGGGA	TATTGGCBA	GGCCCCCTTC	TCGGG--GAGA	GAAGCCCGGA	288	
REVIS. 85P4	TTGG--GGAT	TACTGGCBA	GGCCCCCTTA	TCGGGAAAGG	GATTGGCGGA	272	
Consensus	TTGGDGGGA	TATTGGCBA	GGCCCCCTTC	TCGGGAAAGG	GAAGCCCGGA	300	
IS427. Agrotum	TGCGATGGC	ATCATAGGC	TATTCAGCA	GCATGATGAT	TCCGGACCC	338	
REVIS. 85P4	GGTATGCT	GTCATAGGC	GATTCAGCA	TA--AAAAAT	TTCGGGGGTT	320	
Consensus	KGATGATGCV	ATCATAGGC	KATTCAGCA	GCATGATGAT	TTCGGGSRVY	350	
IS427. Agrotum	ACATGGTCRA	GCARACCGTC	AGCCGCTTG	CCCGTCTGG	CTTTGCCCG	387	
REVIS. 85P4	GGGCCCCAR	TCAGGGNGTC	GGTTGGGGG	GCATCCCCCT	TCCTGCCCTG	370	
Consensus	ASRYSSYCAR	KCARSNATC	AGGTVGKKG	GCATCCCCCT	YVYTGCCCG	400	
IS427. Agrotum	CGGTTAGGCT	DAGCCTGATC	GGGACCCCT	GCCTATCAC	GACCGCCTTG	437	
REVIS. 85P4	CGGTCAATTE	TAAACGGATC	GGGACCCCA	AAACCCCAAC	GGCATTTTIG	420	
Consensus	CGGTAKRYTE	AAACGGATC	GGGACCCCA	AAACCCCAAC	GAACRYKTEG	450	

Figura 5. Alineamiento de las secuencias de la clona p84P4 (REVIS85P4), la cual contiene a un elemento de la familia reiterada D, y la IS427 de *Agrobacterium tumefaciens*. Las regiones sombreadas muestran los nucleótidos comunes.

IS427.Agrotum	ATTTTGGTCG	TGAGCCC GCC	TCGGGACGG	CCGAGACART	GATCTCAT	486													
REVIS.85P4	ATTTTGG	---	---	---	G-TCTCAT	436													
Consensus	ATTTTGGTCG	TGAGCCC GCC	TCGGGACGG	CCGAGACART	GATCTCAT	500													
IS427.Agrotum	-----	-----	-----	-----	CCCCCT	TTTTGCCBTC	502												
REVIS.85P4	CCCCCGGGA	ACGGGCARAGG	GGTTTAAATTT	CCACCCCC	TTT	TTTTGCCBCC	486												
Consensus	CCCCCGGGA	ACGGGCARAGG	GGTTTAAATTT	CCACCCCC	CT	TTTTGCCBYC	550												
IS427.Agrotum	-GCACCC	-T	GTTGGTGGGC	CCGAA	CGAA	ATGC	TGTGA	TCATCTGGAT	549										
REVIS.85P4	GGCACCC	TTT	CTTAAATGGGC	CCGAA	---	AA	TG	GTGA	ATC	---	524								
Consensus	GGCACCC	TTT	CTTAAATGGGC	CCGAA	GGAA	ATGC	TGTGA	TCATCTGGAT	600										
IS427.Agrotum	GTCACCATCA	TAGCCAGCG	TATGGCC	TTC	TATCA	TTCCG	TCCAGACGC	599											
REVIS.85P4	---CACCATC	---	CA	AG	T	A	---	CTC	GA	---	A	---	G	TCT	---	G	---	GA	550
Consensus	GTCACCATCA	TAGCCAGCG	TATGGCC	TTC	TATCA	TTCCG	TCCAGACGM	650											
IS427.Agrotum	CTGCTTTCCG	CCATCG	---	---	TATG	---	---	A	AT	---	---	CGG	TT	627					
REVIS.85P4	TCG	---	TCCG	CCATCG	CTG	GATATG	CGC	CACCA	ATGC	CCTTCCGG	CT	597							
Consensus	YVCTTCCG	CCATCG	CTG	GATATG	CGC	CACCA	ATGC	CCTTCCGM	700										
IS427.Agrotum	GTAGCGGTG	GTTCCG	CGGA	-----	CCA	TG	CG	-----	TTCCG	659									
REVIS.85P4	CCAGCGGTG	AACCGCGGA	AGACACTGTT	-----	CCAD	TG	CG	-----	AACACTTCG	647									
Consensus	SYAGCGGTG	AALCGCGGA	AGACACTGTT	-----	CCAD	TG	CG	-----	AACACTCG	750									
IS427.Agrotum	GCAGATCGCG	CCAGGCGCA	CCGGATCGA	GCAC	CCAGAA	GAT	TCCA	TTC	709										
REVIS.85P4	GCAGATCGCG	CCAGGCGCG	CCGGATC	---	GCAC	---	AA	---	TCCA	---	A	685							
Consensus	GCAGATCGCG	CCAGGCGMR	CCGGATCGA	GCAC	CCAGAA	GAT	TCCA	TTM	800										
IS427.Agrotum	AGCACCGAC	GGTCATCGAC	GCGCGCAC	CCTCTGGCT	TGTGGGTAA	759													
REVIS.85P4	AGCACCGAC	TTCCATAC	ATCGG	---	TGTCGCGGC	GGAC	---	AA	727										
Consensus	AGCACCGAC	KKCATAC	RVCGGCAC	YSTCKVGCY	KLVGGTBA	850													
IS427.Agrotum	AGGAGGTC	ATCCCGGC	ATTCAATC	GGTAGGTC	TATCGGCTC	809													
REVIS.85P4	CCGGTGT	---	GCCTCGTC	ACCAATGAT	---	GTGGTCCG	TATCGGCTC	771											
Consensus	MMGKSTCA	AKCACCGC	AVVCAATC	GGTBAKVS	TAATCGGCTC	900													

Figura 5. Continuación.

IS427.Agrotum	TGCTGGCCT	GATCAGCTT	TTGATCTGTA	ATGGAGCCG	TCATCATTC	859
REV1S.85P4	-C-CACTG	-CTVG-	-T-CCCGA	A- <u>GG</u>	-TCA	794
Consensus	TGCTGGCCT	GAAVYHCTT	TTGAYCAGTA	ATGGAGCCG	TCATCATTC	950
IS427.Agrotum	GGAATCGGA	AGTGGAGGCC	GCGTGCCTCA	AAATCGAGAA	TGCCATARGG	909
REV1S.85P4	G-A- <u>CGA</u>	-----	-----	TTCAAAA	-CA	808
Consensus	GGAATCGGA	AGTGGAGGCC	GCGTGCCTCA	AAATCGAGAA	TGCCATARGG	1000
IS427.Agrotum	TCGATGGAGC	CTCCCGAGC	GCAATCGCC	CGCTCAAAATC	GCTGGCTCCG	959
REV1S.85P4	-----	CC-C-AT	G- <u>ACTGCC</u>	-TCCAAA-	-----	826
Consensus	TCGATGGAGC	CTCCCGAGC	GCAAVCGCC	CGVYCAAAATC	GCTGGCTCCG	1050
IS427.Agrotum	GTCGATCTTC	CTGAGAGTAA	TTAGCATATG	CTGATTTTCA	GCAATGGCGG	1009
REV1S.85P4	-A-----	-GADAGTC-	TT- <u>GAT</u>	CTGA-TTT-	-----	847
Consensus	GTCGATCTTC	CTGAGAGTAA	TTAGCATATG	CTGATTTTCA	GCAATGGCGG	1100
IS427.Agrotum	GGAGGACCG	CTAGCGTCA	GCCGCTTTGG	TTCCAACTCT	ATCCCSCCGA	1059
REV1S.85P4	-----	-----	GCTCA GCC- <u>TTGG</u>	-----GA	ATCCC	866
Consensus	GGAGGACCG	CTAGCGTCA	GCCGCTTTGG	TTCCAACTCT	ATCCCSCCGA	1150
IS427.Agrotum	GGAGCTGCT	CAGATGAGC	GCGACGGAC	TTTCCAGAG	TTTTTCCGG	1109
REV1S.85P4	-AAG	-AG-TC-	C-AC-AC	CGCCTAG-G	-CC-	889
Consensus	GGAGCTGCT	CAGATGAGC	GCGACGGAC	TKVYVAGAG	TTTTTCCGG	1200
IS427.Agrotum	GTTTGTTCGT	GATCGGCGG	AACGGTGGC	GAGAGCGT	GACATTCGAC	1159
REV1S.85P4	-----	TGA- <u>GG</u>	GGTAGCG	G- <u>GAGGTGGT</u>	GTCAT-G-	916
Consensus	GTTTGTTCGT	GATCGGCGG	AACGGTGGC	GAGAGGYSGT	GACATTCGAC	1250
IS427.Agrotum	TTGCGGGAAA	ATCCGCATA	TCCCTGGTC	GCCTTCGACA	TGACACAGCT	1209
REV1S.85P4	-TGC-----	-CGCG--A	T- <u>TGG</u>	GC-	-GATC--T	935
Consensus	TTGCGGGAAA	ATCCGCATA	TCCCTGGTC	GCCTTCGACA	TGACACAGCT	1300
IS427.Agrotum	CGAATCGGA	GAAAGCTCA	GACCGATAGC	GCCATCCTTC	GATGATCGCG	1259
REV1S.85P4	CGAATCGG	-----GCAT	GCCCG-TGGC	G-----C	GATGAA-C	964
Consensus	CGAATCGGA	GAAAGCTCA	GACCGATAGC	GCCATCCTTC	GATGATCGCG	1350

Figura 5. Continuación.

IS427. Agrotum	TGGATT	GAT CGGTCGA	G A	CAAT CA	-----	GT	ATCCG	1292
REVIS. 85P4	---	TTGAT CGGTCGACCG	A	CAATCAAT	CTCTCGCCGC	G	CCGAGCCG	1010
Consensus	TGGATT	GAT CGGTCGACCG	A	CAATCAAT	CTCTCGCCGC	G	CCGAGCCG	1400

IS427. Agrotum	GT	---	ACC	---	A	---	GT	---	TATG	G	---	GT	---	TCAAG	G	1315
REVIS. 85P4	GTCTAGACCT	GTATGCGCGG	GTCCGCTAG	GCTATCGTCG	CGTCCATG	GTG	1060									
Consensus	GTCTAGACCT	GTATGCGCGG	GTCCGCTAG	GCTATCGTCG	CGTCCATG	GTG	1450									

IS427. Agrotum	CGT	---	AAC	AC	AGTACA	-----	-----	-----	1330
REVIS. 85P4	CGTTTGGAC	CCGAGGTTG	GGCACCAC	ATCAGCGAA	CCTATCGCAT	1110			
Consensus	CGTTTGGAC	CCGAGGTTG	GGCACCAC	ATCAGCGAA	CCTATCGCAT	1500			

IS427. Agrotum	-----	-----	---	1330
REVIS. 85P4	TTACAGGGAC	TTGGACTGC	AG	1132
Consensus	TTACAGGGAC	TTGGACTGC	AG	1522

Figura 5. Continuación.

79H4.C.com.seq	ARCTGCATT	GGCTGTCTTC	SATCTTGAG	CCCTGCTTCT	TTATCACC	248
ISRM6	-----	-----	-----	-----	CCCA	14
Consensus	ARCTGCATT	GGCTGTCTTC	SATCTTGAG	CCCTGCYVCH	TTBVCITFCCA	250
79H4.C.com.seq	CGGAAARACT	CTCCATCAT	CCACCGCTGC	TTGTTTCA--	ATCGACAT	295
ISRM6	CGGTAATATT	CTCCAGTTT	GGATTAGAGT	CCGCGCTATT	GAAGGACGAC	64
Consensus	CGGAAABVMT	CTCCBKYVLT	SSBYYRSLGV	YVGVKVCVATT	GAASGACABV	300
79H4.C.com.seq	CTGCCAGGC	GC-CAGTTT	AT-GATGGAA	ATCCCATTC-	GGTGGTFAA-	340
ISRM6	GCATGAGAA	GCAGAGTTT	ACGGAGAGC	AGATTATTGC	GGTROTDAAG	114
Consensus	GVKSMAGRM	GCASGCTV	AVGGAGARM	ARMVYATTGC	GGTRSTFBAAG	350
79H4.C.com.seq	GAGCA--A--	CTAAKGCFAE	GGCTCCGAC	ACCGGTGC-	GT-CGGTIT	382
ISRM6	GAGCAGGAGG	CTGGGCGAAR	GGTGGCGAAC	CTCTGCCGCA	AGCATGGGAT	164
Consensus	GAACAGGAGG	CTVRRGCFAR	GGYKSCGAC	TVCKGVMCA	AGVAVSGGCT	400
79H4.C.com.seq	CGAGGACCT	ACACCCAC	AAATG--AG	CCAAACCC--	CGTCGGGAG	426
ISRM6	TTCCGACCCG	ACATTTTAC	AAATGGAAAG	CCAAATACGG	GGGCATGGAG	213
Consensus	YKMSGACCK	ACCBVYVVMC	AAVTDGAAAG	CCAAVTCGG	SGCKRKGAG	450
79H4.C.com.seq	CCCTTC-TTC	GCGAAGTTT	TGGG-GTTCG-	CCTGATCGTC	GCCTGGCGA	473
ISRM6	GTITCCGAGG	GCGAAGCOT	GAAGGGCCG	AGAGGAGAAC	GCCAGGCTGA	263
Consensus	SVVLTGKWS	GCGSAGVMT	KRMGGVCGA	MSVGRWSRLC	GCCVAGCVBA	500
79H4.C.com.seq	AG---CCCA	GGCAGACGT	TGCCTGCAG-	CGCCGACCG	TCGGCGGC-	517
ISRM6	AGRAGCTGCT	CGCCGACCA	TGCTGGATGC	CGCCGAC--	TCGGCGGCT	311
Consensus	BAAGCYGCA	SGCMSEACG	TGCYKGVVGC	CGCCGACCG	TCBCCGVCCT	550
79H4.C.com.seq	TCGGCAGCT	CGATGTTTG	G-GTGTCCG	GAAC-CAAA	CTGTGTC----	559
ISRM6	TCITGCAAAA	AAATGGTAGG	GCCCGCCGCC	AAACGTAAG	CGGTCCGCGA	361
Consensus	TCYVGBAMW	MAATGGTKG	GCSVGMCGM	AAACVGBAG	CVGTCCGCGA	600
79H4.C.com.seq	TCT--AGGAC	ATCAC--GCA	GTATAGACC	GCGATGAGT	GCAATGTCAT	605
ISRM6	TCTGABGGC	GTCAATGGCC	TCTCGAGCG	GCGG--GCOT	GCCAGATCAT	409
Consensus	TCTGABGGC	ATCAVGGCM	KVLTAGAGCS	GCGGATGMSL	GCMTGATCAT	650

Figura 6. Alineamiento de las secuencias de la clona p79H4 (79H4.C.com.seq), la cual contiene un elemento reiterado de la familia E, y la ISRM6 de *R. meliloti* (ISRM6). Las regiones sombreadas muestran los nucleótidos comunes.

79H4.C.com.seq	GCATGACCGG	CGACAAAGCA	TTGGAAAC-	GTGGGTCCGG	CTGGC-GGGC	651
ISRm6	CTCGGTGAC	CGAAGATGG	TCCGCTATCG	GTACCGCGA	CCGTCTAGGG	459
Consensus	SYMYGMYGMS	CGRMAEAGR	TTCCGCTKCG	GTVRSKCSGR	CTGCTGAGS	700
79H4.C.com.seq	CCACATCATA	AAGAAATCCG	CCATATCCGC	CACTTCGCGA	CCGGCGAACG	701
ISRm6	CCCAACTGCG	AAGAACTTC	CCCACTCCGC	CACTCAGCGA	CCGGCGTTTC	508
Consensus	CCRMAYRYR	BAGAAQVVKC	CCCAVVKCGC	CACTCKMCGA	CCGGCGMYC	750
79H4.C.com.seq	GGCAACCGGC	G--GTT-TCCG	--GGG	GGGGGGGAC	S-CCGTTCGG	742
ISRm6	GGCTACCGCG	GACTGTTTGGT	CCTGCTCAGG	CGGGAGGGAG	AGCCGTCCCG	557
Consensus	GGCAACCGGC	GACTGTTTGGT	CCVCTCAGG	GGGGAGGGAG	VCCCGT/VGG	800
79H4.C.com.seq	GAGTCACCG	CATCTATCCG	CTATATCCGC	AAAGAGGACT	TTCGGTTCGC	792
ISRm6	GAGTCACCG	CATCTATCCG	CTATATCCGC	AAAGAGGACT	TTCGGTTCGC	607
Consensus	GAGTCACCG	CATCTA/SGG	CTATATCCGC	AAAGAGGACT	TTCGGTTCGC	850
79H4.C.com.seq	AAGCAAAAA	CCGGGCGCG	CACTTCGCGC	ACGCTTGCC	CGATCTGGT	842
ISRm6	AAGCGAARA	CCGGGCGCG	TGCGTCCGC	ACG--TGCG	CGATCTGGT	655
Consensus	AAGCRRAAR	SCGGGSGSG	VCAVTCGCGC	ACGCTTGSSC	CGATCTGGT	900
79H4.C.com.seq	CAAGCGAAG	GCCAATGCC	GAAGTCGCT	GAATTCGTT	CAAGACCAIT	892
ISRm6	CAAGCGAAG	GCCAATGCC	GCTGATCGCT	GAATTCGTC	CAAGATCAIT	705
Consensus	CAAGCGAAG	GCCAATGCC	GAAGTCGCT	GAATTCGTY	CAAGACCAIT	950
79H4.C.com.seq	TCGCTTCGG	AAGAGGATT	CGCGTCTCA	ACATGTCFA	AACCTTACC	942
ISRm6	TCGCTTCGG	AAGAGGATT	CGCGTCTCA	ACATGTCFA	TGACCTACC	755
Consensus	TCGCTTCGG	AAGAGGATT	CGCGTCTCA	ACATGTCFA	AACCTTACC	1000
79H4.C.com.seq	CGCGAATGCC	TGGCGCGA	TCCCCTGCAC	CTCGATCTTG	CGTCGTCGG	992
ISRm6	CGCGAATGCC	TGGCGCGA	TCCCCTGCAC	CTCGATCTCC	CGTCGTCGG	804
Consensus	CGCGAATGCC	TGGCGCGA	TCCCCTGCAC	CTCGATCTYS	CGTCGTCGG	1050
79H4.C.com.seq	TTGCTCGC	GAAATGACGG	TCTCTTCGG	AACGGCGCGG	CAAGCCGGG	1042
ISRm6	TTGCTCGC	GAAATGACGG	TCTCTTCGG	AACGGCGCGG	CAAGCCGGG	851
Consensus	TTGCTCGC	GAAATGACGG	YCTCTTCGG	AACGGCGCGG	CAAGCCGGG	1100

Figura 6. Continuación.

79H4.C.com.seq	ATGATGGTCT CCGACATGG CACGGATTTC ACCTCGAATG CCGATCCTTGG	1092
ISRm6	ATGATGGTCT CCGACATGG CACGGATTTC ACCTCGAATG CCGATCCTTGG	901
Consensus	ATGATGGTCT CCGACATGG CACGGATTTC ACCTCGAATG CCGATCCTTGG	1150
79H4.C.com.seq	CAGGTCAAG GATCATATG TCAATTCGCA CCGCATCCG CCGGGAAGC	1141
ISRm6	CAGGTCAAG GATCATATG TCAATTCGCA CCGCATCCG CCGGGAAGC	951
Consensus	CAGGTCAAG GATCATATG TCAATTCGCA CCGCATCCG CCGGGAAGC	1200
79H4.C.com.seq	CGATGCAAA CGDTTATATC GAAAGCTTCA ACGGCAGAT GCGGAGCCAG	1191
ISRm6	CGATGCAAA CGDTTATATC GAAAGCTTCA ACGGCAGAT GCGGAGCCAG	1001
Consensus	CGATGCAAA CGDTTATATC GAAAGCTTCA ACGGCAGAT GCGGAGCCAG	1250
79H4.C.com.seq	ATGDTTAACT AAGCCTGTT CTTGGTCTC GATATGCCC GAGCGCCAT	1241
ISRm6	ATGDTTAACT AAGCCTGTT CTTGGTCTC GATATGCCC GAGCGCCAT	1051
Consensus	ATGDTTAACT AAGCCTGTT CTTGGTCTC GATATGCCC GAGCGCCAT	1300
79H4.C.com.seq	TGC---TGG GCGAGATT AATACCAATT CCGGCCGCAC TCATCGCTCG	1287
ISRm6	TGC---TGG GCGAGATT AATACCAATT CCGGCCGCAC TCATCGCTCG	1101
Consensus	TGC---TGG GCGAGATT AATACCAATT CCGGCCGCAC TCATCGCTCG	1350
79H4.C.com.seq	GA---CAGAC CCGGCACGT GATGCGATA CCATGCCC GACCCGGCTTC	1335
ISRm6	GATATCAGAC CCGGCACGT TATGCGATA CCATGCCC GACCCGGCTTC	1151
Consensus	GATATCAGAC CCGGCACGT TATGCGATA CCATGCCC GACCCGGCTTC	1400
79H4.C.com.seq	AACGCTGGC AAGATGAA--A-----GC TT-----	1358
ISRm6	AACGCTGGC AAGATGAA--A-----GC TT-----	1201
Consensus	AACGCTGGC AAGATGAA--A-----GC TT-----	1450
79H4.C.com.seq	-----	1358
ISRm6	CAAGTGGCGT AATGAAACG CCGAGGCTCT AATCGCCGCT GGATGAAAGT	1251
Consensus	CAAGTGGCGT AATGAAACG CCGAGGCTCT AATCGCCGCT GGATGAAAGT	1500
79H4.C.com.seq	-----	1358
ISRm6	TCAGTGGCAG GTACCCCA	1269
Consensus	TCAGTGGCAG GTACCCCA	1518

Figura 6. Continuación.

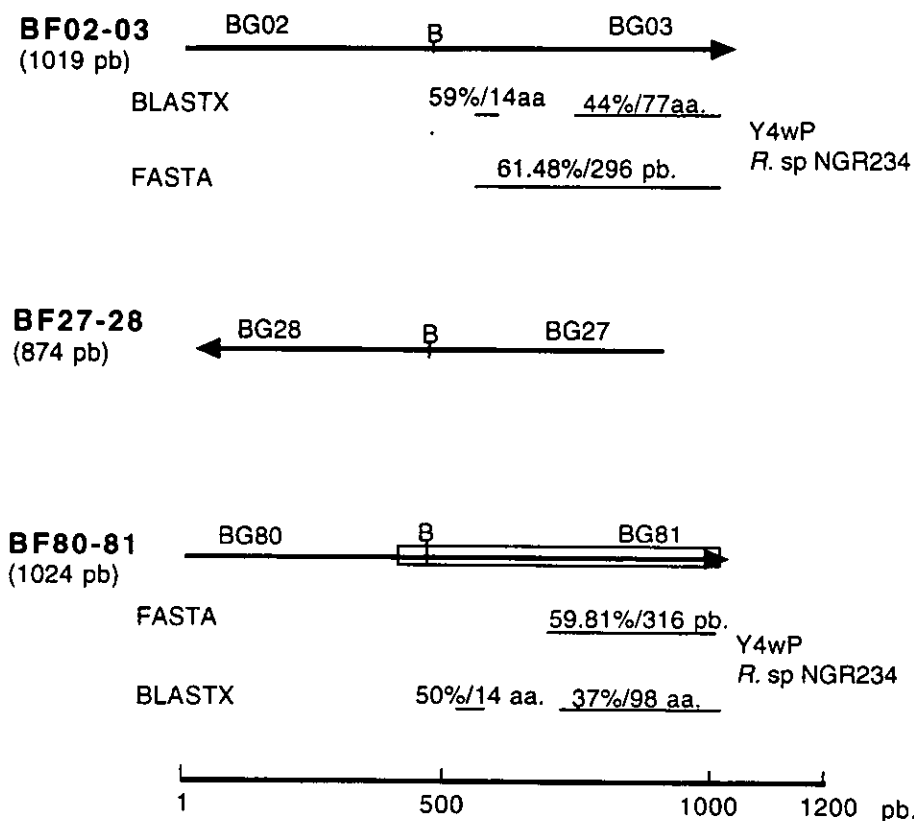


Figura 7. Representación esquemática de los elementos de la familia reiterada BF alineados de acuerdo a sus sitios internos de Bam HI (B). Las líneas gruesas indican la orientación de cada elemento, sobre ellas se muestran los extremos de las bandas de Bam HI en los cuales se encuentra el elemento y debajo se denotan los resultados de los análisis tipo FASTA y BLASTX representados por las líneas delgadas, con los porcentajes de identidad encontrados con el MLA Y4wP de *R. sp.* NGR234. El rectángulo del elemento BF80-81 representa un MLA sin señal de término.

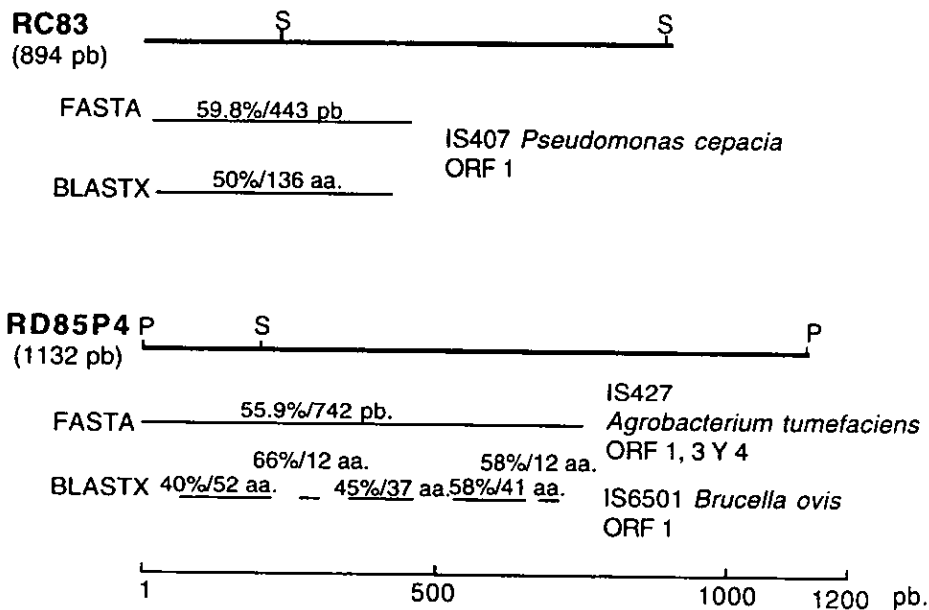


Figura 8. Esquema de las secuencias que contienen los elementos reiterados de las familias C y D1 de las bandas BG83 (RC83) y BG85 (RD85P4) de la genoteca de Bam HI del pSim de *R. etli* CFN42, debajo se muestran los porcentajes de identidad obtenidos por los análisis FASTA y BLASTX además del nombre de las secuencias. Sitios de corte: P, PstI; S, Sal I.

RE79H4
(1358 pb.)

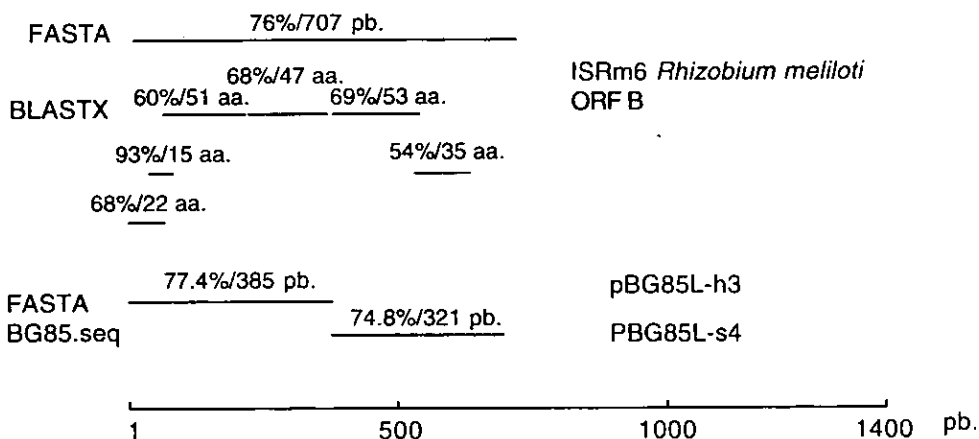


Figura 9. Esquema de la secuencia de la clona p79H4 que contiene un elemento de la familia reiterada E, así mismo se presentan los resultados obtenidos por los programas FASTA y BLASTX y las secuencias de las clonas de la BG85 que mostraron alta identidad con esta secuencia. Sitios de corte: H, Hind III; S, Sal I.

53.33	61.50	58.67	61.33	57.50
60.67	61.83	55.67	52.50	59.67
53.00	54.83	59.33	58.83	64.00
58.00	60.00	56.00	52.00	51.50
56.00	59.00	50.17	58.83	56.17
54.83	62.83	57.44	59.00	59.67
59.33	44.33	58.02	60.85	60.00
52.83	60.37	48.17	53.67	54.83
50.33	50.83	51.67	48.67	52.67
52.33	62.00	63.67	57.33	53.33

Tabla 1. Porcentajes de GC obtenidos de 50 secuencias de 300 pb. del pSim de *R. etli* CFN42 utilizadas para realizar calcular la media muestral de contenido de GC para su comparación con las secuencias de los elementos de las familias BF, C, D1 y E secuenciados. $\sigma = 4.4698$, $X = 56.2166$.

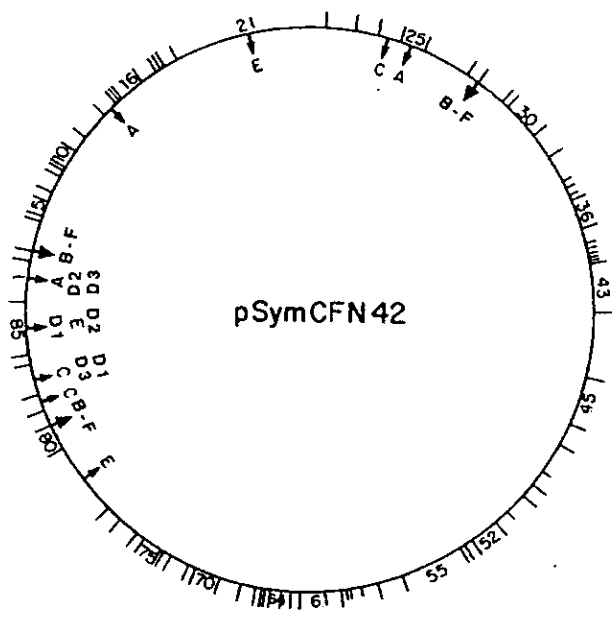


Figura 10. Localización de los elementos de las familias reiteradas BF, C, D1, D2, D3 y E en el mapa de restricción con Bam HI del pSim de *R. etli* CFN42. El resto de los elementos reiterados fueron omitidos.

FAMILIA REITERADA	BANDA	TAMAÑO (pb)	IDENTIDAD
B-F	02-03	1019	Y4wP
	27-28	874	pNGR234a
	80-81	1024	
C	24	5190	IS407
	82	3400	<i>Pseudomonas</i>
	83	874	<i>cepacia</i>
D1	83	7200	IS427
	85	1132	<i>A. tumefaciens</i>
D2	85	12900	?
	01	9400	
D3	01	9400	?
	83	7200	
E	21	33600	ISRm6
	79	1358	<i>R. melliloti</i>
	85	700	

Cuadro 2. Elementos reiterados de las familias BF, C, D1, D2, D3 y E del pSim de *R. etli* CFN42. En números cursivos se muestran el tamaño de los elementos secuenciados hasta el momento.

8. BIBLIOGRAFÍA

- Anderson, R. P. y Roth, J. R. (1977). Tandem Genetic Duplications in Phage and Bacteria. *Ann. Rev. Microbiol.* 31: 473-505.
- Anderson, R. P. y Roth, J. (1981). Spontaneous Tandem Genetic Duplication sin *Salmonella Typhimurium* Arise by Unequal Recombination between rRNA (*rrn*) Cistrons. *Proc. Natl. Acad. Sci. USA* 78(5): 3113-1317.
- Appelbaum, E. R. et al. (1988). *Rhizobium japonicum* USDA 191 Has Two *nodD* Genes That Differ in Primary Structure and Function. *J. Bacteriol.* 170(1):12-20.
- Avise, J. C. (1994). *Molecular Markers, Natural Hystory and Evolution.* Charman & Hall. New York.
- Baldani, J. I. et al. (1997). Recent Advances in BNF Whith Non-Legume Plants. *Soil Biol. Biochem.* 29 (5/6): 911-922.
- Balleza, D. (1998). Análisis de Secuencias de DNA Reiterado y Predicción de Rearreglos Genómicos en *Rhizobium*. Tesis de Maestría en Investigación Biomédica Básica. CIFN. UNAM.
- Barbour, W. M. et al. (1985). Evidence for plasmid and Cromosome-Borne Multiple *nif* Genes in *Rhizobium fredii*. *App. Environ. Microbiol.* 50 (1) : 41-44.
- Barnett, M. J. et al. (1996). Transcription Start Sites for *syrM* and *nodD3* flank an Insertion Sequence Relic in *Rhizobium meliloti*. *J. Bacteriol.* 178(7): 1782-1787.
- Bar-Nir, D. et al. (1994). Reiterated Sequences of *Streptomyces griseus* Plasmid pSG1. *Plasmid* 32: 41-45.
- Birch, A. et al. (1990). Genome Rearrangement and Inestability in *Streptomyces* spp. *J. Bacteriol.* 172(8): 4138-4142.
- Bisercic, M. y Ochman, H. (1993)a. Natural Populations of *E. coli* and *Salmonella typhimurium* Harbor the Same Classes of Insertion Sequences. *Genetics* 133: 449-454.
- Bisercic, M. y Ochman, H. (1993)b. The Ancestry of Insertion Sequences Common to *Escherichia coli* and *Salmonella typhymurium*. *J. Bacteriol.* 175(24): 7863-7868.
- Blattner, F. R. et al. (1997). The Complete Genome Sequence of *Escherichia coli* K-12. *Science* 277: 1453-1462.

- Boivin, C. et al. (1997). Steam Nodulation in Legumes: Diversity, Mechanisms, and Unusual Characteristics. *Crit. Rev. Plant. Sci.* 16(1): 1-30.
- Britten, R. J. y Kohne, D. E. (1968). Repeated Sequences in DNA. *Science* 161: 529-540.
- Brom, S. et al. (1991). High-Frequency Rearrangements in *Rhizobium leguminosarum* bv. phaseoli Plasmids. *J. Bacteriol.* 173(3): 1344-1346.
- Brom, S et al. (1992). Different Plasmids of *Rhizobium leguminosarum* bv phaseoli are Required for Optimal Symbiotic Performance. *J. Bacteriol.* 174(16): 5183-5189.
- Bult, C. J. et al. (1996). Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*. *Science* 273: 1058-1073.
- Campbell, R. C. (1989). *Statistics for Biologists*. 3er. ed. Cambridge University Press.
- Carlson, R. W. et al. (1994). The Biosynthesis of Rhizobial Lipoligosaccharide Nodulation Signal Molecules. *Mol. Plant Interac.* 7(6):684-695.
- Carniel, E. et al. (1996). Characterization of a Large Chromosomal "High-Pathogenicity Island" in Biotipe 1B *Yersenia enterocolitica*. *J. Bacteriol.* 178(23): 6743-6751.
- Charlebois, R. L. y Doolittle, W. F. (1989). Transposables Elements and Genome Structure in Halobacteria. pp. 297-307, en Berg, D. E. y Howe, M. M. *Mobile DNA*. American Society for Microbiology. Washintong D. C.
- Charles, T. C. y Finan, T. M. (1990). Genetic Map of *Rhizobium meliloti* Megaplasmid pRmeSu47b. *J. Bacteriol.* 172(5): 2469-2476.
- Charter, K. F. et al. (1988). Genome Flux in *Streptomyces coelicolor* and other Streptomycetes and its Possible Relevance to Evolution of Mobile Antibiotic Resistance Determinants. pp. 7-42, en Kingsman, A. J. et al. *Transposition*. Cambridge Universtity Press. Cambridge.
- Chien, Y. y Zinder, S. H. (1996). Cloning, Functional Organization, transcript Studies, and Phylogeneic Analysis of the Complete Nitrogenase Structural Genes (*nifHDK2*) and Asociated Genes in the Archeon *Methanosarcina barkeri* 227. *J. of Microbiol.* 178 (1):143-148.
- Cohen-Kupiek, R. et al. (1997). Transcriptional Regulation in Archaea: In vivo demonstration of a Represor site in a Methanogen. *Proc. Natl. Acad. Sci. USA* 94: 1316-1320.

- Colón-lópez, M. S. et al. (1997). Transcriptional and Translational Regulation of Nitrogenase in Ligth-Dark and Continuous-Ligth- Gr own of the Unicelullar Cyanobacterium *Cyanothece* sp Strain ATCC 51142. J. Bacteriol. 179 (13): 4319-4327.
- David, M. et al. (1987). Trancription Patterns of *Rhizobium meliloti* Symbiotyc Plasmid pSym: Identification of *nifA*-Independent *fix* Genes. J. Bacteriol.169(5): 2239-2244.
- Deng, W. et al. (1995). Sequence and Distribution of IS1312: Evidence for Horizontal Transfer for *Rhizobium meliloti* to *Agrobacterium tumefaciens*. J. Bacteriol. 177(9): 2526-2559.
- De Meirsmán, C. et al.(1990). Nucleotide Sequence Analysis of IS427 and Its Target Sites in *Agrobacterium tumefaciens* T37. Plasmid 24: 227-234.
- Díaz, C. L. et al.(1995). Genomic Requirements of Rhizobium for Nodulation of White Clover Hairly Roots Transformed with the Pea Lectin Gene. Mol. Plant-Microb. Interact. 8 (3): 348-356.
- Dilworth, M y Glenn, A. (1984). How Does Legume Nodule Work? Trends in Biochem. Scien. 9: 519-523.
- Dusha, I. et al. (1987). *Rhizobium meliloti* Insertion Element ISRm2 and Its Use for Identification of the *fix* Gene. J. of Bateriaol. 169(4):1403-1409 .
- Fischer, H. M. (1994). Genetic Regulation of Nitrogen Fixation Rhizobia. Microbiol. Rev. 58(3): 352-386.
- Fisher, R. F. y Long, S. R. (1992). *Rhizobium*-Plant Signal Exchange. Nature 357: 655- 659.
- Fleischmann, R. D. et al.(1995). Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd. Science 269: 496-512.
- Flores, M. et al. (1987). Reiterated DNA Sequences in *Rhizobium* and *Agrobacterium* spp. J. Bacteriol.169(12): 5782-5788.
- Flores, M. et al. (1988). Genomic Inestability in *Rhizobium phaseoli*. J. of Bactriol. 170(3): 1191-1196.
- Flores, M. et al.(1993). Gene Amplification in *Rhizobium*: Identification and *in vivo* Cloning of Discrete Amplifiable DNA Regions (Amplicons) for *Rhizobium leguminosarum* biovar *phaseoli*. Proc. Natl. Acad. USA. 90: 4932-4936.
- Fraser, C. M. et al. (1997)a. The Minimal Gene Complement of *Mycoplasma genitalium*. Science 270: 397-403.

- Fraser, C. M. et al. (1997)b. Genomic Sequence of Lyme Disease Spirochaete *Borrelia burgdorferi*. Nature 390: 580-586.
- Freiberg, C. et al. (1997). The Molecular Basis of Symbiosis Between *Rhizobium* and legumes. Nature 387: 394-401.
- Galas, D. J. y Chandler, M. (1989). Bacterial Insertion Sequences. 109-162 pp., en Berg, D. E. y Howe, M. M. (eds), Mobile DNA. American Society for Microbiology. Washington D. C. (1989).
- García, A. S. et al. (1996). *Rhizobium* Plasmids in Bacteria-Legume Interactions. World J. of Microbiol. and Biotech. 12: 119-125.
- Geelen, D. et al. (1995). The *nodD* Locus from *Azorhizobium caulinodans* is Flanked by Two Repetitive Elements. Gene 164: 107-111.
- Girard, M. L. et al. (1991). Structural Complexity of the Symbiotic Plasmid of *Rhizobium leguminosarum* bv. phaseoli. J. Bacteriol. 173(8): 2411-2419.
- Guédon, G. et al. (1995). Characterization and distribution of Two Insertion Sequences, IS1191 and Iso-IS981, in *Streptococcus thermophilus*: Does Intergeneric Transfer of Insertion Sequences Occur in Lactic Acid Bacteria Co-cultures? Mol. Microbiol. 16(1): 69-78.
- Hacker, J. et al. (1997). Pathogenicity Islands of Virulent Bacteria: structure, function and impact on microbial evolution. Mol. Microbiol. 23(6): 1089-1097.
- Hahn, M. y Hennecke, H. (1987). Mapping of *Bradirhizobium japonicum* DNA Region Carrying Genes for Symbiosis and an Asymmetric Accumulation of Reiterated Sequences. Appl. Environ. Microbiol. 53(9): 2247- 2252.
- Halverson, L. J. y Stacy, G. (1986). Signal Exchange in Plant-Microbe Interaction. Microbiol. Rev. 50 (2): 193-225.
- Hanahan, D. (1983). Studies on Transformation of *Escherichia coli* Plasmids. J. Mol. Biol. 166: 557-580.
- Hill, C. W. (1977). Tandem Duplications Resulting From Recombination between Ribosomal RNA Genes in *Escherichia coli*. J. Mol. Biol. 116: 407-428.
- Hill, C. W. (1981). Inversions between Ribosomal RNA Genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 78(11): 7069- 7072.
- Himmelreich, R. et al. (1996). Complete Sequence analysis of the Genome of the Bacterium *Mycoplasma pneumoniae*. Nucl. Acids Res. 24(22): 4420-4449.

- Hirsch, A. M. et al. (1995). Assessing Horizontal Transfer of *nifHDK* Genes in Eubacteria: Nucleotide Sequence of *nifK* from *Frankia* Strain HFPCcl3. *Mol. Biol. Evol.* 12(1): 16-27.
- Homna, M. A. y Ausbel, F. M. (1987). *Rhizobium meliloti* has Three Functional Copies of *nodD* Symbiotic Regulatory Gene. *Proc. Natl. Acad. Sci. USA* 84: 8558-8562.
- Jiménez-Salgado, T. et al (1997). *Coffea arabica* L., a New Host Plant for *Acetobacter diazotrophicus*, and Isolation of other Nitrogen-Fixing Acetobacteria. *App. and Environ. Microbiol.* 63 (9): 3676-3683.
- Jørgensen, P. (1976). A Ribosomal RNA Gene of *Escherichia coli* (*rrD*) on λ daroE Specialized Transducing Phages. *Mol. Gen. Genet.* 146: 303-307.
- Kaluza, K. et al. (1985). Repeated Sequences Similar to Insertion Elements Clustered Around the *nif* Region of *Rhizobium japonicum* Genome. *J. Bacteriol.* 162(2): 535-542.
- Kenerley, M. E. et al. (1977). Characterization of the Hybrid Plasmids Carrying Individual Ribosomal Ribonucleic Acid Transcription Units of *Escherichia coli*. *J. Bacteriol.* 132(30): 931-949.
- Kiss, A. et al. (1977). The Number of rRNA Genes in *Escherichia coli*. *FEBS Lett.* 79(1): 77-79.
- Klenk, H. P., et al. (1997). The Complete Genome Sequence of the Hyperthermophilic, Sulphate-Reducing, Archaeon *Archaeoglobus fulgidus*. *Nature* 390: 364-370.
- Kuns, F. et al. (1997). The Complete Genome Sequence of the Gram-positive Bacterium *Bacillus subtilis*. *Nature* 390:249-256.
- Laberge, S. et al. (1995). Characterization, Nucleotide Sequence, and Conserved Genomic Localizations of Insertion Sequence ISRM5 in *Rhizobium meliloti*. *J. Bacteriol.* 177(11): 3113-3142.
- Lehner, A. F. y Hill, C. W. (1980). Involvement of Ribosomal Ribonucleic Acid Operons in *Salmonella typhimurium* Chromosomal Rearrangements. *J. Bacteriol.* 143(1): 492-498.
- Lehninger, A. L. (1978). *Bioquímica*. 1a. reimp. Omega. Barcelona. 1117 pp.
- Long, S. R. (1989)a. *Rhizobium* Genetics. *Annu. Rev. Genet.* 22: 483-506.
- Long, S. R. (1989)b. *Rhizobium*-Legume Nodulation: Life Together in the Underground. *Cell* 56: 203-214.

- Martínez, E. et al. (1985). Reiteration of Nitrogen Fixation Gene Sequences and Specificity of *Rhizobium* in Nodulation and Nitrogen Fixation in *Phaseolus vulgaris*. J. of General. Microbiol. 131:1779-1786.
- Martínez, E. et al.(1990) The *Rhizobium* Genome. Critical Rev. in Plant Scien. 9(1):59-93.
- Masterson, R. V. y Atherly, A. G. (1986). The Presence of Repeated DNA Sequences and a Partial Restriction Map of the pSym of *Rhizobium fredii* USDA 193. Plasmid 16: 37- 44.
- Mathews, C. K. y Van Holde, K. E. (1990). Biochemistry. The Benjamin/Cummings Publishing Company Inc. Redwood City. 1129 pp.
- Meirsmann, C. de et al. (1989). Identification of Insertion Sequence Element IS427 in pTiT37 Plasmid DNA of an *Agrobacterium tumefaciens* T37. Plasmid 21: 129-137.
- Meirsmann, C. de et al. (1990). Nucleotide Sequence Analysis and its Target Sites in *Agrobacterium tumefaciens* T37. Plasmid 24: 227-234.
- Mellor, R. y Werner, D. (1987). Peribacteroidal Membrane Biogenesis in Mature Legume Root Nodules. Symbiosis 3: 75-100.
- Mergaert, P. et al.(1997). Molecular Mechanisms of Nod Factor Diversity. Mol. Microbiol. 25(5): 811-817.
- Merric, M. J. (1992). Regulation of Nitrogen Fixation Genes in Free-Living and Symbiotic Bacteria. pp 835-876, en G. Stacey, R. H. Burris y H. J. Evans (Eds.), Biological Nitrogen Fixation. Chapman & Hall. New York.
- Mitsui, et al. (1986). Strategy by Which Nitrogen-Fixing Unicellular Cyanobacteria Grow Photoautotrophically. Nature 323 (23):720-722.
- Nap, J. P. y Bisseling, T. (1990). Developmental Biology of Plant-Prokaryote Symbiosis: The Legume Root Nodule. Science 250: 948-954.
- Neilson, J. W. et al. (1994). Frequency of Horizontal Transfer of a Large Catabolic Plasmid (pJP4) in Soil. App. Environ. Microbiol. 60(11): 4053-4058.
- Newton, W. E. (1993). Nitrogenases: Distribution, Composition, Structure and Function, en Palacios et al. New Horizons in Nitrogen Fixation. Kluwer Academic Publishers. Países Bajos.

- Norel, F. y Elmerich, C. (1987). Nucleotide Sequence and Functional Analysis of Two *nifH* Copies of *Rhizobium* ORS571. *J. of General Microbiol.* 133: 1563-1576.
- Noti J. D. et al. (1986). Organization and Characterization of Genes Essential for Symbiotic Nitrogen Fixation from *Bradyrhizobium japonicum* I110. *J. Bacteriol.* 167(3): 774-783.
- Ogawa, J. et al. (1991). Analysis of *Rhizobium meliloti* Mutant WL131: Novel Insertion Sequence ISRM3 in nodG and Altered nod H Protein Product. *J. Bacteriol.* 173(10): 3060-3065.
- Ohtsubo, E. H. et al. (1984). An Evolutionary Analysis of Iso-IS1 Elements from *Escherichia coli* and *Shigella* Strains. *J. Gen. Appl. Microbiol.* 30: 359-376.
- Otten, L. et al. (1992). Evolution of Agrobacteria and Their Ti Plasmids- A Review. *Mol. Plant Microb. Interac.* 5(4): 279-287.
- Ouahrani, S. et al. (1993). Identification and Sequence Analysis of IS6501, an Insertion Sequence in *Brucella* spp.: Relation Between Genomic Structure and the Number of IS6501 Copies. *J. Gen. Genet.* 139: 3265-3273.
- Pau, R. N. (1989). Nitrogenases Without Molybdenum. *Trends in Biochem. Scien.* 14: 183-186.
- Peterson, B. C. y Rownd, R. H. (1985). Recombination Sites in Plasmids Drug Resistance Gene Amplification. *J. Bacteriol.* 164(3): 1359-1361.
- Peterson, S. N. et al. (1993). A Survey of the *Mycoplasma genitalium* Genome by Using Random Sequencing. *J. Bacteriol.* 175(24): 7918-7930.
- Peterson, S. N. et al. (1995). Characterization of Repetitive DNA in *Mycoplasma genitalium* Genome: Possible role in the Generation of Antigenic Variation. *Proc. Natl. Acad. Sci. USA* 92: 11829-11833.
- Perret, X. et al. (1997). Structure and Evolution of NGRRS-1, a Complex, Repeated Element in the Genome of *Rhizobium* sp. Strain NGR234. *J. Bacteriol.* 179(23): 7488-7496.
- Polard, P. y Chandler, M. (1995). Bacterial Transposases and Retroviral Integrases. *Mol. Microbiol.* 15(1): 13-23.
- Priefer, U. B. et al. (1989). ISR1, A Transposable DNA Sequence Resident in *Rhizobium* ClassIV Strains, Show Structural Characteristics of Classical Insertion Elements. *Plasmid* 21: 120-128.

- Quinto, C. et al. (1982). Reiteration of Nitrogen Fixation Gene Sequences in *Rhizobium phaseoli*. *Nature* 299(5885): 724-726.
- Quinto, C. et al. (1985). Nitrogenase Reductase: A Functional Multigene Family in *Rhizobium phaseoli*. *Proc. Natl. Acad. Sci. USA* 82: 1170-1174.
- Ramírez-Romero, M. A. (1997). Sequence, Localization and Characteristics of the Replicator Region of the Symbiotic Plasmid of *Rhizobium etli*. *Microbiol.*143(8): 2825-2831.
- Renalier, M. H. et al.(1987). A New Symbiotic Cluster on the pSym Megaplasmid of *Rhizobium meliloti* 2011 Carries a Functional *fix* Gene Repeat and *nod* Locus. *J. Bacteriol.* 169(5): 2231-2238.
- Robertson, J. G. et al. (1978). Membranes in Lupin Root Nodules.I. The Role of the Golgi Bodies in the Biogenesis of Infection Threads and the Peribacteroidal Membrane. *J. Cell. Sci.* 30: 129-149.
- Rodríguez, S. C. (1992). Amplificación y Delección por Recombinación entre *nifH* de *Rhizobium phaseoli* ¿Un Fenómeno General entre Cepas? Tesis de Maestría en Investigación Biomédica Básica. CIFN. UNAM.
- Rodríguez, S. O. (1992). Estructura y Evolución del Genoma Simbiótico de *Rhizobium phaseoli*. Tesis de Maestría en Investigación Biomédica Básica. CIFN. UNAM.
- Romero, D. (1991). Mecanismos Moleculares Involucrados en la Dinámica de las Familias reiteradas de DNA en el Genoma de *Rhizobium phaseoli*. Tesis Doctoral en Investigación Biomédica Básica. CIFN. UNAM.
- Romero, D. et al. (1995). Discrete Amplifiable Regions (Amplicons) in the Symbiotic Plasmid of *Rhizobium etli* CFN42. *J. Bacteriol.* 177(4): 973-980.
- Romero, D. et al. (1995). The Dynamic Genome of *Rhizobium*. En *Bacterial Genomes: Physical Structural Analysis*. Eds. Bruijin, F. J., Lupski, J. R. y Weinstock, G. Chapman and Hall. New York.
- Romero, D. y Palacios R. (1997). Gene amplification and Genomic Plasticity in Prokaryotes. *Ann. Rev. Genet.* 31: 91-111.
- Ruvkun, G. B. et al. (1982). IS*Rm1*: A *Rhizobium meliloti* Insertion Sequence That Transposes Preferentially into Nitrogen Fixation Genes. *J. of Mol. and App. Genet.* 1: 405-418.
- Sambrook, J. et al. (1989). *Molecular Cloning*. 2a. ed. Cold Spring Harbor Laboratory Press. New York.

- Sapienza, C. et al. (1982). High-Frequency Genomic Rearrangements Involving Archaeobacterial Repeat Sequence Elements. *Nature* 299: 182-185.
- Schmid-Appert, M. et al. (1997). Association of Newly IS Elements with the Dichloromethane Utilization Genes of Methylophilic Bacteria. *Microbiol.* 143: 2557-2567.
- Schofield, P. R. y Watson, J. M. (1986). DNA Sequence of *Rhizobium trifolii* Nodulation Genes Reveals a Reiterated and Potentially Regulatory Sequence Preceding *nodABC* and *nodFE*. *Nucl. Ac. Res.* 14(7): 2891-2903.
- Schofield, P. R. et al. (1987). Evidence for Exchange and Recombination of *Rhizobium* Symbiotic Plasmids in a Soil Population. *Appl. Environ. Microbiol.* 53(12): 2942-2947.
- Schultze, M. et al. (1994). Cell and Molecular Biology of *Rhizobium*-Plant Interaction. *Internat. Rev. of Cytol.* 156: 1-75.
- Selbitschka, W. et al. (1995). Construction and Characterization of a *Rhizobium leguminosarum* biovar *viciae* Strain Designed to Assess Horizontal Gene Transfer in the Environment. *FEMS Microbiol Lett.* 128: 255-263.
- Short, J. M. et al (1988). λ ZAP: a Bacteriophage λ expression vector with *in vivo* excision properties. *Nucl. Acids Res.* 16(15): 7583-7600.
- Sibold, L. y Souillard. (1988). Genetic Analysis of Nitrogen Fixation in Methanogenic Archaeobacteria. pp. 705-710, en H. Bothe, F. J. de Buij in and W. E. Newton (eds.), *Nitrogen Fixation : Hundred Years After Gustav Fisher*, Stuttgart (1988).
- Smith, D. R. et al. (1997). Complete Genome Sequence of *Methanobacterium thermoautotrophicum* Δ H: Functional Analysis and Comparative Genomics. *J. Bacteriol.* 179(22): 7135-7155.
- Soberón, G. et al. (1986). Genetic Rearrangements of a *Rhizobium phaseoli* Symbiotic Plasmid. *J. Bacteriol.* 167(2): 487-481.
- Sobral, B. W. S. et al. (1991). The genomes of Family *Rhizobiaceae*: Size, Stability, and Rarely Cutting Restriction Endonucleases. *J. Bacteriol.* 173(2): 704-709.
- Sonti, R. V. y Roth, J. R. (1989). Role of Gene Duplications in the Adaptation of *Salmonella tiphymurium* to Growth on Limiting Carbon Sources. *Genetics* 123: 19-28.
- Soto, M. J. et al. (1992). Sequence of ISRM4 from *Rhizobium meliloti* strain GR4. *Gene* 120: 125-126.

- Sprent, I. J. (1987). The Ecology of the Nitrogen Cycle. Cambridge University Press. Cambridge. 151 pp.
- Sutton, W. D. et al. (1981). The *Rhizobium* Bacteroid State. 149-177 pp., en Bourne, G. H. y Danielli, J. F. (eds), International Review of Cytology. Academic Press. New York (1981).
- Tomb, J. F. et al. (1997). The Complete Genome Sequence of the Gastric Pathogen *Helicobacter pylori*. Nature 388: 539-547.
- Valdés, M. et al. (1996). La Biología de *Frankia*, Endosimbionte de Árboles Fijadores de Nitrógeno. Ciencia 47: 51-67.
- Van de Putte, P. y Goosen, N. (1992). DNA Inversions in Phages and Bacteria. T. I. G. Rev.8(12): 457-462.
- Waring, M. y Britten, R.J. (1966). Nucleotide Sequence Repetition: a Rapidly Reassociating Fraction of Mouse DNA. Science 154: 791-794.
- Wheatcroft, R. y Watson, R. J. (1988). Distribution of Insertion Sequence ISRM1 in *Rhizobium meliloti* and Other Gram-Negative Bacteria. J. of General Microbiol. 134: 113-121.
- Wheatcroft, R. y Laberge, S. (1991). Identification and Nucleotide Sequence of *Rhizobium meliloti* Insertion Sequence ISRM3: Similarity between the Putative Transposase Encoded by ISRM3 and those Encoded by *Staphylococcus aureus* IS256 and *Thiobacillus ferrooxidans* IST2. J. Bacteriol. 173(8): 2530-2538.
- Wolk, C. P. (1996). Heterocyst Formation. Annu. Rev. Genet. 30: 59-78.
- Wood, M. et al. (1991). IS406 and IS407, Two Gene-Activating Insertion Sequences from *Pseudomonas cepacia*. Gene 105(1): 101-105.
- Young, M. (1984). Gene Amplification in *Bacillus subtilis*. J. Gen. Microbiol. 130: 1613-1621.
- Zecrí, S. y Toro, N. (1996). Identification and Nucleotide Sequence of *Rhizobium meliloti* Insertion Sequence ISRM6, a Small Transposable Element that Belongs to the IS3 Family. Gene 175: 43-48.