

201



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

EVOLUCION CONCERTADA ENTRE MIEMBROS DE UNA FAMILIA MULTIGENICA EN *Rhizobium etli*

T E S I S

QUE PARA OBTENER EL TITULO DE:

B I O L O G O

P R E S E N T A :

EDGARDO ALFREDO SEPULVEDA SANCHEZ HIDALGO



DIRECTOR DE TESIS: DR. DAVID RENE ROMERO CAMARON



2003 FACULTAD DE CIENCIAS SECCION DE TESIS

TESIS CON FALLA DE ORIGEN



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



SECRETARÍA DE EDUCACIÓN PÚBLICA
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Autorizo a la Dirección General de Bibliotecas UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional
 NOMBRE: Edgardo Alfredo Sepúlveda Sánchez Hidalgo
 FECHA: 3/12/02
 FIRMA: [Firma]

DRA. MARÍA DE LOURDES ESTEVA PERALTA
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito: **Evolución Concertada entre Miembros de una Familia Multigénica en Rhizobium etli**

realizado por **Edgardo Alfredo Sepúlveda Sánchez Hidalgo**

con número de cuenta **9853132-1**, quien cubrió los créditos de la carrera de: **Biología**

Dicho trabajo cuenta con nuestro voto aprobatorio.

A t e n t a m e n t e

Director de Tesis **Dr. David René Romero Camarena**
 Propietario

Propietario **Dra. Luisa Alvarina Alba Lois**

Propietario **Dr. Arturo Carlos II Becerra Bracho**

Suplente **Dra. Valeria Francisca Eugenia Leopoldina de María de Guadalupe Souza Saldivar**

Suplente **Dr. Daniel Ignacio Piñero Dalmau**

Consejo Departamental de

[Firma]
 M. en C. **Juan Manuel Rodríguez Chávez**

FACULTAD DE CIENCIAS
 U. N. A. M.



DEPARTAMENTO
 DE BIOLÓGICA

A la memoria del Ing. Octavio Sánchez-Hidalgo

A la Universidad Nacional Autónoma de México

AGRADECIMIENTOS

ACADÉMICOS

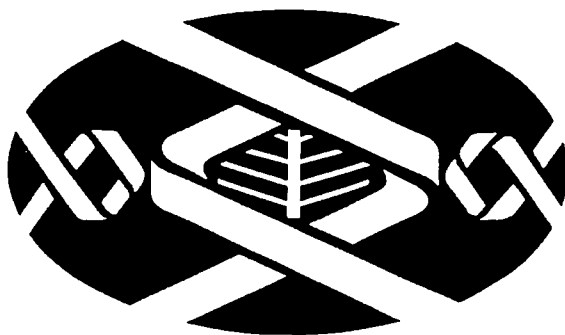
- Al Dr. David Romero por el apoyo, la confianza, la formación y la casi infinita paciencia que mostró tanto en los aspectos académicos como en los no académicos. Gracias por ese único consejo no solicitado.
- A los revisores del trabajo por los comentarios que derivaron en su enriquecimiento, especialmente al Dr. Daniel Piñero, por escribir esos dos libros que alguna vez sorprendieron a un estudiante de segundo de secundaria y a la Dra. Luisa Alba Lois por su guía y amistad.
- A los Profesores del taller de Ciencias Genómicas, especialmente a la Dra. Gina Hernández, directora del centro, por el apoyo que brindó para nuestra integración a la comunidad del CIFN.
- Al Dr. Guillermo Dávila por las facilidades otorgadas para el uso del secuenciador.
- A Claudia Silva y a la Dra. Valeria Souza por facilitarnos las secuencias de las cepas IE.
- Al Dr. Rafael Palacios y a Margarita Flores por proporcionarnos los oligonucleótidos E
- A Rosa Isela Santamaría por su valioso y experimentado apoyo técnico en la secuenciación.
- A José Espíritu por su valiosa ayuda técnica en el aspecto informático.
- Al Biol. Gustavo Santoyo por ser un gran amigo y compañero de proyecto y al M. en C. Manuel Granados por su amistad, confianza y todos los consejos, tips y malas mañan que me ayudaron a sacar adelante el trabajo.
- A los Drs. Jaime "James" Martínez, María de Lourdes Girard, Cesar Rodríguez, Adriana Corvera, Miguel Ramírez y Edgar Valencia por enseñarme (obligarme) a hacer las cosas bien y no como decía Manuel. A Susi Brom por su confianza para lo que hoy es mi proyecto de doctorado. A Araceli, Jacobo, Alejandro, Laura, Nicolás, Alex, Vicky, Mary, Enrique, Javier y Cristina gracias por su compañerismo y amabilidad.
- A todos mis maestros de la UNAM, especialmente a Carlos Candelaria, Víctor Valdés, Lucía Almeida, Tere Valverde, Rafael Lamothe, Miguel Lara, Jesús Caballero, Lupita Vidal, Sigfrido Sierra, Jaime Jiménez, Ramón Espinaza y Julio Prieto que a su manera y en su estilo son el mejor ejemplo del maestro universitario
- Este proyecto fue financiado por el donativo 3753-N de CONACYT y por una beca del PROBETEL.

PERSONALES

- A mis papás, Alfredo y Gloria, por haberme hecho la vida tan fácil e infinitamente feliz. A mi hermano Augusto por los 20 años de compañía y complicidad.
- A Ana Laura, mi mejor amiga, por su amistad, compañía, cariño y sinceridad.
- A mi familia Ojeda, Daniel, Claudia, Oscar y Regina, por ser eso, mi familia.
- A todos los miembros de las familias Sánchez-Hidalgo y Sepúlveda por el cariño que me han tenido.
- A mis amigos de la fac: Brian, German, Erick, Fernando, Sergio, Rodrigo y Paola por su cariño y apoyo. A Pompa por ser mi amiga a su manera.
- A los miembros del Laboratorio de Biología Molecular de la Facultad de Ciencias: Lichi, Víctor, Claudia, Bety, Alfonso y Epi por todo lo que me enseñaron y por compartir conmigo, de manera tan intensa, los momentos dramáticos del 2000.
- Al Lic. Vicente Rincón por la corrección ortográfica del documento y la amistad y formación que me ha dado desde hace 14 años.
- A los integrantes del EBC por compartir con un biólogo molecular su pasión por el trabajo de campo, sus líneas de investigación y sus conocimientos.
- A mis hermanos: Rodrigo, Pablo y Sebas por su cariño y apoyo desde hace más de 14 años y a mi familia putativa, la familia Coria, por todo lo que hemos vivido.
- A mis hermanos de Campestre: Adriana y Amara, mis confidentes, de las que tanto he aprendido; Chenillo y Melisa, mis compañeras universitarias; Dafne, por la física y las matemáticas; Ytzel, porque ya presta sus apuntes; Guillo, mi primer gran amigo; Adela, no se me olvidan los chocolates de sexto de primaria; Cake, las cascaritas en el club; Pomar y García, mis buenos amigos; Carolina, porque me quiere tanto como yo a ella; y a todos los que no menciono pero que igual tengo en mi corazón.
- A todos los maestros que me enseñaron algo durante mi vida, especialmente a Dolores López, Baldor, Natalia, Lety y Julia García.
- A la Facultad de Ciencias, pilar de la ciencia mexicana y al CIFN, mi casa.
- A la UNAM por todo lo que he vivido, por todo lo que me ha dado, por formarme, por la OFUNAM, por el EBC, por los pumas y por permitirme conocer a toda la gente maravillosa que menciono en esta sección.

**ESTE TRABAJO FUE REALIZADO
EN EL:**

**PROGRAMA DE GENÉTICA MOLECULAR DE
PLÁSMIDOS BACTERIANOS**



**CENTRO DE INVESTIGACIÓN SOBRE FIJACIÓN DE
NITRÓGENO**

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

"... los gérmenes que bullen invisibles en la retorta del mundo orgánico, que en el ciclo de sus transformaciones incesantes hacen de toda existencia un medio en que efectuar sus evoluciones, que se emboscan en nuestra fauna, en nuestra flora, en la atmósfera en que estamos sumergidos, en la corriente de agua que se desliza por el suelo, en la corriente de sangre que circula por nuestras venas, y que conspiran con tanto acierto como si fueran seres conscientes, para descomponer toda vida y extraer de la muerte nuevas formas de vida."

Justo Sierra (1910)

ÍNDICE

ÍNDICE	- 1 -
RESUMEN	- 2 -
ANTECEDENTES	- 3 -
TEORÍA NEUTRAL DE LA EVOLUCIÓN MOLECULAR	- 3 -
REITERACIÓN GÉNICA Y EVOLUCIÓN CONCERTADA	- 4 -
CONVERSIÓN GÉNICA	- 5 -
RECOMBINACIÓN HOMÓLOGA	- 7 -
<i>RHIZOBIUM ETLI</i> Y LA FIJACIÓN DE NITRÓGENO	- 11 -
HIPÓTESIS	- 13 -
ESTRATEGIA EXPERIMENTAL	- 14 -
MATERIAL Y MÉTODOS	- 16 -
TÉCNICAS MICROBIOLÓGICAS BÁSICAS	- 16 -
CEPAS	- 16 -
MEDIOS DE CULTIVO	- 16 -
ALMACENAMIENTO DE LAS CEPAS	- 17 -
ANÁLISIS MOLECULARES	- 17 -
PERFIL DE PLÁSMIDOS	- 17 -
EXTRACCIÓN DE DNA	- 17 -
AMPLIFICACIÓN (PCR) DE LAS REGIONES <i>NIFHDK</i>	- 17 -
PURIFICACIÓN DE LAS REGIONES <i>NIFHDK</i>	- 19 -
AMPLIFICACIÓN (PCR) DEL GEN <i>NIFH</i> DE CADA REGIÓN <i>NIF</i>	- 19 -
SECUENCIACIÓN	- 21 -
ANÁLISIS BIOINFORMÁTICOS	- 21 -
BANCO DE SECUENCIAS	- 21 -
ENSAMBLE DE LAS SECUENCIAS	- 21 -
TRADUCCIÓN DE LAS SECUENCIAS	- 22 -
ALINEAMIENTO DE LAS SECUENCIAS	- 22 -
FILOGRAMAS, CÓMPUTO DE DISTANCIAS Y PRUEBAS DE NEUTRALIDAD	- 22 -
ANÁLISIS DE DOMINIOS	- 23 -
RESULTADOS	- 24 -
PERFIL DE PLÁSMIDOS	- 24 -
AMPLIFICACIÓN DE LAS REGIONES <i>NIFHDK</i>	- 25 -
AMPLIFICACIÓN DE <i>NIFH</i>	- 29 -
SECUENCIACIÓN	- 30 -
DISCUSIÓN	- 40 -
REFERENCIAS	- 46 -

Evolución Concertada entre Miembros de una Familia Multigénica en *Rhizobium etli*.

RESUMEN

La duplicación de genes es un proceso que da origen a familias multigénicas. Estas repeticiones de secuencias, sean genes o elementos transponibles, pueden representar de un 10% a un 75% del genoma de un organismo. En bacterias de la familia *Rhizobiaceae* la reiteración génica es común, un ejemplo interesante de esto es la reiteración de los genes estructurales de la (*nifHDK*) en *Rhizobium etli*.

Una de las características de las familias multigénicas, ampliamente estudiada en eucariontes, es el alto grado de homogeneidad entre las secuencias de sus miembros. Frecuentemente las secuencias de los miembros de una familia multigénica son bastante similares en una especie, pero, al compararlas con las de los genes ortólogos en una especie relacionada encontramos mayor divergencia. Esto sugiere que los miembros de una familia no evolucionan independientemente, sino que sufren evolución concertada.

Uno de los mecanismos que han sido propuestos para explicar este fenómeno es la conversión génica. Esta se define como la transferencia no recíproca de información entre dos secuencias homólogas y es uno de los resultados de la recombinación. Este es un evento estocástico que puede atenuar la divergencia entre reiteraciones, ya que una mutación en una de las copias puede ser eliminada por conversión génica con otra de las reiteraciones, o bien puede actuar a la inversa, expandiendo estas mutaciones al resto de las copias. Para que este fenómeno sea eficiente, su tasa de formación debe ser mayor que la de mutación. En el género *Rhizobium*, se ha sugerido que existe evolución concertada debido a la alta identidad de las reiteraciones del gen *nifH* en *R. etli*.

Para estudiar el papel de la conversión génica en *R. etli* hemos obtenido, por PCR específicos, cada una de las copias de *nifH* de 11 cepas de *R. etli* y las secuenciamos. Las secuencias obtenidas fueron alineadas y comparadas en busca de evolución concertada entre las reiteraciones de cada cepa. A pesar de haber encontrado bajos niveles de divergencia entre las secuencias, nuestros resultados revelan un patrón que sugiere la operación de un fenómeno de evolución concertada.

Evolución Concertada entre Miembros de una Familia Multigénica en *Rhizobium etli*.

RESUMEN

La duplicación de genes es un proceso que da origen a familias multigénicas. Estas repeticiones de secuencias, sean genes o elementos transponibles, pueden representar de un 10% a un 75% del genoma de un organismo. En bacterias de la familia *Rhizobiaceae* la reiteración génica es común, un ejemplo interesante de esto es la reiteración de los genes estructurales de la (*nifHDK*) en *Rhizobium etli*.

Una de las características de las familias multigénicas, ampliamente estudiada en eucariontes, es el alto grado de homogeneidad entre las secuencias de sus miembros. Frecuentemente las secuencias de los miembros de una familia multigénica son bastante similares en una especie, pero, al compararlas con las de los genes ortólogos en una especie relacionada encontramos mayor divergencia. Esto sugiere que los miembros de una familia no evolucionan independientemente, sino que sufren evolución concertada.

Uno de los mecanismos que han sido propuestos para explicar este fenómeno es la conversión génica. Esta se define como la transferencia no recíproca de información entre dos secuencias homólogas y es uno de los resultados de la recombinación. Este es un evento estocástico que puede atenuar la divergencia entre reiteraciones, ya que una mutación en una de las copias puede ser eliminada por conversión génica con otra de las reiteraciones, o bien puede actuar a la inversa, expandiendo estas mutaciones al resto de las copias. Para que este fenómeno sea eficiente, su tasa de formación debe ser mayor que la de mutación. En el género *Rhizobium*, se ha sugerido que existe evolución concertada debido a la alta identidad de las reiteraciones del gen *nifH* en *R. etli*.

Para estudiar el papel de la conversión génica en *R. etli* hemos obtenido, por PCR específicos, cada una de las copias de *nifH* de 11 cepas de *R. etli* y las secuenciamos. Las secuencias obtenidas fueron alineadas y comparadas en busca de evolución concertada entre las reiteraciones de cada cepa. A pesar de haber encontrado bajos niveles de divergencia entre las secuencias, nuestros resultados revelan un patrón que sugiere la operación de un fenómeno de evolución concertada.

ANTECEDENTES

Teoría Neutral de la Evolución Molecular

En 1968 Motoo Kimura postuló la teoría de la Teoría Neutral de la Evolución Molecular, en la que propone que la mayor parte de los cambios genéticos son invisibles a la selección natural, por no presentar un fenotipo. Esta neutralidad selectiva causa que sean la deriva génica y la mutación los que más influyen en la fijación o extinción del alelo originado. Por otro lado supone que los cambios genéticos no sinónimos están sujetos a selección, tanto negativa como positiva.

De esto se desprenden 3 predicciones (Kimura y Ohta, 1974):

- 1.- Las tasas de sustitución sinónima deben ser mayores que las de mutación no sinónima, debido a que estas últimas tienen más probabilidades de ser eliminadas por selección.
- 2.- Las regiones génicas de mayor importancia funcional deben evolucionar más lentamente que regiones menos importantes, pues mutaciones en estas últimas tienen más posibilidades de ser neutrales.
- 3.- Las duplicaciones génicas deberán evolucionar a distintas tasas, pues la reiteración puede liberar a una/unas de las copias de la presión de selección.

Estas predicciones fueron complementadas por Ohta, que propuso que la mayoría de la evolución en proteínas es deletérea, aunque sea muy ligeramente (Teoría de los alelos ligeramente deletéreos). Esto explicaría los estrechos rangos de heterocigosis en organismos con diferentes tiempos de generación y tasas de mutación similares, que habían cuestionado la teoría neutral. Por ejemplo se esperaría mayor heterocigosis en los ratones que en los elefantes, ya que los primeros tienen menor tiempo de generación y por tanto acumularían mayor número de mutaciones neutrales en el mismo tiempo. Sin embargo la heterocigosis es similar debido a que, al ser deletéreas las mutaciones, la posibilidad de su fijación es menor debido a que el tamaño de la población de ratones es mucho mayor que la de elefantes. (Gillespie, 1998)

Estas predicciones han sido comprobadas por numerosos estudios (Doyle y Gaut, 2000) por lo que la teoría es utilizada como hipótesis nula en estudios de evolución molecular.

Reiteración Génica y Evolución Concertada

La reiteración genética es un fenómeno que encontramos tanto en genomas eucariontes como procariontes (Romero et al, 1999; Rocha et al, 1999). Estas repeticiones pueden abarcar desde un 10% hasta un 75% de éste (Brown, 1999) y pueden estar constituidas por microsátélites, secuencias de inserción, transposones y genes. Por ejemplo, la reiteración de genes para RNAr se presenta en la mayoría de los procariontes (Liao, 2000). Además existen reiteraciones de los genes de flagelina (*fla*) en *Campylobacter* (Meinersmann y Hiett, 2000) y de los genes para factores de elongación traduccional (*tuf*) en *Salmonella typhimurium* (Lathe y Bork, 2001). Otro ejemplo lo representa la duplicación del gen del regulador transcripcional *fixK* en *R. etli* y en *B. japonicum*. (Girard et al. 2000)

Al obtenerse la secuencia completa del genoma de *Caenorhabditis elegans* se encontró que el 40% de los genes eran miembros de familias multigénicas (Semple y Wolfe, 1999)

Una de las características que las secuencias repetidas confieren a los genomas es la posibilidad de generación de inversiones, deleciones o duplicaciones debido a la recombinación entre estos. (Romero et al, 1991). Este es un evento común e incluso se han identificado enfermedades humanas relacionadas con los rearrreglos ocasionados por repeticiones, como es el caso de la enfermedad de Charcot-Marie-Tooth ocasionada por una deleción debido a la recombinación entre dos repeticiones directas de 24 kb (Hurles, 2001; Shaffer y Lupski, 2000).

Sin embargo, evolutivamente hablando, una de las características importantes de los mecanismos de reiteración génica es el caso en el que dan lugar a familias multigénicas.

Las duplicaciones de genes son consideradas como un importante mecanismo de amortiguamiento genético que permite la estabilización de un fenotipo esencial, así como la diversificación evolutiva por liberación de presión de selección sobre una(s) de las copia(s) (Rutherford, 2000). En el caso de amortiguamiento, este es ejemplificado por la reiteración génica de la nitrogenasa reductasa en *R. etli* que permite que aunque una copia sea incapacitada el fenotipo fijador de nitrógeno se mantenga en cierta proporción (Romero et al, 1988). El fenómeno de diversificación ha sido comprobado en el caso de las hemoglobinas, ya que la beta, la alfa y la mioglobina provienen de la diversificación de duplicaciones de un gen ancestral (Gogarten y Olendzenski, 1999). La misma situación se considera para los genes Hox, en que la reiteración se debe tanto a duplicaciones individuales del gen como a duplicaciones del genoma. (Meyer y Schartl, 1999).

Una de las características más notables de las familias multigénicas es el alto grado de homogeneidad entre las secuencias de sus miembros. Frecuentemente las secuencias de los miembros de una familia multigénica, diversificados o no, en una especie son muy similares entre sí. Pero, al compararlas con las de los mismos genes en otra especie encontramos una mayor divergencia. Esto sugiere que los miembros de la misma familia no evolucionan independientemente sino que sufren una evolución concertada (Dover, 2000). Se han propuesto diversos mecanismos para explicar este fenómeno aunque todos están ligados a la recombinación (Futuyma, 1992):

En el caso de las familias de pseudogenes y transposones, estos son capaces de formar nuevas copias por retrotranscripción a partir de la copia funcional mientras que las copias generadas previamente están sujetas a delección. Si la tasa de delección es mayor a la tasa de mutación entonces la familia será homogénea en secuencia, ya que la mayoría tendrá un origen reciente a partir de un ancestro común, el gen funcional. Al ser objeto de selección neutral las copias tienden a acumular mutaciones pero esto solo se refleja parcialmente por esta relación entre las tasas de delección y sustitución.

Las familias de genes no móviles pueden ser homogeneizadas por recombinación desigual o por conversión génica. La homogenización por recombinación desigual sucede principalmente entre repeticiones en tandem alterando el número de copias. La homogenización de reiteraciones dispersas en el genoma por entrecruzamientos desiguales está menos favorecida debido al riesgo de delección de genes no relacionados.

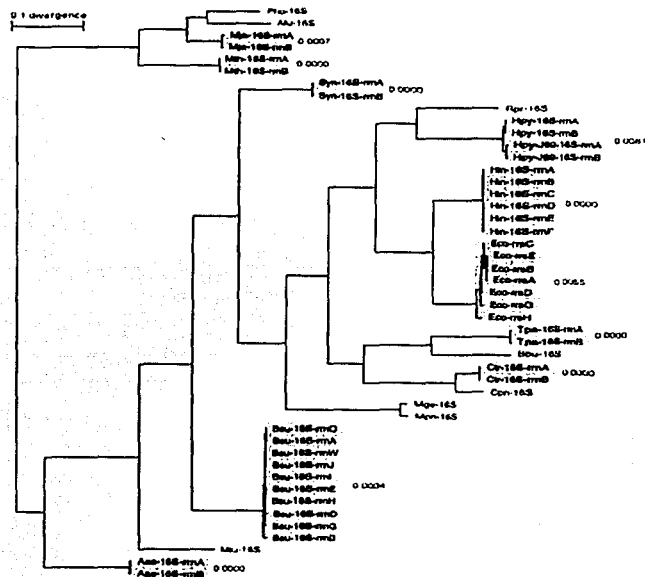
Conversión génica

La conversión génica se define como la transferencia no recíproca de información entre dos secuencias homólogas (Rodríguez y Romero, 1998). Este proceso es uno de los posibles resultados del fenómeno de recombinación genética, actualmente explicado por el modelo de ruptura en doble cadena y reparación. La conversión es flexible con respecto a los substratos participantes, pudiendo ocurrir por apareamiento entre moléculas con una elevada identidad (como cromátidas hermanas o cromosomas homólogos) o por interacción entre reiteraciones localizadas dentro de una misma cromátida o aún entre cromosomas no homólogos (Futuyma, 1992). Este proceso se ha reportado en eucariontes (Drouin et al, 1999).

Recientemente en procariontes se ha reportado conversión génica entre reiteraciones de los genes *fla* de *Campylobacter* (Jordan et al, 2001). Así como en los *tuf* de *S. typhimurium* (Abdulkarim y Hughes, 1996; Meinersmann y Hiatt, 2000), entre los genes de porinas, *hop*, en *H. pylori* (Jordan et al, 2001), y entre los genes 16s ribosomales de varias arqueobacterias y bacterias (Liao, 2000) (Fig. 1).

La conversión génica es un proceso azaroso que puede llevar a retrasar la aparición de divergencia entre reiteraciones, ya que si se da una mutación en una de las copias esta puede ser eliminada al recombinar con otra de las reiteraciones. Para que este proceso sea eficiente, su tasa de aparición debe ser superior a la de mutación. El tracto de conversión tiene un promedio de 500 pb y se ha detectado que existen sesgos en la homogenización de las secuencias pues casi siempre ocurre en la misma dirección, es decir un gen es el que predominantemente convierte a los otros miembros de la familia (Liao, 2000)

Fig. 1- Evolución Concertada Entre las Copias de los Genes de RNAr 16s en Procariontes



La secuencia de las copias de los organismos sombreados mantienen una alta identidad por evolución concertada. Aae: *A. aeolicus*, Afu: *A. fulgidus*, Bsu: *B. subtilis*, Bbu: *B. burgdorferi*, Ctr: *C. trachomatis*, Cpn: *C. pneumoniae*, Eco: *E. coli*, Hin: *H. influenzae*, Hpy: *H. pylori* 26695, Hpy-J99: *H. pylori* J99, Mth: *M. thermoautotrophicum*, Mja: *M. jannaschii*, Mtu: *M. tuberculosis*, Mge: *M. genitalium*, Mpn: *M. pneumoniae*, Pho: *P. horikoshii*, Rpr: *R. prowazekii*, Syn: *Synechocystis* PCC6803, Tpa: *T. pallidum*. (Tomado de Liao, 2000)

Recombinación Homóloga

La recombinación homóloga es un proceso fundamental que rearregla genes dentro y entre replicones, y en eucariontes guía la segregación de cromosomas (Sharples et al, 1999) constituyéndose como una importante fuerza evolutiva que sirve tanto para generar diversidad genética como para conservar identidad. Además, juega un importante papel en la supervivencia celular ya que es parte del mecanismo que repara daños en el DNA, generados durante la replicación (rompimiento de la cadena líder, deslizamientos de la DNA polimerasa, etc..) o por agentes externos como las radiaciones y los rayos ultravioleta. (Michel, 2000)

Fue en 1953, cuando Watson y Crick descifraron la estructura de doble hélice del DNA, haciendo posible desarrollar los primeros modelos para el intercambio genético. En 1964 Robin Holliday, por estudios en hongos, emitió la que se considera la propuesta germinal en cuanto al mecanismo de la recombinación. Esta establece que el primer evento en la recombinación meiótica es la formación de cortes en una de las dos cadenas en cada uno de los duplex participantes, seguido del intercambio de éstas, formando una estructura donde se fusionaban y que según se resolviera podría darse intercambio de los marcadores laterales o no. La característica principal de este modelo es su simetría absoluta, DNA heteroduplex en una molécula es invariablemente acompañada por DNA heteroduplex en la otra (Fig. 2). En este caso la conversión génica se daba por reparación enzimática de los apareamientos erróneos causados por polimorfismos entre los heteroduplex en las dos moléculas. Sin embargo estudios en levaduras, principalmente *Saccharomyces cerevisiae*, mostraron que era difícil encontrar indicios de simetría en las tétradas. (Alani et al, 1994).

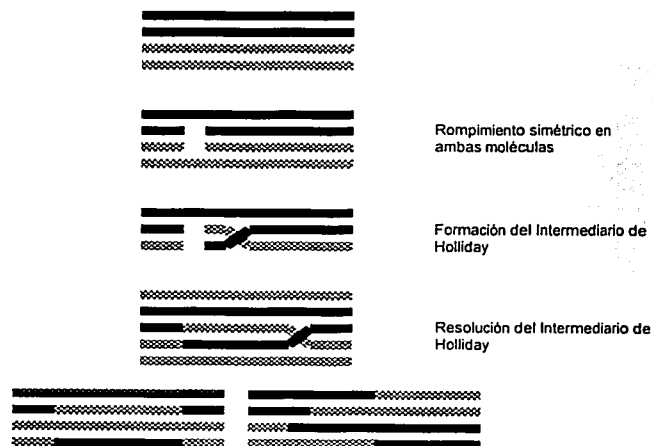
Posteriormente, en 1975, Matt Meselson y Charles Radding propusieron un modelo que inicia con un corte en cadena sencilla, en solamente uno de los duplex participantes. A partir de este ocurre una síntesis de DNA que desplaza a la cadena original. Esta cadena invade entonces una molécula homóloga desplazando la cadena de la misma polaridad que posteriormente es degradada por la acción de una nucleasa formándose una estructura de fusión. Esta estructura es como la propuesta en el modelo de Holliday, por lo que se le conoce como intermediario de Holliday, y aunque es genéticamente asimétrica es estructuralmente simétrica. Posteriormente, se lleva a cabo un desplazamiento de esta estructura del punto de inicio de la recombinación y se resuelve. Nuevamente existe un fenómeno de conversión génica al momento en que se resuelven enzimáticamente los apareamientos erróneos por polimorfismos en el DNA heteroduplex (Fig.3).

Este modelo mostró incongruencia con estudios en *Schizosaccharomyces pombe* y *Neurospora*, ya que en éste se postulaba la transferencia de información únicamente de la cadena invasora y al estudiar sitios promotores de la recombinación en estos organismos se encontró que estos "absorbían" información de la cadena desplazada. (Stahl, 1994)

En 1981 Orr-Weaver et al demostraron que el rompimiento en doble cadena de un fragmento de DNA de levadura contenido en un plásmido estimulaba un evento recombinatorio que incorporaba el plásmido al cromosoma. (Stahl, 1994). Esto comprobaba experimentos de Resnick y Martin (1976) que con ensayos de rayos X habían encontrado que la presencia de rompimientos en doble cadena estimulaban recombinación en levaduras. Así en 1983 Szostak et al publicaron el modelo de recombinación por rompimiento en doble cadena y reparación. Este se ha convertido en el modelo clásico de recombinación y se han identificado en procariontes al menos 25 proteínas involucradas en el proceso (Kowalczykowski, 2000). El proceso comienza con la generación de un corte en doble cadena, este es reconocido por RecBCD, una helicasa con actividad de nucleasa, que degrada y desenrolla ambas cadenas. Este proceso continúa, hasta que la proteína entra en contacto con una secuencia chi, 5'-GCTGGTGG-3' (Eggleston y West, 1997) lo que modifica la subunidad RecD (Dixon y Kowalczykowski, 1994) que resulta en la represión de la actividad de nucleasa en dirección 3' → 5' y la potenciación de la misma actividad en dirección 5' → 3' (Anderson y Kowalczykowski, 1997). A la cadena sencilla en dirección 5' → 3' que se genera a partir de este proceso, y que mantiene el sitio chi intacto, se une entonces la proteína RecA. Catalizada su unión por RecBCD (Anderson y Kowalczykowski, 1997b), RecA lleva a cabo la búsqueda de secuencia homóloga y promueve su invasión por la cadena a la que está adherida, con el consecuente desplazamiento de la homóloga de ésta, y su apareamiento. Esto genera DNA de doble cadena, que es utilizado por una DNA polimerasa para sintetizar la información que había sido degradada, utilizando como templado la información de la molécula invadida, lo que lleva a la formación de dos intermediarios de Holliday. La estructura del intermediario de Holliday es reconocida por la proteína RuvAB que funciona como un motor que extiende el área de DNA involucrada en el evento recombinatorio hasta que la proteína RuvC reconoce una secuencia consenso 5'-(A/T)TT(G/C)-3' (Sharples et al, 1999) y corta, de manera simétrica, dos de las cuatro cadenas resolviendo el intermediario (Fig. 4). En este caso, al haber síntesis de DNA, se produce el fenómeno de conversión génica al resolverse los apareamientos erróneos a favor de la cadena donadora. Recientemente se ha sugerido que en este modelo el fenómeno de conversión génica puede ocurrir cuando, tras la invasión y síntesis de DNA,

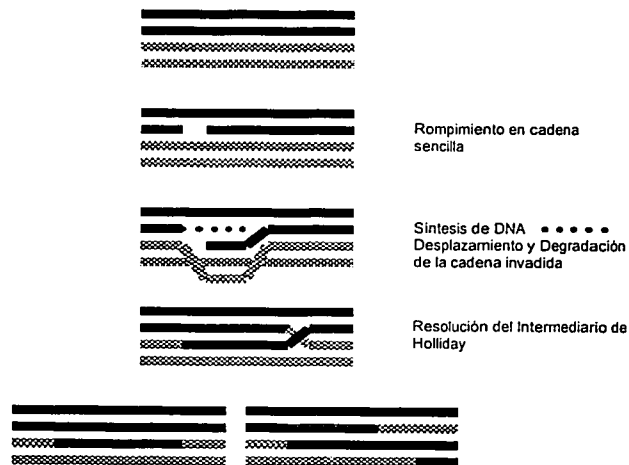
el proceso se interrumpe y la molécula que sufrió los cortes en doble cadena se repara a partir de la información transferida a la cadena invasora por la molécula invadida. (Hunter y Kleckner, 2001).

Fig. 2. Modelo de recombinación de Holliday



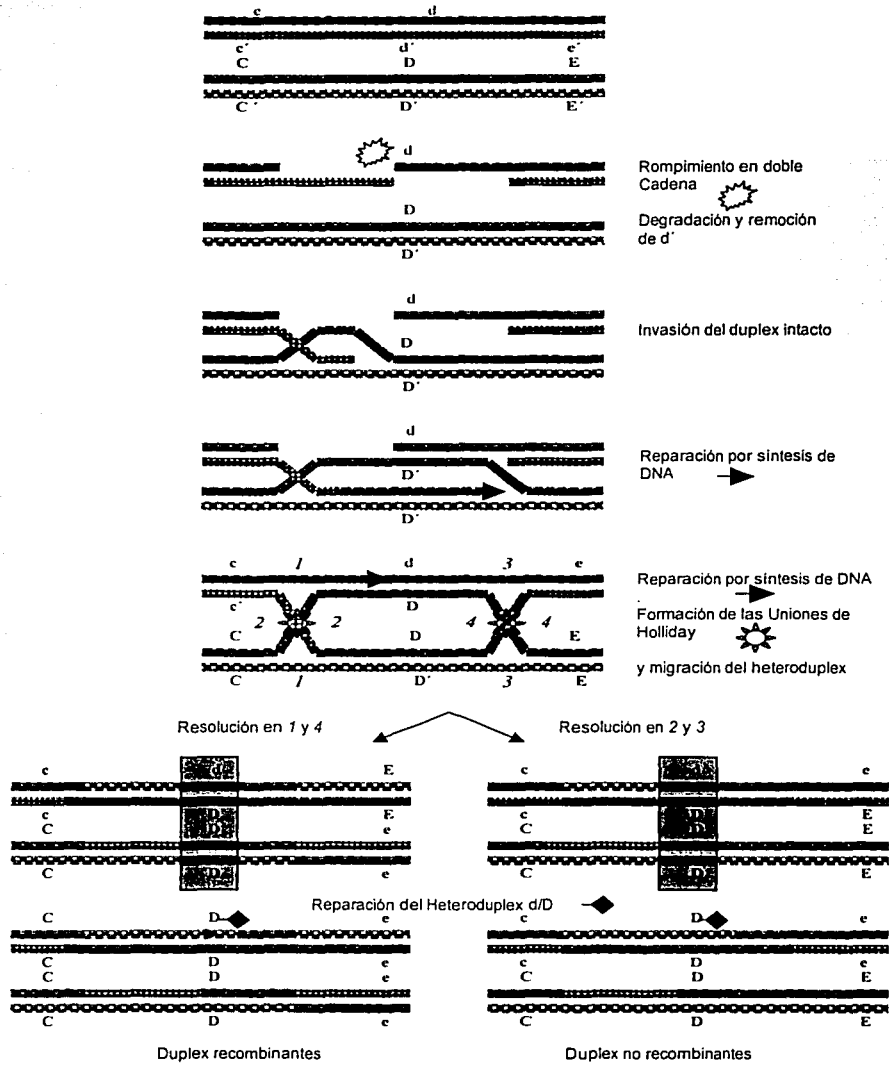
(Modificado de Stahl, 1994)

Fig. 3. Modelo de recombinación de Meselson y Radding



(Modificado de Stahl, 1994)

Fig. 4. Modelo de recombinación por corte en doble cadena y reparación



(Modificado de Stahl, 1994)

***Rhizobium etli* y la fijación de Nitrógeno**

Las rhizobáceas fueron descritas por primera vez por Beijerinck en 1888 a partir de un cultivo de un nódulo de raíz de leguminosa y para 1929, basados en su habilidad y especificidad para nodular se habían descrito seis especies. Posteriormente se les dividió en dos géneros por su velocidad de crecimiento, *Rhizobium* para los de crecimiento rápido y *Bradyrhizobium* para los de crecimiento lento (Wang et al, 2001).

Actualmente la familia *Rhizobaceae* esta integrada por bacterias aeróbicas gram negativas que por secuencia de RNAr 16S se consideran en el grupo de las alfa proteobacterias. Se caracterizan por su habilidad para entablar asociaciones simbióticas con plantas leguminosas en cuyas raíces forman nódulos. En estos, las células bacterianas se diferencian a una nueva forma, conocida como bacteroide, y que es capaz de fijar nitrógeno molecular. Además son capaces de vivir como saprófitos en suelos y se les ha encontrado como endófitos en distintas plantas como el caso de *R. etli* y el maíz. (Gutiérrez-Zamora y Martínez-Romero, 2001). Se han establecido seis géneros dentro de esta familia (*Allorhizobium*, *Azorhizobium*, *Bradyrhizobium*, *Mesorhizobium*, *Sinorhizobium* y *Rhizobium*, Wang et al, 2001)

El género *Rhizobium* está compuesto por bacilos que miden 0.5-1.0 x 1.2-3.0 mm. Se mueven por medio de 6 flagelos que pueden ser peritricos o subpolares. Forman colonias, de color blanco o beige, circulares, convexas, semitranslúcidas u opacas y mucilaginosas. Las cepas del género son del tipo de rápido crecimiento y quimio-organotróficas utilizando gran variedad de ácidos orgánicos y carbohidratos como D-Ribosa, D-mannosa, D-galactosa, fructosa, sucrosa, trehalosa, fumarato, citrato y piruvato, entre otros (Elkan y Bunn, 2002). Algunas cepas requieren de biotina, ácido nicotínico, pantotenato o tiamina como factores de crecimiento.

Hay nueve especies definidas: *R. leguminosarum*, *R. etli*, *R. galegae*, *R. gallicum*, *R. giardinii*, *R. hainanense*, *R. huautlense*, *R. mongolense* y *R. tropici* que nodulan diferentes tipos de leguminosas en zonas tropicales y templadas. (Wang et al, 2001)

La especie *R. etli* está descrita como bastones aeróbicos gram negativos no esporulantes de hasta 3 micrómetros con capacidad de nodular frijol (*Phaseolus vulgaris*). Son incapaces de crecer en medio Luria y en medio PY sin calcio o con carbenicilina, espectinomina, cloranfenicol o rifampin. Son capaces de crecer en medio mínimo con malato como fuente de carbono pero no con arginina, hipoxantina o sorbitol. Su temperatura máxima de crecimiento es 35°C, todas las cepas son resistentes al ácido nalidixico y son agrupadas también por resultados de hibridación DNA-DNA,

ensayos de multilocus y la secuencia de RNAr 16S (Segovia et al, 1993). También es característico de la especie la reiteración de los genes estructurales de la nitrogenasa, la separación de los genes de nodulación en dos unidades estructurales (nodA y nodBC), la presencia del gen de inhibición de polisacáridos, psi (Segovia et al, 1993) y la capacidad de producir melanina. (Wang et al, 2001)

En *Rhizobium etli* la fijación de nitrógeno ocurre por la inducción del complejo enzimático nitrogenasa. Esta enzima lleva a cabo la siguiente reacción:

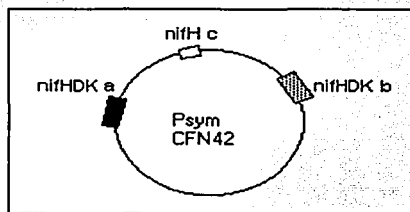


El amonio generado es exportado hacia y asimilado por la planta durante la simbiosis y a cambio ésta provee a las bacterias de fuentes de carbono.

La nitrogenasa está formada por dos proteínas: La proteína Fe- Mo, con un peso molecular aproximado de 240 000 daltons, es un tetrámero $\alpha_2\beta_2$ de dos subunidades codificadas por los genes nifD y nifK. Así como por la proteína Fe, un dímero de dos subunidades idénticas codificadas por el gen *nifH* y que tiene un peso molecular aproximado de 60 000 daltons. (Rees et al, 1993)

En la mayoría de las especies de *Rhizobium* estos 3 genes se encuentran estructurados como un operón único, sin embargo en *Rhizobium etli* CFN42 se encuentran dos copias de éste, además de una región conteniendo solamente el gen de *nifH* (Quinto et al, 1985), todas en el plásmido simbiótico, como puede apreciarse en la figura 5. En *Azorhizobium caulinodans* ORS571 y en *Rhizobium* NGR234 el operón está también duplicado (Badenoch-Jones et al, 1989; Norel y Elmerich, 1987). En el caso de *R. etli* los dos operones son necesarios para una efectividad simbiótica completa, indicando un efecto de dosis génica para estas reiteraciones (Romero et al., 1988). Hallazgos similares han sido reportados en *A. caulinodans* (Norel y Elmerich, 1987).

Fig. 5. Triplicación de *nifH* en *R. etli*

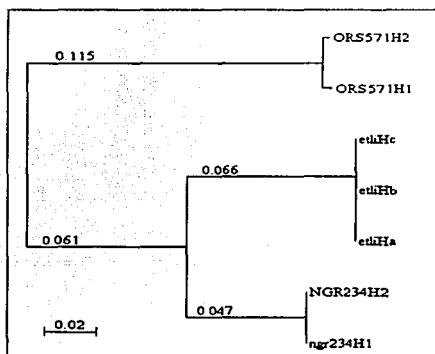


Las reiteraciones a y b se encuentran en un operón que codifica para la nitrogenasa completa; la copia c solo codifica para la subunidad H.

Hipótesis

En el caso del género *Rhizobium* se ha señalado previamente la posibilidad de que ocurriera evolución concertada debido a la alta identidad (100%, Quinto et al, 1985) entre las reiteraciones *nifH* de *R. etli*. Esta alta similitud también se observa en el caso de las reiteraciones de *nifH* en *A. caulinodans* ORS571 (99%, Norel y Elmerich, 1987) y de *Rhizobium* NGR234 (100%, Badenoch-Jones et al, 1989). Esta suposición es reforzada porque al comparar las secuencias de las reiteraciones entre estos organismos su similitud es menor (Fig. 6). Además la conversión génica ha sido demostrada experimentalmente entre los genes *nifH* de *R. etli* con una construcción que demostraba este fenómeno cuando una inserción en *nifH*, que generaba un codón de término, era removida por conversión génica con una copia sin esta inserción, generando un fenotipo de resistencia a Kanamicina (Figura 7); (Rodríguez y Romero, 1998).

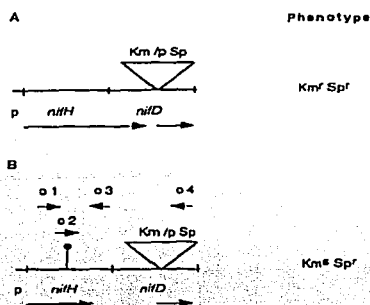
Fig. 6. Filograma de las reiteraciones *nifH* de diferentes rizobáceas



Las secuencias de *nifH* fueron obtenidas de GeneBank; *R. etli* CFN42 *nifHa* (M10587.1) *nifHb* (M15941.1) y *nifHc* (M15942.1); *A. caulinodans* ORS571 *nifH1* (M16709.1) y *nifH2* (M16710.1); y *Rhizobium* NGR234 *nifH1* (AE000101.1) y *nifH2* (AE000105.1). El filograma fue generado con MEGA2 por Neighbor-Joining, a partir de un alineamiento de las secuencias nucleotídicas generado por CLUSTALX. La barra de escala indica 0.02 sustituciones nucleotídicas por sitio. (E. Sepúlveda, datos no publicados)

No obstante que las observaciones anteriores sugieren la operación de evolución concertada, una hipótesis alternativa podría ser que la duplicación de las secuencias haya ocurrido en un lapso evolutivamente corto posterior a la especiación y por lo tanto que su similitud intraespecie se deba a una falta de tiempo para divergir y no por el fenómeno de evolución concertada por conversión génica.

Fig. 7. Construcción para demostrar la existencia de conversión génica en *R. etli*



Cuando B era convertido a A la cepa recuperaba en fenotipo Kanamicina resistente (Tomado de Rodríguez y Romero, 1998)

Para discernir entre estas alternativas nos propusimos hacer la comparación de cada miembro de la familia *nifH* entre cepas de *R. etli*. De esta forma hacíamos más probable que los miembros de la familia sean producto de un mismo evento de duplicación.

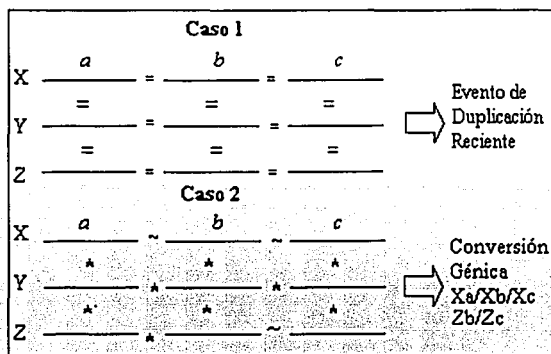
Al comparar las secuencias de las reiteraciones de las cepas de *R. etli* podríamos encontrar distintos escenarios (Fig. 8). Uno es que las secuencias fuesen iguales entre repeticiones y entre cepas, en cuyo caso la explicación más factible sería la de la duplicación reciente. Otro es que la identidad entre los miembros de una cepa fuera mayor que con los correspondientes en las otras cepas, en cuyo caso la hipótesis de conversión génica sería reforzada.

Estrategia experimental

Planeabamos emplear, en la medida de lo posible, genes parálogos. Dos genes ortólogos son aquellos que se encuentran en diferentes organismos que compartieron un ancestro común y dos genes parálogos son aquellos que se encuentran en el mismo organismo. El uso de genes parálogos de diferentes cepas de una misma especie ya ha sido reportado (Meinersmann y Hiatt, 2000) y se justifica por el hecho de que estos proceden del mismo evento de duplicación que dio origen a la familia en la especie.

Para obtener los genes parálogos necesarios se obtuvieron las secuencias de las tres copias de *nifH* presentes en 11 cepas de *R. etli*. Para esto se obtuvo el DNA genómico total de cada cepa y se

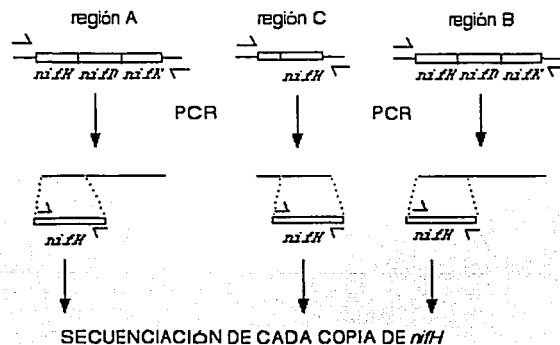
Fig. 8. Posibles escenarios tras la comparación de secuencias



Identidad entre las secuencias de las repeticiones a, b, c de un gen entre las cepas X, Y, Z y su posible explicación.
 = Alta Identidad; ~ Mediana Identidad; * Baja Identidad

amplificó cada región *nif* mediante PCR utilizando oligonucleótidos específicos, correspondientes a los extremos de cada una de las regiones *nif* de la cepa CFN42. Estos oligonucleótidos fueron diseñados por el Dr. Rafael Palacios y han sido eficientes para amplificar cada región *nif* en diferentes cepas de *R. etli* (Flores et al, 2000). El DNA así amplificado se secuenció utilizando oligonucleótidos complementarios al gen *nifH* de *R. etli* CFN42 (Fig. 9). Las secuencias codificantes obtenidas se analizaron computacionalmente. Se ha reportado una heterocigosis del 0.67 en *R. etli*, para genes que codifican, tan solo en cepas correspondientes a territorio mexicano (Piñero et al, 1988) por lo que se esperaba encontrar un cierto grado de variabilidad en las secuencias que permitieran realizar el análisis

Figura 9. Amplificación específica y secuenciación de cada región *nifH* en diferentes cepas.



MATERIAL Y MÉTODOS

TÉCNICAS MICROBIOLÓGICAS BÁSICAS

Cepas

Las cepas de *Rhizobium etli* fueron obtenidas del cepario del Programa de Genética Molecular de Plásmidos Bacterianos del CIFN y se resumen en la Tabla 1.

Tabla 1. Cepas de *Rhizobium etli* Utilizadas en el Estudio

Cepa	Zona Geográfica	Planta Hospedera
CFN 42 ¹	Guanajuato, México	<i>P. vulgaris</i>
BRASIL 5 ¹	Brasil	<i>P. vulgaris</i>
TAL 182 ¹	Hawaii	<i>P. vulgaris</i>
CIAT 894 ¹	Colombia	<i>P. vulgaris</i>
CIAT 895 ¹	Colombia	<i>P. vulgaris</i>
VIKING 1 ¹	Belice	<i>P. vulgaris</i>
KIM 5 ²	Idaho, EUA	<i>P. vulgaris</i>
COC 8 ¹	Morelos, México	<i>P. coccineus</i>
CFN 1 ¹	Guanajuato, México	<i>P. vulgaris</i>
CFN 285 ¹	Morelos, México	<i>P. vulgaris</i>
NITRAGIN ¹	EUA	<i>P. vulgaris</i>
IE4810 ³	México	<i>P. vulgaris</i>
IE4804 ³	México	<i>P. vulgaris</i>
IE1004 ³	México	<i>P. vulgaris</i>

Referencias: 1.- (Piñero et al., 1988); 2.- (Josephson y Pepper, 1984); 3.- (Silva et al, 2002)

Medios de Cultivo

El medio de cultivo usado para crecimiento de *R. etli* fue PY, el cual contiene peptona (0.5% p/v) y extracto de levadura (0.3% p/v). Posteriormente a la esterilización se añadió CaCl_2 a una concentración final de 10mM. Para la preparación de medio sólido se empleó agar bacteriológico a una concentración del 1.5%. Cuando fue necesario, se añadió a los medios ácido nalidixico a una concentración final 20 ug/ml.

Almacenamiento de las Cepas

Las cepas fueron sembradas en PY-Agar Nal 20, se incubaron a 30°C durante tres días y se almacenaron a 4°C, resemebrándose cada mes.

Para la conservación a largo plazo se incubó una azada de cada cepa en 5 ml de PY líquido durante 24 hrs. Posteriormente se transfirieron 0.5 ml del cultivo a 0.5 ml de glicerol previamente esterilizado y se almacenaron a -70°C

ANÁLISIS MOLECULARES

Perfil de Plásmidos

Se incubó cada cepa en 5 ml de PY líquido durante 24 horas y se trató según Eckhardt (Eckhardt, 1978). Se realizó una electroforesis vertical en un gel de agarosa en Tris-Boratos al 0.7% durante 1 hora, a 8 mA corriente constante y posteriormente 4 horas a 40mA, voltaje constante.

Extracción de DNA

Se cada cepa en 3 ml de PY líquido durante 24 hrs., se centrifugó y la pastilla celular fue lavada en TE 50/20. Se lisaron las células utilizando lisozima, pronasa y SDS y se precipitaron las proteínas con Acetato de Amonio saturado. Se centrifugó nuevamente y se precipitó el DNA del sobrenadante con Isopropanol. Se limpió con etanol al 70% y después de evaporar se resuspendió en Te 1/10 con RNAsa. Para verificar la calidad del DNA extraído, se hizo una electroforesis de una muestra de 3 ul del extracto, 2 ul de agua y 1 ul de azul de bromofenol en un gel de agarosa en Tris-Acetatos al 1% durante una hora a 75 volts.

Amplificación (PCR) de las Regiones *nifHDK*

Se utilizaron los pares de oligonucleótidos E201-E202, E215-E216 y E237-E238, específicos para cada región *nifHDK* (Flores, comunicación personal) ,B2U, B2L, U198, L198, U171 y L171. Las reacciones se hicieron con el kit GeneAmp XL (Applied Biosystems) según las instrucciones del fabricante para reacciones de 25 ul y con el programa de temperaturas propuesto para los oligonucleótidos E (Flores et al, 2000), empleando un aparato Mastercycler 5330 (Eppendorf):

94°C 1 minuto

37 ciclos de :

94°C 15 segundos

68°C 1 minuto

72°C 7 minutos.

Para amplificar con los oligonucleótidos B2L, U198, L198, U171 y L171 se diseñó el siguiente programa de temperaturas:

94°C 1 minuto

20 ciclos de:

94°C 30 segundos

60°C 1 minuto

68°C 5 minutos

17 ciclos de:

94°C 30 segundos

60°C 1 minutos

72°C 5 minutos

72°C 10 minutos

Para verificar la obtención del producto deseado se hizo una electroforesis de una muestra de 3 ul de éste, 2 ul de agua y 1 ul de azul de bromofenol en un gel de agarosa en Tris-Acetatos al 1% durante una hora a 95 volts.

Tabla 2. Combinaciones de Oligonucleótidos Usadas para Amplificar las Regiones *nif*

Upper	Lower	Región	Tamaño Esperado
E201	E202	nifHDKa	6618 pb
E215	E216	nifHDc	4668 pb
E237	E238	nifHDKb	7240 pb
B2U	E202	nifHDKa	5360 pb
B2U	E238	nifHDKb	5655 pb
B2U	B2L	nifHDKb	7191 pb
E237	B2L	nifHDKb	8179 pb
U171	L171	nifHDc	7844 pb
U198	L198	nifHDc	7353 pb

Tabla 3. Oligonucleótidos Usados para Amplificar las Regiones *nif*

Oligo	Complementación	Secuencia
E201 ¹	Upper	5' CGGACGCCTTCCATTCCAAAGTAATGTTCA 3'
E202 ¹	Lower	5' CGTCTCTCCTGCAATGGTCTTGATTCTAGC 3'
E215 ¹	Upper	5' GGCAAGGACAAATAATTGGGGTCCGAACAC 3'
E216 ¹	Lower	5' GACAAAGCCGAACCTAGGAGATTGCTGTTC 3'
E237 ¹	Upper	5' GTCTCAGAGCCCGATTCAAAGTTCATTGC 3'
E238 ¹	Lower	5' CGTTCCTGGTTGATATCGAGCCAAGGTGTC 3'
B2U ²	Upper	5' TGTGGTGGTTCGGAGTAACTTTCTG 3'
B2L ²	Lower	5' GGGAGGAATCAAGCGATCCAGGACAC 3'
U198 ²	Upper	5' GAGTAATCCGCACAACAACAGTCT 3'
L198 ²	Lower	5' CTTGACAAAGCCGAACCTAGGAGAT 3'
U171 ²	Upper	5' GCGTGCGTTTTGCTAAGAT 3'
L171 ²	Lower	5' TTCTGGCGTATTTGAATGTTACTGA 3'

Referencias: 1.- (Flores et al., 2000); 2.- (Este trabajo)

Purificación de las Regiones *nifHDK*

El resto del producto de cada una de las regiones amplificadas se separó por electroforesis en un gel de agarosa al 1% en Tris-acetatos a 95 volts por una hora y la banda correspondiente fue eluida por medio del kit GeneClean III de BIO101. Para retirar cualquier traza de la matriz utilizada en la elución (celite), que pudiera interferir con otros experimentos, se trató cada muestra con las columnas de limpieza CENTRISEP (Princeton Separations) y se verificó el producto en un gel de agarosa al 1% en Tris-acetatos a 95 volts por una hora.

Amplificación (PCR) del gen *nifH* de cada región *nif*

Se utilizaron los oligonucleótidos HindU, BamU, BclI y XbaI (Santoyo, 2000) diluidos en TEN. Las reacciones de PCR se establecieron empleando el kit Taq DNA Polymerases (Life Technologies), utilizando las siguientes cantidades para una reacción de 50ul:

H₂O.-25.5 ul

Buffer 10X. - 5 ul

2 mM MgCl₂. - 2 ul

dNTP'S (dATP, dGTP, dTTP, dCTP, 2.5mM c/u).- 8 ul

DMSO.- 3 ul

Oligo 1 (5uM).- 2.5 ul

Oligo 2 (5uM).- 2.5 ul

DNA.- 1 ul

Taq pol.- 0.5 ul

Las reacciones se realizaron con el siguiente patrón de temperaturas:

30 ciclos:

96°C 1 minuto

56°C 1 minuto

72°C 1 minuto

72°C 10 minutos

Para verificar los productos se hizo una electroforesis de una muestra de 3ul de cada uno, 2 ul de agua y 1 ul de azul de bromofenol en un gel de agarosa en Tris-Acetatos al 1% durante una hora a 75 volts.

Tabla 4. Oligonucleótidos Usados para Amplificar el gen *nifH*

Oligo	Complementación	Secuencia
HindU ¹	Upper <i>nifH</i> base 12	5' AGGAAGCTTATATGTCAGATTTGC 3'
BamU ¹	Upper <i>nifH</i> base 135	5' GGATCCGACCCGAAAGCC 3'
BclL ¹	Lower <i>nifH</i> base 831	5' CTTGATCATGCCGAAGTCGAG 3'
XbaL ¹	Lower <i>nifH</i> base 911	5' TCCTCTAGACAGCGGCAGTTAT 3'

Referencias: 1.- (Santoyo, 2000)

Tabla 5. Combinaciones de Oligonucleótidos Usadas para Amplificar *nifH*

Upper	Lower	Región	Tamaño
HindU	BclL	<i>nifH</i>	819 pb
HindU	XbaL	<i>nifH</i>	900 pb
BamU	BclL	<i>nifH</i>	696 pb
BamU	XbaL	<i>nifH</i>	776 pb

Secuenciación

Los productos de PCR de *nifH* de cada cepa fueron tratados nuevamente con columnas CENTRISEP para eliminar impurezas. Se utilizaron los oligonucleótidos KasL, BamU, BclI, ApaU (Santoyo, 2000), HI388 y Hu765 (este trabajo), en una mezcla que incluía 100 ng de DNA y 5 pmol de Oligo llevados a 12 ul en agua, por reacción. Las reacciones fueron secuenciadas en el Programa De Evolución Molecular del CIFN utilizando el kit DYEnamic ET Dye Terminator (Amersham Biosciences), donde se completaron y analizaron en un secuenciador MEGABACE 1000 de Amersham Biosciences.

Tabla 6. Oligonucleótidos Usados para Secuenciar el Gen *nifH*

Oligo	Complementación	Secuencia
ApalU ¹	Upper	5' GTGCACATGACGATGTCCACT 3'
Hu765 ¹	Lower	5' GACCATTCCGACCCCGATTAC 3'
KasL ¹	Upper	5' GCGCCCGTTACAGATCAG 3'
HL388 ¹	Lower	5' AAAGCCACCGCACACCACATC 3'

ANÁLISIS BIOINFORMÁTICOS

Banco de Secuencias

Las secuencias que se utilizaron en este trabajo y que ya se encontraban depositadas en el Genebank son las siguientes, (el número de acceso está entre paréntesis):

R. etli CFN42 *nifHa* (M10587.1) *nifHb* (M15941.1) y *nifHc* (M15942.1); *A. caulinodans* ORS571 *nifH1* (M16709.1) y *nifH2* (M16710.1); y *Rhizobium* NGR234 *nifH1* (AE000101.1) y *nifH2* (AE000105.1). *R. gallicum* bv. *Gallicum* FL27 *nifH* (M55226)

Ensamble de las Secuencias

Las secuencias fueron anotadas utilizando el programa Phred (Ewing et al, 1998) y depuradas manualmente, editándolos con CONSED (Gordon et al, 1998). Cada secuencia fue depurada independientemente, sin tener acceso a las otras secuencias de manera que no se afectara la objetividad de la depuración. Los archivos de secuencia depurada fueron convertidos en archivos de secuencia de LaserGene (agrupados por región y por cepa) y ensamblados utilizando el

LaserGene Genomic Project Manager con las variables preestablecidas en éste. En los sitios traslapantes de las secuencias en que se presentaban polimorfismos se resolvió por calidad de la anotación y por la base anotada en la mayoría de las secuencias. En los polimorfismos en que no se podía aplicar objetivamente este criterio se volvió a secuenciar la región hasta que se pudiera discernir a que bases correspondían. Los genes ensamblados fueron guardados como archivos de secuencia de LaserGene y como archivos de texto en formato FASTA.

Traducción de las Secuencias

Las secuencias nucleotídicas fueron traducidas en los 3 marcos de lectura en la cadena codificante, con BioEdit (Hall, 1999), usando la tabla de codones estándar. Se consideró al ORF mas largo como el codificante y se guardaron como archivos de texto en formato FASTA.

Alineamiento de las Secuencias

Las secuencias, tanto nucleotídicas como de aminoácidos, fueron alineadas con CLUSTALX (Thompson et al, 1997) utilizando los parámetros preestablecidos. Todos los polimorfismos detectados en los alineamientos fueron revisados nuevamente con CONSED, y en caso de persistir la duda se resecuenció la región.

Filogramas, Cómputo de Distancias y Pruebas de Neutralidad

Los filogramas fueron generados en MEGA 2.1 (Kumar et al, 2001) por el método de Neighbor-Joining usando el modelo de dos parámetros de Kimura y aplicando un bootstrap de 1000 repeticiones sembrado a partir del 24054.

Las distancias entre grupos y dentro de los grupos fueron calculadas utilizando MEGA2 por número total de diferencias y por el modelo de dos parámetros de Kimura.

Los números de sustituciones sinónimas y no sinónimas fueron calculados con MEGA2 usando el método modificado de Nei-Gojobori y agrupando las secuencias por locus (nifHa, nifHc y nifHb), la tasa dN/dS fue calculada manualmente a partir de estos datos.

Análisis de dominios

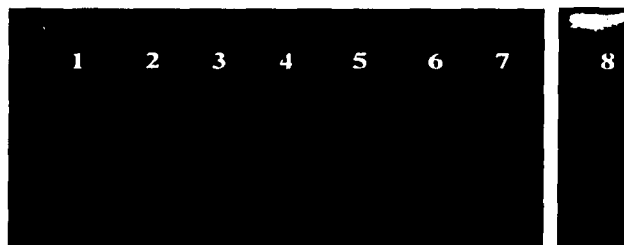
Los sitios polimórficos que permanecieron tras la segunda vuelta de depuración fueron analizados con PsiPred, utilizando las variables preestablecidas, (Jones, 1999) para revisar si afectaban el dominio de estructura secundaria del que forman parte. Además se hizo un BLAST, también con las variables preestablecidas, a partir del cual se hizo una búsqueda de la base de datos de dominios conservados (Marchler et al, 2002) para revisar la incidencia de estas en genes parálogos.

RESULTADOS

Perfil de Plásmidos

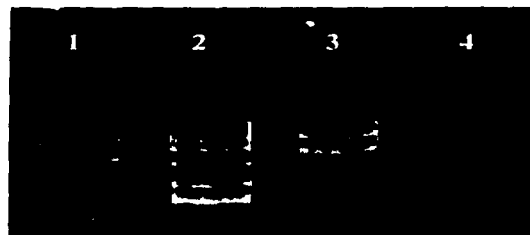
El perfil de plásmidos es un método utilizado ampliamente en la subdivisión de especies (Vandamme et al, 1996) y fue realizado para asegurar que las cepas utilizadas eran las que se querían utilizar y que no hubiesen sufrido una alteración que podría alterar el modelo. El perfil de plásmidos coincidió con lo reportado previamente en las cepas CFN 42, BRASIL 5, TAL 182, CIAT 894, CIAT 895, VIKING 1, COC 8, CFN 1, NITRAGIN 8251 (Rodríguez, 1992), KIM 5 (Vinuesa et al, 1999) IE4810, IE4804 y IE1004 (Silva et al, 2002). En el caso de la cepa CFN 285 se encontraron alteraciones en el patrón pues se había reportado que los plásmidos b y c migraban juntos mientras que en esta ocasión uno de ellos migró a la altura del plásmido d. Sin embargo esta cepa es conocida por el alto índice de rearreglos que presenta y se le utilizó con reserva en caso de encontrar un resultado fuera de lo común en análisis posteriores (Figs. 10, 11 y 12).

Figura 10. Perfil de Plásmidos de Cepas de *R. etli*



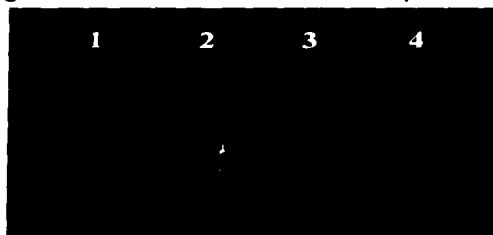
1-CFN 42(d), 2-CFN285, 3-CFN1, 4-TAL182, 5-NITRAGIN8251, 6-BRASIL5(c) , 7-CIAT894 (c), 8 CIAT895(b)

Figura 11. Perfil de Plásmidos de Cepas de *R. etli*



1-COC8 (b), 2-CFN42, 3-KIM5, 4-VIKING1

Figura 12. Perfil de Plásmidos de Cepas de *R. etli*



1-IE4810, 2-IE4804, 3-IE1004, 4-CFN42

Amplificación de las regiones *nifHDK*

A partir de una dilución 1:10 de DNA total de cada cepa, utilizando los pares de oligonucleótidos E201-E202, E215-E216 y E237-E238 se intentó amplificar las regiones *nifHDKa*, *nifHDc* y *nifHDKb* de cada una. En la figura 13 carriles 1, 3 y 4 podemos apreciar el tamaño de los productos de cada una de estas reacciones en la cepa CFN42.

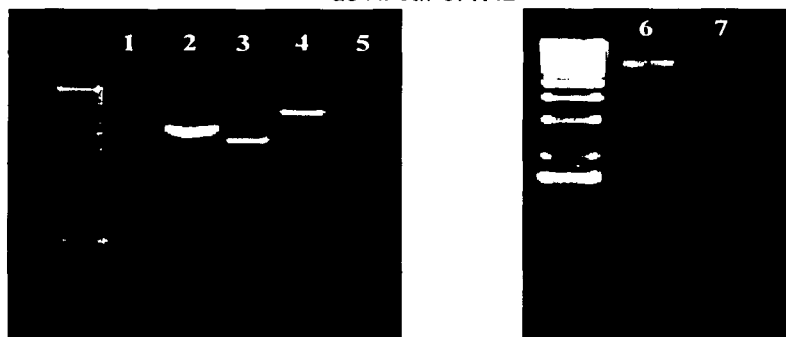
Con los oligonucleótidos indicados, fue posible amplificar solamente las regiones a, b y c de tres de catorce cepas (CFN42, TAL182 y CIAT894). Con esta combinación fue mas frecuente (6 de 14 cepas) obtener producto de las regiones a y c, pero no de la b (BRA5, CIAT895, VIKING1, KIM5, COC8 e IE1004). En 2 de los casos (IE4810 e IE2804) solo se pudo amplificar la región a, mientras que en tres cepas (CFN1, CFN285 y NITRAGIN) no fue posible amplificar ninguna de las regiones (Tabla 7).

Estos resultados sugieren la existencia de polimorfismos en secuencia en sectores aledaños a las regiones *nif*, particularmente la región b. En ese sentido, revisando la secuencia liberada del plásmido simbiótico de *R. etli* CFN 42 (<http://www.cifn.unam.mx/retlidb/>), además se

detectó que el oligo E237 es complementario a una región identificada como una probable IS que puede estar expuesta a inserciones y deleciones que afectarían esta complementariedad. Así mismo se encontró que el oligo E215 es complementario a una región no codificante, la cual pudiera tener una mayor tolerancia a mutaciones. Por esto, se diseñaron nuevos oligonucleótidos para amplificar las regiones *nifHDKb* y *nifHD*c*. En el primer caso se diseñaron B2U, que es complementario con la región regulatoria de *nifHb* y *nifHa* y B2L que es complementario a *nifN*. También se diseñaron dos nuevos pares para amplificar la región *nifHDKc*: U198-L198 en los que el primero es complementario a *noIO* y el segundo a *nifQ* y U171-L171 complementarios también, en diferente lugar, a *noIO* y *nifQ* respectivamente (Fig 14). Utilizando diferentes combinaciones de estos oligos (Tabla 8) se logró amplificar la región b de la mayoría (9 de 11) de las cepas en las que ésta no se había podido obtener, la región c de dos cepas (IE1804 y NITRÁGIN) y la a de dos cepas más (CFN285 y NITRAGIN).

En resumen se amplificaron 35 copias diferentes de las regiones *nif*, estas corresponden a las tres copias (a, b y c) de 10 cepas (CFN42, BRASIL5, TAL182, CIAT894, CIAT895, VIKING, KIM5, IE1004, IE4804 y COC8) y a las copias a y b de las cepas IE4810 y NITRAGIN. La cepa CFN285, de la que solo se obtuvo la región a, y la cepa CFN1, de la que no se pudo amplificar ninguna copia, fueron desechadas ante la imposibilidad de obtener genes parálogos para realizar el estudio.

Figura 13. Amplificación de las regiones *nifHDK* de *R. etli* CFN42



1-*nifHDKa* E-201/E-202 , 2-*nifHDKa* B2U/E-202, 3-*nifHDKc* E-215/E-216 4-*nifHDKb* E-237/E-238 , 5-*nifHDKb* B2U/E-238 , 6-*nifHD*c* U178/L178 , 7- *nifHD*c* U198/L198 . Gel Vertical de agarosa en TA al 1% teñido con Bromuro de etidio. Todos los productos fueron limpiados con el kit de GenClean, con excepción del carril 2 para ejemplificar el producto en bruto, y se añadieron 4 ul de muestra por carril

Tabla 7. Amplificación de las Regiones *nif* con Oligos E

Cepa	<i>nif</i> HDKa	<i>nif</i> HDc	<i>nif</i> HDKb
CFN 42	+	+	+
BRASIL 5	+	+	x
TAL 182	+	+	+
CIAT 894	+	+	+
CIAT 895	+	+	x
VIKING 1	+	+	x
KIM 5	+	+	x
COC 8	+	+	x
CFN 1	x	x	x
CFN 285	x	x	x
NITRAGIN	x	x	x
IE4810	+	x	x
IE4804	+	x	x
IE1004	+	+	x

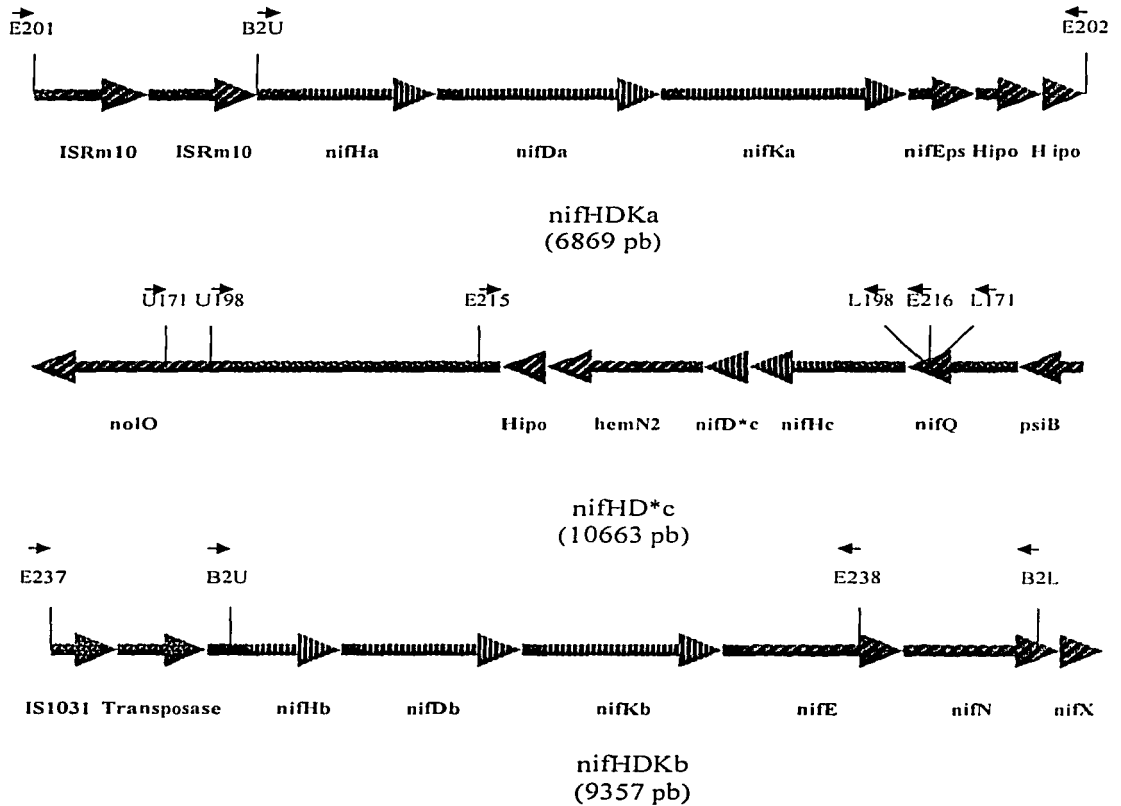
+ se obtuvo; X no se obtuvo. Oligos: *nif*HDKa E-201/E202, *nif*HDKc E-215/E-216, *nif*HDKb E-237/E-238

Tabla 8. Amplificación de las Regiones *nif*

Cepa	<i>nif</i> HDKa	<i>nif</i> HD*c	<i>nif</i> HDKb
BRASIL 5	+	+	B2U/E238
CIAT 895	+	+	B2U/E238
VIKING 1	+	+	B2U/E238
KIM 5	+	+	B2U/E238
COC 8	+	+	B2U/E238
CFN 1	x	x	x
CFN 285	B2U/E202	x	x
NITRAGIN	B2U/E202	x	B2U/E238
IE4810	+	x	B2U/B2L
IE4804	+	U198/198	E237/B2L
IE1004	+	+	E237/B2L

X no se obtuvo; + previamente obtenido

Figura 14. Anotación de las regiones HDK



Las flechas indican la dirección de los oligonucleótidos señalados. Anotación realizada en GeneConstruction Kit con información obtenida de RETLIDB (RETLIDB, 2002)

Amplificación de *nifH*

Los productos obtenidos fueron separados por electroforesis en un gel de agarosa al 1% y eluidos por medio del Kit GeneClean III de BIO101 y se pasaron por una columna de Sephadex. De esta forma se eliminó el DNA genómico que podría dar un falso positivo y los residuos de la limpieza que podrían interferir las siguientes reacciones de PCR. En la figura 15 carril 2 se puede apreciar el producto antes de la limpieza y en los demás carriles como quedan los productos después de esta.

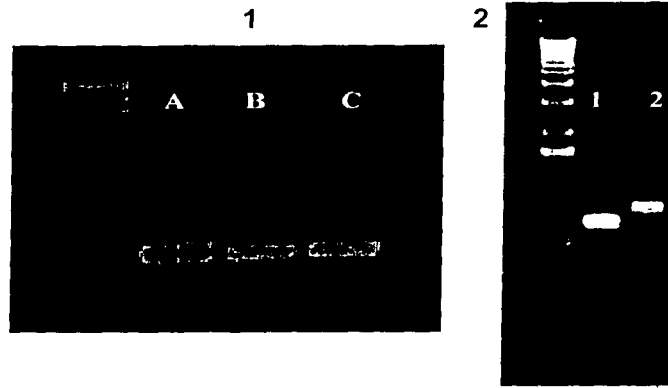
A partir de las regiones purificadas se hicieron reacciones de PCR con los oligos HindU y XbaI (Figs. 15 y 16) para amplificar el gen *nifH* de cada región, sin embargo no todas las reacciones dieron producto. Solo se obtuvieron las tres copias de *nifH* de cuatro cepas (CFN42, BRASIL5, TAL182 y VIKING1) y la copia de la cepa CIAT894 (Tabla 9). Esto sugiere la existencia de polimorfismos en las zonas aledañas a *nifH*, por lo que, para tratar de amplificar las copias de *nifH* que no se pudieron obtener y cuyas regiones estaban disponibles, se probaron las combinaciones HinU-BclI y BamHU-XbaI (Figuras 15 y 16), logrando amplificar las tres copias de 9 cepas (CFN42, BRASIL5, TAL182, VIKING1, CIAT894, CIAT895, KIM5, IE1004 y COC8) y las copias a y b de dos cepas (IE4810 y NITRAGIN). La cepa IE4804 fue desechada al no poder obtener al menos dos genes parálogos para el estudio.

Tabla 9. Amplificación de *nifH*

1				2			
Cepa	<i>nifHa</i>	<i>nifHc</i>	<i>nifHb</i>	Cepa	<i>nifHa</i>	<i>nifHc</i>	<i>nifHb</i>
CFN 42	+	+	+	CIAT 894	*	++	++
Brasil 5	+	+	+	CIAT 895	++	++	++
TAL 182	+	+	+	KIM 5	++	++	++
CIAT 894	+	x	x	COC 8	++	++	++
CIAT 895	x	x	x	CFN 1	o	o	o
VIKING 1	+	+	+	CFN 285	++	o	o
KIM 5	x	x	x	NITRAGIN	++	o	++
COC 8	x	x	x	IE4810	++	o	++
CFN 1	o	o	x	IE4804	++	x	x
CFN 285	x	o	o	IE1004	++	++	++
NITRAGIN	x	o	x				
IE4810	x	o	x				
IE4804	x	x	x				
IE1004	x	x	x				

(1) HindU y XbaI; (2) HinU-BclI y BamHU-XbaI; + se obtuvo; X no se obtuvo; o No disponible; * Ya obtenido; En (2) ++ Obtenido con las dos combinaciones

Figura 15. Amplificación de *nifH*

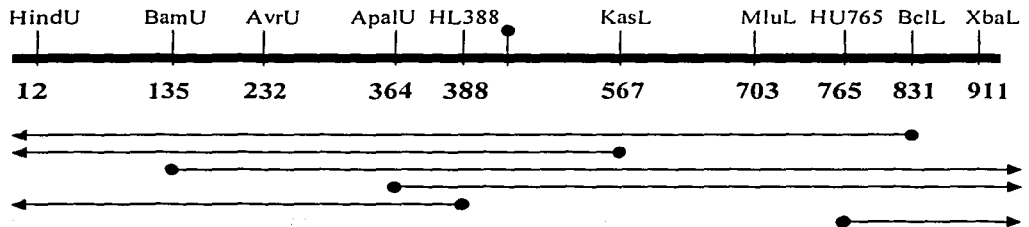


(1) HindU/ XbaL A-*nifHa*; B-*nifHc*; C-*nifHb* CFN42. (2) 1- HindU/BclL *nifHa* CFN42; 2 BamHU/XbaL *nifHa* CFN42. Gel de agarosa en TA al 1% teñido con bromuro de etidio.

Secuenciación

Los productos obtenidos, de las cepas de las que se había podido amplificar por lo menos dos reiteraciones de *nifH*, fueron secuenciados en ambas cadenas utilizando los oligonucleótidos HL388, HU765, KasL, BamU, BclL y ApaU (Fig. 16) en 3 ocasiones.

Figura 16. Oligonucleótidos Complementarios a *nifH* y Estrategia de Secuenciación



(Modificado de Santoyo y Romero)

Una vez depuradas y ensambladas las secuencias se obtuvieron 31 ORFs de 895 pb. Las copias disponibles de los genes en las cepas CFN42, CIAT894, CIAT895, NITRAGIN2851,

BRASIL5, TAL182 resultaron ser idénticas. En la cepa Viking1 se encontró un polimorfismo en *nifHc*. En la cepa KIM5 se encontraron cuatro cambios en cada copia, aunque *nifHc* no compartía dos de ellos con las otras copias. En la cepa COC8 se encontraron cuatro cambios en la copia *nifHa* y cinco en *nifHb* y ambas copias compartían 4 cambios con *nifHa* y *nifHb* de KIM5; de manera interesante, la copia *nifHc* resultó idéntica a las copias de CFN42. La cepa IE4810 tiene cinco cambios compartidos en ambas copias disponibles y 2 más en *nifHb*, 3 de estos compartidos con las cepas COC8, KIM5 y IE1004. Por último, en la cepa IE1004 las tres copias resultaron ser idénticas entre sí presentando 8 cambios (Fig. 17).

La secuencia codifica para un péptido de 297 aminoácidos con 18 sitios variables casi todos se encuentran en tercera base de los codones (Tabla 13 y Figura 19), codificando para el mismo aminoácido, es decir son sinónimos y comúnmente se consideran neutrales. El cambio en la base 80 de Kim5 *nifHc* representa una sustitución en la posición 27 de una alanina por una valina, ambos aminoácidos no polares. El cambio en la base 185 de COC8 *nifHb* provoca una sustitución en la posición 62 de una leucina, un aminoácido no polar, por una arginina, un aminoácido con carga positiva. En la cepa IE4810 en la base 119 hay un cambio que provoca una sustitución de un ácido aspártico por el aminoácido polar serina, en ambas copias. De esta cepa, la copia *nifHa* presenta dos cambios no neutrales más. Uno en la base 309 que provoca un cambio de una isoleucina por una valina, ambos aminoácidos no polares, en la posición 105 y otro en la base 386 que provoca el cambio de una glicina, un aminoácido no polar, por un ácido aspártico en la posición 129 (Fig.18). Un análisis de predicción de estructura secundaria mostró que estas mutaciones no alteran los dominios de Hélice de los que forman parte (datos no mostrados).

Figura 17. Alineamiento de las Secuencias obtenidas de *nifH*

Al ser idénticas en sus 3 copias las cepas CFN42, CIAT894, CIAT895, NITRAGIN2851, BRASIL5, TAL182 se resumen con la copia *nifHa* de CFN42, las tres copias de la cepa IE1004 se resumen con la *nifHa* y la copia *nifHb* de Viking con la *nifHa*

	10	20	30	40	50	50	70
CFN42Ha	ATGTCAGATT	TGCGTCAAAT	CGCATTTTAC	GGCAAAGGGG	GGATCGGCAA	GTCCACCACC	TCCCAAATA
IE4810Ha
IE4810Hb
IE1004Ha
COCHa
COCHb
COCHc
KIMHa
KIMHc
VIKHc

	80	90	100	110	120	130	140
CFN42Ha	CGCTCGCAGC	GCTTGTGCAC	CTCGGGCAGA	AGATCCTGAT	CGTCGGATGC	GACCCGAAG	CCGACTCCAC
IE4810Ha	C
IE4810Hb	C
IE1004Ha
COCHa
COCHb
COCHc
KIMHa
KIMHc	T
VIKHc

	150	160	170	180	190	200	210
CFN42Ha	CCGGCTGATC	CTGAACGCCA	AAGCACAGGA	CACGGTCTG	CATCTGGCAG	CGCAGGAAGG	TTCGGTGGAA
IE4810Ha
IE4810Hb
IE1004Ha
COCHa	T	G
COCHb	T	G
COCHc
KIMHa	T	G
KIMHc	G
VIKHc

	220	230	240	250	260	270	280
CFN42Ha	GACCTTGAGC	TCGAGGACGT	GCTCAAGGCC	GGCTACAAAG	GCATCAAGTG	CGTGGAGTCC	GGCGGTCCGG
IE4810Ha
IE4810Hb
IE1004Ha	G	T
COCHa
COCHb
COCHc
KIMHa
KIMHc
VIKHc

	290	300	310	320	330	340	350
CFN42Ha	AACCGGGCGT	CGGCTGCGCC	GGGCGCGGCG	TCATCACCTC	GATCAATTTC	CTTGAAGAGA	ACGGTGCATA
IE4810Ha	T	G	C	T
IE4810Hb	C	T
IE1004Ha	C	T
COCHa	C	T
COCHb	C	T
COCHc
KIMHa	C	T
KIMHc	C	T
VIKHc

	360	370	380	390	400	410	420
CFN42Ha	TGACGATGTC	GACTACGTCT	CCTATGACGT	GCTCGGCGAT	GTGGTGTGCG	GTGGCTTTGC	GATGCCGATC
IE4810Ha	A	C
IE4810Hb	C
IE1004Ha	C
COCHa	C
COCHb	C
COCHc
KIMHa	C
KIMHc	C
VIKHc

	430	440	450	460	470	480	490
CFN42Ha	CGTGAGAACA	AGGCCCCAGGA	GATCTACATC	GTGATGTCCG	GCGAGATGAT	GGCGCTCTAT	GCCGCCAACA
IE4810Ha
IE4810Hb
IE1004Ha	A.....
COCHa
COCHb
COCHc
KIMHa
KIMHc
VIKHC

	500	510	520	530	540	550	560
CFN42Ha	ACATCGCCAA	GGGCATCCTG	AAATATGCC	ATTCCGGCGG	CGTGCGGCTC	GGCGGCCTGA	TCTGTAACGA
IE4810Ha
IE4810Hb
IE1004Ha
COCHa
COCHb
COCHc
KIMHa
KIMHc
VIKHC

	570	580	590	600	610	620	630
CFN42Ha	GCGCCAGACG	GACCCGCGAGC	TCGACCTCTC	CGAGGCGCTG	GCTGCCAGGC	TCAATTCCAA	GCTCATCCAC
IE4810Ha
IE4810Hb
IE1004Ha
COCHa
COCHb
COCHc
KIMHa
KIMHc
VIKHC

	640	650	660	670	680	690	700
CFN42Ha	TTTGTGCCGC	GTGACAACAT	CGTCCAGCAC	GCCGAGCTCA	GGAAGATGAC	GGTGATCCAG	TACGCGCCCG
IE4810Ha
IE4810Hb
IE1004Ha
COCHa
COCHb
COCHc
KIMHa
KIMHc
VIKHC

	710	720	730	740	750	760	770
CFN42Ha	ACTCCAAGCA	GGCCGGGGAA	TATCGGGCGC	TAGCCGAGAA	GATCCATGCC	AATTCGGGCC	AAGGGACCAT
IE4810Ha
IE4810Hb
IE1004Ha
COCHa
COCHb
COCHc
KIMHa
KIMHc
VIKHC

```

              780      790      800      810      820      830      840
CFN42Ha      TCCGACCCCG ATTACCATGG AAGAGCTCGA AGACATGCTG CTCGACTTCG GCATCATGAA GAGCGACGAG
IE4810Ha    .....
IE4810Hb    .....
IE1004Ha    .....
COCHa       .....
COCHb       .....
COCHc       .....
KIMHa       .....
KIMHc       .....
VIKHc       .....

```

```

              850      860      870      880      890
CFN42Ha      CAGATGCTGG CCGAACTACA GGCCAAGGAG TCAGCGGTGG TTGCGGCTCA ATAA
IE4810Ha    .....
IE4810Hb    .....
IE1004Ha    .....
COCHa       .....
COCHb       .....
COCHc       .....
KIMHa       .....
KIMHc       .....
VIKHc       .....

```

Alineamiento generado con CLUSTALX usando los parámetros por default. Tamaño total del alineamiento 894 pb.

Figura 18. Alineamiento de las Secuencias de Aminoácidos de las copias de *nifH*

Al ser idénticas en sus 3 copias las cepas CFN42, CIAT894, CIAT895, NITRAGIN2851, BRASIL5, TAL182 se resumen con la copia *nifHa* de CFN42, las tres copias de la cepa IE1004 se resumen con la *nifHa* y la copia *nifHb* de Viking con la *nifHa*

```

              10      20      30      40      50
CFN42Ha      MSDLRQIAFY GKGIGKSTT SONTLALVD LGQKILIVGC DPKADSTRLI
CocHb       .....
CocHc       .....
KimHa       .....
KimHc       .....
VikHc       .....
VikHa       .....
IE1004a     .....
IE4810Hb    .....
IE4810Ha    .....

              60      70      80      90      100
CFN42Ha      LNAKAQDTVL HLAAQEGSVE DLELEDV LKA GYKGIKCVES GGPEPGVGCA
CocHb       .....
CocHc       .....
KimHa       .....
KimHc       .....
VikHc       .....
VikHa       .....
IE1004Ha    .....
IE4810Hb    .....
IE4810Ha    .....

              110     120     130     140     150
CFN42Ha      GRGVITSINF LEENGAYDDV DYVSYDVLGD VVCGGFAMPI RENKAQEIYI
CocHb       .....
CocHc       .....
KimHa       .....
KimHc       .....
VikHc       .....
VikHa       .....
IE1004Ha    .....
IE4810Hb    .....
IE4810Ha    .....

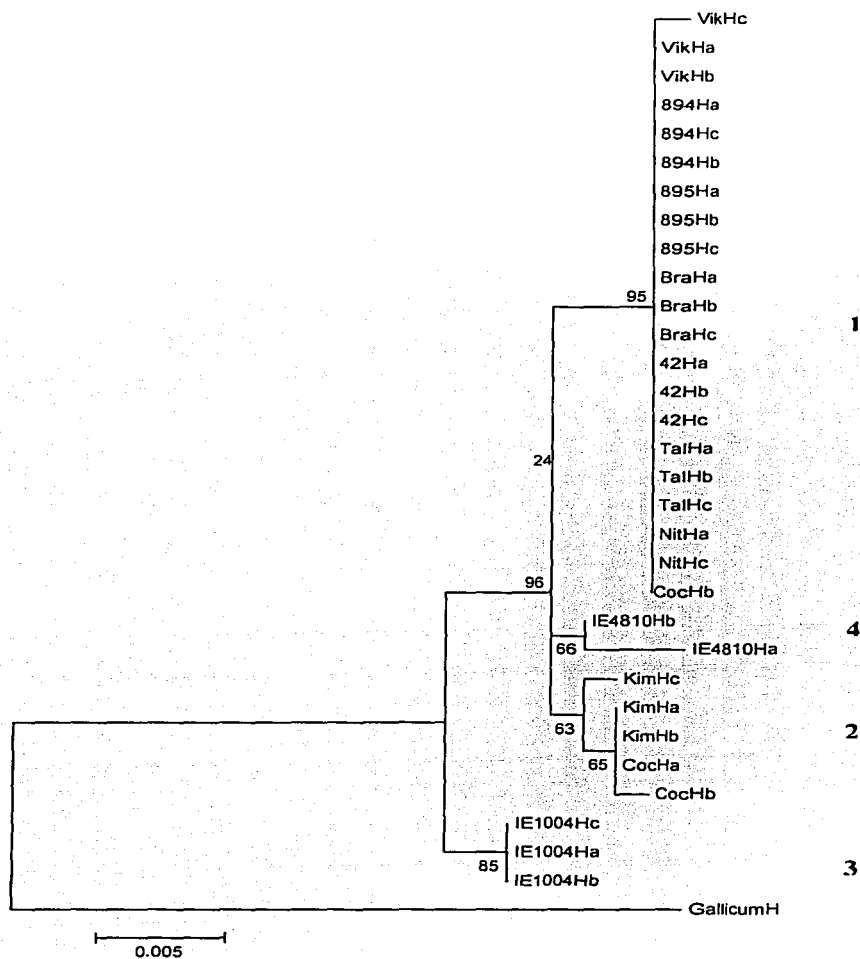
```

	160	170	180	190	200
CFN42Ha	VMSGEMMALY	AANNIAKGIL	KYAHSGGVRL	GGLICNERQT	DRELDLSEAL
CocHb
CocHc
KimHa
KimHc
VikHc
VikHa
IE1004Ha
IE4810Hb
IE4810Ha
	210	220	230	240	250
CFN42Ha	AARLNSKLIH	FVPRDNIVQH	AELRKMTVIQ	YAPDSKQAGE	YRALAEKIHA
CocHb
CocHc
KimHa
KimHc
VikHc
VikHa
IE1004Ha
IE4810Hb
IE4810Ha
	260	270	280	290	
CFN42Ha	NSGQGTIPTP	ITMEELEDML	LDGIMKSDE	QMLAELQAKE	SAVVAAQ
CocHb
CocHc
KimHa
KimHc
VikHc
VikHa
IE1004Ha
IE4810Hb
IE4810Ha

A partir de los alineamientos se generó un filograma en que se puede apreciar mejor la relación de secuencia entre las reiteraciones de las cepas. Para poder dar polaridad el filograma fue enraizado utilizando la secuencia disponible en GeneBank de *Rhizobium gallicum* FL27, una especie aislada en Francia del género *Rhizobium* al que pertenece *R etli*, y se realizó un bootstrap para comprobar la fortaleza de éste. Para facilitar el análisis de los datos el filograma fue dividido en 4 grupos (Fig. 19).

De lo que se puede apreciar en el árbol llama la atención la poca divergencia que hay entre los grupos, a lo mucho 8 pares de bases, lo que parece indicar un origen reciente de la triplicación de *nifH*. En el caso de las cepas IE4810, KIM5 y COC8 se puede observar cierta divergencia en alguna de las copias. Esto podría interpretarse como que las copias de estas cepas que no están bajo el mismo tipo de selección y por lo tanto que no evolucionan de forma concertada. Para descartar esta posibilidad se obtuvo la tasa de sustituciones no sinónimas sobre sinónimas de manera que se observó que las tres copias de *nifH* se encuentran bajo selección purificadora (Tabla 10). Esta conclusión es reforzada si consideramos la alta similitud que existe a nivel de secuencia aminoácidos entre las copias de diferentes cepas, e incluso la que existe con respecto a otras

Figura 19. Filograma de las Reiteraciones de *nifH* en 11 Cepas de *R. etli*



Filograma generado con MEGA 2.1 usando el método Neighbor-Joining. Las primeras 3 letras indican Nit: NITRAGIN8251, 895: CIAT895, Tal: TAL182, 894: CIAT894, Bra: BRASIL5, 42: CFN42, Coc: COC8, Vik: VIKING1, Kim: KIM5, IE1004: IE1004, IE4810: IE4810, Gal: *Rhizobium gallicum* FL27. Las últimas dos letras indican Ha: *nifHa*, Hc: *nifHc*, Hb: *nifHb* excepto en *R. gallicum* en que solo una copia se encuentra disponible (M55226). El bootstrap se realizó con 1000 reiteraciones. Los números indican los grupos en que se dividió el filograma para su discusión.

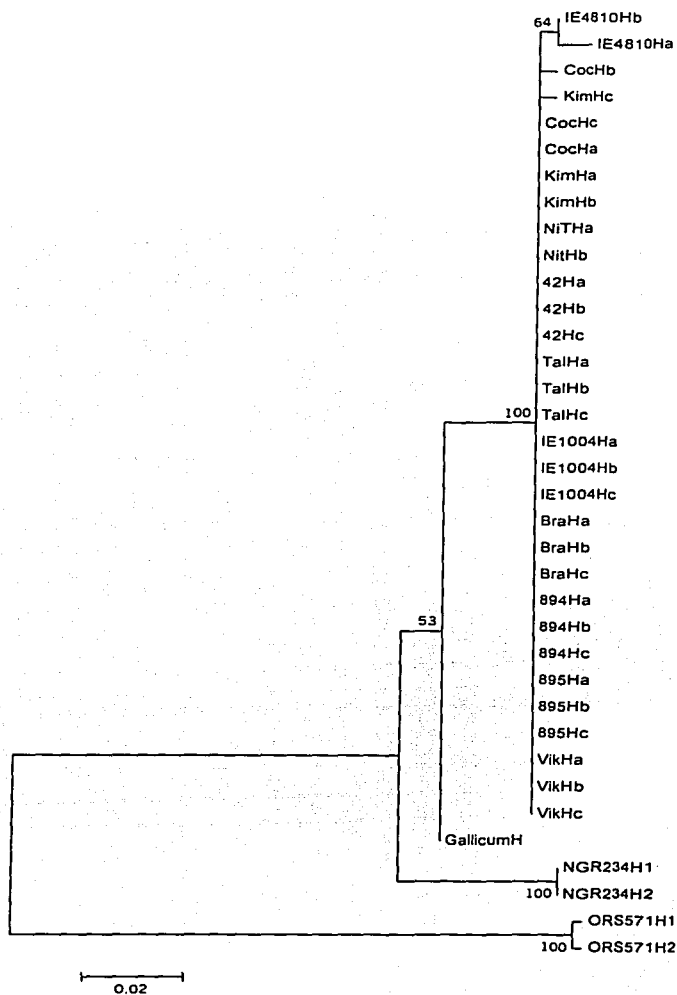
rhizobaceas (Fig. 20). Sin embargo, si estas 3 cepas se analizan de forma independiente, los resultados muestran que los genes de la cepa IE4810 se encuentran bajo selección diversificadora (Tabla 10).

Esta selección purificadora, a nivel de la proteína, también podría estar actuando a nivel de secuencia nucleotídica, según el modelo de nacimiento-muerte. Este modelo asume que nuevos genes son creados por duplicación génica repetida y que algunos de éstos son mantenidos en el genoma por un largo periodo de tiempo, mientras que otros son deletados o se vuelven no funcionales. En este caso y en ausencia de conversión génica las diferencias en secuencia se observarían principalmente en los sitios sinónimos, siendo estos tan numerosos entre las reiteraciones de una especie como entre las reiteraciones de diferentes especies. En el caso de que existiese evolución concertada en presencia de selección purificadora, se esperaría que el número de diferencias sinónimas fuese menor entre las copias dentro de una especie que entre las copias de diferentes especies (Piontkivska et al, 2002). Para discernir entre estas posibilidades realizamos comparaciones del número de sustituciones sinónimas que hay entre todas las copias de *Rhizobium etli* (Tabla 11) y con respecto a las de otras rhizobaceas (Tabla 12). Los resultados de estas muestran que, con excepción de la cepa COC8 de *Rhizobium etli*, a nivel de sustituciones sinónimas hay menor variabilidad entre las copias de una cepa que si comparamos las diferencias que hay entre las secuencias de diferentes cepas. Como se mencionó previamente la copia c de la cepa COC8 se localizó en el grupo 1, mientras que las copias a y b se localizan en el grupo 2; esto constituye una situación particular que se analizará hacia el final de la discusión.

Tabla 10. Tasa de Sustituciones No Sinónimas y Sinónimas

	Dn	ds	dN/dS
nifHa	0.54545	3.45455	0.157893
nifHb	0.22222	2.72222	0.081632
nifHc	0.36364	3.27273	0.111112
IE4810	2	1	2
COC8	0.66667	3.33333	0.200001
Kim5	0.66667	0.66667	1

Figura 20. Filograma de la Secuencia de Aminoácidos de las Reiteraciones de *nifH* en 11 Cepas de *R. etli*



Filograma generado con MEGA 2.1 usando el método Neighbor-Joining. Las primeras 3 letras indican Nit: NITRAGIN8251, 895: CIAT895, Tal: TAL182, 894: CIAT894, Bra: BRASIL5, 42: CFN42, Coc: COC8, Vik: VIKING1, Kim: KIM5, IE1004: IE1004, IE4810: IE4810, Gal: *Rhizobium gallicum* FL27, ORS571: *A. caulinodans* ORS571, NGR234: *Rhizobium* sp. NGR234 Las últimas dos letras indican Ha: *nifHa*, Hc: *nifHc*, Hb: *nifHb* excepto en *R. gallicum* en que solo una copia se encuentra disponible (M55226). El bootstrap se realizó con 1000 reiteraciones. Los números indican los grupos en que se dividió el filograma para su discusión.

Tabla 11. Número de diferencias Sinónimas *ps* (x100) entre los miembros de la familia multigénica *nifH* en diferentes cepas y especies de rhizobaceas

Rhizobacea	ps	Err. Std.
<i>R. etli</i> IE1004	0.000	0.000
<i>R. etli</i> COC8	1.300	0.600
<i>R. etli</i> Kim5	0.300	0.200
<i>R. etli</i> CFN42	0.000	0.000
<i>R. etli</i> IE4810	0.400	0.400
<i>Rhizobium</i> sp. NGR234	0.000	0.000
<i>A. caulinodans</i> ORS571	2.000	0.800

El error standard se calculó en base a un bootstrap con 1000 reiteraciones.

Tabla 12. Número de diferencias Sinónimas *ps* (x100) en los miembros de la familia multigénica *nifH* entre diferentes cepas y especies de rhizobaceas

	IE1004	COC8	KIM5	CFN42	IE4810	NGR234
IE1004						
COC8	2.8 (0.9)					
KIM5	2.6 (0.9)	0.7 (0.3)				
CFN42	3.1 (1.0)	1.3 (0.5)	1.8 (0.8)			
IE4810	2.5 (0.9)	1.5 (0.6)	1.2 (0.6)	1.8 (0.7)		
NGR234	31.2 (2.7)	31.7 (2.7)	31.5 (2.7)	32.0 (2.8)	31.8 (2.7)	
ORS571	41.9 (2.8)	42.7 (2.8)	42.6 (2.8)	42.7 (2.8)	42.5 (2.8)	34.8 (2.7)

ps (x100) y entre paréntesis el error standard calculado en base a un bootstrap con 1000 reiteraciones. Las cinco primeras filas representan *Rhizobium etli* cepas IE1004, COC8, KIM5, CFN42 e IE4810; La sexta y séptima fila representan *Rhizobium* sp. NGR234 y *A. caulinodans* ORS571

DISCUSIÓN

Según los postulados básicos de la teoría de la evolución neutral, la reiteración de un gen implicaría la liberación de la presión selectiva de al menos una de las copias, lo que permitiría que se genere divergencia a nivel de secuencia entre éstas, al no haber selección negativa de mutaciones espontáneas. Pero incluso si se mantuviese una presión de selección en todas las copias, éstas presentarían cierto grado de divergencia por la aparición al azar de mutaciones sinónimas, aunque manteniendo la misma tasa de mutación. La divergencia entre las cepas de *R. etli* se refleja tanto por los variados patrones de plásmidos observados como por la dificultad para amplificar las regiones *nifHDK* y *nifH*, sugiriendo la existencia de polimorfismos a nivel de secuencia. Es así que tratándose de una familia reiterada se esperaría encontrar variabilidad también en cuanto a la secuencia de *nifH*. Sin embargo, la variación es inexistente tanto en la cepa control (CFN42) como en 6 cepas de origen geográfico distinto (Grupo 1). Las reiteraciones de la cepa IE1004 parecen ser más cercanas en secuencia al ancestro común del gen *nifH* de *R. gallicum* y *R. etli*, a partir de ahí se dan nuevos fenómenos de divergencia de los cuales se distinguen los grupos 1, 2 y 4. En el resto de las cepas se encontraron solo unos cuantos cambios, algunos de ellos compartidos y casi todos en la región correspondiente a los primeros 500 pb del gen.

Existen tres posibles explicaciones para interpretar nuestros resultados:

1) La triplicación de los genes *nifH* es un evento reciente.- La absoluta identidad que existe entre las copias de la cepa IE1004 y entre las copias del grupo uno nos permiten descartar la explicación del origen reciente de las reiteraciones pues implicaría que esta se dio en dos ocasiones, sin embargo si podemos asumir que la divergencia entre los grupos es un evento evolutivamente reciente.

2) Las tres copias de *nifH* se encuentran bajo una intensa selección purificadora.- Nuestros análisis de la secuencia de los genes demostraron que, efectivamente, estos se encuentran bajo selección purificadora, (con excepción de la cepa IE4810 que parece estar diversificando). Esto lo podemos comprobar al observar que, a pesar de la variación en secuencia nucleotídica, la secuencia de la proteína es prácticamente igual en las de los 4 grupos.

A pesar de que nuestras comparaciones de sustituciones sinónimas descartan la ocurrencia de selección bajo el modelo de nacimiento-muerte no podemos concluir fehacientemente la ausencia de

selección inclusive en sitios sinónimos. Este fenómeno se podría explicar en el caso en que la secuencia codificase para un RNA antisentido (situación no descrita para *nifH*) o si existiera un fuerte sesgo en el uso de codones. En general, la selección sobre sitios sinónimos no se ha podido establecer claramente en bacterias, pero si ha sido identificada en genes de mamíferos y aunque no se le ha considerado con la misma intensidad que la selección sobre sitios no sinónimos (Jordan et al, 2002 y Hurst y Pál, 2001) no tenemos evidencia que nos permita descartar esta posibilidad. También se podría especular una selección ligada al sistema agrícola al que se encuentran asociadas las cepas epidémicas del grupo 1 y las variantes de los grupos 2, 3 y 4. Sin embargo podemos descartar esta posibilidad ya que sabemos que, aunque las cepas IE se encuentran asociadas a un sistema en específico, una gran mayoría de las cepas aisladas en este sistema resultaron poseer una secuencia idéntica a la de las cepas del grupo 1 (C. Silva, comunicación personal y Silva et al, 2002) y que la cepa COC8 fue aislada a partir de suelo de un sistema similar al que se asocian las cepas del grupo 1 (E. Martínez-Romero, comunicación personal).

3) La evolución concertada juega un papel fundamental para conservar la identidad entre los miembros de esta familia multigénica.- Como se recordará, la evolución concertada puede ser explicada por fenómenos de transposición, recombinación desigual y conversión génica. En *R. etli* solo existen tres copias de *nifH*, ninguna contenida en un transposón o flanqueada por IS's. La recombinación desigual ocurre preferencialmente entre genes en tandem, con alto número de copias, y además no se detectaron los rearrreglos característicos de este fenómeno en los perfiles de plásmidos de nuestras cepas. La hipótesis de conversión génica es reforzada por la distribución de los polimorfismos, preferentemente entre las bases 199 y 420, lo que parece ser una regionalización de los cambios de manera que se forma un tracto, además de que se cuenta con el importante antecedente de la comprobación experimental de conversión génica entre estos genes (Rodríguez y Romero, 1998). Adicionalmente, resultados obtenidos recientemente en nuestro laboratorio (Gustavo Santoyo) parecen concordar en el hecho de que hay una cierta región del gen que tiende a sufrir de conversión génica en las zona de los primeros 500 pb de *nifH*.

En realidad resulta difícil poder distinguir si la identidad en las secuencias de *nifH* se debe a una intensa selección purificadora o a un fenómeno de conversión génica pues la poca divergencia de nuestra familia le resta resolución a nuestros datos. Los resultados del análisis de nuestras secuencias en conjunto con la evidencia experimental que nuestro grupo ha generado parecen

inclinarse hacia esta última. Sin embargo no podemos dejar de considerar que la selección purificadora juega un papel preponderante en la conservación de la secuencia y que para poder llegar a distinguir cual es la principal fuerza responsable del fenómeno es necesario que se amplíe el conocimiento sobre el impacto que esta pueda llegar a tener sobre los sitios sinónimos de la secuencia nucleotídica. Además también se debe conocer más sobre la historia evolutiva y ecológica de las cepas y la especie de manera que puedan identificarse otras particularidades sobre estas y sobre los genes.

Si bien la recombinación es un proceso genético bien estudiado (Kowalczykowski, 2000) aún quedan por entender los mecanismos que derivan en la evolución concertada. Con este trabajo hemos estudiado el posible papel que este fenómeno ha jugado en la historia evolutiva de una familia multigénica de tres copias en una misma unidad replicativa. ¿Pero que pasa en familias con diferente número de copias? Probablemente con dos copias habría un efecto similar con el que se encontró con tres copias, aunque en nuestro laboratorio se ha demostrado que la tasa de conversión de *nifH* desciende con el número de copias (Rodríguez y Romero, 1998). A mayor número de copias debería ser más factible que algunas copias se escapen del fenómeno y tiendan a divergir. En cuanto a esta última suposición, se sabe que los genes de RNA ribosomal 16S que están reiterados hasta 8 veces muestran poca o nula de divergencia en diferentes bacterias. (Liao, 2000). Pero es necesario aclarar que se trata de genes que se encuentran en la misma unidad replicativa, el cromosoma bacteriano. Esto da pie a una segunda pregunta: ¿Cómo se alterarían los patrones de evolución concertada por conversión génica en familias multigénicas cuyos miembros se encuentren en diferentes unidades replicativas? La respuesta a esta pregunta debería de tomar en cuenta el tamaño de las moléculas involucradas y el porcentaje de homología entre ellas, ambas características necesarias para el apareamiento de secuencias homólogas. En nuestro laboratorio Cesar Rodríguez encontró que la tasa de conversión en su modelo (descrito en la introducción de estas tesis) no cambiaba significativamente si las copias se encontraban en diferentes unidades replicativas (Rodríguez y Romero, 1998). Por otro lado, Gustavo Santoyo encontró que con 1% de divergencia la tasa de recombinación conjugativa entre secuencias de aproximadamente 1 kb se reduce 100 veces. Un trabajo similar al de esta tesis podría ampliar el conocimiento sobre la conversión de reiteraciones en diferentes unidades replicativas. Para llevar esto a cabo propongo dos familias multigénicas para ser estudiadas: La de los genes *fixNOQP* que se encuentran el

plásmido simbiótico, p42d, y el plásmido p42f de *R. etli* y la de el gen *fnrN* que se encuentra en el cromosoma y el plásmido simbiótico, p42d, también de *R. etli* (López et al, 2001).

Las reiteraciones de ambas familias son más divergentes que los de *nifH*, poseen una identidad del 80%, lo que nos permitiría aplicar una prueba estadística. El método de Sawyer estudia para una familia de genes parálogos la longitud de regiones similares de sitios polimórficos silenciosos delineados entre regiones no similares de éstos, esperando que en el caso que haya conversión génica entre los miembros, las regiones similares sean de mayor tamaño que las que se encontrarían por azar (Sawyer, 1989; Sawyer, 1999). Para que el análisis funcione, se necesita que se pueda delimitar zonas similares entre zonas variables, por lo que realizar el análisis con secuencias idénticas, como en el caso de las copias de *nifH* daría un resultado positivo pero poco convincente.

En el caso de *fixNOQP* solo la reiteración del plásmido simbiótico es esencial para la fijación de nitrógeno en simbiosis (Girard et al, 2000) por lo que se podría asumir una pérdida de presión selectiva en la otra copia.

Al comparar la topología de los árboles construidos con la de los que aquí presentamos podríamos identificar las diferencias en la actividad de conversión génica y su reflejo en los patrones de evolución concertada de genes contenidos en la misma unidad replicativa (*nifH*), en dos distintas con gran diferencia de tamaño (*fnrN*) y en dos distintas de tamaño similar (*fixNOQP*). En el caso que estos no fueran claros contaríamos con una metodología estadística que nos permitiría detectar la existencia del fenómeno.

Adicionalmente existen familias reiteradas en otras rhizobaceas que podrían ser utilizadas como modelo para ayudar a resolver estas preguntas (Tabla 13).

A continuación quisiera referirme al caso de la cepa COC8 cuyas copias de *nifH* se ubicaron en distintos grupos dentro del filograma. Esta particularidad se vió reflejada cuando se compararon las distancias de sustituciones sinónimas inter e intracepa, pues estas son similares lo que parecería indicar la existencia de selección bajo el modelo de nacimiento-muerte. Sin embargo esto implicaría que a partir de mutaciones al azar una de las copias haya llegado por convergencia a una secuencia idéntica a la de las cepas del grupo uno, por lo que mas bien parecería un fenómeno de polimorfismos ancestrales. Una explicación alternativa para este hallazgo podría ser la transferencia horizontal. Este fenómeno es responsable de la propagación de características genéticas entre un amplio espectro de especies bacterianas y a partir de análisis de genomas completos ha sido

Tabla 13. Genes reiterados con alta identidad en Rhizobiaceae

CEPAS	GENES	IDENTIDAD	BASES	CARACTERISTICAS
<i>Sinorhizobium meliloti</i> 1021	<i>flaA</i> y <i>flaB</i>	85%	1191	2 copias en Tandem
<i>Sinorhizobium meliloti</i> RU10406	<i>flaA</i> y <i>flaB</i>	93%	1185	2 copias en Tandem
<i>Sinorhizobium meliloti</i> 2011	<i>fixK</i>	100%	636	2 copias en el plásmido Sim.
<i>Sinorhizobium meliloti</i> 1021	<i>groEL</i>	97%	1638	2 copias en el plásmido Sim.
<i>Rhizobium tropici</i> CFN299	<i>csA</i>	96%	1213	Una copia en cromosoma y otra en pSim.
<i>Rhizobium. leguminosarum</i> bv. viceae VF39	<i>fixN</i>	94%	1623	Una copia en el plásmido c y otra en el plásmido d.

Modificada de Rodríguez, 1998

reconocido como importante fuerza generadora de diversidad bacteriana (Eisen, 2000 y Garcia-Vallvé et al, 2000). Sin embargo también se considera que ciertos genes no son factibles de ser transferidos, por ejemplo se cree que genes con homólogos funcionales en un organismo tendrían poco éxito permaneciendo dentro de éste (Lawrence, 1999). ¿Pero que pasaría si entre éstos dos genes se diera un evento de recombinación? En nuestro grupo de trabajo, Gustavo Santoyo ha realizado cruza de *E. coli* (conteniendo una copia de *nifH* con mutaciones que generan un sitio de corte para enzima de restricción cada 100 pb en la primera mitad del gen) con una cepa de *R. etli* con un plásmido con una sola copia de *nifH* (con mutaciones del mismo estilo en la última mitad del gen). Como resultado se ha detectado que hay recombinación entre estas dos regiones de homología de los plásmidos (a una frecuencia de 1×10^{-7}), suceso al que va ligado un evento de conversión génica (98% de los casos) en el que invariablemente la copia del plásmido externo a *R. etli* transfiere información a la copia endógena. Así entonces el patrón observado en la cepa COC8 podría ser explicado por la conversión de *nifHc* por una copia exógena que hubiese penetrado ya sea por transformación o por conjugación. Si tomamos en cuenta que, como se mencionó previamente, esta cepa parece compartir localización geográfica con las cepas del grupo 1 estaríamos hablando de un evento particular pero posible, pues se ha reportado movilización del plásmido simbiótico en el género *Rhizobium* (Wernegreen y Riley, 1999) y entre cepas de *R. etli* (Valdés y Piñero, 1992). Además *Rhizobium etli* CFN42 posee un plásmido (p42a) que es automovilizable a una tasa de 1×10^{-2} y que por cointegración puede movilizar al plásmido simbiótico (p42d o pSym) con una frecuencia de 2×10^{-6} (Brom et al, 2000). Incluso recientemente, el

grupo del Dr. Juan Sanjuan (Granada, España), con el cual nuestro grupo mantiene una colaboración ha encontrado evidencia de que el p42d es automovilizable. Evidentemente al ser idéntica la copia de *nifHc* con las del grupo 1 debemos suponer que es un evento evolutivamente reciente y probablemente en un número de generaciones la copia c será convertida nuevamente a similitud de las copias a y b o incluso se puede dar el caso que la conversión sea en sentido contrario lo que podría también explicar la singular situación de las cepas del grupo 1.

En el caso de las mutaciones no sinónimas de las cepas IE4810 y COC8 por medio de los análisis de dominio que realizamos inferimos que estas no modificaban la estructura secundaria, sin embargo sería interesante realizar ensayos de fijación de nitrógeno para detectar posibles diferencias entre estas cepas y las que no muestran cambios en su secuencia de aminoácidos.

Por último creo que es necesario aumentar los esfuerzos por amplificar los copias faltantes de la cepa IE4810 e IE4804, que por su variación en secuencia (Silva et al, 2002 y este trabajo) con respecto a las demás cepas contienen valiosa información que podría apuntalar nuestras conclusiones.

REFERENCIAS

1. -Abdulkarim, F., y D. Hughes. 1996. Homologous recombination between the *tuf* genes of *Salmonella typhimurium*. J. Mol. Biol. 260: 506-522.
2. -Alani, E., R. Reenan y R. D. Kolodner. 1994. Interactions between mismatch repair and genetic recombination in *Saccharomyces cerevisiae*. Genetics. 137:19-39
3. -Anderson, D. G. y S. C. Kowalczykowski. 1997. The recombination hotspot x is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. Genes Dev. 11:571-581
4. -Anderson, D. G. y S. C. Kowalczykowski. 1997b. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a x-regulated manner. Cell. 90: 77-86
5. -Badenoch-Jones, J., T. A. Holton, C. Morrison, K. F. Scott y J. Shine. 1989. Structural and functional analysis of nitrogenase genes from the broad-host-range *Rhizobium* strain ANU240. Gene 77: 141-153.
6. -Brom, S., A. García-de los Santos, L. Cervantes, R. Palacios y D. Romero. (2000). In *Rhizobium etli* symbiotic plasmid transfer, nodulation competitiveness and cellular growth require interaction among different replicons . Plasmid. 44: 34-43
7. -Brown, T. 1999. Genomes. Bios Scientific Publishers. 1a ed.
8. -Dixon D. A., y S. C. Kowalczykowski. 1994. Reversible inactivation of the Escherichia coli RecBCD by the recombination hotspot x in vitro: Evidence for functional inactivation or loss of the RecBCD subunit. PNAS. 91: 2980-2984
9. -Dover, G. 2000. How genomic and developmental dynamics affect evolutionary processes. BioEssays 22: 1153-1159
10. -Doyle, J. y B. Gaut. 2000. Evolution of genes and taxa: a primer. Plant Mol. Biol. 42: 1-23
11. -Drouin, G., F. Prat, M. Ell y G. D. P. Clarke. 1999. Detecting and characterizing gene conversions between multigene family members. Mol. Biol. Evol. 16: 1369-1390
12. -Eckhardt, T. 1978. A rapid method for the identification of plasmid desoxyribonucleic acid in bacteria. Plasmid. 1(4): 584-8
13. -Eggleston, A. K., y S. C. West. 1997. Recombination initiation: Easy as A, B, C, D...x?. Curr. Biol. 7: R745-R749

14. -Eisen, J. A. 2000. Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Curr. Op. Genet. Dev.* 10:606-611
15. -Elkan, G. H. y C. Bunn. 2002. The Rhizobia. en *The prokaryotes*. M. Dworkin. Springer-Verlag.
16. -Ewing G. B., P. Green. 1998. Base calling of automated sequencer traces using phred II error-probabilities. *Genome Research*: 8:186-194
17. -Flores, M., V. González, S. Brom, E. Martínez, D. Piñero, D. Romero, G. Dávila y R. Palacios. 1987. Reiterated DNA sequences in *Rhizobium* and *Agrobacterium*. *J. Bacteriol.*, 169: 5782-5788.
18. -Flores, M., Mavingui P., Perret X., Broughton W.J., Romero D., Hernandez G., Davila G., Palacios R. 2000. Prediction, identification, and artificial selection of DNA rearrangments in *Rhizobium*: Toward a natural genomic design. *Proc. Natl. Acad. Sci. U.S.A.* 97(16): 9138-9143
19. -Futuyma, D. 1986. *Evolutionary Biology*. Sunderland, Massachusetts. Sinauer
20. -García-Vallvé, S., A. Romeu, J. Palau. 2000. Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Research*. 10:1719-1725
21. -Gillespie, J. 1998. *Population genetics, a concise guide*. John Hopkins University Press: 78-82
22. -Girard, L., S. Brom, A. Dávalos, O. López, M. Soberón y D. Romero. 2000. Differential regulation of *fixN* reiterated genes in *Rhizobium etli* by a novel *fixL-fixK* cascade. *Mol. Plant-Microbe Interact.* 13: 1283-1292
23. -Gogarten, P. y L. Olendzenski. 1999. Orthologs, paralogs and genome comparisons. *Curr. Op. in Gen. and Dev.* 9:630-636
24. -Gordon, D., C. Abajian y P. Green. 1998. Consed: a graphical tool for sequence finishing. *Genome Research*. 8: 195-202
25. -Gutierrez-Zamora, M. L. y E. Martinez-Romero. 2001. Natural endophytic association between *Rhizobium etli* and maize (*Zea mays* L.). *J. of Biotech.* 91: 117-126
26. -Hall, T. A. 1999. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* 41: 95-98

27. -Hunter, N y N. Kleckner. 2001. The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. *Cell*. 106:59-70
28. -Hurles, M. 2001. Gene conversion homogenizes the CMT1A paralogous repeats: *BMC Genomics*. 2:11
29. Hurst, L. D. y C. Pál. 2001. Evidence for purifying selection acting on silent sites in BRCA1. *TRENDS Genetics*. 17(2): 62-65
30. -Jones, D. J. 1999. GenThreader: an efficient and reliable protein folds recognition method for genomic sequences. *J. Mol. Biol.* 287: 797-815
31. -Jordan, I. K, K. Makarova, Y. Wolf y E. Koonin. 2001. Gene conversion in genes encoding outer-membrane proteins in *H. pylori* and *C. pneumoniae*. *TRENDS Genetics*. 17(1): 7-10
32. -Jordan, I. K., I. B. Rogozin, Y. I. Wolf y E. Koonin. 2002. Microevolutionary Genomics of Bacteria. *Theoretical Population Biology*. 61: 435-447
33. -Josephson K. L. y I. L. Pepper. 1984. Competitiveness and effectiveness of strains of *Rhizobium phaseoli* isolated from the Sonoran desert. *Soil. Biol. Biochem.* 16(6): 651-655
34. -Kimura M. y T. Ohta. 1974. On some principles governing molecular evolution. *PNAS* 71(7): 2848-2852
35. -Kowalczykowski, S., 2000. Initiation of genetic recombination and recombination-dependent replication. *TIBS*. 25.
36. Kumar, S., K. Tamura, I. B. Jakobsen, y M. Nei (2001). MEGA2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, Arizona, USA
37. -Lathe, W y P. Bork. 2001. Evolution of *tuf* genes: ancient duplication, differential loss and gene conversion. *FEBS Letters*. 502: 113-116
38. -Lawrence, J.G. 1999. Gene transfer, speciation, and the evolution of bacterial genomes. *Curr. Op. in Microbiol.* 2: 519-523
39. -Lewin, B. 2000. *Genes 2000*. Oxford Univ. Press. 7a ed.
40. -Liao, D. 2000. Gene conversion drives within genic sequences: Concerted evolution of ribosomal RNA genes in bacteria and archaea. *J Mol Evol.* 51:305-317
41. -Lopez, O., C. Morera, J. Miranda-Rios, L. Girard, D. Romero y M. Soberón. 2001. Regulation of Gene expression in response to oxygen in *Rhizobium etli*: Role of *FnrN* in *fixNOQP* expression and in symbiotic nitrogen fixation. *J. Bact.* 183(24): 6999-7006

42. -Marchler, B., A. Panchenko, A. R. Shoemaker, B. A. Thiessen, P. A. Geerly y S. H. Bryant. 2002. CDD a database of conserved domain alignments with links to domain three-dimensional structure. *Nucl. Acids Res.* 30: 281-283
43. -Meinersmann, R y K. Hiatt. 2000. Concerted evolution of duplicate *fla* genes in *Campylobacter*. *Microbiology.* 146: 2283-2290
44. -Meyer, A. y M. Schartl. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Op. Cell Biol.* 11: 699-704
45. -Michel, B. 2000. Replication fork arrest and DNA recombination. *TIBS.* 25:173-178
46. -Nei, M. y S. Kumar. 2000. Molecular evolution and filogenetics. Oxford University Press:260-262
47. -Norel, F. y C. Elmerich. 1987. Nucleotide sequence and functional analysis of the two *nifH* copies of *Rhizobium* ORS571. *J. Gen. Microbiol.* 133: 1563-1576.
48. Piontkivska, H., A. P. Rooney y M. Nei. 2002. Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol. Biol. Evol.* 19(5):689-697
49. -Piñero, D., E. Martínez y R. K. Selander. 1988. Genetic diversity and relationships among isolates of *Rhizobium leguminosarum* biovar *phaseoli*. *Appl. Environ. Microbiol.* 54: 2825-2832.
50. -Quinto, C., H. de la Vega, M. Flores, J. Leemans, M.A. Cevallos, M.A. Pardo, R. Azpiroz, M.L. Girard, E. Calva y R. Palacios. 1985. Nitrogenase reductase: a functional multigene family in *Rhizobium phaseoli*. *Proc. Natl. Acad.Sci. U.S.A.* 82: 1170-1174.
51. -Rees, D. C. et al. 1993. Structures and functions of the nitrogenase proteins. En *New horizons on nitrogen fixation*. R. Palacios et al. (Eds.). Kluwer Academic Publishers: 83-88
52. -Resnick, M. A., y P. Martin. 1976. The repair of double strand breaks in the nuclear DNA of *Saccharomyces cerevisiae* and its genetic control. *Mol. Gen. Genet.* 143:119-129
53. -Rocha, E. P. C., A. Danchin y A. Viari. 1999. Analysis of long repeats in bacterial genomes revealed alternative evolutionary mechanisms in *Bacillus subtilis* and other competent prokaryotes. *Mol. Biol. Evol.* 16: 1219-1230.
54. -Rodríguez, C. 1992. Amplificación y delección por recombinación entre *nifH* de *Rhizobium phaseoli* ¿Un fenómeno general entre cepas?. Tesis de maestría en Ciencias. Facultad de Ciencias, UNAM.

55. -Rodríguez, C. 1998. Mecanismos de variación concertada entre secuencias reiteradas en el genoma de *Rhizobium etli* CFN42. Tesis de Doctorado en Investigación Biomédica Básica. CIFN, UNAM
56. -Rodríguez, C. y D. Romero. 1998. Multiple recombination events maintain sequence identity among members of the nitrogenase multigene family in *Rhizobium etli*. *Genetics*. 149: 785-794
57. -Romero, D., J. Martínez-Salazar, E. Ortiz, C. Rodríguez y E. Valencia-Morales. 1999. Repeated sequences in bacterial chromosomes and plasmids: a glimpse from sequenced genomes. *Res. Microbiol.* 150: 735-743.
58. -Romero, D., P. Singleton, L. Segovia, E. Morett, B. Bohlool, R. Palacios y G. Davila. 1988. Effect of naturally occurring *nif* reiterations on symbiotic effectiveness in *Rhizobium phaseoli*. *Appl. Env. Microbiol.* 54(3): 848-850
59. -Romero, D., S. Brom, J. Martínez-Salazar, M. Girard, R. Palacios y G. Dávila. 1991. Amplification and deletion of a nod-nif region in the symbiotic plasmid of *Rhizobium phaseoli*. *J. Bacteriol.* 173(8): 2435-2441
60. Rozas, J. y R. Rozas. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15: 174-175.
61. -Rutherford, S. 2000. From genotype to phenotype: buffering mechanisms and the storage of genetic information. *Bioessays*. 22: 1095-1105
62. -Sawyer, S. A. 1989. Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* 6: 526-538
63. -Sawyer, S. A. 1999. GENECONV: a computer package for the statistical detection of gene conversion. Distribuido por el autor, Dept. of Mathematics, Washington University in St. Louis, disponible en <http://www.math.wustl.edu/~sawyer> (PC)
http://bioag.byu.edu/zoology/crandall_lab/programs.htm (Versión Macintosh)
64. -Segovia, L., P. W. Young, E. Martínez-Romero. 1993. Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* type 1 strain as *Rhizobium etli* sp. nov. *Int. J. of Syst. Bact.* 43(2): 374-377
65. -Semple, C. y K. Wolfe. 1999. Gene duplication and gene conversion in the *Caenorhabditis elegans* genome. *J. Mol. Evol.* 48: 555-564

66. -Shaffer, L. G. y J. R. Lupski. 2000. Molecular mechanisms for constitutional chromosomal rearrangements in humans. *Annu. Rev. Genet.* 34: 297-329
67. -Sharples, G. S. Ingleston y R. Lloyd. 1999. Holliday junction processing in bacteria: insights from the evolutionary conservation of RuvABC, RecG and RusA. *J. of Bact.* 181(18): 5543-5550
68. -Silva, C., P. Vinuesa, L. E. Eguiarte, E. Martinez-Romero y V. Souza. 2002. Population genetics and biogeographic implications of *Rhizobium etli* and *Rhizobium gallicum* nodulating common bean in a traditionally managed milpa plot in México. *Appl. Env. Microbiol.* En revisión.
69. -Stahl, F. 1994. The Holliday Junction on its thirtieth anniversary. *Genetics.* 138: 241-246
70. -Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin y D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* 25(24): 4876-4882
71. -Valdés, A. M. y D. Piñero. 1992. Phylogenetic estimation of plasmid exchange in bacteria. *Evolution.* 46(3): 641-656
72. -Vandamme P, B. Pot, M. Gillis, P. de Vos, K. Kersters, J. Swings. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60(2): 407-438
73. -Vinuesa, P., B. Reuhs, C. Breton y D. Werner. 1999. Identification of a plasmid-borne locus in *Rhizobium etli* KIM5 involved in lipopolisaccharide O-chain biosynthesis and nodulation of *Phaseolus vulgaris*. *J. Bacteriol.* 181(18): 5606-5614
74. -Wang, E. T., J. Martínez-Romero y Lopez-Lara I. M., 2001. *Rhizobium* y su destacada simbiosis con plantas. en *Microbios en línea.* E. Martínez-Romero y J. Martínez-Romero. UNAM: 131-156 (<http://biblioweb.dgsca.unam.mx/libros/microbios/index.html>)
75. -Wernergreen, J. J. y M. A. Riley. (1999). Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages. *Mol. Biol. Evol.* 16(1): 98-113.