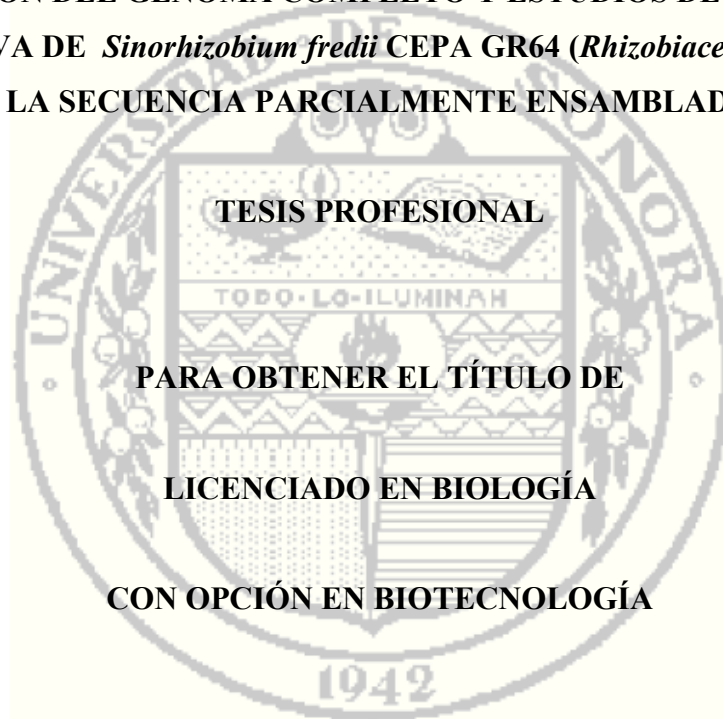


UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS BIOLÓGICAS Y DE LA SALUD

DEPARTAMENTO DE INVESTIGACIONES CIENTÍFICAS Y TECNOLÓGICAS

OBTENCIÓN DEL GENOMA COMPLETO Y ESTUDIOS DE GENÓMICA
COMPARATIVA DE *Sinorhizobium fredii* CEPA GR64 (*Rhizobiaceae*) APARTIR DE
LA SECUENCIA PARCIALMENTE ENSAMBLADA



TESIS PROFESIONAL
PARA OBTENER EL TÍTULO DE
LICENCIADO EN BIOLOGÍA
CON OPCIÓN EN BIOTECNOLOGÍA

PRESENTA:

DIANA PATRICIA GALLEGO BRISEÑO

Hermosillo, Sonora

Julio, 2014

Universidad de Sonora

Repositorio Institucional UNISON



"El saber de mis hijos
hará mi grandeza"



Excepto si se señala otra cosa, la licencia del ítem se describe como openAccess

FORMATO DE APROBACIÓN

Los miembros del Comité de Tesis designado para revisar la Tesis de **Diana Patricia Gallego Briseño** la han encontrado satisfactoria y recomiendan que sea aceptada como requisito para obtener el Título de Licenciado en Biología con Opción en Biotecnología.

Dra. Susana Brom Klanner
Director de Tesis

Dr. Marco Antonio López Torres
Sinodal Secretario

Dr. Luis Enrique Gutiérrez Millán
Sinodal

Dr. Alejandro Varela Romero
Suplente

DEDICATORIA

A mis padres.

Padres, una palabra que generalmente hace referencia a un padre y una madre. Yo tuve la fortuna de tener como padres a un abuelo, Don Epitacio Gallego, una abuela, Doña Amparo Briseño, y una madre, Adriana Gallego. Todo lo bueno que existe en mi persona, que he hecho y haré, se lo debo a ellos.

AGRADECIMIENTOS

A todas aquellas personas que a lo largo del camino me tendieron la mano, me escucharon, me animaron e incluso me alimentaron cuando lo necesité, todas aquellas cosas buenas que han hecho por mí, pequeñas y grandes, sumadas me abrieron las puertas. Familia, personas que se convirtieron en amigos, amigos que se convirtieron en familia.

A mi madre, Adriana, por el gran amor y ejemplo que me ha dado. El amor es el principal ingrediente de la felicidad, por eso ser feliz no cuesta ni dinero, ni trabajo, eso no lo aprendí en la escuela, lo aprendí contigo mamá. No hubiera logrado nada sin tu apoyo y ese amor.

A mis abuelos, Epitacio y Amparo, por su amor y paciencia, gracias a ustedes tengo altas expectativas de mi misma y ganas de ser una buena persona.

A mi hermano Adrián, por venir al mundo a cambiar mi vida para bien. Sin ti no sería quien soy hoy, tú eres lo más valioso que tengo.

A mi tía Diana, que desde que tengo memoria ha estado a mi lado, desde los momentos más felices, hasta los más difíciles, brindándome su apoyo y amor en todo tiempo.

A mi tía Paty y la familia López Gallego, mi tío Jaime y prima Maru, por aceptarme en su hogar y adoptarme durante este proceso, haciéndome sentir como en casa, le estaré eternamente agradecida.

A mis amigos, Nacho, Emiliano, Thelma y Alejandra, por hacer la vida académica más llevadera y divertida. Gracias por aguantar mis días de control freak, por todas las risas, llantos y momentos de estrés y alegrías que compartimos en nuestro camino para convertirnos en biólogos, nuestra aventura más fascinante.

A mi gran amigo y amor, Nacho, y a toda su familia, por todo el apoyo, dedicación, cariño y amor invertido en mí. Espero que te quedes a mi lado para seguir compartiendo logros y alegrías.

A todos mis maestros de la licenciatura, por el conocimiento que nos compartieron en las aulas y fuera de ellas. Especialmente a mis sinodales, el Dr. Marco López, el Dr. Alejandro Varela y el Dr. Luis Enrique Gutiérrez, a mis maestros el Dr. Francisco Molina Freaner, el Dr. Carlos Peña, y al Dr. Alf Meling López, por compartir algo más que conocimiento, cada uno con su estilo muy particular, y al Dr. Fernando Enríquez y la M.C. Gerardina Nubes por la revisión de este trabajo.

A mis asesores de tesis, la Dra. Susana Brom y mi mentor y Sensei el Dr. Gonzalo Torres, por la paciencia, el apoyo, el cariño y sobre todo, por el valioso conocimiento que me brindaron.

Al Centro de Ciencias Genómicas de la UNAM, por el apoyo brindado, y a todos mis compañeros y amigos del Laboratorio de Ingeniería Genómica, que no sólo me compartieron conocimiento, sino que también me ofrecieron su amistad.

Y finalmente, pero no menos importante a la Universidad de Sonora, al Departamento de Investigaciones Científicas y Tecnológicas y a la División de Ciencias Biológicas de la Salud por el apoyo económico brindado para poder ir a realizar este trabajo.

CONTENIDO

| | PÁGINA |
|---|--------|
| I. INTRODUCCIÓN | 1 |
| II. ANTECEDENTES | |
| II.1. Fijación Biológica del Nitrógeno y Nodulación | 3 |
| II.2. Genes de simbiosis: Islas y plásmidos simbióticos | 4 |
| II.3. Rizobios | 4 |
| II.3.1. Evolución y filogenia de Rizobios | 5 |
| II.3.1.1 Transferencia horizontal de genes: Papel en la evolución de los genomas de Rizobios | 6 |
| II.3.2. Características de los genomas de Rizobios | 6 |
| II.4. <i>Rhizobium etli</i> CFN42 | 7 |
| II.5. <i>Sinorhizobium fredii</i> GR64 | 8 |
| II.5.1. Secuenciación del plásmido pSfrGR64a de <i>S. fredii</i> GR64 | 8 |
| II.5.2. Pirosecuenciación del genoma de <i>S. fredii</i> GR64 | 9 |
| II.6. Secuenciación de genomas | 9 |
| II.6.1. Tecnologías de secuenciación | 10 |
| II.6.1.1. Terminación cíclica reversible | 11 |
| II.6.1.2. Pirosecuenciamiento: Roche 454 | 11 |
| II.6.2. Programas bioinformáticos de ensamble y finalización de genomas | 12 |
| II.6.2.1. IMAGE (Iterative Mapping and Assembly for Gap Elimination) | 12 |
| II.6.2.2. SSPACE (Scaffolding of Pre-Assembled Contigs after Extension) | 13 |
| II.6.2.3 Consed | 13 |
| II.6.3. Herramientas bioinformáticas de genómica comparativa | 14 |
| II.6.3.1. R2Cat (Related Reference Contig Arrangement Tool) | 15 |
| II.6.3.2. Mauve | 15 |

| | PÁGINA |
|--|--------|
| VI.4.2.1.2. Caminata por integración genómica | 30 |
| VI.4.2.2. Extracción y purificación de bandas | 31 |
| VI.4.2.3. Secuenciación mediante el método de Sanger | 32 |
| VI.5. Ensamblado final mediante el programa Consed | 32 |
| VI.6. Estudio de sintenia del genoma de <i>Sinorhizobium fredii</i> GR64 | 32 |
| VII. RESULTADOS | |
| VII.1. Determinación del orden e identidad de los scaffolds mediante R2cat | 33 |
| VII.2. Finalización bioinformática de gaps | 41 |
| VII.3. Estrategias para la obtención de las secuencias de los Gaps remanentes | |
| VII.3.1. Obtención de Cepas | 42 |
| VII.3.2. Oligonucleótidos | 44 |
| VII.3.3. Obtención de la secuencia completa del cromosoma | 46 |
| VII.3.4. Obtención de la secuencia completa del megaplásmido pSfrGR64c | 47 |
| VII.3.5. Obtención de la secuencia completa del plásmido simbiótico pSfrGR64b | |
| VII.3.5.1. Amplificación del gap por PCR | 50 |
| VII.3.5.2. Caminata genómica | 50 |
| VII.3.6. Verificación del ensamble del plásmido pSfrGR64a | 54 |
| VII.4. Obtención de las Lecturas | 55 |
| VII.5. Re-ensamblado de la Secuencias Mediante Phred-Phrap/Consed | 57 |
| VII.6. Estudio de Sintenia del Genoma de <i>S. fredii</i> GR64 | |
| VII.6.1. R2cat | |
| VII.6.1.1. Cromosoma de <i>S. fredii</i> GR64 | 58 |
| VII.6.1.2. Megaplásmido pSfrGR64c | 60 |
| VII.6.1.3. Plásmido simbiótico pSfrGR64b | 61 |
| VII.6.2. Mauve | |
| VII.6.2.1. Cromosoma de <i>S. fredii</i> GR64 | 62 |
| VII.6.2.2. Megaplásmido pSfrGR64c | 65 |

| | PÁGINA |
|--|--------|
| VII.6.2.3. Plásmido simbiótico pSfrGR64b | 67 |
| VIII. DISCUSIÓN | 69 |
| IX. CONCLUSIONES | 75 |
| X. RECOMENDACIONES | 76 |
| XI. LITERATURA CITADA | 77 |

LISTA DE TABLAS

| | | PÁGINA |
|----------|---|--------|
| Tabla I | Cepas y plásmidos utilizados | 22 |
| Tabla II | Oligonucleótidos utilizados en este trabajo | 45 |

LISTA DE FIGURAS

| | | PÁGINA |
|-------------|--|--------|
| Figura 1 | Filogenia de <i>Sinorhizobium fredii</i> GR64 | 33 |
| Figura 2.A. | Colinealidad entre los scaffolds de <i>Sinorhizobium fredii</i> GR64 y cromosomas de bacterias filogenéticamente cercanas (incluye comparaciones con las bacterias: <i>Sinorhizobium fredii</i> HH103. | 34 |
| Figura 2.B. | Colinealidad entre los scaffolds de <i>Sinorhizobium fredii</i> GR64 y cromosomas de bacterias filogenéticamente cercanas (incluye comparaciones con las bacterias: <i>S. fredii</i> USDA257 y <i>Rhizobium</i> sp. NGR234) | 35 |
| Figura 2.C. | Colinealidad entre los scaffolds de <i>Sinorhizobium fredii</i> GR64 y cromosomas de bacterias filogenéticamente cercanas (incluye comparaciones con las bacterias: <i>S. medicae</i> WSM419 y <i>S. meliloti</i> 1021) | 36 |
| Figura 2.D. | Colinealidad entre los scaffolds de <i>Sinorhizobium fredii</i> GR64 y cromosomas de bacterias filogenéticamente cercanas (incluye comparaciones con las bacterias: <i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM2304 y <i>R. leguminosarum</i> bv. <i>viciae</i> 3841) | 37 |
| Figura 2.E. | Colinealidad entre los scaffolds de <i>Sinorhizobium fredii</i> GR64 y cromosomas de bacterias filogenéticamente cercanas (incluye comparaciones con las bacterias: <i>Rhizobium etli</i> CFN42) | 38 |
| Figura 3 | Colinealidad entre los scaffolds de <i>Sinorhizobium fredii</i> GR64 y megaplásmidos de bacterias filogenéticamente cercanas | 39 |
| Figura 4 | Colinealidad entre los scaffolds de <i>Sinorhizobium fredii</i> GR64 y genes <i>nodABC</i> de <i>Rhizobium</i> sp. (No. de acceso X73362.1) | 40 |
| Figura 5 | Colinealidad entre los scaffolds de <i>Sinorhizobium fredii</i> GR64 y el plásmido transferible pSfGR64a (Cervantes et al., 2011) | 41 |
| Figura 6 | Geles de lisis <i>in situ</i> donde se muestran los perfiles plasmídicos de las cepas obtenidas | 43 |

| | PÁGINA | |
|-----------|--|----|
| Figura 7 | Verificación de la transferencia del pSfrGR64c::pK18mob:: <i>oriT</i> al fondo genómico de GMI9023 mediante amplificación del fragmento clonado en el vector pK18mob | 44 |
| Figura 8 | Verificación de la posición de las regiones 16S ribosomal en el cromosoma y amplificación del gap 1 | 46 |
| Figura 9 | Estructura del megaplásmido pSfrGR64c predicha por el programa bioinformático SSPACE | 47 |
| Figura 10 | Estructura del megaplásmido pSfrGR64c confirmada mediante PCRs combinatorias | 48 |
| Figura 11 | Estructura del gap 8 del megaplásmido pSfrGR64c predicha por el programa bioinformático SSPACE | 48 |
| Figura 12 | Amplificación de gaps del megaplásmido pSfrGR64c | 49 |
| Figura 13 | Amplificación del gap del plásmido simbiótico pSfrGR64b | 50 |
| Figura 14 | Ligación del vector pK18mob con el fragmento amplificado del plásmido simbiótico, fragpSym::pK18mob | 51 |
| Figura 15 | Integración del pK18mob:: fragpSym en el plásmido simbiótico pSfrGR64b, GR64 pSfrGR64c::pK18mob::fragpSym | 52 |
| Figura 16 | Geles obtenidos en el procedimiento de caminata genómica del plásmidos simbiótico pSfrGR64b | 52 |
| Figura 17 | Religaciones pK18mob::fragpSym-XEnzima de restricción | 53 |
| Figura 18 | Oligonucleótidos para verificar el ensamble del pSfrGR64a | 54 |
| Figura 19 | Verificación del ensamble del plásmido transferible pSfrGR64a | 55 |
| Figura 20 | Gaps amplificados | 56 |
| Figura 21 | Replicones de <i>Sinorhizobium fredii</i> GR64 y GC skew | 57 |
| Figura 22 | Análisis de sintenia entre el cromosoma de <i>Sinorhizobium fredii</i> GR64 y cromosomas de bacterias filogenéticamente cercanas con el programa R2cat | 59 |
| Figura 23 | Análisis de sintenia entre el megaplásmido de <i>Sinorhizobium</i> <i>fredii</i> GR64 pSfrGR64c y replicones de bacterias filogenéticamente cercanas con el programa R2cat | 60 |

| | PÁGINA | |
|--------------|---|----|
| Figura 24 | Análisis de sintenia entre el plásmido simbiótico pSfrGR64b de <i>Sinorhizobium fredii</i> GR64 y plásmidos simbióticos de bacterias filogenéticamente cercanas con el programa R2cat | 62 |
| Figura 25.A. | Comparaciones entre el cromosoma de <i>Sinorhizobium fredii</i> GR64 y cromosomas de bacterias filogenéticamente cercanas con el programa Mauve (incluye comparaciones con las bacterias: <i>S. fredii</i> HH103, <i>Rhizobium</i> sp. NGR234, <i>S. fredii</i> USDA 257 y <i>S. medicae</i> WSM 419) | 63 |
| Figura 25.B. | Comparaciones entre el cromosoma de <i>Sinorhizobium fredii</i> GR64 y cromosomas de bacterias filogenéticamente cercanas con el programa Mauve (incluye comparaciones con las bacterias: <i>S. meliloti</i> 1021, <i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM2304, <i>R. leguminosarum</i> bv. <i>viciae</i> 3841 y <i>R. etli</i> CFN42) | 64 |
| Figura 26 | Comparaciones entre el megaplásmido pSfrGR64c de <i>Sinorhizobium fredii</i> GR64 y replicones de bacterias filogenéticamente cercanas con el programa Mauve | 66 |
| Figura 27.A. | Comparaciones entre el plásmido simbiótico pSfrGR64b y plásmidos simbióticos de bacterias filogenéticamente cercanas con el programa Mauve (incluye comparaciones con los plásmidos simbióticos: pNGR234a de <i>Rhizobium</i> sp. NGR234, pRetCFN42d de <i>Rhizobium etli</i> CFN42 y el pSym A de <i>S. meliloti</i> 1021) | 67 |
| Figura 27.B. | Comparaciones entre el plásmido simbiótico pSfrGR64b y plásmidos simbióticos de bacterias filogenéticamente cercanas con el programa Mauve (incluye comparaciones con los plásmidos simbióticos: pSMEDO3 de <i>S. medicae</i> WSM419, pUSDA257 de <i>S.fredii</i> USDA257 y pSMEDO2 de <i>S. medicae</i> WSM419) | 68 |

| | PÁGINA |
|---|--------|
| Figura 28 | 74 |
| Evento de recombinación intragenómica (escisión o integración de una región) en el cromosoma de <i>Sinorhizobium fredii</i> . | |

RESUMEN

La bacteria *Sinorhizobium fredii* generalmente establece simbiosis con la leguminosa *Glycine max*, pero la cepa GR64, interactúa con *Phaseolus vulgaris*, mas no con *G. max*. Uno los plásmidos de la GR64, el pSfrGR64a, es una molécula quimérica originada a partir de la transferencia horizontal de un replicón de *Rhizobium etli* y eventos de recombinación entre dicho replicón y el cromosoma de *S. fredii*. Debido a las peculiares características de GR64, su genoma completo fue secuenciado, ya que los datos obtenidos a partir de la secuenciación de genomas clarifican los patrones en el flujo genes. En el borrador de la secuencia del genoma completo de GR64 se obtuvieron 216 contigs (>500 pb), de los cuales 146 fueron localizados en 24 scaffolds. Los contigs, presentan huecos o gaps entre ellos, y aunque los programas bioinformáticos disminuyen la cantidad de gaps, no son capaces de dilucidar la secuencia completa del genoma, por lo que es necesario determinarla mediante secuenciación por el método de Sanger. El objetivo de este trabajo fue determinar la secuencia completa de GR64, obteniendo las secuencias de los gaps mediante estrategias bioinformáticas y de biología molecular, con el propósito de realizar estudios de genómica comparativa. Dichos estudios revelaron que el genoma de *S. fredii* GR64 exhibe las características de los genomas de rizobios, con un cromosoma conservado que contiene bloques de genes sinténicos al compararlos con otras cepas de la familia *Rhizobiaceae*, y plásmidos que presentan sintenia sólo con algunos genes; además, sugieren que los plásmidos pSfrGR64a y pSfrGR64b tienen un origen reciente, reforzando la teoría de que ambos fueron adquiridos mediante un evento de transferencia horizontal. El megaplásmido pSfrGR64c presenta características que lo distinguen de los otros plásmidos, que pudieran catalogarlo como un cromosoma secundario.

I. INTRODUCCIÓN

La fijación biológica del nitrógeno constituye la principal entrada de formas reducidas de nitrógeno, desde la atmósfera hacia sistemas biológicos (Romero y Brom, 2004), y la capacidad para fijar nitrógeno, la diazotrofia, es exclusiva de los (Raymond et al., 2004). Los organismos diazótrofos se encuentran tanto en vida libre como en simbiosis (Lloret y Martínez-Romero, 2005). Los diazótrofos simbióticos, aunque pueden crecer en condiciones de vida libre, generalmente fijan nitrógeno sólo cuando establecen una interacción simbiótica con plantas, habitualmente con aquellas de la familia *Fabaceae*. Esta simbiosis ocurre principalmente con bacterias de la familia *Rhizobiaceae* (Phylum Proteobacteria) (Romero y Brom, 2004), y generalmente dicha interacción simbiótica es específica.

La simbiosis conduce al desarrollo de un órgano especial en la planta llamado nódulo, mediante la producción de compuestos por parte de la bacteria, llamados factores de nodulación, los cuales juegan un papel importante en la determinación de la especificidad de huésped. Los factores de nodulación son sintetizados a partir del producto de los llamados genes de nodulación o genes *nod*. En los rizobios la mayoría de los genes *nod* y *nif* (genes de fijación de nitrógeno) están localizados en un plásmido, llamado plásmido simbiótico (pSym). Sólo en algunas especies los genes simbióticos *nod* están localizados en el cromosoma (Sullivan et al., 1995).

La transferencia horizontal de genes es la transferencia de información genética entre células bacterianas y es diferente a la transferencia de ADN de la célula madre a la célula hija. Es un factor integral en la generación de variación genética (Kelly et al., 2009), contribuyendo a la diseminación y diversificación de la información genética, particularmente la que está localizada en plásmidos (Cervantes et al., 2011). Debido a que en numerosas especies de rizobios los genes de simbiosis se encuentran en plásmidos, se dice que estos genes forman parte del paquete genético flexible de los rizobios (Lloret y Martínez-Romero, 2005).

Rhizobium etli CFN42 es una bacteria Gram-negativa de la familia *Rhizobiaceae* capaz de establecer simbiosis y fijar nitrógeno con el frijol, *Phaseolus vulgaris*. Tanto el frijol, como su simbionte *R. etli* son originarios de Mesoamérica y a principios siglo XVI el frijol fue introducido en Europa y otras partes del mundo. Estudios realizados en suelos de Granada,

España, analizaron la población bacteriana aislada de nódulos de *Phaseolus vulgaris*, donde encontraron una cepa (GR64) de la especie *Sinorhizobium fredii* capaz de nodular frijol y no soya. Usualmente *S. fredii* tiene la capacidad de nodular soya y no frijol, lo cual sugirió que la capacidad de nodular frijol se debía a un evento de transferencia horizontal de genes, desde *R. etli*. Para probar esta hipótesis, uno de los replicones de *S. fredii* GR64, el plásmido pSfrGR64a, fue secuenciado. Este estudio reveló que el pSfrGR64a fue originado a partir de la transferencia horizontal de un replicón de la bacteria *R. etli* a una cepa de *S. fredii*, seguido de eventos de recombinación entre los plásmidos de *R. etli* y el cromosoma de *S. fredii* (Cervantes et.al., 2011).

Debido a las peculiares características de *S. fredii* GR64, la interacción simbiótica con *P. vulgaris* y la estructura genómica del plásmido pSfrGR64a, se decidió secuenciar su genoma, ya que los datos que se obtienen a partir la secuenciación de genomas permiten verificar los patrones en el flujo genes (Bentley, 2009). Para obtener la secuencia completa de un genoma es necesario ensamblar las millones de lecturas que se obtienen. El armado de estas lecturas se limita a un set de secuencias contiguas o contigs, las cuales presentan huecos o gaps entre ellas. Esto puede ser atribuido a la presencia de elementos repetitivos en el genoma, y a los errores de los métodos de secuenciación (Boetzer et al., 2011). Existen programas bioinformáticos que disminuyen la cantidad de gaps de los ensambles, sin embargo, usualmente no son capaces de dilucidar la secuencia completa del genoma. Por este motivo, es necesario determinar las secuencias de los gaps, que los programas bioinformáticos no pueden obtener, mediante secuenciación por el método de Sanger.

El draft de las secuencias del genoma completo de *S. fredii* GR64 fue obtenido mediante secuenciación por Illumina y pirosecuenciación Roche 454, realizando un ensamble híbrido de las lecturas. Este ensamble resultó en 216 contigs grandes (>500 pb) con un tamaño total de 6,959,392 bases. Un total de 146 de estos contigs fueron localizados en 24 scaffolds (Torres-Tejerizo et al., 2012).

El objetivo de este trabajo fue determinar la secuencia completa del genoma de la bacteria *S. fredii* GR64 mediante estrategias bioinformáticas y de biología molecular, con el fin de realizar estudios de genómica comparativa, determinando la sintenia de los replicones de la bacteria *Sinorhizobium fredii* cepa GR64 con respecto a otras especies de bacterias de la Familia *Rhizobiaceae*.

II. ANTECEDENTES

II.1. Fijación Biológica del Nitrógeno y Nodulación

El nitrógeno es un elemento esencial para la vida y aunque es abundante en la atmósfera como nitrógeno molecular, sólo es útil para los organismos cuando es convertido a formas reducidas, como el amonio, en las cuales está biológicamente disponible. A este proceso de conversión se le denomina fijación del nitrógeno y a pesar de que existen procesos abióticos que fijan nitrógeno, la fijación biológica del nitrógeno constituye la principal entrada de formas reducidas de nitrógeno al ecosistema (Romero y Brom, 2004). La capacidad para fijar nitrógeno, la diazotrofia, es exclusiva de los procariontes (Raymond et al., 2004) y radica en el complejo enzimático de la nitrogenasa (Lloret y Martínez Romero, 2005).

Los diazótrofos están distribuidos en grupos parafiléticos, incluyendo aerobios, anaerobios, autótrofos, heterótrofos y metanótrofos y se encuentran tanto en vida libre como en simbiosis (Lloret y Martínez Romero, 2005). Los diazótrofos simbióticos, aunque pueden crecer en condiciones de vida libre, generalmente fijan nitrógeno sólo cuando establecen una interacción simbiótica con plantas, habitualmente con aquellas de la familia *Fabaceae*. Esta simbiosis ocurre principalmente con bacterias del orden *Rhizobiales* (Phylum Proteobacteria). Las interacciones simbióticas entre leguminosas y bacterias fijadoras de nitrógeno generalmente es específica, sin embargo, ha sido demostrado que en algunos casos puede ser una interacción más flexible (Romero y Brom, 2004).

La simbiosis entre rizobios y leguminosas conduce al desarrollo de un órgano especial llamado nódulo, de diversa morfología, fisiología y desarrollo. Los nódulos permiten al microsimbionte fijar nitrógeno atmosférico, el cual comparte con la planta a cambio de fotosintatos (Pappas y Cevallos, 2011). Las bacterias inducen la formación de nódulos después de una compleja comunicación molecular, donde es esencial que las mismas produzcan compuestos llamados factores de nodulación (Jones et al., 2007). Los factores de nodulación se componen de una estructura básica y modificaciones que son específicas en cada especie o cepa bacteriana, las cuales juegan un papel importante en la determinación de la especificidad de huésped. Estos compuestos son sintetizados a partir del producto de los llamados genes de

nodulación o genes *nod* (Romero y Brom, 2004). La transcripción de los genes *nod* se activa en respuesta a moléculas señal presentes en los exudados de la raíz, iniciando la síntesis y secreción de los factores de nodulación, los cuales se encuentran formados por oligómeros de N-acetil-D-glucosamina (Perret et al., 2000). Una vez dentro de las células de la raíz, las bacterias responden a la baja concentración de oxígeno transcribiendo los genes de fijación de nitrógeno o genes *nif* (*nitrogen fixation*), que codifican para el complejo de la nitrogenasa.

II.2. Genes de Simbiosis: Islas y Plásmidos Simbióticos

En los rizobios la mayoría de los genes implicados en la nodulación y en la fijación de nitrógeno (*nod* y *nif*) están localizados en plásmidos, llamado plásmidos simbióticos (pSym). La presencia de un pSym es común en especies de rápido crecimiento de los géneros *Rhizobium* y *Sinorhizobium*, con un tamaño variable entre las especies (180 kb-1.35 Mb). En aquellas especies de lento y mediano crecimiento, de los géneros *Bradyrhizobium*, *Azorhizobium* y *Mesorhizobium*, generalmente los genes simbióticos están localizados en el cromosoma. Hay especies, como *Mesorhizobium loti*, en las que los genes simbióticos están localizados en elementos conjugativos integrables, denominados islas simbióticas (Sullivan y Ronson, 1998).

Tanto los plásmidos como las islas simbióticas han sido descritos como elementos genéticos móviles, que contienen secuencias de ADN que codifican proteínas importantes para la mediación de la transferencia horizontal de genes entre genomas y desempeñan un papel integral en la evolución de genomas bacterianos (Frost et al., 2005).

II.3. Rizobios

Las bacterias fijadoras de nitrógeno que son capaces de inducir el desarrollo de nódulos son llamadas bacterias rizobias o rizobios y se distribuyen en familias de la Clase α -Proteobacteria y α -Proteobacteria. Están incluidas en ocho familias, siete en la Clase α -Proteobacteria, dentro del Orden *Rhizobiales*: *Rhizobiaceae*, *Methylobacteriaceae*, *Xantobacteriaceae*,

Brucellaceae, *Hiphomicrobiaceae*, *Bradyrhizobiaceae* y *Phyllobacteriaceae*; y también en la Clase β -Proteobacteria, dentro del Orden *Burkholderiales*: la familia *Burkholderiaceae*, con dos géneros, *Burkholderia* y *Cupriavidus* (Sprent, 2008).

De estas, la familia *Rhizobiaceae* contiene el mayor número de géneros que nodulan leguminosas, incluyendo a *Rhizobium*, *Allorhizobium* y *Sinorhizobium*, las cuales son habitantes comunes del suelo y la rizósfera. El rango de hospederos de las bacterias rizobias varía ampliamente. Algunas especies nodulan una o pocas leguminosas, mientras que otras nodulan especies diferentes. Así también, una leguminosa puede ser nodulada por una especie bacteriana específica, mientras otras pueden ser noduladas por varias (Pappas y Cevallos, 2011).

II.3.1. Evolución y filogenia de los rizobios

La duplicación de genes antes de la separación de dos linajes ha sido utilizada para ubicar su tiempo de divergencia (Iwabe et al., 1989). El gen ancestral de la glutamino sintasa (GS) se duplicó antes de la separación de procariontes y eucariontes (Pesole et al., 1995), y la comparación entre los dos genes parálogos GSI y GSII ha permitido utilizarlos como relojes moleculares para calibrar el tiempo de divergencia entre los linajes que los contienen.

Con esa pareja de genes se ha estimado el tiempo de divergencia de los linajes de rizobios de α -Proteobacteria, Clase *Rhizobiales*. La GSII ubica como linajes más recientes a *Rhizobium* y *Sinorhizobium* y *Mesorhizobium* junto con *Bradyrhizobium* como los linajes más antiguos. Con la GSI la separación de *Rhizobium-Sinorhizobium-Mesorhizobium* ocurre al mismo tiempo, y *Bradyrhizobium* se ubica como el linaje más antiguo.

Los tiempos calculados a partir de las GSs y de genes ortólogos sinténicos (genes ortólogos compartidos), datan la separación de los linajes de los rizobios más antiguos (*Bradyrhizobium*) y más recientes (*Sinorhizobium*) entre 553 y 203 m.a., mucho tiempo antes de la aparición de las leguminosas hace 70 m.a., incluso de las angiospermas hace más de 150 m.a. (Martin et al., 1989). Ambos miembros de la pareja de simbiosis han coevolucionado hace al menos 70 a 100 m.a., desde que las leguminosas se originaron (Lloret y Martínez-Romero, 2005).

II.3.1.1. Transferencia horizontal de genes:

Papel en la evolución de los genomas de Rizobios

La transferencia horizontal de genes es la transferencia de información genética entre células bacterianas, y es diferente a la transferencia de ADN de la célula madre a la célula hija (Madigan et al., 2004). Es un factor integral en la generación de variación genética y evolución en bacterias, permitiendo una rápida adaptación a los cambios ambientales (Kelly et al., 2009). La THG involucra la movilización de elementos genéticos móviles (plásmidos, ADN extracelular, fagos y transposones conjugativos), y los principales mecanismos son la transformación, la transducción y la conjugación (Raz y Tannenbaum, 2010).

Se considera que la transferencia conjugativa es el mecanismo de transferencia horizontal más relevante que contribuye a la diseminación y diversificación de la información genética, particularmente la localizada en plásmidos (Cervantes et al., 2011). Como se mencionó anteriormente, los genes de simbiosis se encuentran en plásmidos e islas simbióticas, que funcionan como vehículos de transferencia genética que pueden perderse o adquirirse, formando parte de un paquete genético flexible (Lloret y Martínez-Romero, 2005). La selección diferencial de su permanencia lleva a la divergencia de estilos de vida del simbionte (Ochman, 2001). El genoma flexible de los rizobios no necesariamente refleja las relaciones filogenéticas del genoma básico de la bacteria, sino más bien la relación con su hospedero (Lloret y Martínez-Romero, 2005).

II.3.2. Características de los genomas de Rizobios

Se encuentran disponibles 14 genomas completos de rizobios, pertenecientes a α -proteobacterias y uno de una β -proteobacteria. Estas especies tienen genomas grandes, que van desde 5.36 Mpb (*Azorhizobium caudilans* ORS572) a 9.1Mpb (*Bradyrhizobium japonicum* USDA 110). Muchos de los genomas de rizobiales son multipartitos, compuestos por un cromosoma circular y varios plásmidos, algunos de los cuales alcanzan un tamaño de 2 Mpb. El tamaño del genoma, número de plásmidos y el contenido de genes de los rizobiales varía enormemente, incluso entre cepas de la misma especie (Gonzalez et al., 2010).

Los genomas bacterianos se dividen en el genoma core (del inglés núcleo), que incluye los genes presentes en todas las cepas de las especies y que generalmente son requeridos en el metabolismo básico (housekeeping genes) y el genoma accesorio, que está compuesto por los genes presentes sólo en algunas cepas de una especie, que generalmente son dispensables y tienden a estar presentes en plásmidos adquiridos mediante transferencia horizontal. La suma del genoma *core* y el genoma accesorio constituyen el pangenoma de una especie. Los plásmidos simbióticos son un importante componente del genoma accesorio de los rizobios, dispensables bajo ciertas circunstancias, pero esenciales para la simbiosis. Los plásmidos de los rizobios pertenecientes a las α -proteobacterias, con algunas excepciones, están bajo el control del operón de replicación/partición *repABC*. Estos operones no son exclusivos de los rizobios, existen en un gran rango de α -proteobacterias que viven en nichos ecológicos diferentes. Otra característica de los genomas de rizobios es la abundancia de secuencias de inserción, agrupados en regiones o compartimentos genómicos específicos y la presencia de secuencias largas y repetidas dispersas en plásmidos y cromosomas (Pappas y Cevallos, 2011).

II.4. *Rhizobium etli* CFN42

Rhizobium etli CFN42 es una bacteria Gram-negativa capaz de establecer simbiosis y fijar nitrógeno con el frijol, *Phaseolus vulgaris*. Su genoma está completamente secuenciado, cuenta con un cromosoma y seis plásmidos: pRetCFN42a, pRetCFN42b, pRetCFN42c, pRetCFN42d, pRetCFN42e y pRetCFN42f. El plásmido pRetCFN42a es un plásmido autotransferible a alta frecuencia (10^{-2} transconjugantes por célula donadora) y el plásmido pRetCFN42d es un plásmido simbiótico que se transfiere a baja frecuencia (10^{-6} transconjugantes por célula donadora) cuando se cointegra por el pRetCFN42a (Brom et al., 2000). Tanto el frijol, como su simbiote *R. etli* son originarios de Mesoamérica. A principios del siglo XVI el frijol fue introducido en Europa y otras partes del mundo (Cervantes et al., 2011).

II.5. *Sinorhizobium fredii* GR64

Estudios realizados en suelos de Granada, España, donde se analizó la población bacteriana aislada de nódulos de *Phaseolus vulgaris*, mostraron cepas de cinco diferentes especies de *Rhizobiales*: *Rhizobium etli*, *R. gallicum*, *R. giardinii*, *R. leguminosarum* y *Sinorhizobium fredii* (Herrera-Cervera et al., 1999). De estas especies, *S. fredii* comúnmente tiene la capacidad de nodular soya y no frijol, lo cual sugirió que la capacidad de nodular frijol en *S. fredii* GR64 se debe a un evento de transferencia horizontal de genes, desde *R. etli* (Cervantes et al., 2011). Además, GR64 no tiene la capacidad de nodular soya.

El genoma de *S. fredii* comprende un cromosoma y tres plásmidos: pSfGR64a, pSfGR64b y pSfGR64c. El plásmido pSfGR64a es autotransferible a alta frecuencia y puede movilizar al plásmido pSfGR64b, que es el plásmido simbiótico y el plásmido pSfGR64c es un megaplásmido del cual no se han estudiado sus propiedades conjugativas (Cervantes et al., 2011).

II.5.1. Secuenciación del plásmido pSfrGR64a de *Sinorhizobium fredii* GR64

El plásmido pSfrGR64a fue secuenciado previamente mediante la purificación del mismo por gradientes de CsCl y la realización de una biblioteca genómica. El análisis de dicho plásmido, encontró 166 marcos de lectura abiertos (ORFs) y que está compuesto por tres grandes segmentos de origen evolutivo diferente. El primer segmento presenta 38 ORFs, los cuales son altamente similares a genes del cromosoma de *Rhizobium* sp. NGR234. El segundo segmento contiene 51 ORFs que son altamente similares a genes encontrados en el plásmido pRetCFN42d de *R. etli* CFN42, incluyendo genes de replicación, más no los de simbiosis. El tercer segmento contiene 36 ORFs similares a genes localizados en el plásmido pSfrGR42a de *R. etli* CFN42, 20 de ellos involucrados en la transferencia conjugativa (Cervantes et al., 2011).

Esto sugiere que el plásmido pSfrGR64a se originó a partir de la transferencia de un replicón de *R. etli* a una cepa de *S. fredii*, seguido de eventos de recombinación entre los plásmidos de *R. etli* y el cromosoma de *S. fredii* (Cervantes et al., 2011).

II.5.2. Pirosecuenciación del genoma de *Sinorhizobium fredii* GR64

El genoma completo de *S. fredii* GR64 fue determinado mediante secuenciación por Illumina y por pirosecuenciación Roche 454. Se secuenciaron dos bibliotecas mediante Illumina, una de 2 kb, obteniendo 8,163,266 lecturas de 40 Mb y otra de 350 pb, obteniendo 8,555,556 lecturas de 760 Mb. Las lecturas de Illumina fueron ensambladas mediante el método SOAP *de novo*, resultando en 188 contigs (Torres Tejerizo et al., 2012).

También se secuenció una biblioteca de 3 kb mediante pirosecuenciación 454, obteniendo 467, 525 lecturas de alrededor de 89.96 Mb de la información total secuenciada. Se realizó un ensamble híbrido de las lecturas del pirosecuenciación 454 y secuenciación Illumina utilizando el gsAssembler implementado en Newbler Software (Roche), el cual resultó en 216 contigs grandes (>500 pb) con un tamaño total de 6,959,392 bases. Un total de 146 de estos contigs fueron localizados en 24 scaffolds (Torres Tejerizo et al., 2012).

II.6. Secuenciación de Genomas

La información biológica presente en el DNA se encuentra codificada en un orden de nucleótidos preciso dentro de la molécula de DNA. La secuenciación de genomas consiste en determinar el ordenamiento de nucleótidos específico de un genoma particular (Rodríguez-Martínez, 2004). El genoma secuenciado de una especie incrementa el valor de muchos estudios biológicos, por ejemplo, al comparar el papel de los organismos en el ecosistema en diferentes niveles, evolución microbiológica, metabolismo y ecología (DeLong y Karl, 2005). Los genomas completamente secuenciados son considerados superiores a fragmentos secuenciados del genoma porque proveen la única referencia precisa para interpretar metagenomas y transcriptomas (Kisand y Lettieri, 2013).

II.6.1 Tecnologías de secuenciación

La técnica para secuenciar ADN más ampliamente utilizada durante los últimos años es conocida como el método de terminadores en cadena, y fue desarrollada por Frederick Sanger y colaboradores en 1977 (Sanger et al., 1977). La técnica se basa en el empleo de la ADN polimerasa, la cual extiende una cadena de ADN polimerizando nucleótidos en el extremo 3'. Para esto se necesita esencialmente el ADN que sirve de molde, un oligonucleótido que sirve como iniciador de la reacción y los cuatro nucleótidos. La clave de la técnica consiste en adicionar nucleótidos modificados que actúan como terminadores de la cadena, por lo cual al final de la reacción se obtienen una mezcla de moléculas de diferente tamaño, a partir de las cuales se puede deducir la posición en que termina la polimerización, mediante electroforesis en geles de poliacrilamida. En el método original, se llevaban a cabo cuatro reacciones independientes, para cada uno de los nucleótidos que componen el ADN, hasta que en 1986 aparecieron los terminadores marcados con colorantes fluorescentes, lo cual permite que se lleve a cabo una sola reacción. Otra de las modificaciones que ha sufrido la técnica es el desarrollo de secuenciadores basados en electroforesis capilar para la separación de las reacciones, que permiten el procesamiento simultáneo de un número mucho mayor de reacciones (Rodríguez-Martínez, 2004).

Sin embargo, de 1000 bases unas 800 como máximo son de alta calidad, para obtener una molécula mayor a 800 bases es necesario emplear iniciadores separados entre sí y sólo se puede realizar una reacción de secuenciación una vez que se llevó a cabo la anterior, lo cual lo hace un proceso lento y no automatizable. Debido a estas limitaciones, se han desarrollado nuevos métodos de secuenciación considerados “tecnología de nueva generación” (Metzker, 2010).

Entre estas tecnologías, se encuentran comercialmente disponibles la secuenciación Roche/454, Illumina/Solexa, Life/APG y Helicos-Bioscience, las cuales incluyen métodos similares entre ellas, como preparación del molde o plantilla (ADN recombinante constituido de una región conocida y la secuencia blanco), secuenciación y ensamblado y análisis de datos. La combinación única de protocolos específicos distinguen a una tecnología de las otras (Metzker, 2010).

II.6.1.1. Terminación cíclica reversible: Illumina/Solexa

Como su nombre lo indica, la secuenciación de terminación cíclica reversible utiliza terminadores reversibles en un método cíclico en el cual se incorporan nucleótidos y se obtienen señales fluorescentes (Wheatcroft y Watson). La clave de la TCR es el terminador reversible, el cual está marcado con material fluorescente, al que luego de la detección de la señal se le remueve el grupo protector; esto permite la adición de un nuevo ciclo de bases (Metzker, 2010).

Actualmente el Analizador de Genomas Illumina/Solexa es la tecnología de secuenciación mediante TCR que domina el mercado. Utiliza la TCR acoplada a un método de amplificación clonal del ADN molde. Las señales son detectadas mediante Fluorescencia total de reflexión interna o TIRF (Total Internal Reflection Fluorescence), utilizando dos láseres.

II.6.1.2. Pirosecuenciamiento: Roche 454

La pirosecuenciación es un método bioluminiscente que no utiliza electroforesis, el cual mide la liberación de pirofosfato inorgánico al convertirlo proporcionalmente en luz visible utilizando una serie de reacciones enzimáticas (Ronaghi et al., 1998). A diferencia de otros métodos que utilizan nucleótidos modificados para terminar la síntesis de ADN, la pirosecuenciación manipula la ADN polimerasa por la simple adición de dNTP's en cantidades limitadas. Mediante la incorporación de dNTP's complementarios, la ADN polimerasa extiende el oligonucleótido y se detiene cuando la concentración de estos baja. La síntesis de ADN es reiniciada siguiendo la adición de dNTP's en un ciclo de dispensación. El orden y la intensidad de los picos de luz son registrados revelando la secuencia de ADN. La primera plataforma que integra la pirosecuenciación es el instrumento comercial Roche 454. Utiliza ADN molde obtenido de emPCR (PCR de emulsificación) (Metzker, 2010).

II.6.2. Programas Bioinformáticos de Ensamble y Finalización de Genomas

Una vez obtenidas las secuencias del genoma fragmentado o lecturas, estos fragmentos se ensamblan mediante programas bioinformáticos hasta reconstruir la molécula original, proceso denominado secuenciación *de novo*. Las lecturas deben aportar una cantidad de secuencia equivalente a varias veces la molécula a reconstruir, ya que gracias a la redundancia se pueden identificar los errores que se producen en las secuencias individuales.

Siempre que el genoma no contenga secuencias repetitivas se puede llevar a cabo un fraccionamiento al azar del genoma y los programas requeridos para su ensamble son relativamente sencillos. Los problemas prácticos de esta técnica provienen de las secuencias repetidas presentes en los genomas y de las desviaciones del azar que se producen durante la preparación de los fragmentos del genoma (Rodríguez-Martínez, 2004).

La reducción de los costos de la secuenciación de genomas está comprometida a que las lecturas obtenidas de las reacciones de secuenciación son típicamente cortas y contienen errores de secuenciación. Esto conlleva a la necesidad de programas computacionales más eficaces que sean capaces de ensamblar la molécula original con el menor error posible.

Por lo tanto, el rendimiento de los programas de ensamblaje *de novo* depende de la longitud de la secuencia, la cantidad de la cobertura de secuencia (equivalentes de genoma, o la cobertura de repeticiones), tamaño de los fragmentos de las plantillas que se secuenciaron y los tipos de errores de secuencia.

Un enfoque para reducir estos problemas es combinar la información de la secuencia obtenida con diferentes tecnologías (Tsai et al., 2010). Existen pocos programas capaces de ordenar, o determinar los scaffolds (andamios), de las secuencias contiguas (contigs) pre-ensambladas (Boetzer et al., 2011), a continuación se detallan algunos de ellos.

II.6.2.1. IMAGE (Iterative Mapping and Assembly for Gap Elimination)

IMAGE (por sus siglas en inglés, Iterative Mapping and Assembly for Gap Elimination) como su nombre lo dice, es un programa bioinformático de ensamble y mapeado de secuencias mediante cálculos iterativos, que tiene como propósito eliminar gaps. Este programa es utilizado para elevar la calidad de los borradores de secuenciación de genomas, sin

intervención manual, mediante ensamblajes locales de lecturas (búsqueda de similaridad local entre secuencias) de las regiones gap. El enfoque de IMAGE es utilizar el gran número de secuencias que un analizador de Illumina Genome produce y las lecturas que corresponden a lagunas o zonas dudosas se identifican y se vuelven a ensamblar localmente antes de incorporarse de nuevo en el ensamble final. Una ventaja de un ensamblaje local en lugar de uno *de novo* es que el número de lecturas utilizado es sólo una fracción del total de lecturas disponibles. Esto reduce la complejidad de las regiones a ser ensambladas, así como el tiempo y la memoria de computación requerida (Tsai et al., 2010).

II.6.2.2. SSPACE (Scaffolding of Pre-Assembled Contigs after Extension)

SSPACE (por sus siglas en inglés, Scaffolding of Pre-Assembled Contigs after Extension) es un programa para ordenar secuencias pre-ensambladas (dilucidar scaffolds) producidas por cualquier otro programa o herramienta de ensamble de genomas. La entrada en el programa son archivos de lecturas pareadas de Illumina, un grupo de secuencias contiguas (contigs). El programa tiene la facilidad de que se puede especificar si las lecturas pareadas serán utilizadas para determinar scaffolds (orden de secuencias) o también para extensión de los contigs. SSPACE reduce el número de contigs y el número de gaps del ensamble (Boetzer et al., 2011).

II.6.2.3 Consed

Consed es un programa de procesamiento de datos de secuencias que es utilizado para la edición y ensamble de secuencias. Este programa fue desarrollado para trabajar en conjunto con otros programas como el programa base-calling Phred, desarrollado por Ewing *et. al.* en 1998, el ensamblador Phrap, desarrollado por Green y el visualizador de ensambles de alto nivel Phrapview. Estos programas se caracterizan por tener un criterio muy objetivo para medir la precisión de las secuencias y ensambles. En particular, Phred utiliza parámetros para producir la probabilidad de error asociada a cada base de lectura y Phrap, utiliza esta probabilidad junto con los alineamientos de las lecturas para adjuntar una probabilidad de

error a cada base a la secuencia inferida o secuencia consenso. Esto incrementa la precisión del ensamblado y de la secuencia consenso; permitiendo identificar regiones que tienen una alta contribución de errores esperados, obtener datos adicionales, o editar dichas regiones para reducir el número de errores esperados en la secuencia consenso. El propósito de Consed es implementar la estrategia de Phred-Phrap. Consed utiliza la probabilidad de error como principal herramienta para guiar el proceso de edición de secuencias. Consed es una herramienta útil para la edición final de un ensamblado de secuencia, ya que produce una secuencia final que posee un número de errores esperados debajo de un nivel predeterminado (Gordon et al., 1998).

II.6.3. Herramientas bioinformáticas de genómica comparativa

Mediante la genómica comparativa es posible estudiar semejanzas y diferencias entre genomas de diferentes organismos. La información proporcionada por secuencias completas permite estudiar procesos del funcionamiento y procesos evolutivos que actúan sobre los genomas. Existen herramientas que permiten realizar comparaciones entre secuencias, predecir genes, dilucidar la estructura de los genomas, dilucidar filogenia y clasificación y una gran cantidad de estudios. Una de las herramientas bioinformáticas de genómica comparativa más sencillas y utilizadas es el BLAST (Basic Local Alignment Search Tool). Existen muchas herramientas bioinformáticas de genómica comparativa y son utilizadas de acuerdo al objetivo del estudio a realizar (Claverie y Notredame, 2007).

II.6.3.1. R2Cat (Related Reference Contig Arrangement Tool)

El program r2cat (por sus siglas en inglés, Related Reference Contig Arrangement Tool) es una herramienta que alinea genomas o secuencias de ADN de acuerdo a su sintenia. La sintenia es la co-localización física de un *loci* genético en el genoma de una misma o varias especies y se refiere como la conservación del orden de bloques entre dos sets de cromosomas o genomas que son comparados el uno con el otro. Esta herramienta es de gran utilidad para cerrar genomas, ya que permite un rápido alineamiento de un conjunto de contigs con un genoma relacionado, ordena los contigs de acuerdo a las regiones pareadas y exhibe el resultado en un gráfico de sintenia interactivo. Sin embargo, la alineación no está restringida a contigs, ya que el programa también puede ser utilizado para visualizar la sintenia de dos genomas completos, permitiendo realizar estudios de genómica comparativa (Husemann y Stoye, 2010).

II.6.3.2. Mauve

Mauve es un programa de alineación de genomas que integra el análisis de eventos evolutivos a gran escala con una alineación múltiple de secuencias tradicional. Mediante la integración de estos pasos de análisis, Mauve ofrece facilidad de uso y sensibilidad frente a otros programas cuando se comparan genomas con una reordenación significativa. Al igual que otros programas de alineación, Mauve utiliza el anclaje (anchoring) como una heurística para acelerar la alineación. A diferencia de otros sistemas de alineación múltiple de genomas, el método de selección de anclaje de Mauve relaja el supuesto de que los genomas en estudio son colineales. En su lugar, identifica y alinea las regiones de colinealidad locales llamados bloques localmente colineales (Locally collinear blocks, LCBs). Cada bloque localmente colineal es una región homóloga de la secuencia compartida por dos o más de los genomas en estudio y no contiene ningún reordenamiento de secuencia homóloga (Darling et al., 2004).

II.7. Consideraciones sobre la secuenciación de genomas completos

Al ensamblar un genoma, la tarea es empalmar millones de lecturas de secuencias cortas. El armado de estas lecturas se limita a un set de secuencias contiguas o contigs, aunque en teoría el número total de secuencias puede ser más alto. Esto puede ser atribuido a la presencia de elementos repetitivos en el genoma y a los errores de los métodos de secuenciación (Boetzer et al., 2011). Como se mencionó anteriormente, existen programas bioinformáticos que disminuyen la cantidad de huecos o gaps de los ensamblajes, sin embargo, no son capaces de dilucidar la secuencia completa del genoma. Por este motivo, es necesario realizar secuenciamiento por el método Sanger para determinar las secuencias de los gaps que los programas bioinformáticos no pueden obtener. Esto se logra mediante varios métodos moleculares, tales como la reacción en cadena de la polimerasa (PCR) y la caminata genómica (DNA walking) (Rodríguez-Martínez, 2004).

III. JUSTIFICACIÓN

La bacteria *Sinorhizobium fredii* generalmente establece una interacción simbiótica con la leguminosa *Glycine max* (soya), pero se ha encontrado que la cepa GR64 de *S. fredii*, aislada de suelos de Granada (España), es capaz de interactuar con *P. vulgaris* (frijol) pero no con *G. max*. La secuenciación de uno de los plásmidos de la cepa GR64, el pSfrGR64a, reveló que el mismo fue originado a partir de la transferencia horizontal de un replicón de *R. etli*, rizobio simbiote de *P. vulgaris*, a una cepa de *S. fredii*, seguido de eventos de recombinación entre los plásmidos de *R. etli* y el cromosoma de *S. fredii* (Cervantes et al., 2011). Debido a las peculiares características de esta cepa, la interacción simbiótica con *P. vulgaris* y la interesante estructura genómica del plásmido pSfrGR64a, se decidió secuenciar el genoma completo de la GR64. Los datos que se obtienen a partir la secuenciación de genomas permiten clarificar los patrones en el flujo genes (Bentley, 2009), por lo cual la obtención de la secuencia completa del genoma de la bacteria *S. fredii* cepa GR64 será una valiosa herramienta para el entendimiento de la evolución de genomas de rizobios.

IV. HIPÓTESIS

Una característica de los genomas de rizobios es la tendencia a tener cromosomas con un mayor contenido de genes conservados y ordenados que los plásmidos, estos elementos extracromosomales tienden a ser heterogéneos en tamaño y contenido genético. La secuenciación del plásmido pSfrGR64a de *S. fredii* GR64, reveló que fue originado a partir de la transferencia horizontal de un replicón de la bacteria *Rhizobium etli* a una cepa de *S. fredii*, seguido de eventos de recombinación entre los plásmidos de *R. etli* y el cromosoma de *S. fredii* (Cervantes et al., 2011). Debido a la naturaleza quimérica de dicho plásmido y a las características conocidas de los genomas de rizobios, se espera que el estudio de genómica comparativa, mediante la obtención de la secuencia completa del genoma de *S. fredii* GR64, revele que el cromosoma contiene bloques de genes sinténicos al compararlos con otras cepas de la familia *Rhizobiaceae*, mientras que los plásmidos presenten sintenia sólo con algunos genes, que son conservados entre plásmidos de rizobios e incluso con genes cromosomales de otras especies, debido a fenómenos de recombinación y transposición de genes causados por los eventos de transferencia horizontal de genes.

V. OBJETIVO GENERAL

Obtener el genoma completo de la bacteria *Sinorhizobium fredii* cepa GR64 (*Rhizobiaceae*) a partir de la secuencia parcialmente ensamblada para realizar estudios de genómica comparativa.

V.1. Objetivos Específicos

Determinar las secuencias entre los scaffolds del draft de la secuenciación del genoma de la bacteria *Sinorhizobium fredii* cepa GR64 mediante estrategias bioinformáticas y de biología molecular.

Determinar la sintenia de los replicones de la bacteria *Sinorhizobium fredii* cepa GR64 con respecto a otras especies de bacterias de la Familia *Rhizobiaceae* mediante herramientas bioinformáticas.

VI. METODOLOGÍA

VI.1. Genómica comparativa de replicones utilizando el programa R2Cat

Recientemente se ha publicado un draft del genoma completo de *S. fredii* GR64, donde se obtuvieron 216 contigs grandes (>500 pb) con un tamaño total de 6, 959, 392 bases, de los cuales 146 fueron localizados en 24 scaffolds (Torres Tejerizo et al., 2012). Con el fin de obtener el orden de los 24 scaffolds se realizó un alineamiento de los mismos utilizando el programa R2cat y ordenándolos en referencia con otros genomas de rizobios.

VI.2. Cerrado de gaps mediante los programas IMAGE y SSPACE

El orden de los scaffolds, obtenido mediante el mapeado con el programa R2cat, fue alineado nuevamente en los programas de ensamble IMAGE y SSPACE, de esta manera se logró caminar sobre los gaps, es decir, que nuevas lecturas se alinearan en los extremos de los scaffolds ordenados. La utilización de este recurso de manera iterativa permitió completar la mayoría de los gaps.

Debido a que con estas herramientas bioinformáticas no fue posible cerrar completamente el genoma, se llevaron a cabo procedimientos microbiológicos y de biología molecular para obtener las secuencias de los gaps resultantes.

VI.3. Procedimientos microbiológicos

VI.3.1. Cepas y plásmidos utilizados

Usualmente, la presencia de gaps se debe a la presencia de secuencias repetitivas. Para disminuir el error que provocan estas secuencias en el genoma, los replicones de *S. fredii*

GR64 fueron analizados de forma independiente. Las cepas utilizadas para esto y para realizar otros procedimientos de biología molecular, se muestran en la tabla I.

Tabla I. Cepas y plásmidos utilizados.

| Cepas y plásmidos | Características relevantes | Referencia |
|-----------------------------|---|-----------------------------|
| <i>Sinorhizobium fredii</i> | | |
| GR64 | Cepa silvestre, Nal ^R | |
| GR64 pSfrGR64c (pJB3) | Derivada de la transferencia conjugativa del pJB3 de S17-1, Rifa-Nm ^R | (Este trabajo) |
| <i>Agrobacterium</i> | | |
| GMI 9023 | Rifa ^R | |
| GMI 9023 (pSfrGR64a) | Derivada de la transferencia conjugativa del pSfrGR64a, Rifa-Nm ^R | (Este trabajo) |
| GMI 9023 (pSfrGR64b) | Derivada de la transferencia conjugativa del pSfrGR64b (pSym), Rifa-Nm ^R | (Cervantes et al., 2011) |
| GMI 9023 (pSfrGR64c) | Derivada de la transferencia conjugativa del pSfrGR64c (pJB3), Rifa-Nm ^R | (Este trabajo) |
| <i>Escherichia coli</i> | | |
| DH5a | Receptora para transformación, Nal ^R | (Taylor, 1993) |
| S17-1 | Receptora para transformación, Km ^R | (Bierman et al., 1992) |
| Plásmidos | | |
| pK18mob | Vector suicida en rizobios, Km ^R | (Schäfer et al., 1994) |
| pRK600 | Plásmido helper, Km ^R | (Keen et al., 1988) |
| pRK2013 | Plásmido helper, Km ^R | (Figurski y Helinski, 1979) |
| pJB3 | Plásmido helper, Tc ^R | (Figurski et al., 1976) |

VI.3.2. Medios de cultivo

Las cepas de *S. fredii* y *Agrobacterium* GMI 9023 se cultivaron a 30°C en medio PY (peptona de caseína 0.5%, extracto de levadura 0.3%, CaCl₂ 7 µM). Las cepas de *E. coli* se cultivaron a 37°C en medio Luria Bertani (LB, peptona de caseína 1%, extracto de levadura 0.5%, NaCl 1%).

Para realizar la selección de transconjugantes se adicionaron antibióticos a los medios ya esterilizados en las concentraciones siguientes: Ácido nalidíxico (Nal), 20 µg/ml; Rifampicina (Rifa), 100 µg/ml; Neomicina (Nm), 30 µg/ml, y Tetraciclina (Tc), 3.3 µg/ml para *S. fredii* y *Agrobacterium*. Para *E. coli* se utilizó Kanamicina (Km), 30 µg/ml, y Tetraciclina (Tc), 10 µg/ml.

VI.3.3. Estrategias para obtención de cepas

Las cepas se obtuvieron mediante transferencia conjugativa de plásmidos, los cuales se marcaron con un transposón que confiere una resistencia a antibiótico para facilitar la selección positiva de los mismos y transformación con células quimiocompetentes (explicado a detalle en la sección de VI.4.1. Métodos de biología molecular: VII.4.1.3. Transformación con células quimiocompetentes y VI.4.1.6. Marcaje de plásmidos).

VI.3.3.1. Métodos de conjugación biparentales

Las cepas fueron inoculadas en 2 ml de medio de cultivo (PY para rizobios y LB para *E. coli*) e incubadas toda la noche. Posteriormente se mezclaron 150 µl de la cepa receptora con 75 µl de la cepa donadora, se goteó la mezcla sobre una caja de medio PY y se incubó toda la noche a 30°C. El crecimiento celular fue resuspendido en 1 ml de solución estéril de MgSO₄ 10mM, 0.01% Tween 40. Se realizaron diluciones de 10⁻¹ a 10⁻³ y se sembraron 100 µl de cada una de las diluciones en el medio selectivo apropiado para la obtención de transconjugantes, incubando por un periodo de 2 a 3 días a 30°C.

VI.3.3.2. Métodos de conjugación triparentales

Este procedimiento fue utilizado para transferir el de plásmido pSfrGR64c (megaplásmido) de *S. fredii* GR64 a *Agrobacterium* GMI9032. Se utilizó la cepa GR64 con el plásmido pSfrGR64c marcado con resistencia a neomicina como donadora, los plásmidos pJB3, pRK600 u pRK2013 (En *E. coli* S17-1) como plásmidos colaboradores y la cepa GMI9023 como receptora.

Las cepas fueron inoculadas en 2 ml de medio de cultivo e incubadas toda la noche. Posteriormente se mezclaron 150 μ l de la cepa receptora, 75 μ l de la cepa donadora y 75 μ l de la cepa colaboradora. Se colocó la mezcla sobre una caja de medio PY y se incubó por 12 horas a 30°C. El crecimiento celular fue resuspendido en 1 ml de solución estéril de MgSO₄ 10mM, 0.01% Tween 40. Se realizaron diluciones de 10⁻¹ a 10⁻³ y se sembraron 100 μ l de cada una de las diluciones en el medio selectivo apropiado para la obtención de transconjugantes, incubando por un periodo de 2 a 3 días a 30°C.

VI.4. Procedimientos moleculares

VI.4.1. Métodos de biología molecular

Los métodos de biología molecular utilizados en este trabajo se citan a continuación.

VI.4.1.1. Reacción en cadena de la polimerasa (PCR)

Se utilizó ADN total, ADN plasmídico o directamente una colonia de bacteria como molde de PCR. Para realizar PCRs con bacteria como molde, las colonias se tomaron con un palillo de madera estéril, se resuspendieron en 60 μ l de agua estéril y se calentó durante media hora en agua hirviendo. Luego se centrifugó y se tomaron 5 μ l como molde de ADN para la PCR.

Las reacciones de PCR se llevaron a cabo con la enzima Taq polimerasa y para la amplificación de fragmentos a clonarse en vectores se utilizó la Pfu polimerasa (la cual deja extremos romos). Para cada reacción se añadió: Buffer (10X) a una concentración final de 1X, dNTP's (dATP, dTTP, dCTP y dGTP) 100 μ M c/u, oligonucleótidos 0.25 μ M, 1 U de Taq polimerasa y el ADN molde (~100 ng).

Se utilizó un termociclador Thermal Cycler C1000. Se emplearon de 30-35 ciclos de amplificación, cada ciclo: 94°C, 0:20-1:00 minuto (desnaturalización del ADN); 56-60 °C, 0:20-1:00 minuto (alineación de oligonucleótidos); y 65-72°C, 1:00-5:00 minutos (elongación de la cadena de ADN).

Para verificar clones, se llevó a cabo una PCR utilizando los oligonucleótidos universales M13-40 Fw y M13-40 Rv, con 35 ciclos de amplificación, cada ciclo: 94°C por 0:20 minutos (desnaturalización del ADN); 56 °C por 0:20 minutos (alineación de oligonucleótidos); y 72°C por 0:40 minutos (elongación de la cadena de ADN) (Sambrook et al., 1989).

VI.4.1.2. Reacciones de digestión con enzimas de restricción y ligación

Las reacciones de digestión con enzimas de restricción se realizaron según las recomendaciones de los proveedores y se llevaron a cabo en tubos eppendorf de 0.5 ml. Para cada reacción de digestión se adicionaron 2 μ l de buffer (el correspondiente a cada enzima), 1 μ l de la enzima de restricción, ~5-8 μ l de fragmento de ADN a digerir (dependiendo de su concentración) y se aforó a 20 μ l con agua. Los tubos de reacción se dejaron a 37°C por 16 horas. La inactivación de las enzimas se realizó colocándolas a 75°C durante 15 minutos.

Las reacciones de ligación se llevaron a cabo en tubos eppendorf de 0.5 ml. Se utilizó el vector pK18mob, el cual es un vector suicida en rizobios (no replica en rizobios, sólo puede permanecer si se integra al genoma), que fue previamente digerido con la enzima de restricción *SmaI*, la cual corta y deja extremos romos. Para cada reacción de ligación se adicionaron 2 μ l de buffer, 1 μ l de la enzima ligasa T4, ~5 μ l del vector pK18mob digerido con *SmaI* (inactivada previamente), ~5-7 μ l de fragmento de ADN a ligar (dependiendo de su concentración) y se aforó a 20 μ l con agua. El tubo de reacción se dejó a 22°C por 16 horas.

Para reacciones de re-ligación se aforó a 50 µl con agua, y se corrigió la cantidad de buffer (Sambrook et al., 1989).

VI.4.1.3. Transformación con células quimiocompetentes

Se utilizaron las células quimiocompetentes (preparadas con CaCl₂ según Sambrook *et. al.* 1989) *E. coli* DH5α y *E. coli* S17-1. Para cada transformación, se descongelaron las células quimiocompetentes en hielo, se añadieron de 4-25 µl de ADN (plásmidos o ligaciones) transformante y se dejó reposar en hielo durante 30 minutos. Se dio un choque térmico a las células, colocándolas a 42°C durante 2 minutos y transfiriéndolas enseguida a hielo durante 5 minutos. Se añadió 1 ml de medio LB líquido y se incubó durante 1 hora a 37°C con agitación. La mezcla se sembró en cajas con medio selectivo adecuado (antibiótico kanamicina, y en las cajas de selección de ligaciones con pK18mob: X-gal, 40mg/ml disuelto en DMSO), plateando 100 µl y centrifugado (1 min, 14000 rpm) y se incubó 16 horas a 37°C (Sambrook et al., 1989).

VI.4.1.4. Purificación de plásmidos

La purificación de ADN de plásmidos se realizó utilizando el High Pure Plasmid Isolation Kit de Roche, siguiendo las indicaciones del proveedor. Las cepas de interés se inocularon en 3 ml de medio de cultivo y se incubaron 12 horas en agitación. Se centrifugó 1.5 ml de cultivo a velocidad máxima (14000 rpm) durante 3 minutos, se retiró por completo el sobrenadante utilizando una jeringa de 10 ml, la pastilla celular se resuspendió en 250 µl de la solución 1 fría (Buffer de suspensión) y se añadieron 250 µl de la solución 2 (Buffer de lisis). Se dejó reposar a temperatura ambiente durante 5 minutos, se añadieron 350 µl de la solución 3 fría (Buffer binding) y se incubó en hielo durante 5 minutos. Se centrifugó 10 minutos a 14000 rpm, el pellet contiene los restos celulares. Se transfirieron ~ 800 µl de sobrenadante a una columna de sílica acoplada a un tubo de plástico de 1.5 ml y se centrifugó por 1 minuto a 14000 rpm. Se descartó el filtrado y se añadieron 700 µl de la solución 5 (Buffer de lavado II)

y se centrifugó por 1 minuto a 14000 rpm. Nuevamente, se descartó el filtrado y se cambió la columna de sílica a un tubo colector, eppendorf de 1.5 ml.

Se añadieron 80 µl de la solución 6 (Buffer de elución) y se centrifugó por 1 minuto a 14000 rpm, colectando los plásmidos purificados en el tubo eppendorf de 1.5 ml debidamente etiquetados.

VI.4.1.5. Electroforesis en geles de agarosa

La separación de fragmentos amplificados por PCR, purificaciones de plásmidos y digestiones con enzimas de restricción, se realizó mediante electroforesis en geles de agarosa. Se utilizaron geles de agarosa al 1% en buffer TAE 1X (40mM Tris-acetato, 1mM EDTA, pH 8.0). La cantidad de ADN aplicada a cada carril fue de 4 a 10 µl mezclado con 2 µl de colorante M/M (0.25% azul de bromofenol, 0.25% de xilen-cianol, 15 % ficoll).

La electroforesis se realizó sometiendo a los geles a 100-120 volts durante 40- 60 minutos, en cámaras con buffer TAE 1X. Los geles fueron teñidos durante 5 minutos con bromuro de etidio (5 µg/ml), lavados con agua bidestilada durante 10 minutos y fotografiados sobre un transiluminador UVP con lámpara de luz ultravioleta, mediante el Gel Logic 100 Imaging System y el Programa Kodak Mi.

VI.4.1.6. Marcaje de plásmidos

Se diseñaron oligonucleótidos para amplificar por PCR una región de ~400-700 pb del plásmido a marcar. Los amplicones obtenidos fueron clonados al vector pK18mob (el cual tiene genes de resistencia a kanamicina/neomicina) mediante una reacción de ligación con ligasa T4. Las reacciones fueron transformadas en *E. coli* DH5α, y las células transformadas fueron seleccionadas en LB suplementado con Km y Xgal. Dicha selección permite obtener colonias blancas y azules, en las colonias blancas se clonó el fragmento deseado en el vector (las colonias azules son capaces de degradar X-gal, las blancas son incapaces ya que el fragmento clonado en el vector interrumpe el gen *lacZ*).

Se seleccionaron 5 colonias blancas y se llevó a cabo una PCR con los oligonucleótidos M13-40 Fw y M13-40 Rv para verificar la presencia del vector pK18mob con el inserto. Los plásmidos fueron purificados con el GeneJET Gel Extraction Kit de Thermo Scientific (siguiendo el procedimiento del proveedor, descrito a detalle en la sección VII.4.1.4. Purificación de plásmidos) y fueron transformados en *E. coli* S17-1. Los clones con el vector suicida pk18mob se transfirieron por conjugación, de S17-1 a GR64 wt, y se seleccionaron los transconjugantes en medio selectivo PY suplementado con Nal y Nm.

VI. 4.1.7. Geles de lisis *in situ*

Para verificar la transferencia conjugativa se llevaron a cabo electroforesis de geles de lisis *in situ* (Eckhardt, 1978), en los cuales se observaron los perfiles plasmídicos de las cepas. Las cepas fueron inoculadas en medio de cultivo e incubadas por 16 horas. Se midió la DO del cultivo, a λ 620 nm, y se añadieron 0.5 unidades de DO en un tubo eppendorf de 1.5 ml. Se centrifugó durante 3 minutos, a 4°C, 14 000 x g, y se descartó el sobrenadante. Se resuspendió la pastilla celular con 1 ml de agua fría. Se transfirieron 120 μ l a otro tubo eppendorf de 1.5, se centrifugó durante 3 minutos, a 4°C, 14 000 x g y se descartó el sobrenadante. La pastilla celular se resuspendió en 500 μ l de sarcosilato de sodio 0.3% (P/V), se centrifugó durante 3 minutos, a 4°C, 14 000 rpm, se descartó el sobrenadante y los tubos con la pastilla celular fueron colocados a -20°C durante 10 minutos.

Se utilizaron geles de agarosa al 0.75%, al que se añadió dodecilsulfato sódico (SDS) al 10% p/v (cada 100 ml de agarosa), disueltos en buffer TBE 1X. Se preparó una solución de lisis (10mg de lizosima, 960 μ l de sacarosa 20%, 40 μ l de RNasa, 10mg/ml), de la cual se agregó 30 μ l para resuspender la pastilla celular y se cargó en el gel.

La electroforesis se realizó sometiendo los geles a 5 volts durante 1 hora, y se subió el voltaje a 80 volts durante 5-8 horas, en cámaras con buffer TBE 1X. Los geles fueron lavados con agua bidestilada durante 10 minutos, posteriormente teñidos durante 5 minutos con bromuro de etidio (5 μ g/ml), lavados nuevamente con agua bidestilada durante 10 minutos y fotografiados sobre un transiluminador UVP con lámpara de luz ultravioleta.

VI.4.1.8. Extracción de ADN total

Se llevó a cabo la técnica de extracción de ADN total con fenol-cloroformo (Sambrook et al., 1989). Se inocularon 5 ml de medio de cultivo con las cepas GR64, GMI (pGR64a), GMI (pSymp-GR64b), GMI (pGR64c), incubando toda la noche a 30°C con agitación. Las células fueron centrifugadas a 14000 x g a temperatura ambiente por 10 minutos, se desechó el sobrenadante y fueron lavadas con 1 ml de solución amortiguadora TE 50/20 (pH 8.0, Tris 80mM, EDTA 20 mM pH8, el EDTA es un agente quelante de cationes divalentes). Se centrifugó a 14000 x g por 3 minutos y se decantó el sobrenadante.

La pastilla celular se resuspendió en 550 µl de TE50/20 utilizando un vórtex. Se adicionaron 20 µl de lizosima (20 mg/ml en agua), se mezcló e incubó a 37°C por 5 minutos. Se agregaron 50 µl de pronasa (5 mg/ml en TE 50/20, predigerida 1 hora a 37°C), se mezcló e incubó a 37°C por 5 minutos. Se agregaron 50 µl de SDS al 10% (diluido en H₂O), se mezcló e incubó a 37°C por 10 minutos. Luego, se incubaron las muestras a 80°C durante 10 minutos y se dejaron enfriar a temperatura ambiente. Se agregaron 3 µl de RNasa (10 mg/ml en TE 10/1, pretratada 10 minutos a 95°C), se mezcló e incubó a 37°C por 10 minutos. Se añadieron 220 µl de acetato amónico (saturado), se mezcló vigorosamente y se incubó en hielo por 5 minutos. Se centrifugó 5 minutos a 14000 rpm, el pellet contiene los restos celulares y proteínas precipitadas.

Se transfirió el sobrenadante a un tubo eppendorf (~800 µl) y se agregaron 500 µl de fenol-cloroformo-isoamílico (24:24:1), se mezcló vigorosamente y se centrifugó por 3 minutos a 14000 x g. Se recuperó la fase acuosa y se transfirió a otro tubo eppendorf (~800 µl), se añadieron 500 µl de cloroformo-isoamílico (42/1), se mezcló vigorosamente y se centrifugó por 3 minutos a 14000 x g. Se transfirió el sobrenadante a otro tubo eppendorf, se agregaron 600 µl de isopropanol y se mezcló varias veces.

Cuando hubo formación de hebras de ADN, éstas se transfirieron a un tubo eppendorf conteniendo etanol al 70%. Se centrifugó el pellet conteniendo el ADN durante 3 minutos a 14000 x g y se descartó el sobrenadante. Se lavó dos veces con 0.5 ml de etanol al 70%, se retiró el sobrenadante y se secó la pastilla de ADN en un evaporador centrífugo. Finalmente se resuspendió la pastilla de ADN en 100 µl TE 10/1 (Tris 10 mM, EDTA 1mM, pH 8.0) adicionando 2 µl de RNasa (20mg/ml), y se almacenaron a -20°C (Sambrook et al., 1989).

VI.4.2. Obtención de las secuencias de los *gaps* resultantes

VI.4.2.1. Estrategias para la obtención de secuencias de *gaps* resultantes

Para determinar las secuencias de los *gaps* resultantes, uniendo así los contigs y scaffolds, se llevaron a cabo dos estrategias, la reacción en cadena de la polimerasa (PCR) y la caminata por integración genómica.

VI.4.2.1.1. Reacción en cadena de la polimerasa (PCR)

Para ordenar los scaffolds resultantes y cerrar los *gaps* se llevaron a cabo PCRs, diseñando oligonucleótidos en los extremos de cada uno de los scaffolds resultantes. Las reacciones fueron realizadas utilizando un termociclador Bio-Rad. Se utilizaron diferentes ADN polimerasas, según tamaño del fragmento a amplificar, la Taq polimerasa de Invitrogen (con una temperatura óptima de extensión de 72°C), la Pfu polimerasa (con una temperatura óptima de extensión de 65° C) y la LongAmp polimerasa de New England Biolabs (con una temperatura óptima de extensión de 65° C). Las PCRs con la LongAmp polimerasa fueron preparadas en hielo y los tubos fueron colocados en el termociclador cuando este alcanzaba una temperatura de 105°C (shock térmico) (siguiendo indicaciones del proveedor).

Los dNTPs fueron utilizados 2.5 mM en reacciones con Taq y Pfu polimerasas y a 10 mM para reacciones con LongAmp polimerasa (según las indicaciones de los proveedores). Los oligonucleótidos (inciadores) se utilizaron a una concentración stock de 5 pmol/μl. El molde utilizado en las PCRs para amplificar *gaps* fue el ADN total de las cepas GR64 wt, GMI (pGR64a) y GMI (pSym), según la localización de la secuencia del *gap* a amplificar.

VI.4.2.1.2. Caminata por integración genómica

Esta estrategia se utilizó para cerrar *gaps* en los cuales la reacción de PCR no resultó satisfactoria. Para ello se amplificó un fragmento de 400 pb en los extremos de los *scaffolds* y

se clonó en el vector pK18mob, mediante una reacción de ligación (1 µl de ADN ligasa, 2 µl de buffer, ~ 4 µl de ADN molde y aforando a 20 µl con H₂O). Dichas ligaciones fueron transformadas en *E. coli* DH5α, las colonias blancas se evaluaron por PCR (Oligos M13 fw-M13 rv, Taq polimerasa, dNTPs 2.5 mM, Buffer Pfu 10X), las positivas se purificaron mediante el High Pure Plasmid Isolation Kit y nuevamente se transformaron en *E. coli* S17-1. Con los vectores resultantes se realizaron integraciones sitio específicas y a partir de los mutantes obtenidos se aisló ADN total, se digirió con enzimas de restricción, para luego realizar una religación y una posterior transformación en *E. coli* DH5α. El vector recuperado con los fragmentos que flanqueaban la región blanco (*gap*) fue secuenciado desde sus extremos para completar la secuencia de la zona en cuestión.

VI.4.2.2. Extracción y purificación de bandas

Las bandas con los fragmentos de ADN, correspondientes a los amplicones obtenidos a partir de las PCRs y a los vectores con las regiones buscadas, fueron cortadas de los geles de agarosa visualizados en un transiluminador UVP con lámpara de luz ultravioleta, utilizando un bisturí estéril. Dichas bandas fueron colocadas en tubos eppendorf de 1.5 ml y purificadas mediante el Kit GeneJET Gel Extraction de Thermo Scientific.

Se añadió la solución 1 (Buffer de lavado) hasta cubrir por completo fragmento de gel, y se colocó a 65°C durante 10 minutos, esta solución disminuye la temperatura de fusión de la agarosa, acelerando su fundición. La mezcla se agitó utilizando vórtex. Se transfirieron 800 µl a una columna de sílica acoplada a un tubo de plástico de 1.5 ml y se centrifugó 1 minuto a 14000 x g. Se descartó el líquido filtrado. Para productos de secuenciación se añadieron 100 µl de la solución 1 y se centrifugó 1 minuto a 14000 rpm, se descartó el líquido filtrado y se añadieron 700 µl de la solución 2 (Buffer binding). Se centrifugó 1 minuto a 14000 x g y se nuevamente se descartó el líquido filtrado.

Se repitió otra centrifugación 1 minuto a 14000 y se cambió la columna de sílica a un tubo colector, eppendorf de 1.5. Se añadieron de 30-50 µl de solución 3 (buffer de elución), se dejó reposar por 1 minuto y se centrifugó por 1 minuto a 14000 rpm colectando el ADN purificado en los tubos eppendorf debidamente etiquetados.

VI.4.2.3. Secuenciación mediante el método de Sanger

Los amplicones obtenidos a partir de las PCRs y los vectores con las regiones buscadas fueron secuenciados mediante el método de Sanger. Las muestras fueron enviadas a la empresa Macrogen ubicada en Seul, Corea del Sur, para su secuenciación.

Cada muestra fue secuenciada con los mismos oligonucleótidos (forward y reverse) que fueron utilizados para amplificar el fragmento. Después de cada reacción, lecturas de ~700-800 de pb, se diseñaron nuevos oligonucleótidos mediante el programa Primer 3 (disponible en línea: <http://bioinfo.ut.ee/primer3-0.4.0/>), los cuales fueron sintetizados por Macrogen, utilizados en la siguiente reacción de secuenciación, hasta obtener la secuencia faltante de los scaffolds.

VI.5. Ensamblado final mediante el programa Consed

Las secuencias de las lecturas obtenidas mediante Sanger fueron alineadas con el pre-ensamble, para verificar que los extremos de las secuencias correspondieran entre sí. Una vez que se verificó que los extremos alinearan, se realizó el ensamblado final insertando las lecturas de los gaps en el pre-ensamble utilizando el programa Consed. Una vez cerradas las secuencias de los replicones de *S. fredii* GR64, las secuencias fueron anotadas automáticamente mediante el programa en línea RAST.

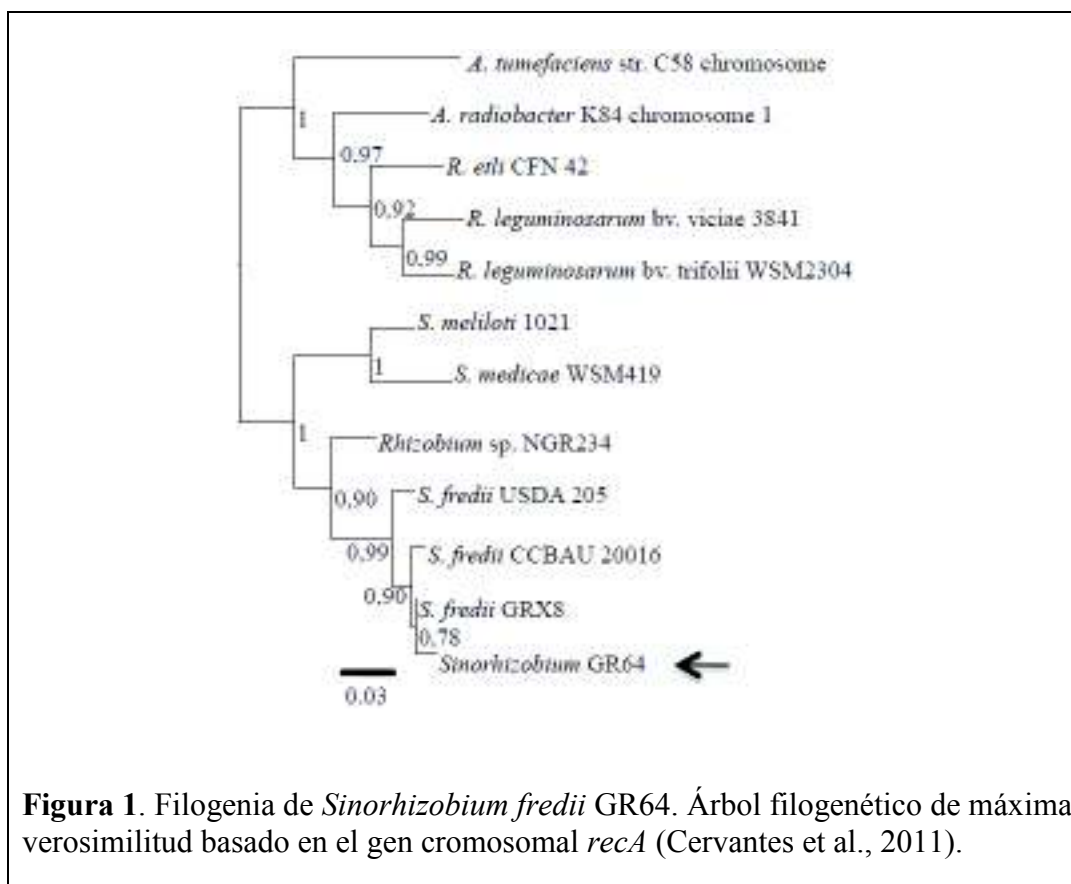
VI.6. Estudio de Sintenia del Genoma de *Sinorhizobium fredii* GR64

Se realizó un análisis de sintenia utilizando los programas R2Cat y Mauve, en el cual se comparó la secuencia completa del genoma de *S. fredii* GR64 con los genomas de otras bacterias de la familia *Rhizobiaceae*, cepas incluidas en el árbol filogenético de máxima verosimilitud basado en el gen cromosomal *recA* del trabajo de Cervantes et al. (2011). Se determinó la sintenia de cada replicón independiente, identificando bloques sinténicos en cada uno.

VII. RESULTADOS

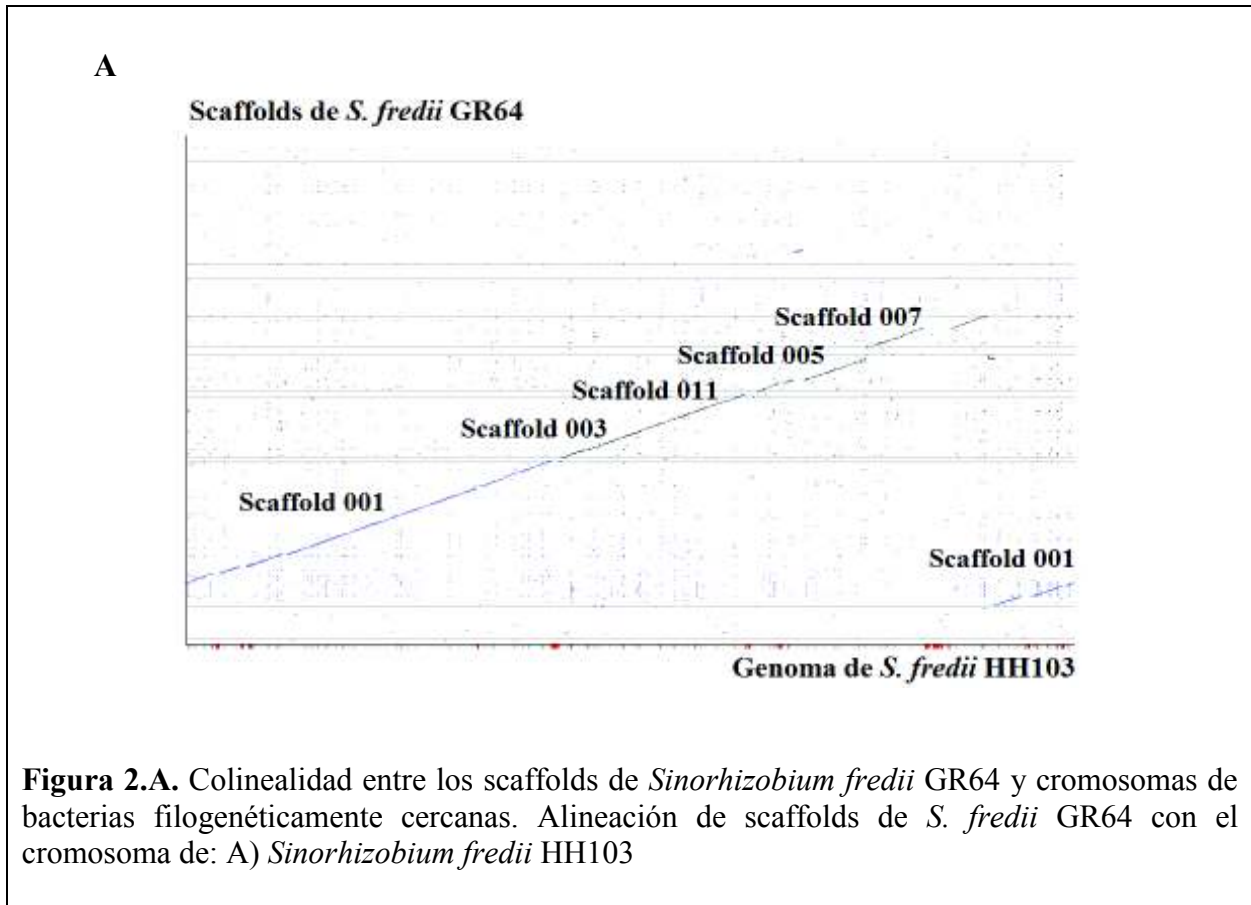
VII.1. Determinación del Orden e Identidad de los Scaffolds Mediante R2cat

Para realizar el análisis de colinealidad entre los scaffolds de *S. fredii* GR64, obtenidos previamente por secuenciación masiva (Torres Tejerizo et al., 2012), y los cromosomas de bacterias filogenéticamente cercanas, se utilizaron algunas de las cepas incluidas en el árbol filogenético de máxima verosimilitud basado en el gen cromosomal *recA* del trabajo de Cervantes et al. (2011) (Figura 1), y la cepa *S. fredii* HH103. Se realizó el análisis de colinealidad con *S. fredii* HH103, *S. fredii* USDA257, *Rhizobium* sp. NGR234, *S. medicae* WSM419, *S. meliloti* 1021, *R. leguminosarum* bv. *trifolii* WSM2304, *R. leguminosarum* bv. *viciae* 3841 y *R. etli* CFN42.

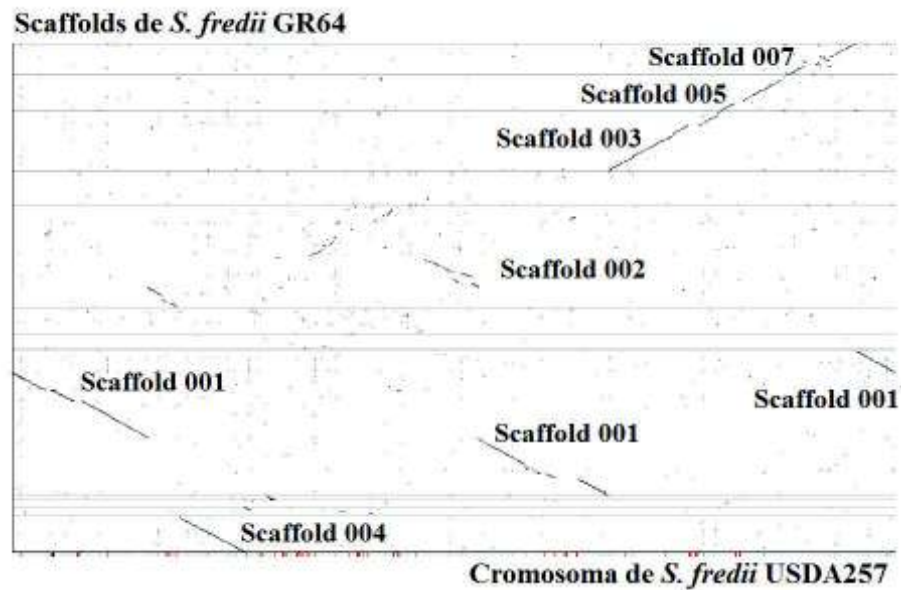


La colinealidad varió entre las cepas, sin embargo, el orden preponderante fue: scaffold 001, scaffold 003, scaffold 005 y scaffold 007 (Ver figuras 2.A-2.E.). La colinealidad entre el cromosoma de *S. fredii* HH103 y los scaffolds de *S. fredii* GR64 reveló mayor grado de sintenia, con el orden: scaffold 001, scaffold 003, scaffold 011, scaffold 005 y scaffold 007. La colinealidad con los cromosomas de *S. meliloti* 1021, *R. leguminosarum* bv. *viciae* 3841 y *R. etli* CFN42, también ordenó al scaffold 011 entre los scaffolds 003 y 005.

El orden consenso de los scaffolds de *S. fredii* GR64 correspondiente a la secuencia del cromosoma fue: scaffold 001, scaffold 003, scaffold 011, scaffold 005 y scaffold 007.



B



C

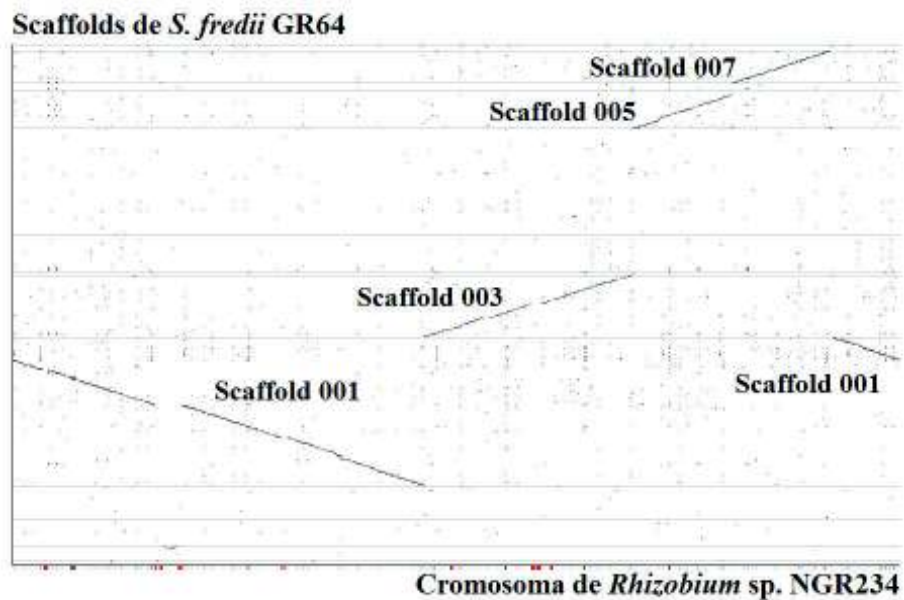
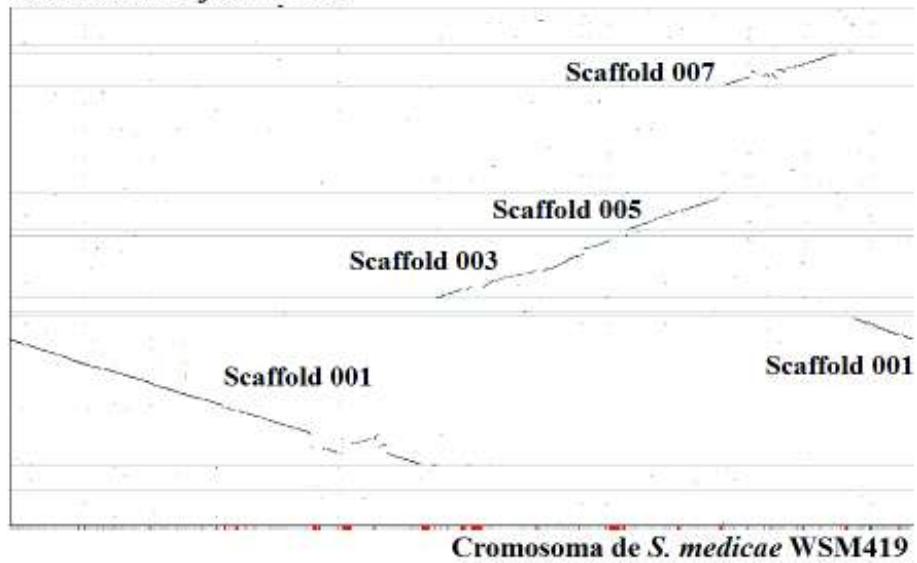


Figura 2.B. Colinealidad entre los scaffolds de *Sinorhizobium fredii* GR64 y cromosomas de bacterias filogenéticamente cercanas. Alineación de scaffolds de *S. fredii* GR64 con el cromosoma de: B) *S. fredii* USDA257 y C) *Rhizobium* sp. NGR234.

D

Scaffolds de *S. fredii* GR64



E

Scaffolds de *S. fredii* GR64

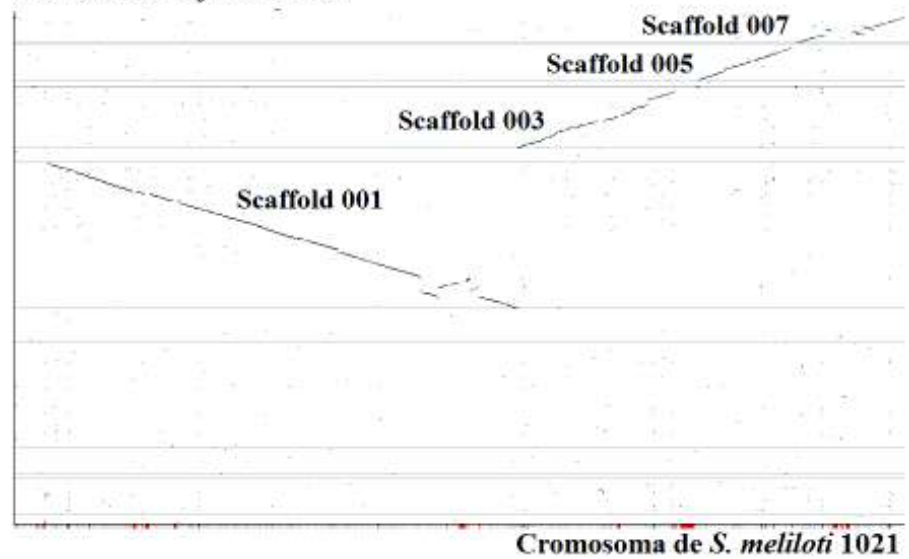
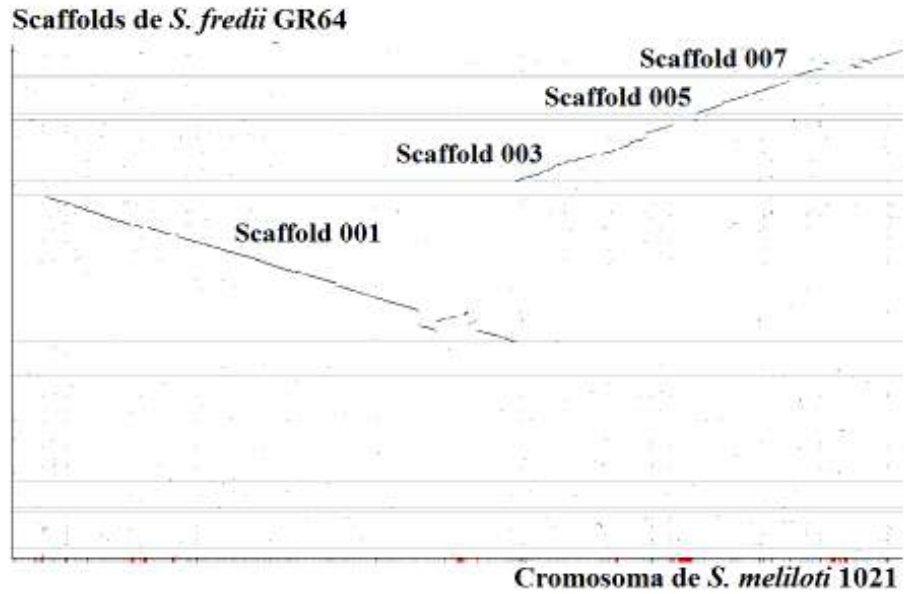


Figura 2.C. Colinealidad entre los scaffolds de *Sinorhizobium fredii* GR64 y cromosomas de bacterias filogenéticamente cercanas. Alineación de scaffolds de *S. fredii* GR64 con el cromosoma de: D) *S. medicae* WSM419 y E) *S. meliloti* 1021.

F



G

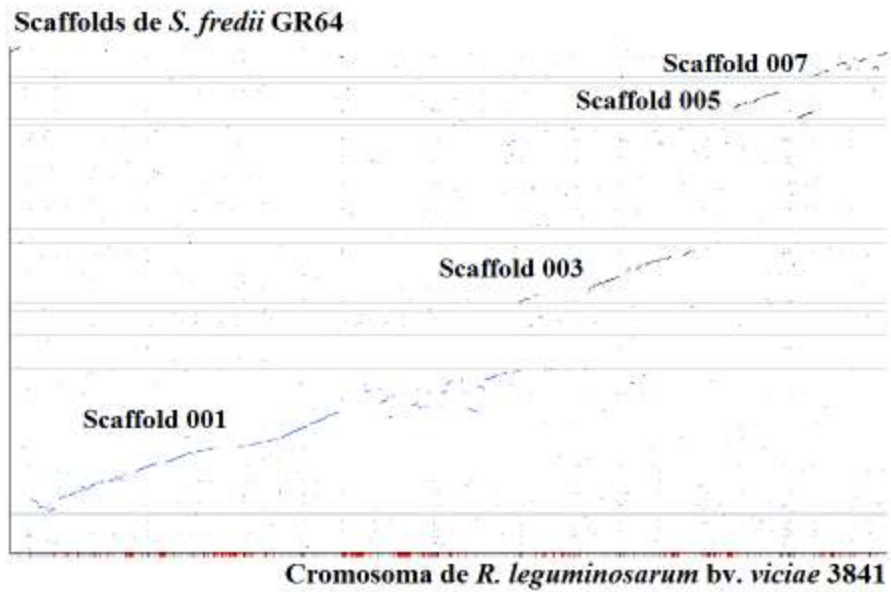


Figura 2.D. Colinealidad entre los scaffolds de *Sinorhizobium fredii* GR64 y cromosomas de bacterias filogenéticamente cercanas. Alineación de scaffolds de *S. fredii* GR64 con el cromosoma de: F) *Rhizobium leguminosarum* bv. *trifolii* WSM2304 y *R. leguminosarum* bv. *viciae* 3841.

H

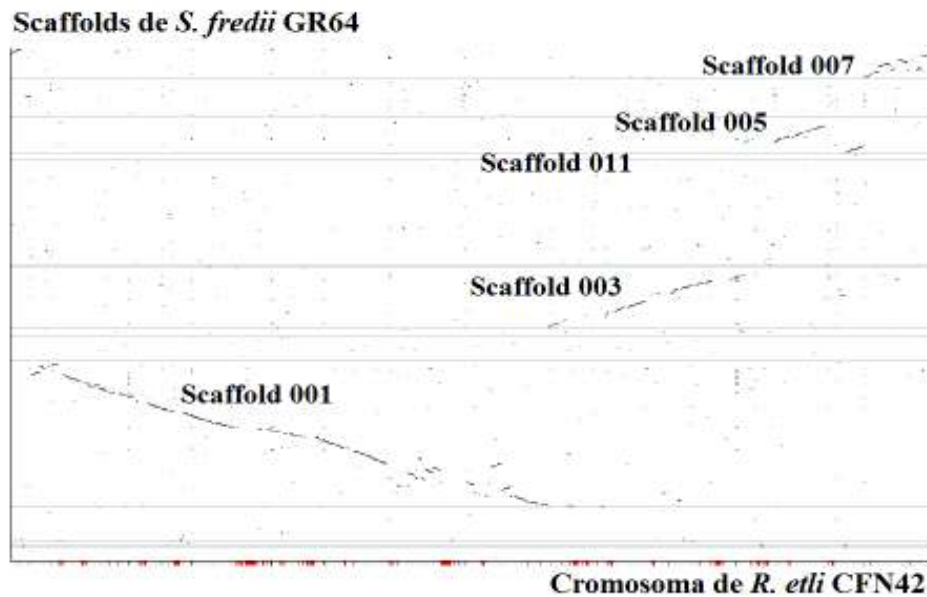
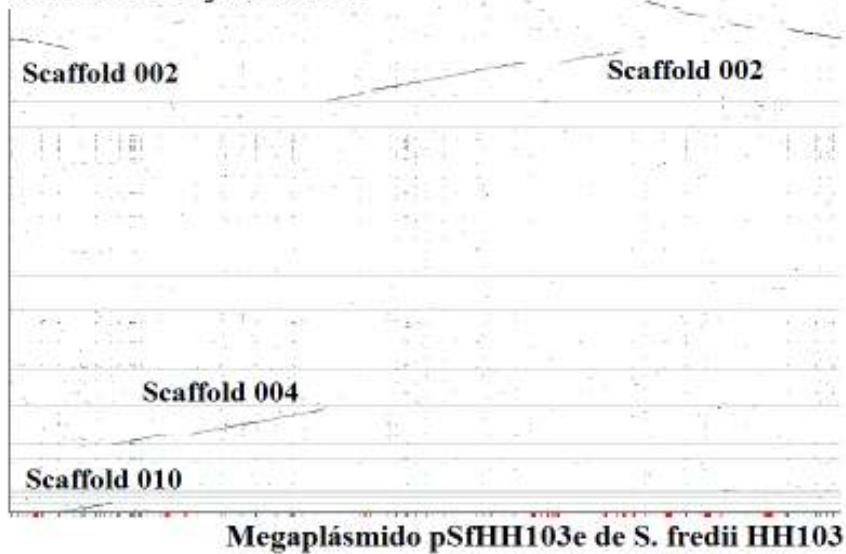


Figura 2.E. Colinealidad entre los scaffolds de *Sinorhizobium fredii* GR64 y cromosomas de bacterias filogenéticamente cercanas. Alineación de scaffolds de *S. fredii* GR64 con el cromosoma de: H) *R. etli* CFN42.

Para determinar la identidad y el orden del megaplásmido pSfrGR64c se realizó un análisis de colinealidad entre los scaffolds de *S. fredii* GR64 y megaplásmidos de las bacterias filogenéticamente cercanas, el pSfHH103e de *S. fredii* HH103 y el pNGR234b de *Rhizobium* NGR234 (figura 3). El orden consenso de los scaffolds fue: 002, 004 y 010.

A

Scaffolds de *S. fredii* GR64



B

Scaffolds de *S. fredii* GR64

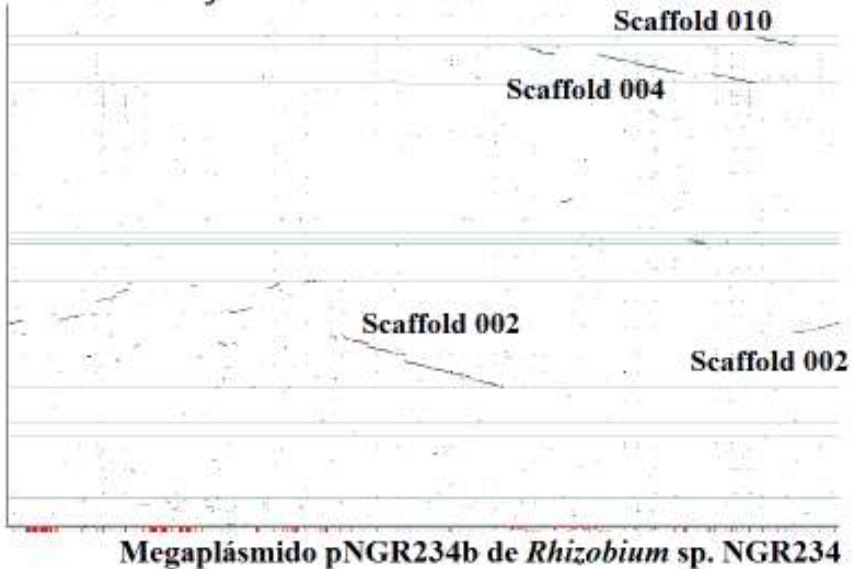
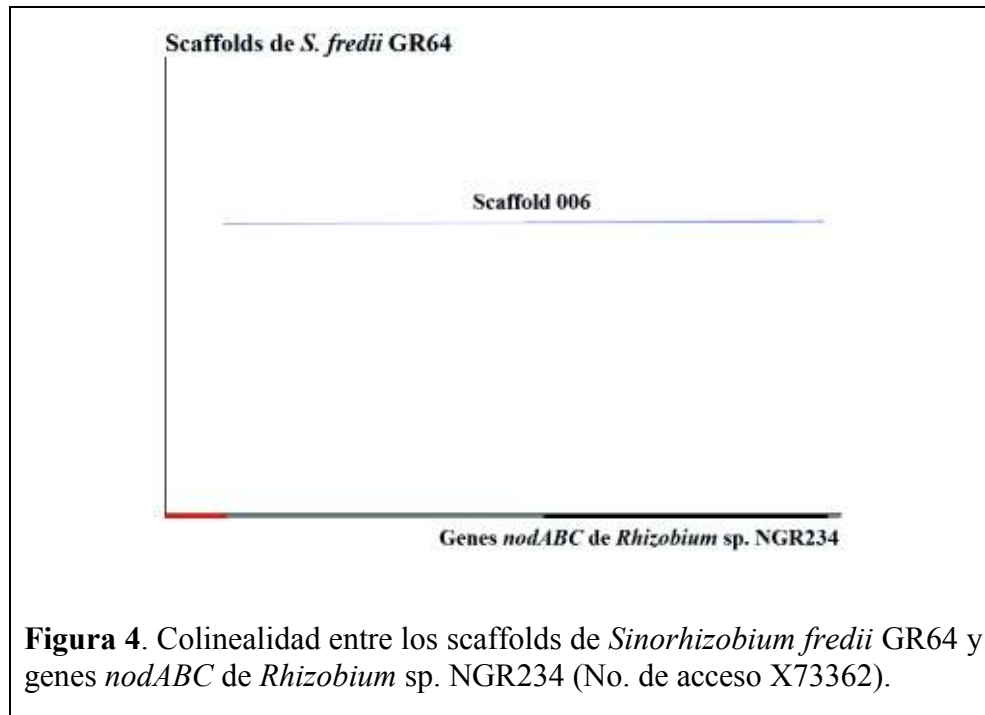
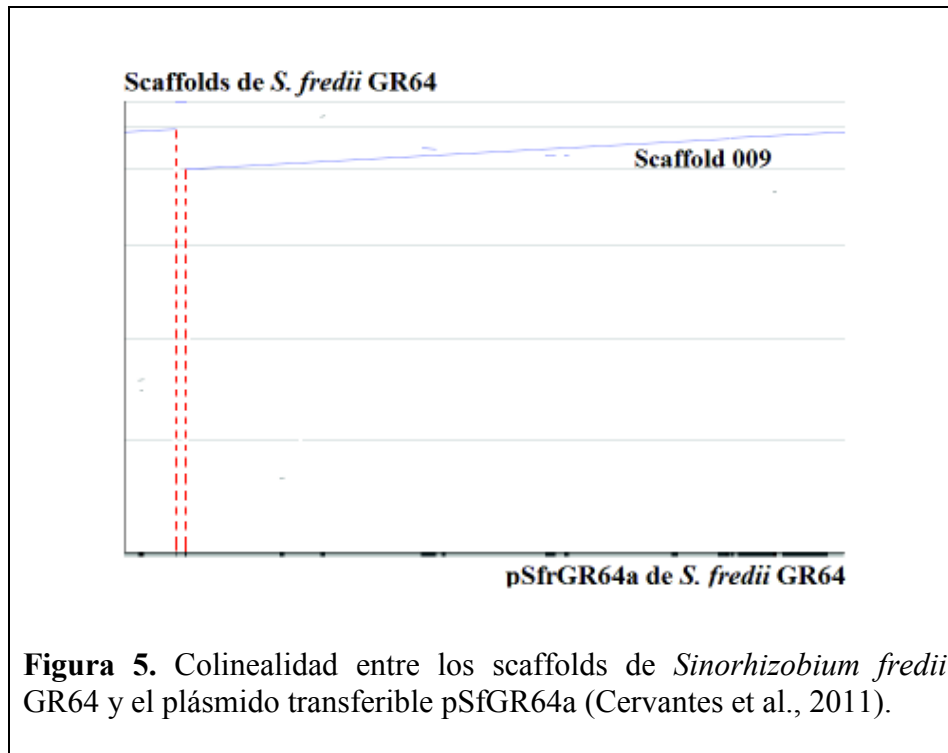


Figura 3. Colinealidad entre los scaffolds de *Sinorhizobium fredii* GR64 y megaplásmidos de bacterias filogenéticamente cercanas. Alineación de los scaffolds de *S. fredii* GR64 con los megaplásmidos: A) pSfHH103e de *S. fredii* HH103 y B) pNGR234b de *Rhizobium* NGR234.

Para determinar la identidad del plásmido simbiótico se analizó la colinealidad con los genes simbióticos *nodABC* de *Rhizobium* sp. NGR234 (número de acceso en genebank X73362) y los scaffolds de *S. fredii* GR64 (figura 4). El análisis reveló que el plásmido simbiótico pSfrGR64b corresponde al scaffold 006 de *S. fredii* GR64.



Para determinar la identidad del plásmido transferible pSfrGR64a de *S. fredii* GR64, se analizó la colinealidad entre los scaffolds de *S. fredii* GR64 y el pSfrGR64a secuenciado por Cervantes en el 2011 (figura 5), obtenido mediante la purificación del mismo por gradientes de CsCl y la realización de una biblioteca genómica. El análisis de colinealidad reveló que el scaffold 009 corresponde al plásmido pSfrGR64a, y que, además, existe una región adicional al ensamble original, la cual tiene un tamaño de ~10 kpb. Por este motivo es necesario verificar si dicha región es parte de la secuencia del pSfrGR64a.



VII.2. Finalización bioinformática de gaps

El orden de los scaffolds obtenidos fue incorporado a los programas IMAGE y SSPACE, y mediante iteraciones de esos programas se completaron los gaps remanentes. Así, se obtuvo un total de 6 contigs. Un contig correspondiendo al cromosoma, otro al plásmido simbiótico pSfrGR64b, otro al plásmido transmisible pSfrGR64a, y tres al megaplásmido pSfrGR64c.

VII.3. Estrategias para la Obtención de Secuencias de los Gaps Remanentes

VII.3.1. Obtención de Cepas

La cepa utilizada como receptora de los replicones de *S. fredii* GR64 fue la cepa *Agrobacterium tumefaciens* GMI9023, la cual no posee plásmidos. La cepa GMI9023 con el plásmido simbiótico pSfrGR64b utilizada fue obtenida del laboratorio de Ingeniería Genómica del Centro de Ciencias Genómicas de la UNAM. GMI9023 con los plásmidos pSfrGR64a y pSfrGR64c fueron obtenidos mediante métodos de transferencia conjugativa.

GMI9023 (pSfrGR64a) se obtuvo mediante el marcaje del plásmido pSfrGR64a con el vector pK18mob (contiene un *oriT*), y la transferencia conjugativa biparental desde *S. fredii* GR64 hacia GMI9023. Se seleccionó positivamente con el antibiótico neomicina y se realizó un gel de lisis *in situ* para verificar la transferencia del pSfrGR64a::pK18mob mediante la visualización de los perfiles plasmídicos (figura 5, carriles 1, 2 y 3).

Para obtener la cepa GMI9023 (pSfrGR64c) se realizaron conjugaciones triparentales desde una cepa de GR64 cuyo plásmido pSfrGR64c fue marcado con un pK18mob (contiene un *oriT*) y con los plásmidos helper pRK600, pRK2013 y pJB3, sin obtener resultados satisfactorios. Para omitir la conjugación triparental, y realizar simplemente una biparental, en primer lugar se transfirió el plásmido helper pJB3 a *S. fredii* GR64 (pSfrGR64c::pK18mob) mediante una conjugación biparental, se seleccionó positivamente con el antibiótico tetraciclina y se realizó un gel de lisis *in situ* para verificar la transferencia del plásmido pJB3 (figura 6, carriles 4, 5 y 6).

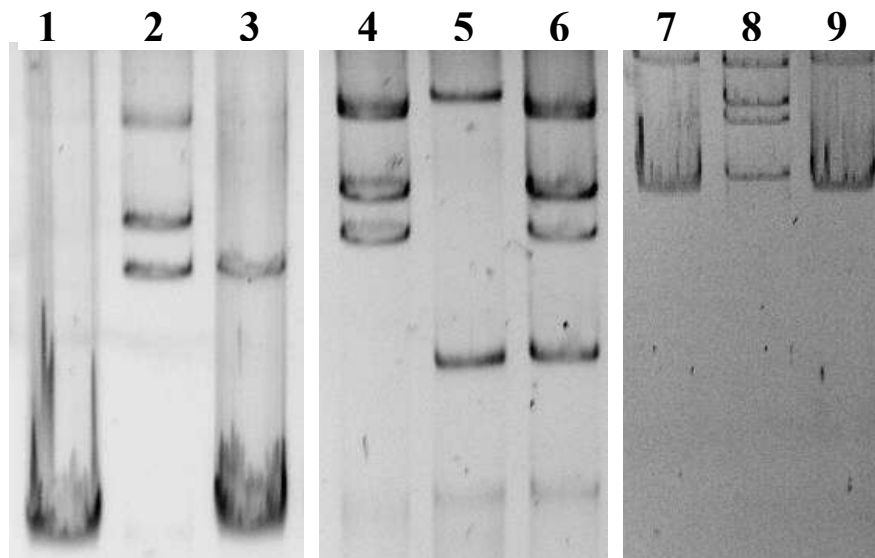


Figura 6. Geles de lisis *in situ* donde se muestran los perfiles plasmídicos de las cepas obtenidas. Carril 1 y 7: *Agrobacterium* GMI9023, Carril 2: *S. fredii* GR64 (pSfrGR64a marcado), Carril 3: GMI9023 (pSfrGR64a::pK18mob), Carril 4: *S. fredii* GR64 (pSfrGR64c::pK18mob), Carril 5: *E. coli* (pJB3), Carril 6 y 8: *S. fredii* GR64 (pSfrGR64c::pK18mob, pJB3), Carril 9: GMI9023 (pSfrGR64c:: pK18mob).

Luego se transfirió el megaplásmido pSfrGR64c desde *S. fredii* GR64 (pSfrGR64c::pK18mob, pJB3) hacia GMI9023 mediante una conjugación biparental. Se seleccionó positivamente con el antibiótico neomicina y se realizó un gel de lisis *in situ* para verificar la transferencia del megaplásmido pSfrGR64c (figura 6, carriles 7, 8 y 9). No fue posible visualizar el plásmido pSfrGR64c en los geles de lisis *in situ*.

Para verificar la transferencia del pSfrGR64c se realizó una PCR para amplificar el fragmento que se clonó en el vector pK18mob. Se utilizó como ADN molde las colonias directas, seleccionadas positivamente con antibióticos. El ciclado utilizado fue: Inicio a 94°C por 5 minutos; 35 ciclos de, desnaturalización a 94°C por 0:15 minutos, alineación a 56°C por 0:15 minutos y extensión a 72°C por 0:40 minutos; y una extensión final a 72°C por 1:20 minutos. El gel con los fragmentos amplificados se muestran en la figura 7. Todos los clones analizados fueron positivos, amplificando el fragmento clonado en el vector pK18mob.

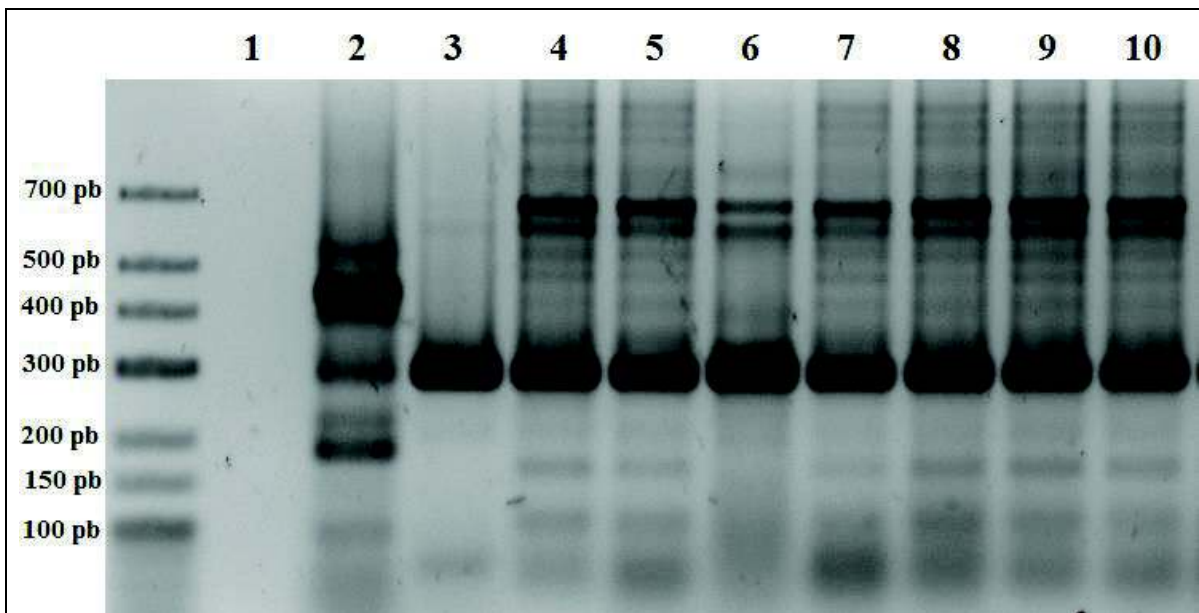


Figura 7. Verificación de la transferencia del pSfrGR64c::pK18mob::oriT al fondo genómico de GMI9023 mediante amplificación del fragmento clonado en el vector pK18mob. Carril 1: Control negativo, Carril 2: GMI9023 (control), Carril 3: *S. fredii* pSfrGR64c::pK18mob (control positivo), Carriles 4-10: GMI9023(pSfrGR64c::pK18mob).

Se aisló el ADN total de estas cepas, el cual se utilizó como ADN molde en las PCRs realizadas para amplificar los gaps remanentes.

VII.3.2. Oligonucleótidos

Los oligonucleótidos utilizados en este trabajo se listan en la tabla II. Se diseñaron oligonucleótidos para amplificar los gaps remanentes, para verificar la posición de regiones grandes repetidas en el genoma y para verificar la estructura de replicones.

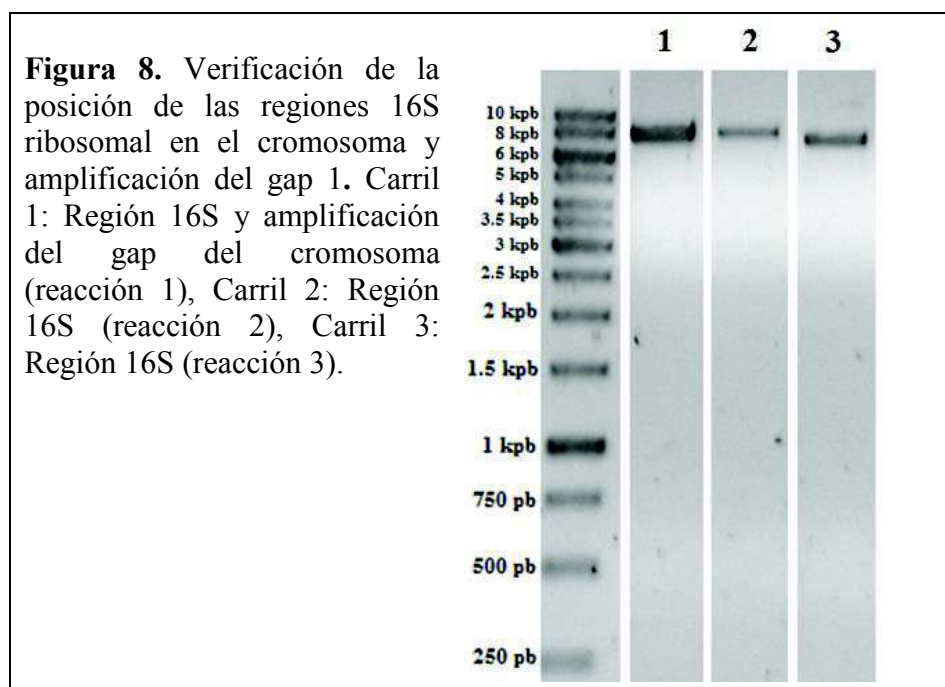
Tabla II. Oligonucleótidos utilizados en este trabajo.

| Ubicación | Nombre | Secuencia |
|---------------------------|----------------|------------------------------|
| Cromosoma | crGR64_fw 1 | 5'GCACCCTTTTGCGAACA-3' |
| | crGR64_rv1 | 5'GCGGAGGAGCCTAAAAC-3' |
| Simbiótico pSfrGR64b | pSymGR64_head | 5'-CCACCCCACTAGCATCAC-3' |
| | pSymGR64_tail | 5'-TTGGATTTTTATCCTATG-3' |
| | pSym-head-fw | 5'-GGAGGTGAGCTGAGCTAGAT-3' |
| | pSymGR64_tail2 | 5'-GAATGATGCTAAGGCGG-3' |
| Megaplásmido pSfrGR64c | mega1_tail | 5'-ACACTCCCCGTTATTCGT-3' |
| | mega2_head | 5'-GCATACGATTGAGATTTTCAAA-3' |
| | mega2_tail | 5'-CCAGTTCGTGCTTCATATGTAA-3' |
| | mega3_head | 5'-TGAAGGCGTCCAGTATATATTT-3' |
| | c144_tail | 5'-ACAATCGTTCCGGTGCTCTATC-3' |
| | c144_head | 5'-GCATCAGGCTTCCTCGTTT-3' |
| | c142_tail | 5'- ATCGCTCCAAAGGTCGTG-3' |
| | c142_head | 5'-TTCCTTCTCGCTCGCTTG-3' |
| pSfrGR64a | p64aN_head | 5'-TTTGAGCAAGTGGACTGG-3' |
| | p64aN_tail | 5'-CCGTCCGATGAATCATT-3' |
| | p64aV_head | 5'-TCACCTTGAAAGCAAGTAGTA-3' |
| | p64aV_tail | 5'-GCGGCCAGATCACACTAT-3' |
| Vector pK18mob | M13-40 fw | 5'-GTAAAACGACGGCCAGT-3' |
| | M13-40 rv | 5'-GGAAACAGCTATGACCATG-3' |

VII.3.3. Obtención de la secuencia completa del cromosoma

En el cromosoma circular de *S. fredii* GR64 quedó un gap remanente, denominado gap 1. Además era necesario verificar las posiciones de las regiones 16S ribosomales, repetidas tres veces en el genoma de *S. fredii* GR64. Dichas regiones fueron posicionadas en el ensamble de acuerdo con los *loci* de las mismas en bacterias filogenéticamente cercanas, utilizando la herramienta BLAST (disponible en la página: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Se realizaron tres PCRs para verificar la posición de cada una de las regiones 16S ribosomales, reacciones 1, 2 y 3. La reacción 1, además, comprende la secuencia del gap 1.

Las reacciones 1, 2 y 3 se amplificaron utilizando la pareja de primers crGR64_fw1-crGR64_rv1, crGR64_fw2-crGR64_rv2 y crGR64_fw3-crGR64_rv3, respectivamente. Se utilizó la enzima LongAmp polimerasa, las condiciones de ciclado fueron similares, excepto en la temperatura de alineación: Inicio de 94°C por 5 minutos; 36 ciclos de, desnaturalización a 94°C por 20 segundos, alineación por 20 segundos y extensión final a 65°C por 10 minutos. La temperatura de alineación de la reacciones 1 y 3 fue 60°C y de la reacción 2 fue 54°C. Las bandas amplificadas fueron de ~8 kpb en las reacciones 1 y 2, y ~7 kpb en la reacción 3 (figura 8).



VII.3.4. Obtención de la secuencia completa del megaplásmido pSfrGR64c

El programa bioinformático SSPACE predijo la estructura del megaplásmido pSfrGR64c mostrada en la figura 9, formado por los contigs 1, 2 y 3. El extremo cabeza del contig 1 y el extremo cola del contig 3 comparten una región de ~30 kpb, lo cual indica que los contigs están superpuestos. Sin embargo, quedan por dilucidar dos gaps. Se diseñaron oligonucleótidos en los extremos: cola del contig 1, el extremo cabeza del contig 3, y ambos extremos del contig 2, para cerrar los dos gaps remanentes de la estructura predicha.

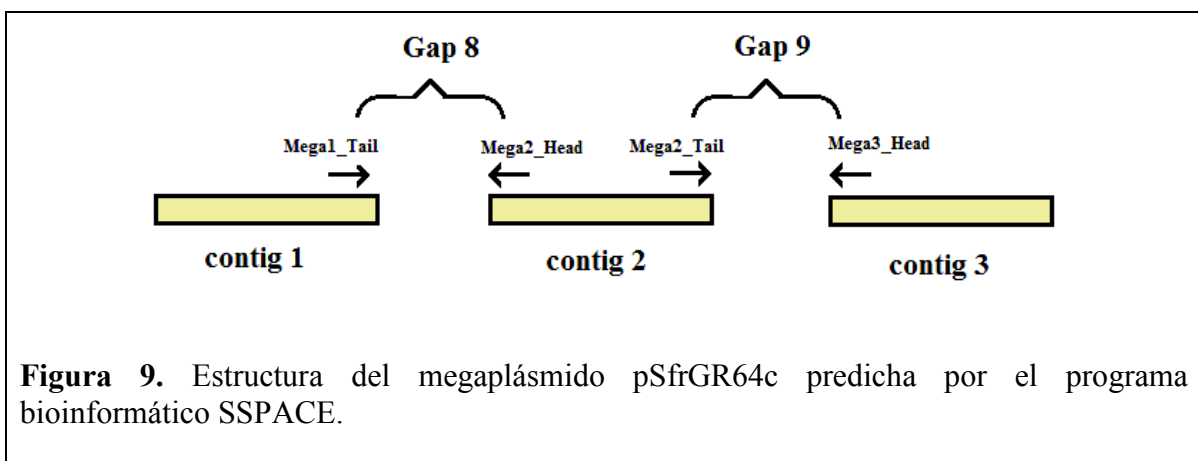
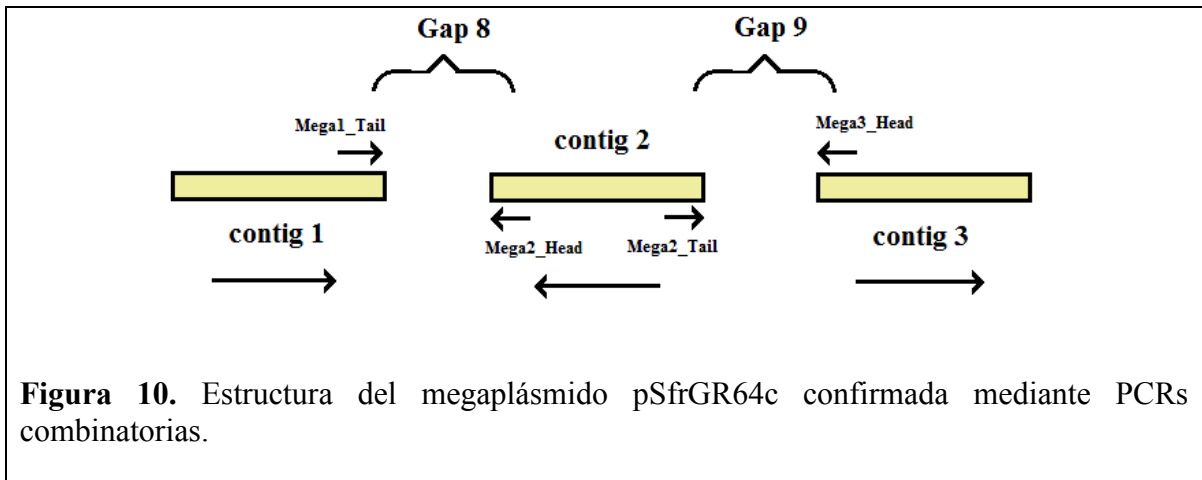


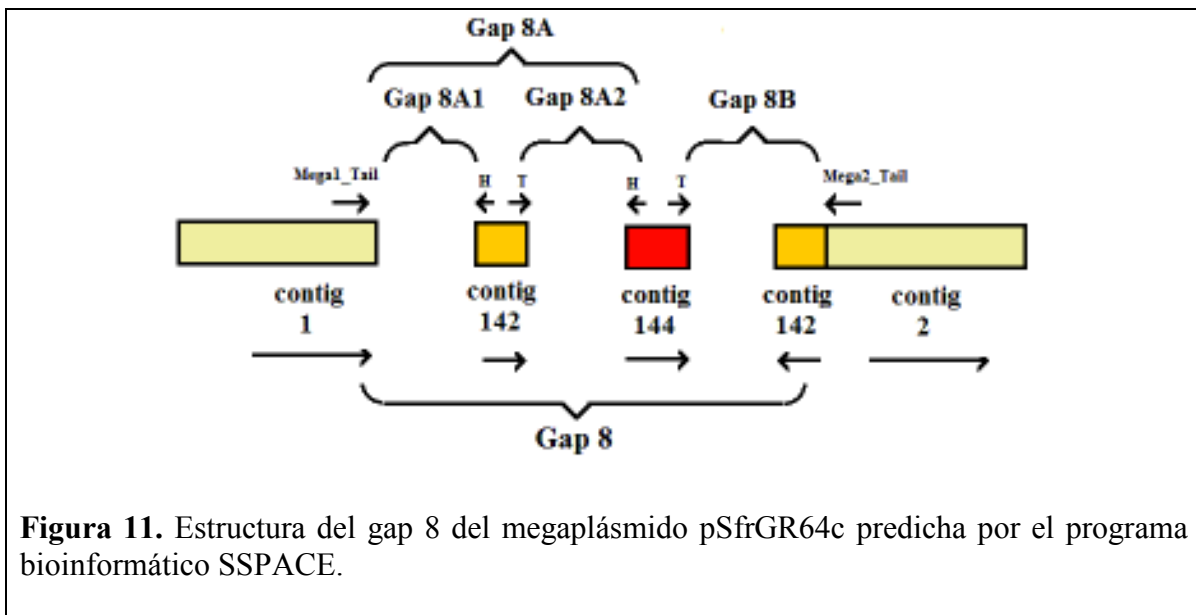
Figura 9. Estructura del megaplásmido pSfrGR64c predicha por el programa bioinformático SSPACE.

El gap entre el contig 1 y el contig 2 se denominó reacción o gap 8, y se utilizaron los oligonucleótidos Mega1_tail y Mega2_head, y la enzima LongAmp polimerasa. El gap entre el contig 2 y el contig 3 se denominó reacción o gap 9, y se utilizaron los oligonucleótidos Mega2_tail y Mega3_head, y la enzima LongAmp polimerasa. Se realizaron más de 10 reacciones de PCR, variando las condiciones, sin obtener resultados satisfactorios. Debido a esto, se procedió a realizar PCRs combinatorias entre los oligonucleótidos Mega1_tail, Mega2_head, Mega2_tail y Mega3_head, y se lograron amplificar bandas a partir de las combinaciones Mega1_tail-Mega2_tail y Mega2_head-Mega3_head, la última de ellas amplificando una banda de gran tamaño. Esto indicó que el contig 2 estaba invertido, como se muestra en la figura 10.



El gap 9 se amplificó utilizando los oligonucleótidos Mega2_head y Mega3_head, y la enzima LongAmp polimerasa, mediante el siguiente ciclado: Inicio de 94°C por 5 minutos; 37 ciclos de desnaturalización a 94°C por 20 segundos, alineación a 52°C por 20 segundos, y extensión a 65°C por 10 minutos. La banda amplificada fue de ~ 8 kpb (figura 12, carril 4).

No fue posible obtener resultados satisfactorios de las PCRs realizadas para amplificar el gap 8. Se utilizó el programa bioinformático SSPACE para caminar en los gaps del megaplásmido invirtiendo el contig 2. Se obtuvo una nueva estructura, la cual se muestra en la figura 11.



El programa bioinformático SSPACE predijo a los contigs 142 y 144 dentro del gap 8, y al contig 142 invertido en el extremo del contig 2. En esta nueva estructura el gap 8 se compone de tres gaps, a los que se denominó gap 8A1, gap 8A2 y gap 8B. Se diseñaron oligonucleótidos flanqueando dichos gaps y se realizaron PCRs para amplificarlos.

Los gaps 8A1, 8A2 y 8B se amplificaron utilizando la pareja de oligonucleótidos mega1_tail -c142_head, c142_tail-c144_head y c144_tail-mega2_tail, respectivamente. Se utilizó la enzima Taq polimerasa, y el ciclado fue similar al utilizado para amplificar el gap 9, excepto que la temperatura de alineación fue 62°C, para los gaps 8A1 y 8A2, y 60°C para el gap 8B. Las bandas amplificadas fueron de ~ 3.3 kpb, ~ 2.5 kpb y ~ 5.5 kpb, para el gap 8A1, 8A2 y 8B respectivamente (figura 12, carril 1, 2 y 3).

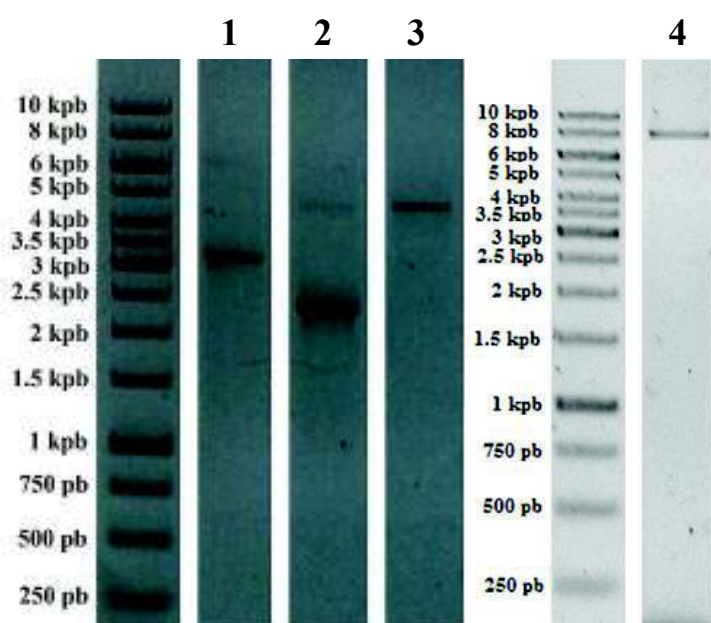


Figura 12. Amplificación de gaps del megaplásmido pSfrGR64c. Carril 1: Gap 8A1, Carril 2: Gap 8A2, Carril 3: Gap 8B, Carril 4: Gap 9.

VII.3.5. Obtención de la secuencia completa del plásmido simbiótico pSfrGR64b

VII.3.5.1. Amplificación del gap por PCR

El gap del plásmido simbiótico pSfrGR64b se denominó reacción o gap 7, y se utilizaron los oligonucleótidos pSymGR64_tail y pSymGR64_head, y la enzima LongAmp polimerasa. Se realizaron más de 10 reacciones de PCR, variando las condiciones, sin obtener resultados satisfactorios.

El extremo tail del gap 7 contiene una gran cantidad de secuencias repetidas, por esta razón el oligonucleótido pSymGR64_tail se diseñó fuera la región repetitiva (~5.5 kpb de distancia del extremo tail). Debido a que no se obtuvieron resultados satisfactorios, se diseñó un segundo oligonucleótido, el pSymGR64_tail2, en una región más cercana a la repetitiva (~2.9 kpb de distancia del extremo tail), con el cual se logró amplificar el gap 7. Mediante el siguiente ciclado: Inicio de 94°C por 5 minutos; 37 ciclos de, desnaturalización a 94°C por 20 segundos, alineación a 54.5°C por 20 segundos, y extensión a 65°C por 10 minutos. La banda amplificada fue de ~ 8 kpb (figura 13).

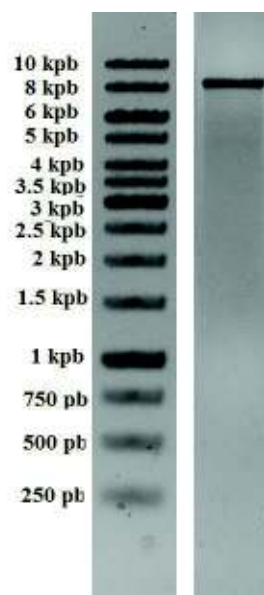
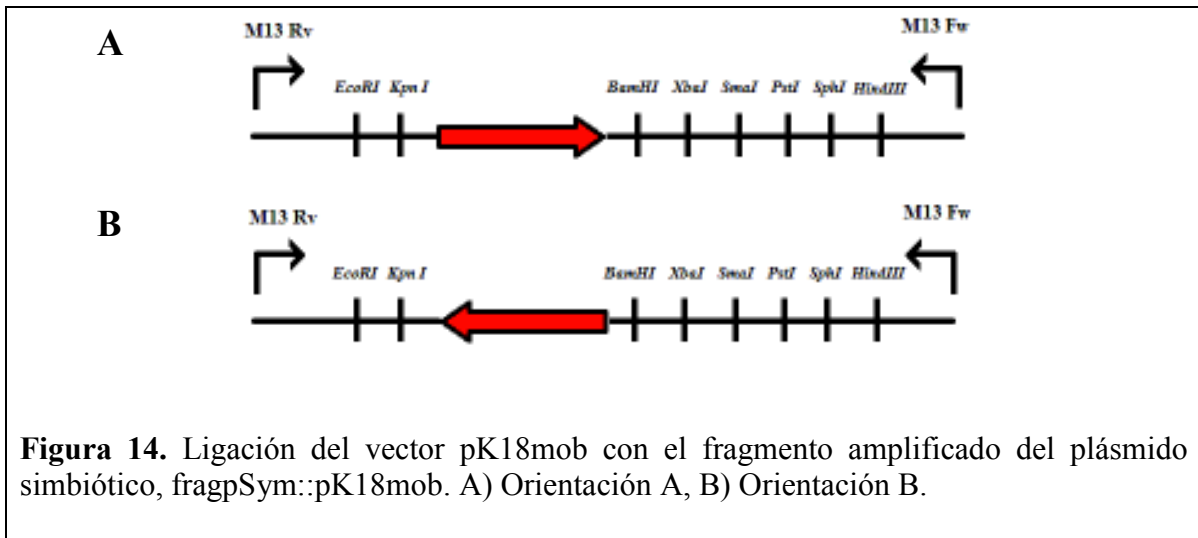


Figura 13. Amplificación del gap del plásmido simbiótico pSfrGR64b.

VII.3.5.2. Caminata genómica

Un fragmento del extremo del contig del plásmido simbiótico pSfrGR64b, se amplificó a partir de los oligonucleótidos pSymGR64_head y pSymGR64_head-fw. La banda obtenida fue de ~ 450 pb y se muestra en la figura 16 (carril 1). El fragmento, denominado “fragpSym”, fue clonado en el vector pK18mob digerido con la enzima *SmaI*. y se verificó la orientación de

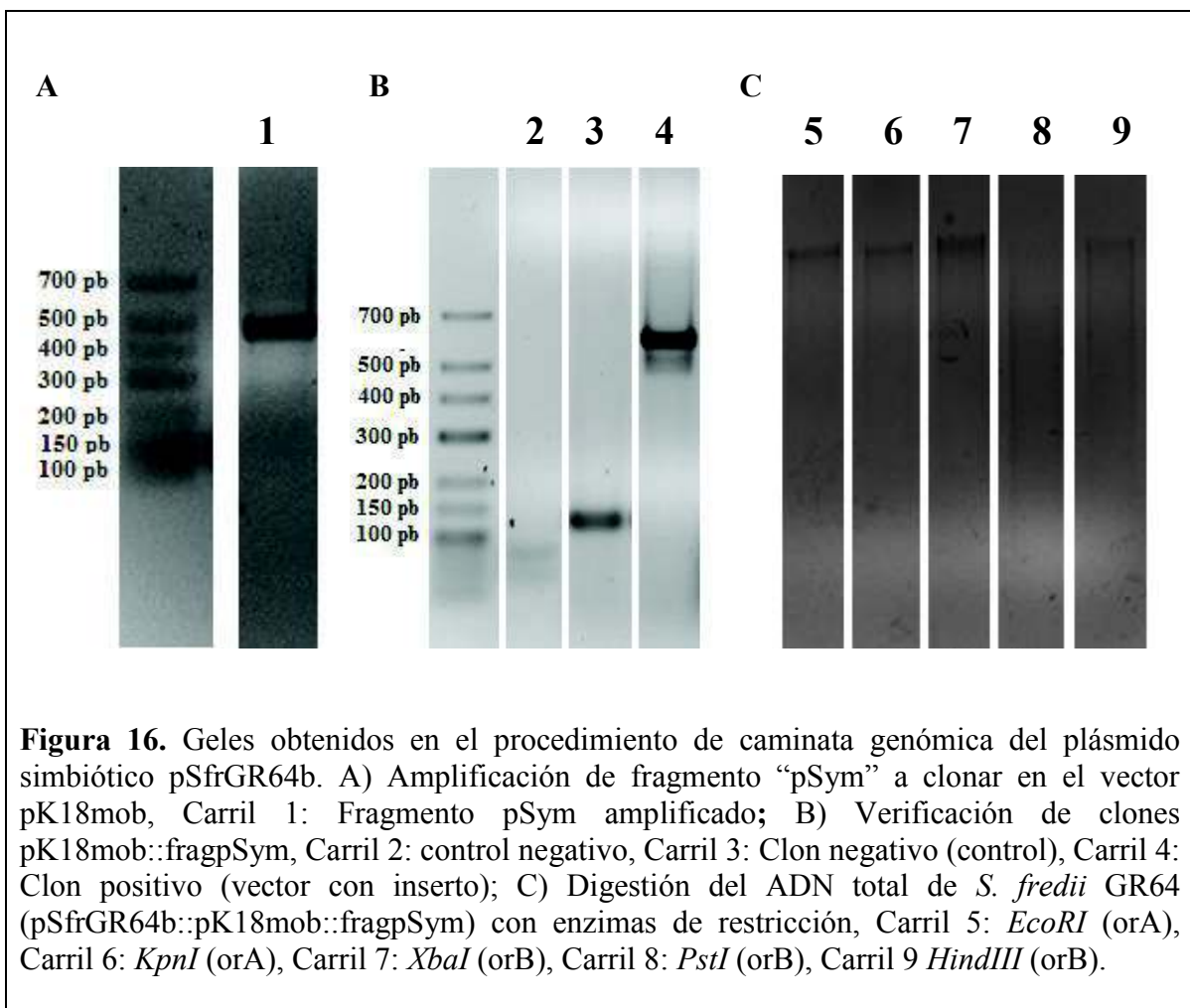
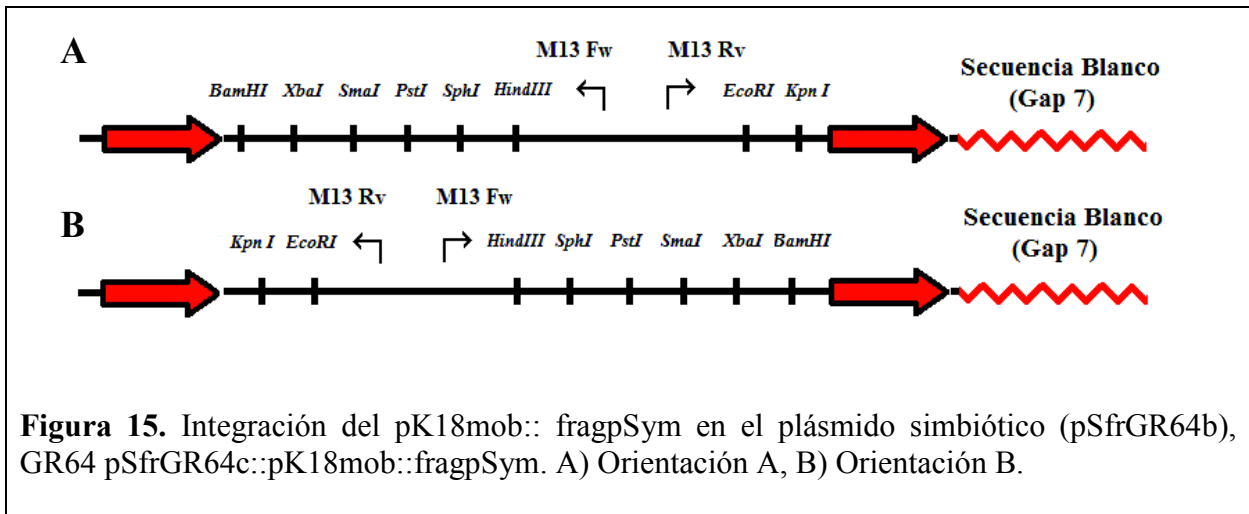
dicho fragmento (Figura 16), mediante digestión con las enzima de restricción *XhoI*. En la orientación A, la enzima corta originando dos bandas, una de 366 pb y otra de 217 pb. En la orientación B, la enzima corta originando dos bandas, una de 329 pb y otra de 254 pb.



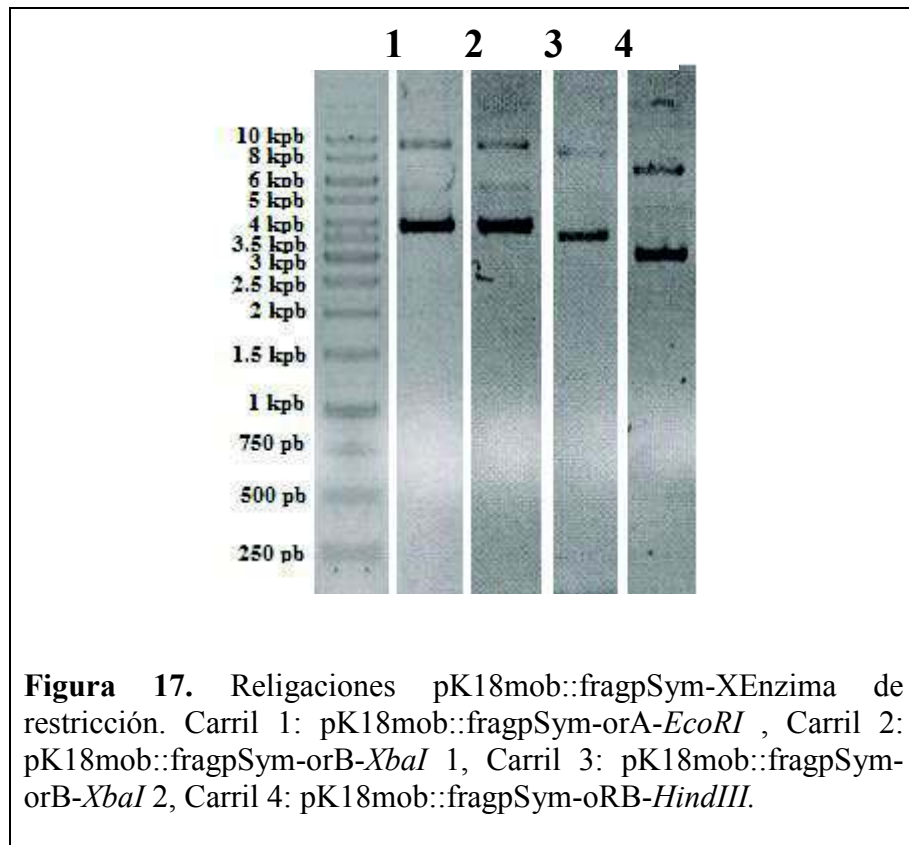
Las ligaciones pK18mob::fragpSym, fueron transformadas en *E. coli* S17-1. Posteriormente este vector fue transferido a una GR64 mediante una conjugación biparental, desde *E. coli* S17-1. Se seleccionó positivamente la integración del vector suicida con el antibiótico neomicina, y se verificó mediante la amplificación por PCR del fragmento fragpSym (Figura 16, carriles 2-4).

Se integró tanto la orientación A como la orientación B del pK18mob::fragpSym (figura 15). En la orientación A, los sitios de restricción cercanos a la secuencia blanco, el gap 7, son sitios *EcoRI* y *KpnI*, y en la orientación B, son sitios *HindIII*, *SphI*, *PstI*, *SmaI*, *XbaI* y *BamHI*.

El ADN total de *S. fredii* con la integración pK18mob::fragpSym (orA) se digirió con las enzimas de restricción *EcoRI* y *KpnI* (figura 16, carriles 5 y 6 respectivamente). El ADN total de *S. fredii* con la integración pK18mob::fragpSym (orB) se digirió con las enzimas de restricción *HindIII*, *PstI* y *XbaI* (figura 16, carriles 7, 8 y 9 respectivamente).

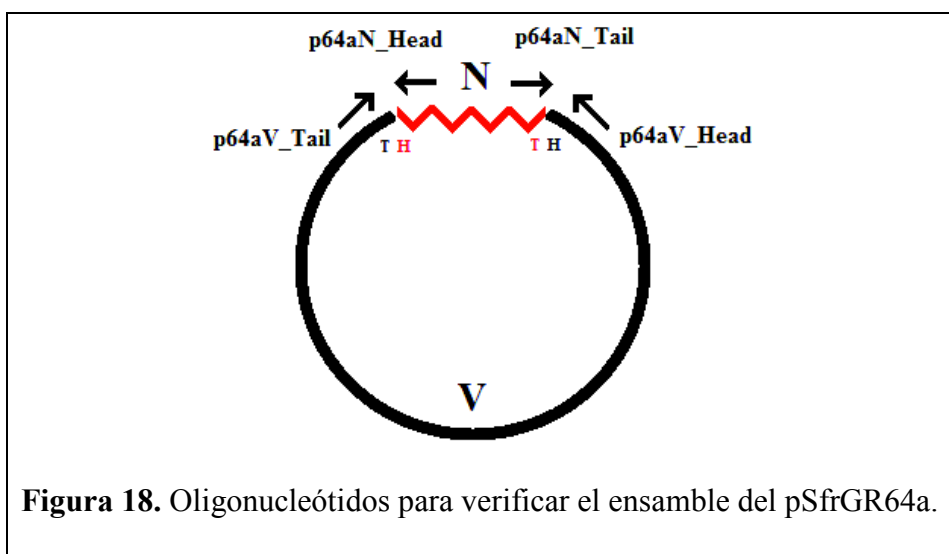


Las digestiones fueron religadas, transformadas en *E. coli* DH5 α y se seleccionaron las religaciones del vector con la secuencia blanco (*gap7*) mediante el antibiótico kanamicina. Se obtuvo la religación pK18mob::fragpSym-OrA-*EcoRI* (~3.5 kpb), pK18mob::fragpSym-OrB-*XbaI* 1 (~3.5 kpb), pK18mob::fragpSym-OrB-*XbaI* 2 (~3.7 kpb) y pK18mob::fragpSym-OrB-*HindIII* (~2.5 kpb) (figura 17). Debido a que los plásmidos obtenidos solamente permitieron recuperar pequeñas regiones genómicas (no mayores a 4 kpb), y el amplicón obtenido por PCR fue de ~8 kpb, los plásmidos no fueron secuenciados ya que la información no nos permitiría avanzar más que el producto de PCR.

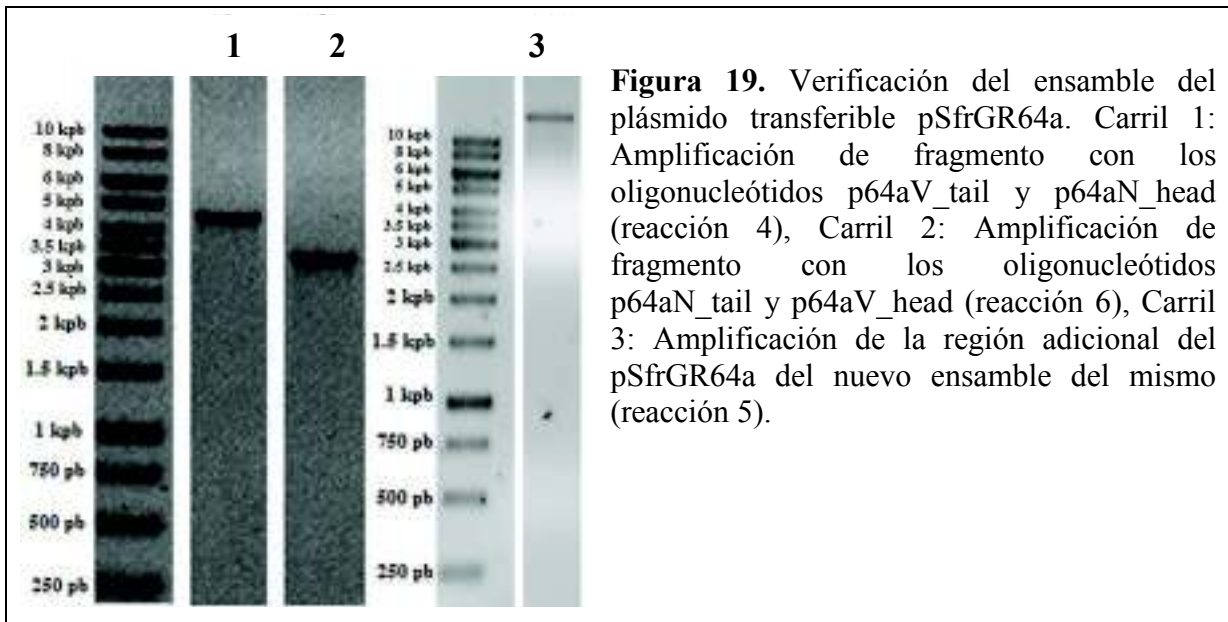


VII.3.6. Verificación del ensamblaje del plásmido transferible pSfrGR64a

Para la verificación del ensamblaje del pSfrGR64a se diseñaron oligonucleótidos en los extremos de la región adicional o región “Nueva” y en los extremos del ensamblaje original o “Viejo”, como se muestra en la figura 18. Se realizaron tres reacciones de PCR: La reacción 4, 5 y 6. La reacción 4 corresponde a la región entre los oligonucleótidos p64aV_tail y p64aN_head. La reacción 5 corresponde a la región nueva, entre los oligonucleótidos p64aV_head y p64aV_tail. Y la reacción 6 corresponde a la región entre los oligonucleótidos p64aN_tail y p64aV_head.



De esta forma, si las reacciones 4 y 6 no amplifican, y la reacción 5 amplifica un fragmento de ~4 kpb, la región adicional de 10 kpb no pertenece a la secuencia del pSfrGR64a. Si las reacciones 4 y 6 amplifican, y la reacción 5 amplifica un fragmento mayor a 10 kpb, la región adicional pertenece a la secuencia del pSfrGR64a, y el ensamblaje obtenido por pirosecuenciamiento es correcto. Se lograron amplificar las regiones 4, 5 y 6 (figura 19).



Las reacciones se amplificaron utilizando la enzima LongAmp polimerasa, mediante el siguiente ciclado: Inicio de 94°C por 5 minutos; 36 ciclos de: desnaturalización a 94°C por 20 segundos, alineación a 60°C por 20 segundos, y extensión a 65°C por 10 minutos. La banda de la reacción 4 amplificada fue de ~ 4.5 kpb (figura 19, carril 1), la de la reacción 6 fue de ~ 3 kpb (figura 19, carril 2) y la reacción 5 que incluye la región adicional del pSfrGR64a fue de ~ 15 kpb (figura 19, carril 3).

VII.4. Obtención de las Lecturas.

Las bandas de gel de los gaps amplificados, el gap 1 (cromosoma), 8A1 (megaplásmido pSfrGR64c), 8A2 (megaplásmido pSfrGR64c), 8B (megaplásmido pSfrGR64c), 9 (megaplásmido pSfrGR64c), 4 (plásmido transferible pSfrGR64a), y 7 (plásmido simbiótico pSfrGR64b) fueron purificadas y enviadas a Macrogen, en Seul Korea, para su secuenciación (figura 20).

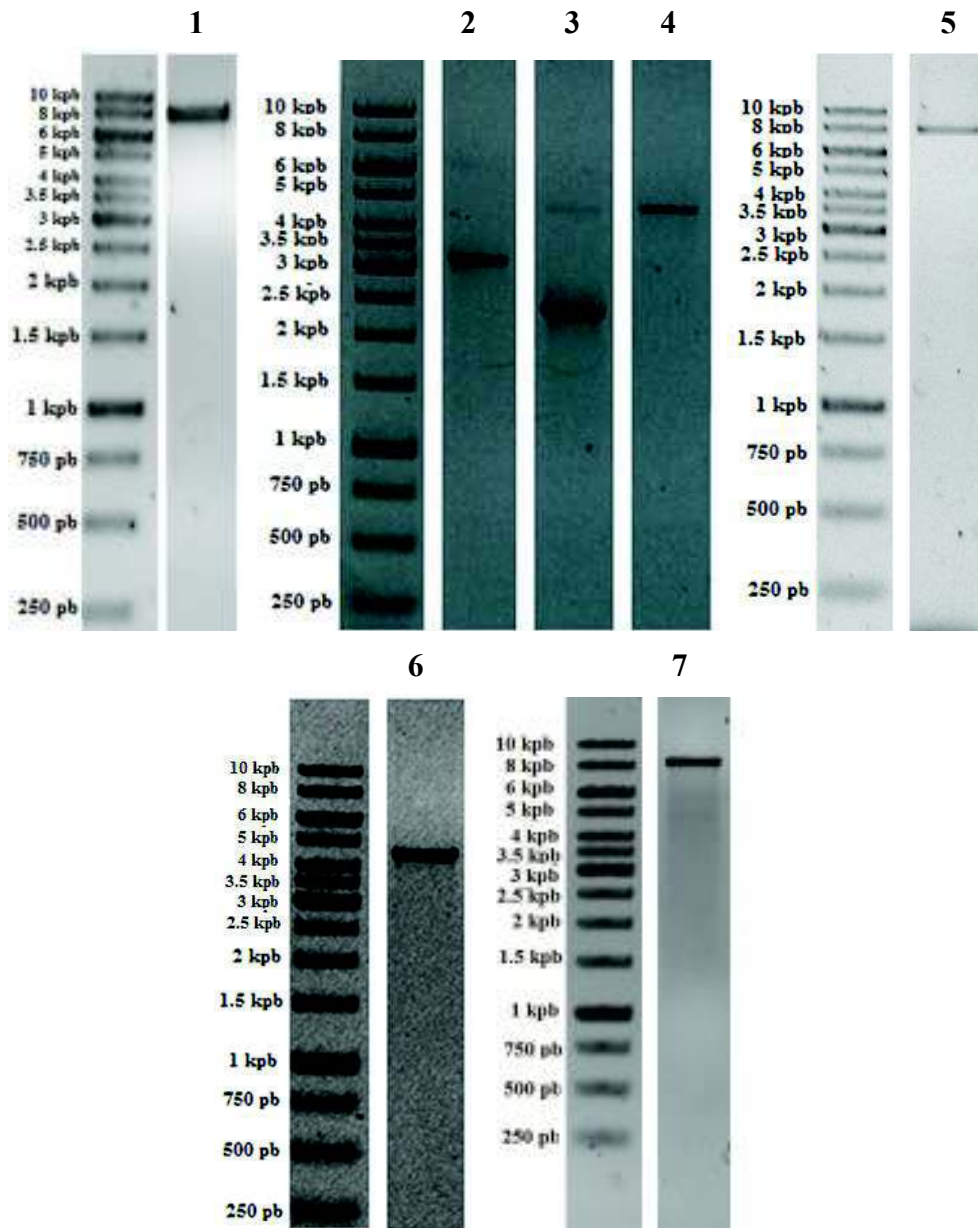


Figura 20. Gaps amplificados. Carril 1: Gap 1 (cromosoma), Carril 2: Gap 8A1 (megaplásmido pSfrGR64c), Carril 3: Gap 8A2 (megaplásmido pSfrGR64c), Carril 4: Gap 8B (megaplásmido pSfrGR64c), Carril 5: Gap 9 (megaplásmido pSfrGR64c), Carril 6: Gap 4 (plásmido transferible pSfrGR64a), Carril 7: Gap 7 (plásmido simbiótico pSfrGR64b).

VII.5. Re-ensamblado de la Secuencias Mediante Phred-Phrap/Consed

La calidad de las lecturas obtenidas fue visualizada en el programa Contig Express dentro del paquete Vector NTI, y posteriormente fueron re-ensambladas mediante el software Phred-Phrap/Consed. Se lograron completar las cuatro moléculas: el cromosoma, el megaplásmido pSfrGR64c, el plásmido simbiótico pSfrGR64b y el plásmido pSfrGR64a (figura 21).

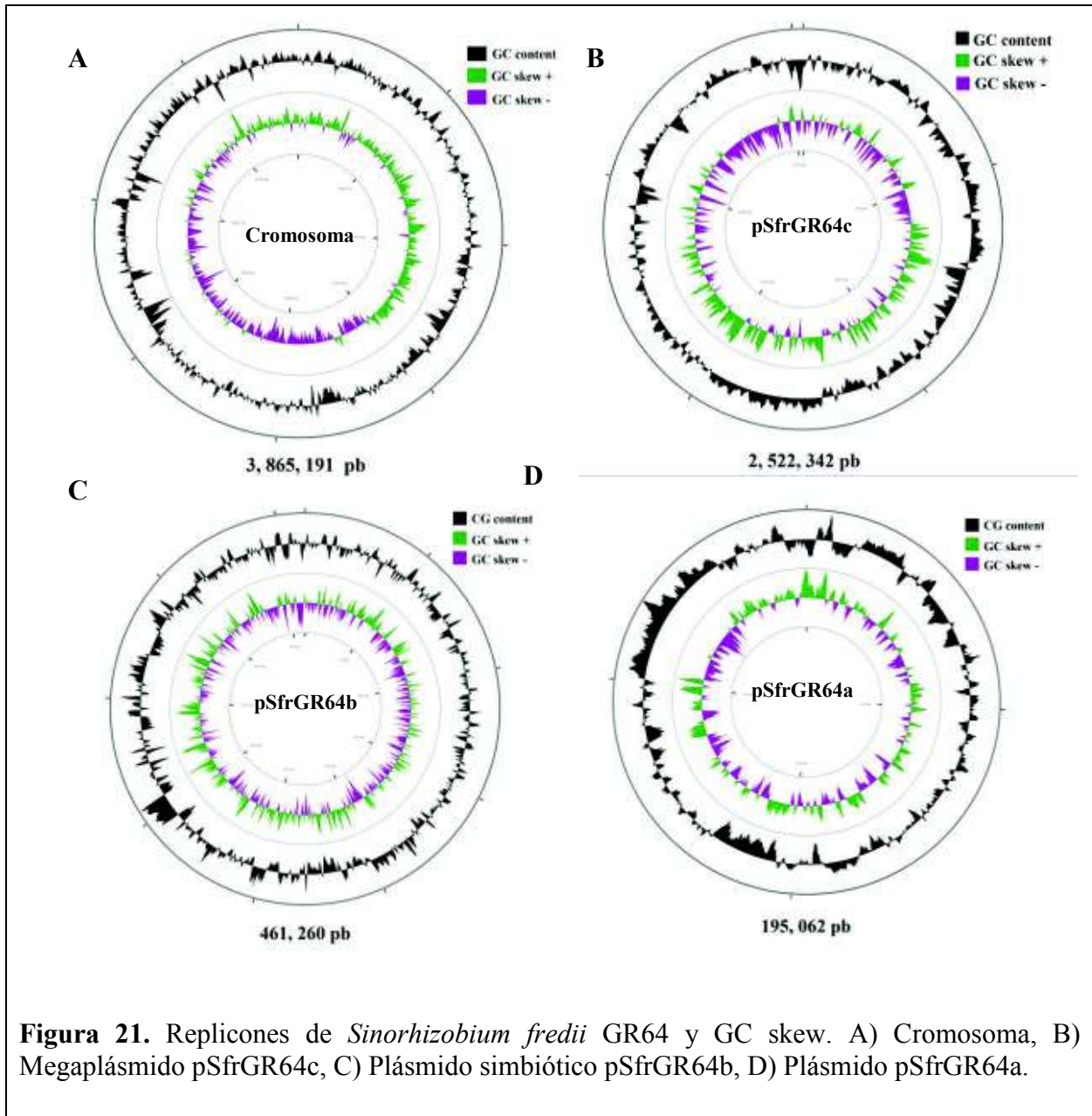


Figura 21. Replicones de *Sinorhizobium fredii* GR64 y GC skew. A) Cromosoma, B) Megaplásmido pSfrGR64c, C) Plásmido simbiótico pSfrGR64b, D) Plásmido pSfrGR64a.

Las moléculas fueron analizadas en los programas en línea CG View Server del Stothard Research Group (disponible es: http://stothard.afns.ualberta.ca/cgview_server/), para visualizar el CG skew y contenido de CG de cada replicón (figura 20), y Genomatix Software Suite (disponible en: <http://www.genomatix.de/cgi-bin//tools/tools.pl?s=21f71472f58539a31c45b18eb46722da;TASK=statistics>), para calcular el porcentaje de CG. El genoma de *Sinorhizobium fredii* GR64 tiene un tamaño de 7,043,855 pb y un contenido de CG de 62.2%. El cromosoma tiene un tamaño de 3,865,191 pb y un contenido de CG de 62.9%. El megaplásmido pSfrGR64c tiene un tamaño de 2,522,342 pb y un contenido de CG de 61.9%. El plásmido simbiótico pSfrGR64b tiene un tamaño aproximado de 465,260 pb y un contenido de CG de 59.1%. Y finalmente, el plásmido pSfrGR64a tiene un tamaño de 191,062 pb y un contenido CG de 59.1 %.

VII.6. Estudio de Sintenia del Genoma de *S. fredii* GR64

VII.6.1. R2cat

VII.6.1.1. Cromosoma de *S. fredii* GR64

La figura 22 muestra la colinealidad entre el cromosoma de *S. fredii* GR64, con los cromosomas de las bacterias filogenéticamente cercanas, *S. fredii* USDA257, *S. fredii* HH103, *Rhizobium* sp. NGR234, *S. meliloti* 1021, *S. medicae* WSM419, *R. leguminosarum* bv. *trifolii* WSM2304, *R. leguminosarum* bv. *viciae* 2841 y *R. etli* CFN42.

Se observó una relación colineal o sintenia, entre el cromosoma de *S. fredii* GR64 y todos los cromosomas comparados, siendo los más colineales o sinténicos el cromosoma de *S. fredii* HH103 y el de *Rhizobium* sp. NGR234, las bacterias filogenéticamente más cercanas a *S. fredii* GR64.

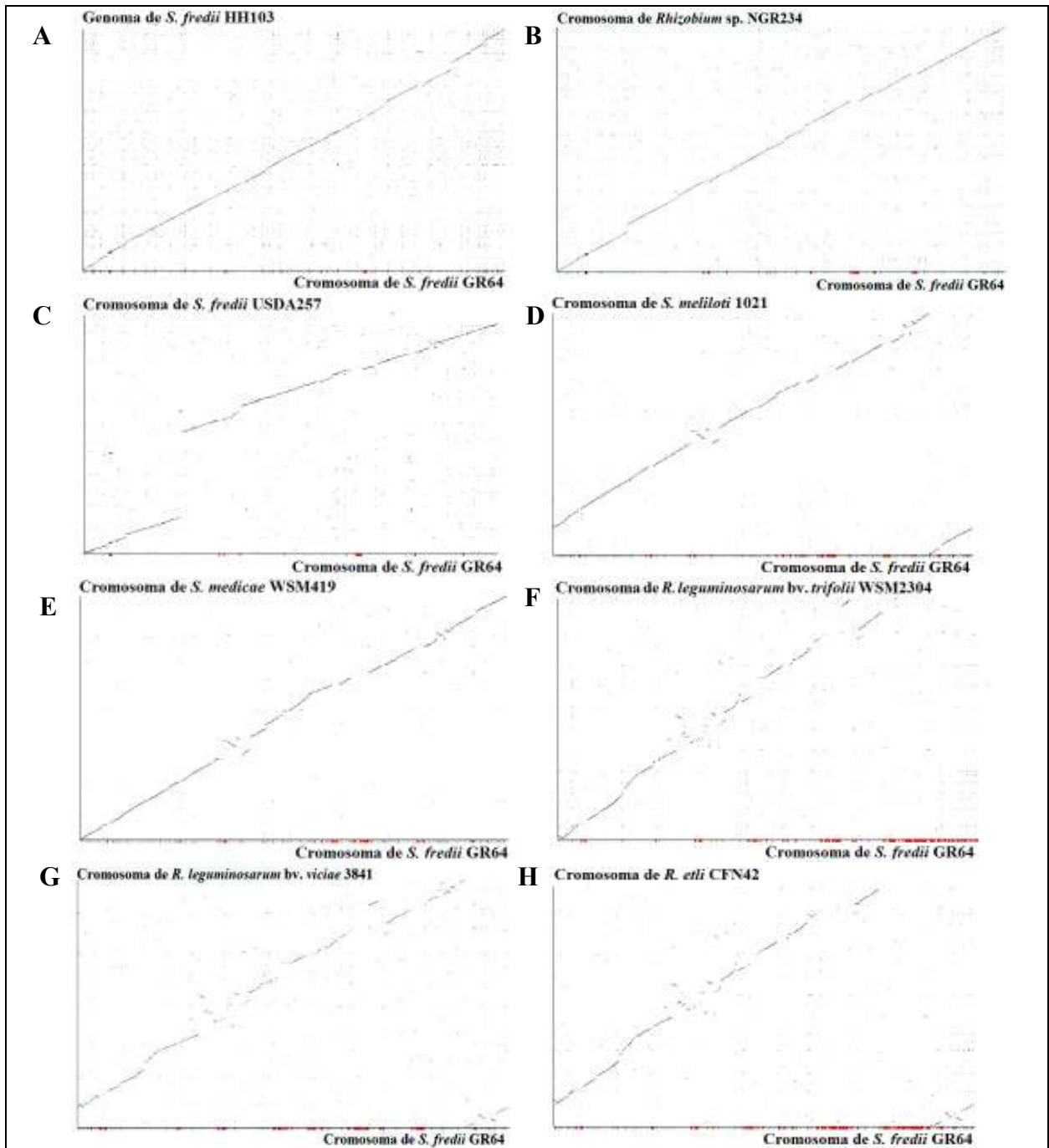


Figura 22. Análisis de sintenia entre el cromosoma de *Sinorhizobium fredii* GR64 y cromosomas de bacterias filogenéticamente cercanas. A) *S. fredii* HH103, B) *Rhizobium* sp. NGR234, C) *S. fredii* USDA257, D) *S. meliloti* 1021, E) *S. medicae* WSM419, F) *R. leguminosarum* bv. *trifolii* WSM2304, G) *R. leguminosarum* bv. *viciae* 2841, H) *R. etli* CFN42.

VII.6.1.2. Megaplásmido pSfGR64c

La figura 23 muestra la colinealidad entre el megaplásmido pSfrGr64c de *S. fredii* GR64, con megaplásmidos de bacterias filogenéticamente cercanas, el pHH103e de *S. fredii* HH103, pNGR234b de *Rhizobium* sp. NGR234, pSMEDO1 de *S. medicae* WSM419 y el pSym B de *S. meliloti* 1021.

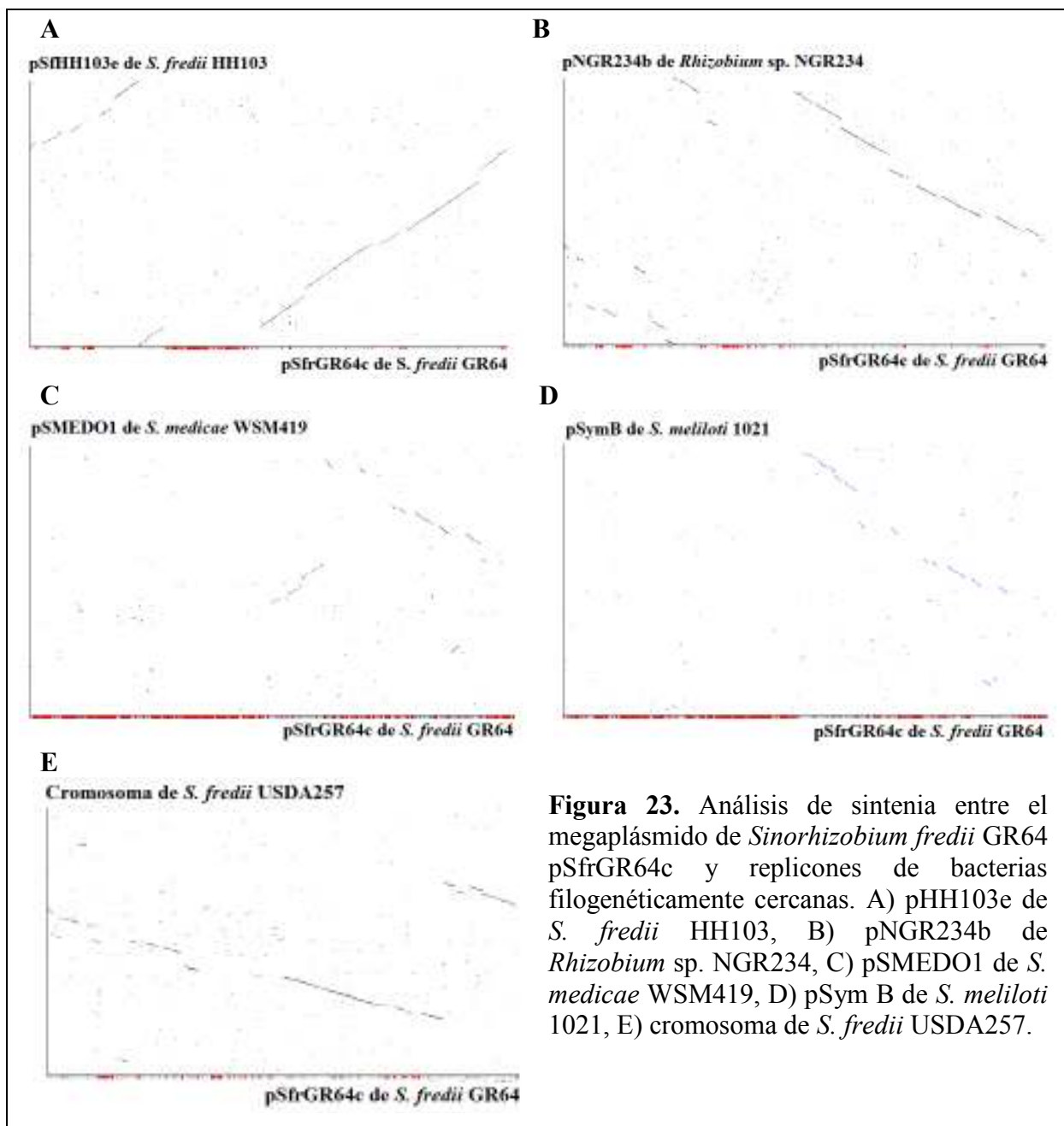


Figura 23. Análisis de sintenia entre el megaplásmido de *Sinorhizobium fredii* GR64 pSfrGR64c y replicones de bacterias filogenéticamente cercanas. A) pHH103e de *S. fredii* HH103, B) pNGR234b de *Rhizobium* sp. NGR234, C) pSMEDO1 de *S. medicae* WSM419, D) pSym B de *S. meliloti* 1021, E) cromosoma de *S. fredii* USDA257.

El análisis reveló regiones colineales entre el megaplásmido pSfrGR64c y otros megaplásmidos analizados. El pSfrGR64c presenta un mayor grado de colinealidad o sintenia, con los plásmidos pSfHH103e y pNGR234b. Mientras que los plásmidos pSMEDO1 y pSymB presentan sólo algunas regiones colineales con el megaplásmido pSfrGR64c.

Se realizó un análisis de colinealidad entre el megaplásmido pSfrGR64c y los cromosomas utilizados para analizar la colinealidad con el cromosoma de *S. fredii* GR64: *S. fredii* USDA257, *S. fredii* HH103, *Rhizobium* sp. NGR234, *S. meliloti* 1021, *S. medicae* WSM419, *R. leguminosarum* bv. *trifolii* WSM2304, *R. leguminosarum* bv. *viciae* 2841 y *R. etli* CFN42. Se observó una región colineal entre el megaplásmido y el cromosoma de *S. fredii* USDA257 (figura 23, imagen E).

VII.6.1.3. Plásmido simbiótico pSfrGR64b

La figura 24 muestra la colinealidad entre el plásmido simbiótico pSfrGr64b de *S. fredii* GR64, con plásmidos simbióticos de bacterias filogenéticamente cercanas, el pUSDA257 de *S. fredii* USDA257, pNGR234a de *Rhizobium* sp. NGR234, pSMEDO2 de *S. medicae* WSM419, pSymA de *S. meliloti* 1021, pRLG201 de *R. leguminosarum* bv. *trifolii* WSM2304 y el pRetCFN42d de *R. etli* CFN42.

No se observó colinealidad o sintenia, entre el plásmido simbiótico de *S. fredii* GR64, pSfrGR64b y todos los plásmidos simbióticos analizados. También se realizó un análisis de colinealidad entre el plásmido simbiótico pSfrGR64b y los cromosomas utilizados para analizar la colinealidad con el cromosoma de *S. fredii* GR64, antes mencionados. No se observó colinealidad, ni regiones colineales, entre el plásmido simbiótico y los cromosomas.

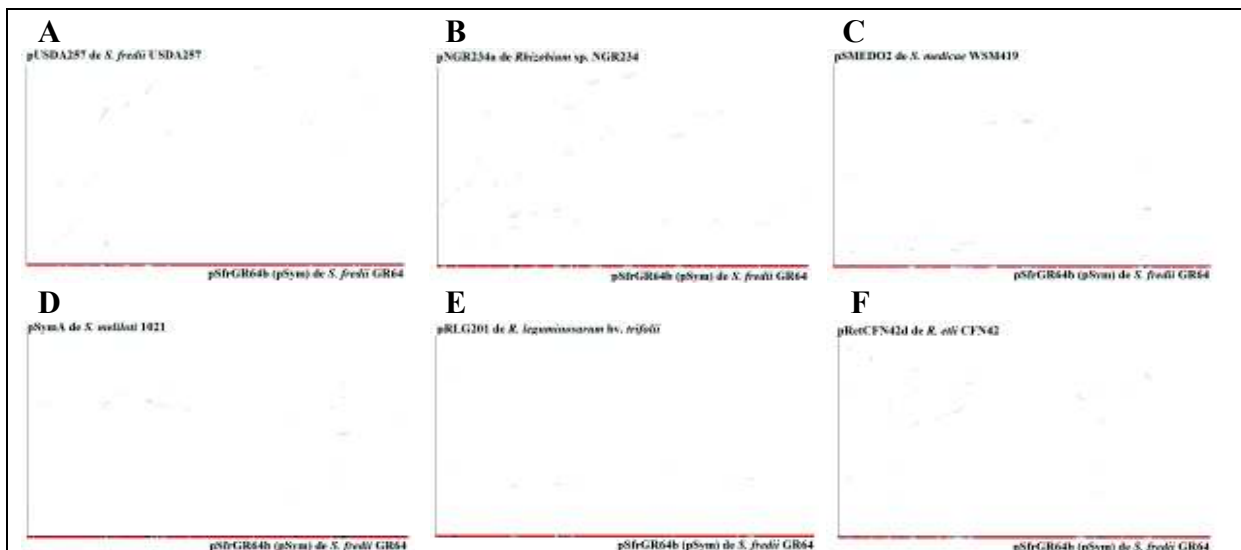
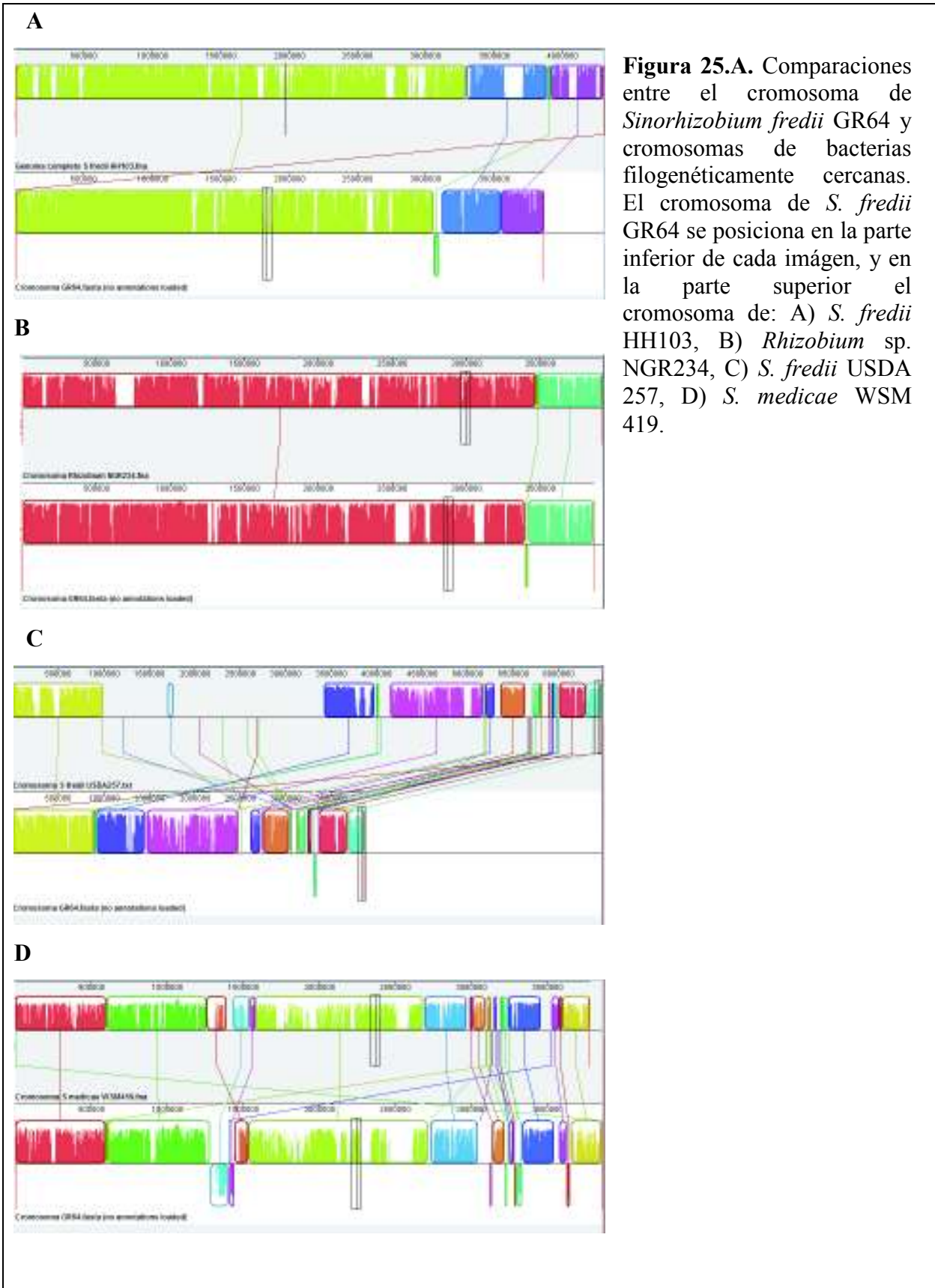


Figura 24. Análisis de sintenia entre el plásmido simbiótico pSfrGR64b de *Sinorhizobium fredii* GR64 y plásmidos simbióticos de bacterias filogenéticamente cercanas. Alineación del pSfrGR64b con los plásmidos: A) pUSDA257 de *S. fredii* USDA257, B) pNGR234a de *Rhizobium* sp. NGR234, C) pSMEDO2 de *S. medicae* WSM419, D) pSymA de *S. meliloti* 1021, E) pRLG201 de *R. leguminosarum* bv. *trifolii* WSM2304, F) pRetCFN42d de *R. etli* CFN42. **No se observó colinealidad o sintenia, entre el plásmido simbiótico de *S. fredii* GR64, pSfrGR64b y todos los plásmidos simbióticos analizados.**

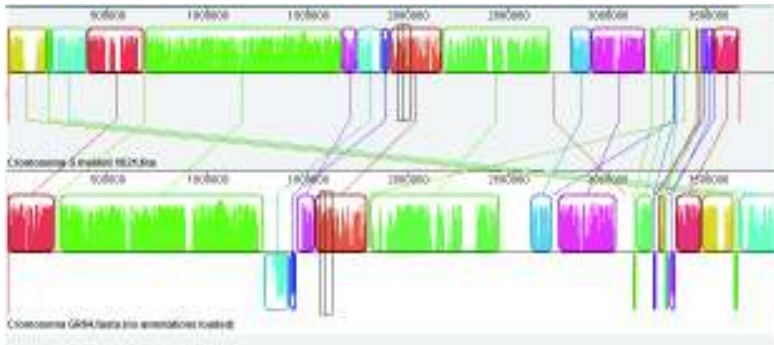
VII.6.2. Mauve

VII.6.2.1. Cromosoma de *S. fredii* GR64

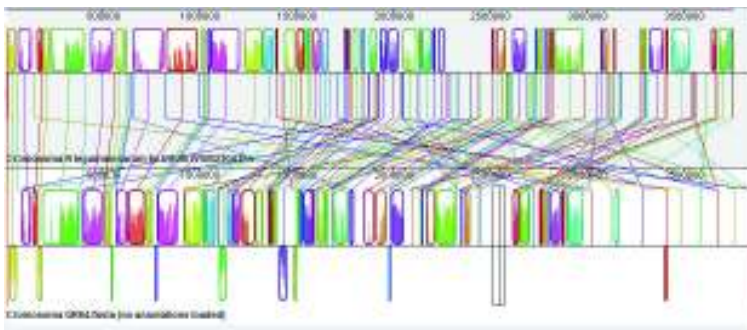
Las figuras 25.A y 25.B muestran la colinealidad y los bloques sinténicos del cromosoma de *S. fredii* GR64, al compararlo con los cromosomas de las bacterias filogenéticamente cercanas, *S. fredii* USDA257, *S. fredii* HH103, *Rhizobium* sp. NGR234, *S. meliloti* 1021, *S. medicae* WSM419, *R. leguminosarum* bv. *trifolii* WSM2304, *R. leguminosarum* bv. *viciae* 2841 y *R. etli* CFN42.



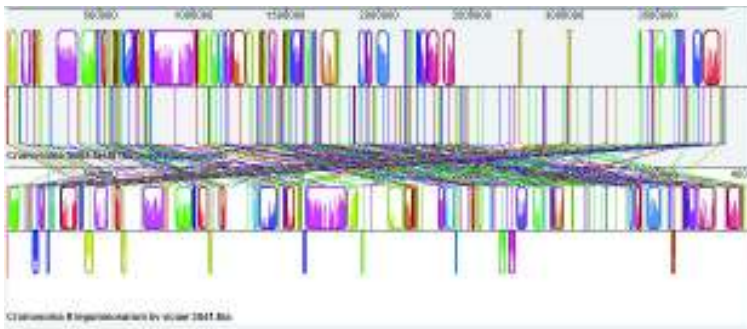
E



F



G



H

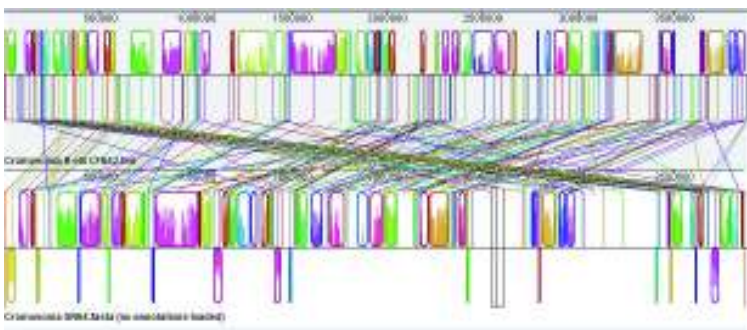


Figura 25.B. Comparaciones entre el cromosoma de *Sinorhizobium fredii* GR64 y cromosomas de bacterias filogenéticamente cercanas. El cromosoma de *S. fredii* GR64 se posiciona en la parte inferior de cada imagen, y en la parte superior el cromosoma de: E) *S. meliloti* 1021, F) *R. leguminosarum* bv. *trifolii* WSM2304, G) *R. leguminosarum* bv. *viciae* 3841, H) *R. etli* CFN42.

Los cromosomas de *S. fredii* HH103 y *Rhizobium* sp. NGR234 exhiben el mayor grado de sintenia con el cromosoma de *S. fredii* GR64, conformados por grandes bloques sinténicos, seguido por los de *S. fredii* USDA257, *S. medicae* WSM 419, *S. meliloti* 1021, con un menor grado de sintenia y mayor cantidad de bloques sinténicos. Finalmente, los cromosomas de *R. leguminosarum* bv. *trifolii* WSM2304, *R. leguminosarum* bv. *viciae* 3841, y *R. etli* CFN42, presenten el menor grado de sintenia, con respecto al cromosoma de *S. fredii* GR64, de todos los cromosomas analizados, exhibiendo una gran cantidad de bloques sintéticos que conforman una estructura mosaico.

VII.6.2.2. Megaplásmido pSfrGR64c

La figura 26 muestra la colinealidad y bloques de sintenia del megaplásmido pSfrGR64c de *S. fredii* GR64, al compararlo con replicones de bacterias filogenéticamente cercanas, el pHH103e de *S. fredii* HH103, pNGR234b de *Rhizobium* sp. NGR234, pSym B de *S. meliloti* 1021, pSMEDO1 de *S. medicae* WSM419 y el cromosoma de *S. fredii* USDA257.

Los megaplásmidos comparados con el megaplásmido pSfrGR64c presentan un mayor grado de sintenia con respecto a las comparaciones entre plásmidos simbióticos, pero mucho menor con respecto a las comparaciones entre cromosomas. Los megaplásmidos analizados, exhiben una gran cantidad de bloques sintéticos, los cuales están dispersos sin orden conformando una estructura mosaico. El megaplásmido pSfrHH103 de *S. fredii* HH103, es el plásmido que exhibe mayor sintenia con respecto al pSfrGR64c, presentando una menor cantidad de bloques sinténicos, los cuales son más largos y presentan más orden que los bloques sintéticos exhibidos por los otros plásmidos analizados.

En análisis de la región colineal entre el pSfrGR64c y el cromosoma de *S. fredii* USDA257 reveló que dicha región está conformada por una gran cantidad de bloques sinténicos cortos, los cuales no están alineados en orden, sino dispersos en la región sinténica. Muchos de estos bloques sinténicos, también están invertidos.

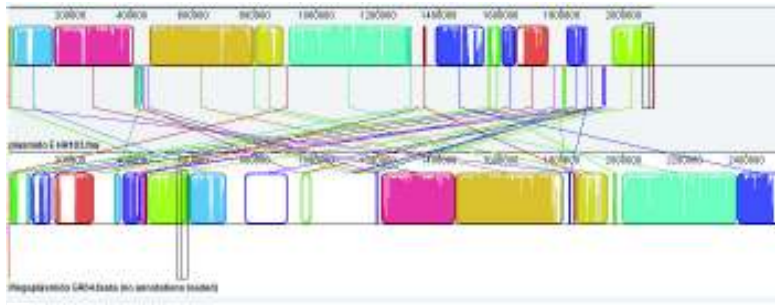
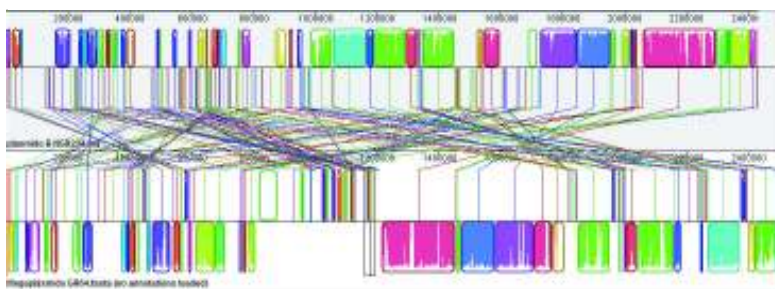
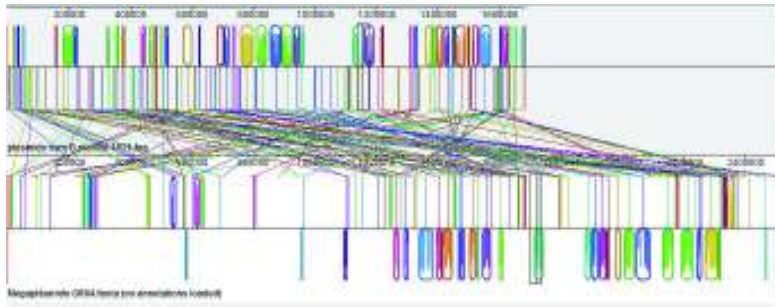
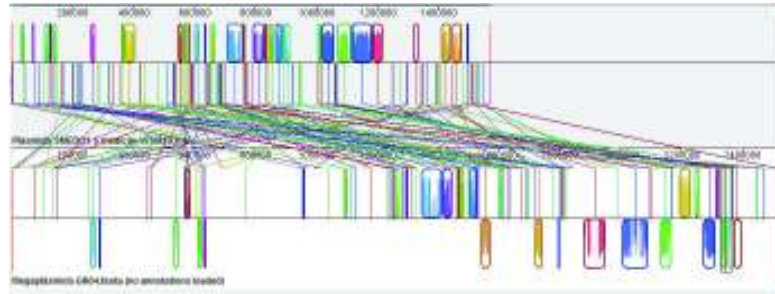
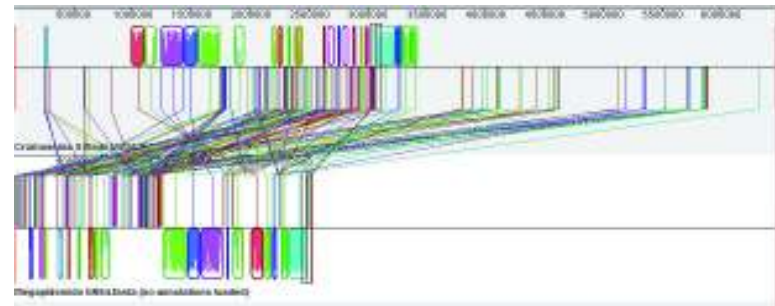
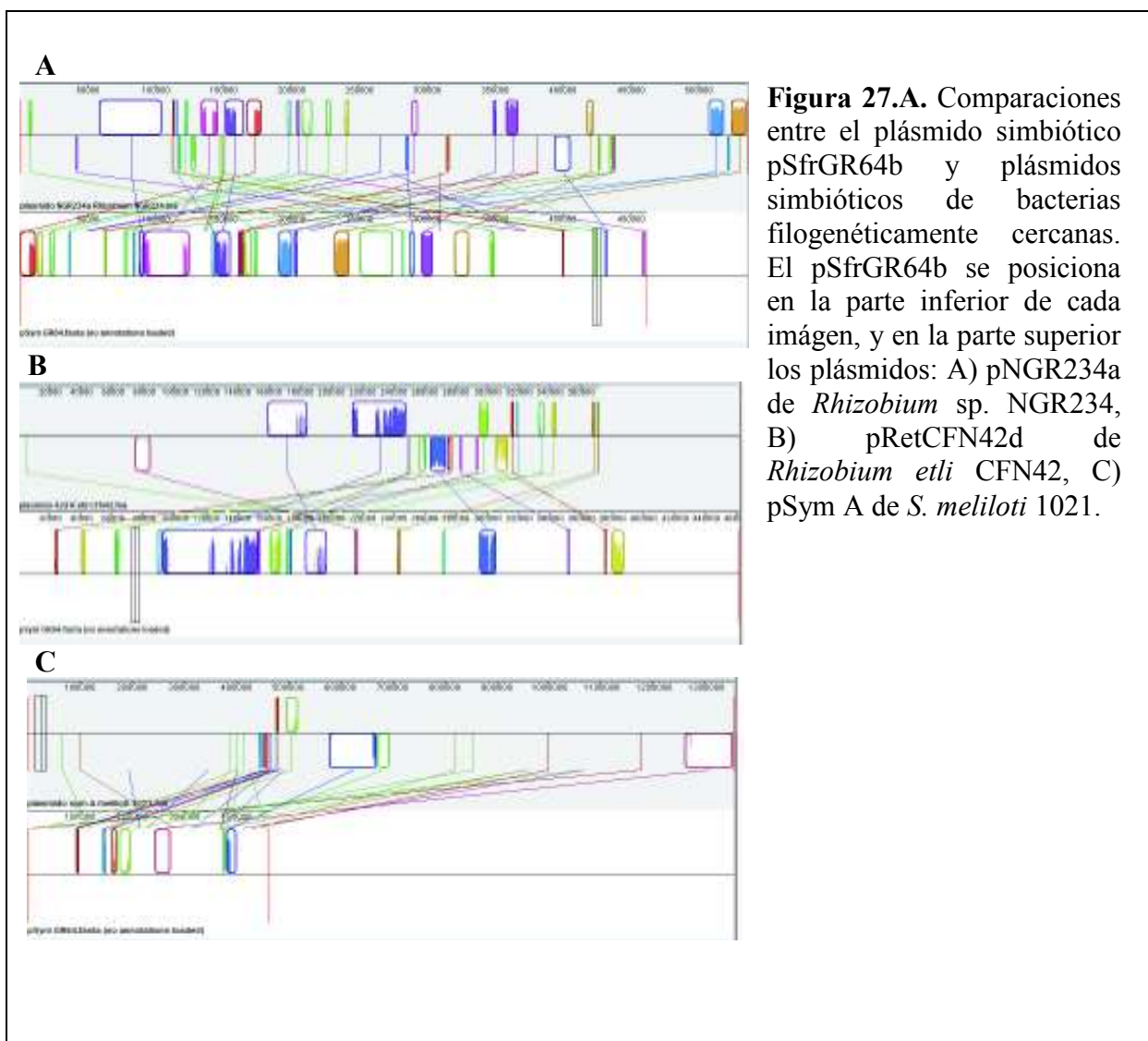
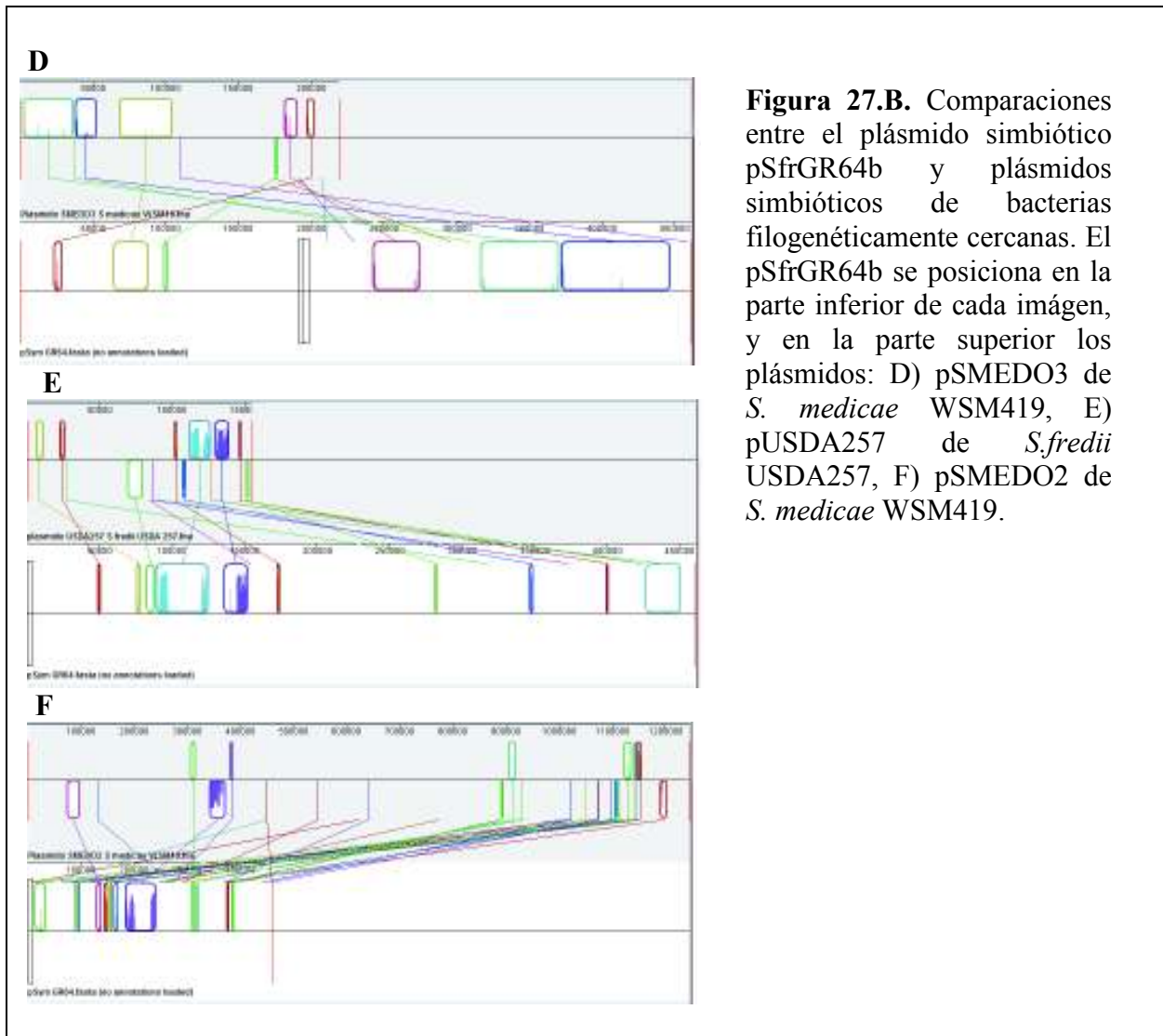
A**B****C****D****E**

Figura 26. Comparaciones entre el megaplásmido pSfrGR64c de *Sinorhizobium fredii* GR64 y replicones de bacterias filogenéticamente cercanas. El pSfrGR64c se posiciona en la parte inferior de cada imagen, y en la parte superior se posicionan los replicones: A) pHH103e de *S. fredii* HH103, B) pNGR234b de *Rhizobium* sp. NGR234, C) pSym B de *S. meliloti* 1021, D) pSMEDO1 de *S. medicae* WSM419, E) cromosoma de *S. fredii* USDA257.

VII.6.2.3. Plásmido simbiótico pSfrGR64b

La figuras 27.A y 27.B muestran la colinealidad y bloques de sintenia del plásmido simbiótico pSfrGR64b de *S. fredii* GR64, al compararlo con plásmidos simbióticos de bacterias filogenéticamente cercanas, el pNGR234a de *Rhizobium* sp. NGR234, pRetCFN42d de *Rhizobium etli* CFN42, pSym A de *S. meliloti* 1021, pSMEDO3 de *S. medicae* WSM419, pUSDA257 de *S.fredii* USDA257 y pSMEDO2 de *S. medicae* WSM419.





Los plásmidos analizados no muestran sintenia con respecto al plásmido simbiótico pSfrGR64b, exhibiendo sólo algunos bloques sinténicos dispersos. Los plásmidos que exhiben una mayor cantidad de bloques sinténicos, son los plásmidos pNGR234b y el pRetCFN42d de *Rhizobium* sp. NGR234 y *Rhizobium etli* CFN42 respectivamente, sin embargo, dichos bloques están dispersos sin orden en dichos plásmidos.

VIII. DISCUSIÓN

Existen varios problemas prácticos al ensamblar genomas, generalmente ocasionados por las secuencias repetitivas en el genoma, la baja cobertura de secuencias y errores en las reacciones de secuenciación (Rodríguez-Martínez, 2004). Frente a estos problemas, los programas bioinformáticos son incapaces de obtener moléculas sin gaps. Programas como SSPACE e IMAGE son capaces de cubrir la mayor cantidad de gaps posibles, sin embargo, quedan regiones gaps que deben obtenerse mediante métodos moleculares. Al llevar a cabo los procedimientos moleculares deben tenerse presentes dichos problemas prácticos.

En ocasiones, cuando regiones repetitivas son procesadas por programas de ensamblado, estos son incapaces de integrar dichas regiones en el ensamblado, lo cual no sólo genera gaps, sino que no permite obtener el orden de las regiones repetidas. Esto representa un problema para obtener el orden correcto de regiones tales como la región 16S ribosomal, que en nuestro caso está repetida tres veces en el genoma de la bacteria. Para solucionar esta complicación, basta con apegarse al orden de dicha región en genomas de otras bacterias. Dichas regiones codifican para el ARN 16S ribosomal, una molécula con varias funciones vitales para todos los seres vivos, por lo cual está ampliamente distribuida. Siendo este un componente de un sistema auto-replicativo, se aísla fácilmente, por lo cual su secuencia ha cambiado muy poco a través del tiempo, es decir, es una región altamente conservada (Woese y Fox, 1977). Para solucionar este inconveniente, verificamos mediante PCR y secuenciación que dichas regiones se encontraban en el orden predicho.

Se determinó que la secuencia de plásmido pSfrGR64a, publicado por Cervantes et al. en el 2011, carece de una región de ~10 kbp. La secuencia de 10 kbp extra y las regiones que lo flanquean no son secuencias repetitivas, por lo cual el error probablemente puede deberse a una baja cobertura en la secuenciación previa del plásmido o errores generados debido al uso de un transposon Tn5 para marcar dicho plásmido y a la pérdida de dicha región por la integración del transposon.

Otra dificultad que se enfrentó fue dilucidar la secuencia completa del megaplásmido pSfrGR64c. Como ya se mencionó anteriormente, la movilidad del megaplásmido pSfrGR64c no había sido estudiado, pero era de esperarse que por su tamaño este plásmido no pudiera

movilizarse por sí mismo. Por esta misma razón, aun utilizando plásmidos helper, se complicó la obtención de la cepa GMI9023 (pSfrGR64c), al igual que la verificación de su transferencia hacia GMI9023. No fue posible visualizarlo en los perfiles plasmídicos de los geles de lisis *in situ*, esto debido a que usualmente los plásmidos de gran tamaño son difíciles de visualizar ya que se rompen por fuerzas mecánicas (Eckhardt, 1978).

Se enfrentaron dificultades para obtener los gaps del megaplásmido pSfrGR64c, mayormente debido a que el tamaño de los mismos era mayor al esperado. Existen programas que predicen la orientación de los contigs, sin embargo, en ocasiones la predicción de estos programas no es certera cuando los contigs están muy separados entre sí. Entonces, es necesario verificar la orientación de los contigs mediante PCRs combinatorias. De esta forma se confirmó que el contig 2 del megaplásmido pSfrGR64c está en sentido inverso al predicho por el programa SSPACE. Sin embargo, después de realizar diversas reacciones de PCR variando las condiciones, no se obtuvieron resultados satisfactorios para amplificar el gap 8. Al posicionar el contig 2 en el sentido correcto, SSPACE fue capaz de predecir una nueva estructura con la cual fue posible obtener la secuencia completa del gap 8, la cual no había podido ser amplificada ya que tiene un tamaño de ~ 18.5 kpb (11.3 kpb de los gaps 8A1, 8A2 y 8B, 2.2 kpb del contig 144 y 2.5 kpb del contig 142 el cual esta repetido).

Las secuencia repetitivas presentes en el plásmido simbiótico pSfrGR64b fueron motivo de conflicto para obtener la secuencia del gap en dicho plásmido (gap 7). El extremo cola del contig del plásmido simbiótico es una región altamente repetitiva, por esta razón, se diseñó el oligonucleótido, pSymGR64_tail, fuera de la dicha región repetitiva, a una distancia de ~5.5 kpb del final del contig. Sin embargo, no fue posible amplificar el gap tras realizar varias PCRs modificando las condiciones, por lo cual se diseñó un segundo oligonucleótido más cercano al final del extremo cola, el pSymGR64_tail2, estaba a una distancia de ~2.9 del final, con el cual fue posible amplificar el gap, con un tamaño de ~8 kpb, de los cuales ~4 kpb corresponden a la secuencia del gap. Con el primer oligonucleótido en el extremo cola, GR64pSym_Tail, el producto final sería de ~10.5 kpb, tamaño mayor al que se tenía considerado. Por otro lado, mediante la caminata genómica no fue posible recuperar regiones genómicas mayores a 1 kpb de secuencia blanco (gap 7), cuando era necesario más de 4 kpb para obtener la secuencia del gap. Lo cual puede deberse a que sitios de restricción, de las enzimas utilizadas, en la región blanco se encuentran muy cerca del vector integrado.

Obtenidas las secuencias completas de los replicones, se procedió a su análisis, en el cual se exhiben las principales características de los genomas de rizobios. El contenido de CG del cromosoma (62.9%) fue mayor que el calculado para el genoma completo (62.2%). Como era de esperarse los plásmidos, pSfrG64a (59.1%) y pSfrGR64b (59.1%), tuvieron un contenido de CG menor al del calculado para el genoma completo, aunque el pSfrGR64c (61.9%) muestra un contenido de CG mayor que los otros plásmidos, casi asimilándose al calculado para el genoma completo. El GC skew del cromosoma exhibe el patrón típico de los cromosomas bacterianos (Lobry, 1996), debido a que son moléculas altamente conservadas, la mitad de la molécula presenta un GC skew positivo y la otra mitad un GC skew negativo, definiendo el origen de replicación del cromosoma. También puede observarse un patrón de GC skew de una molécula conservada en el megaplásmido pSfrGR64c, la mitad de la molécula presenta un GC skew positivo y la otra mitad un GC skew negativo, aunque no se observan un origen de replicación tan definido como en el cromosoma. Dicho patrón no se observa en los plásmidos pSfrGR64a y pSfrGR64b, en los cuales las regiones negativas y positivas de CG skew no se disponen en dos mitades del replicón, sino que se encuentran dispersas, lo cual sugiere que dichos replicones tienen un origen reciente (Lobry, 1996), reforzando la teoría de que ambos fueron adquiridos mediante un evento de transferencia horizontal.

Las comparaciones entre el cromosoma de *S. fredii* GR64 y cromosomas de otras bacterias revelaron la gran conservación de esta molécula, no sólo en el contenido de genes sinténicos, sino en el orden de los mismos. Esta es una de las principales características de los genomas de rizobios, los cuales tienden a tener un mayor contenido de genes conservados y en orden que otros replicones (Pappas y Cevallos, 2011).

Se observó un mayor grado de conservación entre cromosomas de bacterias filogenéticamente más cercanas, *S. fredii* HH103 y *Rhizobium* sp. NGR234, con una estructura conformada por bloques sinténicos de gran tamaño y orden. Se observó que la sintenia disminuyó con respecto a mayores distancias filogenéticas entre las bacterias, hasta exhibir una estructura conformada por una gran cantidad de bloques sintéticos que conforman una estructura mosaico.

Al comparar al plásmido simbiótico pSfrGR64b con otros plásmidos simbióticos no se observó sintenia, exhibiendo sólo algunos bloques sinténicos los cuales están muy dispersos.

Otra de las características de los genomas de rizobios, es que los plásmidos tienen pequeños bloques de genes sinténicos y abundantes genes accesorios (Pappas y Cevallos, 2011). Aunque la fijación de nitrógeno es un carácter complejo, requiere la participación de alrededor de 60 genes y de forma interesante, se ha encontrado que los sectores relevantes no son contiguos, sino que están dispersos en los pSyms de los rizobios. Los grandes tamaños de pSyms les permiten codificar para otras funciones además de la simbiosis (Romero y Brom, 2004). Interesantemente, los plásmidos que exhiben una mayor cantidad de bloques sinténicos son los plásmidos pNGR234b y pRetCFN42d de las bacterias *Rhizobium* sp. NGR234 y *Rhizobium etli* CFN42, siendo *Rhizobium* sp. NGR234 filogenéticamente cercana, y *Rhizobium etli* CFN42 la bacteria analizada filogenéticamente más alejada de *S. fredii* GR64. Como se ha mencionado anteriormente, se considera que los plásmidos forman el paquete genético flexible en rizobios, por lo cual estos no necesariamente reflejan la filogenia de los rizobios, sino la relación con su hospedero (Lloret y Martínez-Romero, 2005). La región sinténica entre *S. fredii* GR64 y *R. etli* CFN42 seguramente corresponde a genes de simbiosis, por lo cual ambas bacterias son capaces de nodular a la misma leguminosa, *Phaseolus vulgaris*, independientemente de su relación filogenética. Estos genes debieron ser adquiridos por *S. fredii* GR64, sin embargo, los resultados no muestran evidencia de un evento de transferencia horizontal reciente, ya que los bloques sinténicos están muy dispersos y desordenados; pero sugieren eventos de recombinación intragenómica, en los cuales permanecieron dichos bloques.

Por último, las comparaciones entre el megaplásmido pSfrGR64c y otros replicones revelaron cosas interesantes. Se observó un grado de sintenia considerable entre el megaplásmido y otros megaplásmidos analizados, cuando lo que se esperaba es que no exista sintenia entre plásmidos. Los genomas de rizobios se caracterizan por tener plásmidos de gran tamaño, para los cuales se vuelve conflictivo ser categorizados como plásmidos, ya que no cumplen con una característica de estos, que es ser dispensables. Muchos autores denominan como “cromosomas secundarios” a plásmidos de rizobios que son recalcitrantes a la curación, tales como el pSymB de *S. meliloti* y p42e de *Rhizobium etli* CFN42, sugiriendo que codifican componentes esenciales para la viabilidad de la célula (Pappas y Cevallos, 2011). Por lo cual, la adquisición de genes esenciales es probablemente el pre-requisito para el origen de cromosomas secundarios. Los cromosomas secundarios surgen cuando sets de genes,

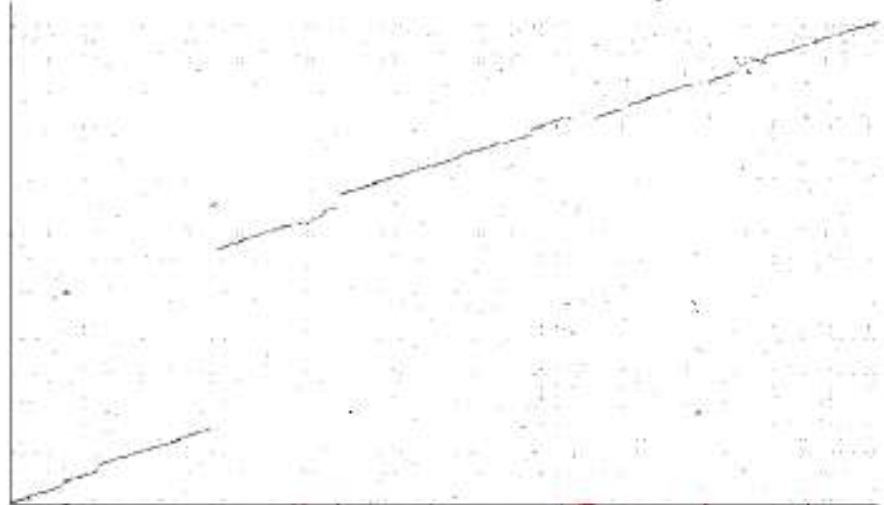
involucrados en funciones básicas del mantenimiento celular, incluyendo *minCDE*, *hutIGU* y *pcaGHID*, se transfieren en un evento de transferencia intragenómica a un replicón *repABC* ancestral, el sistema de replicación/partición conservado en genomas de rizobios (Slater et al., 2009). Una de las características de los cromosomas secundarios es que especies cercanas contienen más bloques de genes sinténicos que los de especies alejadas (Pappas y Cevallos, 2011), tal como se observó en los resultados del análisis del pSfrGR64c.

Por otro lado, también fue interesante encontrar una región colineal entre el megaplásmido pSfrGR64c y el cromosoma de *S. fredii* USDA257. Interesantemente, esta región no es compartida entre el pSfrGR64c y el cromosoma de GR64. El análisis de sintenia entre los cromosomas de *S. fredii* GR64 y *S. fredii* USDA257, revela que existe una región de gran tamaño presente en *S. fredii* USDA257 y ausente en *S. fredii* GR64, la cual tiene un tamaño similar a la región compartida entre el pSfrGR64c y el cromosoma de *S. fredii* USDA257 (figura 28).

Esto sugiere un evento de recombinación intragenómica, la cual pudo ocurrir en el fondo genómico de *S. fredii* GR64, cuando dicha región se escindió del cromosoma de *S. fredii* GR64 y se integró en el megaplásmido pSfrGR64c, o en el fondo de *S. fredii* USDA257, cuando la región mencionada se integró en el cromosoma de *S. fredii* USDA257. Dicha región está conformada por una gran cantidad de bloques sinténicos cortos, los cuales no están alineados en orden, sino dispersos en la región sinténica. Esto refleja la enorme plasticidad del genoma de los rizobios. Esta plasticidad es efecto de algunas de sus características: la presencia de secuencias largas y repetidas dispersas en plásmidos y cromosomas de las especies de este grupo, que en su mayoría son secuencia de inserción (SI) y genes duplicados y de la versatilidad del sistema de replicación/partición *repABC* (Pappas y Cevallos, 2011). Las SI y los genes duplicados son buenos sustratos para la recombinación homóloga, que cuando sucede en el mismo replicón, puede llevar a cabo eliminaciones o inversiones de las secuencias; y cuando sucede en dos replicones que comparten al menos dos secuencias de ADN, pueden intercambiar secuencias (Alberts et al., 2002). Estas propiedades permiten la translocación de grupos de genes entre replicones en un genoma.

A

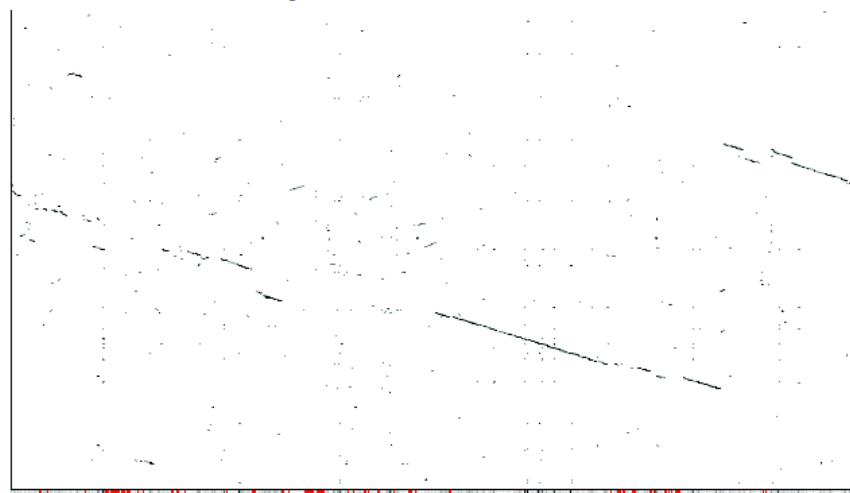
Cromosoma de *S. fredii* USDA257



Cromosoma de *S. fredii* GR64

B

Cromosoma de *S. fredii* USDA257



pSfrGR64c de *S. fredii* GR64

Figura 28. Evento de recombinación intragenómica (escisión o integración de una región) en el cromosoma de *S. fredii*. A) Sintenia entre cromosoma de *S. fredii* USDA257 y cromosoma de *S. fredii* GR64. Se observa un salto en la colinealidad, que refleja a una región presente en el cromosoma *S. fredii* USDA257 y ausente en el cromosoma de *S. fredii* GR64. B) Sintenia entre cromosoma de *S. fredii* USDA257 el megaplásmido pSfrGR64c. Se observa una región colineal entre ambos replicones, la cual no es compartida entre el megaplásmido de pSfrGR64c y el cromosoma de *S. fredii* GR64.

IX. CONCLUSIONES

Mediante el presente trabajo se logró determinar la secuencia completa del genoma de la bacteria rizobia noduladora de frijol *S. fredii* cepa GR64 (*Rhizobiaceae*), de la cual se ha propuesto un origen evolutivo de sus plásmidos por THG y recombinación, determinando las secuencias entre los scaffolds del draft mediante estrategias bioinformáticas y de biología molecular. Esta caracterización genética tiene una enorme importancia y valor para ayudar a desentrañar la historia evolutiva de los rizobios y sus mecanismos de diversificación y evolución.

También se logró realizar estudios de genómica comparativa de la secuencia completa del genoma de *S. fredii* GR64, determinando la sintenia de los replicones de la misma con respecto a otras especies de bacterias de la Familia *Rhizobiaceae*. El genoma de *S. fredii* GR64 exhibe las características de los genomas de rizobios: El cromosoma de *S. fredii* GR64 contiene bloques de genes sinténicos al compararlos con otras cepas de la familia *Rhizobiaceae*, mientras que los plásmidos presentan sintenia sólo con algunos genes, que están conservados entre plásmidos de rizobios. El plásmido pSfrGR64c presenta varias características que lo distinguen de los otros plásmidos, las cuales podrían catalogarlo como cromosoma secundario, además comparte una región colineal con el cromosoma de *S. fredii* USDA257, del mismo tamaño que una región ausente en el cromosoma de *S. fredii* GR64 y presente en *S. fredii* USDA257, sugiriendo un evento de recombinación intragenómica de integración o escisión de dicha región, lo cual refleja la enorme plasticidad del genoma de los rizobios. El patrón observado en el GC skew de los plásmidos pSfrGR64a y pSfrGR64b sugiere que ambos tienen un origen reciente, reforzando la teoría de que ambos fueron adquiridos mediante un evento de transferencia horizontal.

X. RECOMENDACIONES

Las interesantes características del megaplásmido pSfrGR64c observadas en el análisis, su gran tamaño y su alto grado de conservación, evidenciado en su patrón de GC skew y en los análisis de sintenia mediante los programas R2cat y mauve, sugieren que éste podría ser un cromosoma secundario. Se recomienda realizar experimentos dirigidos a corroborar dicha hipótesis, tales como tratar de curar el plásmido pSfrGR64c de la cepa *Sinorhizobium fredii* GR64, para observar si el mismo es un replicón indispensable para la viabilidad de la célula. A su vez, se deberían realizar mediante herramientas bioinformáticas, alineamientos de genes vitales de rizobios sobre el pSfrGR64c, para observar si los mismos están presentes en la secuencia del megaplásmido.

Algo que será de vital importancia es la anotación de la secuencia de este genoma. Además de la anotación automática, realizar una anotación no automatizada, reduciendo de esta forma errores provocados por los programas de anotación automática. Con el genoma anotado será más fácil identificar los genes localizados en la región localizada en el megaplásmido pSfrGR64c que también se localizan en el cromosoma de *S. fredii* USDA257, identificando si estos son genes vitales para la viabilidad de la célula.

La secuencia del genoma de *S. fredii* GR64 será de gran utilidad, y podrá ser utilizada para numerosos trabajos de investigación, contribuyendo al entendimiento de los genomas de rizobios y su evolución.

XI. LITERATURA CITADA

- Alberts, B., A. Johnson y J. Lewis. 2002. DNA Replication, Repair, and Recombination. 845-856 p. En: B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, y P. Walter (Eds.), *Molecular Biology of the Cell*. Garland Science. Nueva York.
- Bentley, S. 2009. Sequencing the species pan-genome. *Nature Reviews Microbiology* 7(4): 258-259.
- Bierman, M., R. Logan, K. O'Brien, E. T. Seno, R. Nagaraja Rao y B.E. Schoner. 1992. Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* 116: 43-49
- Boetzer, M., C. V. Henkel, H. J. Jansen, D. Butler y W. Pirovano. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4): 578-579.
- 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(1): 34-43.
- Cervantes, L., P. Bustos, L. Girard, R. I. Santamaria, G. Davila, P. Vinuesa, et al. 2011. The conjugative plasmid of a bean-nodulating *Sinorhizobium fredii* strain is assembled from sequences of two *Rhizobium* plasmids and the chromosome of a *Sinorhizobium* strain. *BMC Microbiology* 11: 149.
- Claverie, J. M. y C. Notredame. 2007. *Bioinformatics for dummies*. Wiley Publishing Inc. Indianapolis, Estados Unidos.
- Darling, A. C., B. Mau, F. R. Blattner y N. T. Perna. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14(7): 1394-1403.
- DeLong, E. F., y D. M. Karl. 2005. Genomic perspectives in microbial oceanography. *Nature* 437(7057): 336-342.
- Eckhardt, T. 1978. A rapid method for the identification of plasmid desoxyribonucleic acid in bacteria. *Plasmid* 1(4): 584-588.
- Figurski, D. H., R. Meyer, D. S. Miller y D. R. Helinski. 1976. Generation in vitro of deletions in the broad host range plasmid RK2 using phage Mu insertions and a restriction endonuclease. *Gene* 1: 107-119
- Figurski, D. H. y D. R. Helinski. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proceedings of the National Academy of Sciences* 76(4):1648-1652.
- Frost, L. S., R. Leplae, A. O. Summers y A. Toussaint. 2005. Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology* 3(9): 722-732.
- Gonzalez, V., J. L. Acosta, R. I. Santamaria, P. Bustos, J. L. Fernandez, I. L. Hernandez-Gonzalez, et al. 2010. Conserved symbiotic plasmid DNA sequences in the multireplicon pangenomic structure of *Rhizobium etli*. *Applied and Environmental Microbiology* 76(5): 1604-1614.
- Gordon, D., C. Abajian y P. Green. 1998. Consed: a graphical tool for sequence finishing. *Genome Research* 8(3): 195-202.

- Herrera-Cervera, J. A., J. Caballero-Mellado, G. Laguerre, H. V. Tichy, N. Requena, N. Amarger, et al. 1999. At least five rhizobial species nodulate *Phaseolus vulgaris* in a Spanish soil. *Federation of European Microbiological Societies Microbiology Ecology* 30(1): 87-97.
- Husemann, P., y J. Stoye. 2010. R2cat: synteny plots and comparative assembly. *Bioinformatics* 26(4): 570-571.
- Iwabe, N., K. Kuma, M. Hasegawa, S. Osawa y T. Miyata. 1989. Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proceedings of the National Academy of Sciences* 86(23): 9355-9359.
- Jones, K. M., H. Kobayashi, B. W. Davies, M. E. Taga y G.C. Walker. 2007. How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. *Nature Reviews Microbiology* 5(8): 619-633.
- Keen, N. T., S. Tamaki, D. Kobayashi y D. Trollinger. 1988. Improved broad-host-range plasmids for DNA cloning in gram-negative bacteria. *Gene* 70: 191-197.
- Kelly, B. G., A. Vespermann, y D. J. Bolton. 2009. The role of horizontal gene transfer in the evolution of selected foodborne bacterial pathogens. *Food and Chemical Toxicology* 47(5): 951-968.
- Kisand, V., y T. Lettieri. 2013. Genome sequencing of bacteria: sequencing, de novo assembly and rapid analysis using open source tools. *BMC Genomics* 14: 211.
- Lobry, J. R. 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular Biology and Evolution* 13(5): 660-665.
- Lloret, L. y E. Martínez-Romero. 2005. Evolución y filogenia de *Rhizobium*. 43-63, *Revista Latinoamericana de Microbiología* 47: 43-63.
- Madigan, M. T., J. M. Martinko y J. Parker. 2004. Brock: Biología de los Microorganismos. Pearson Prentice Hall. Madrid, España.
- Metzker, M. L. 2010. Sequencing technologies - the next generation. *Nature Reviews Genetics* 11(1): 31-46.
- Ochman, H. 2001. Lateral and oblique gene transfer. *Current Opinion in Genetics and Development* 11(6): 616-619.
- Pappas, K. M. y M. A. Cevallos. 2011. Plasmids of the Rhizobiaceae and Their Roles in Interbacterial and Transkingdom Interactions. 295-325 p. En: G. Witzany (Ed.), *Biocomunicación in Soil Microorganisms*. Springer. Berlin Heidelberg.
- Perret, X., C. Staehelin y W. J. Broughton. 2000. Molecular basis of symbiotic promiscuity. *Microbiology and Molecular Biology Reviews* 64(1): 180-201.
- Pesole, G., C. Gissi, C. Lanave y C. Saccone. 1995. Glutamine synthetase gene evolution in bacteria. *Molecular Biology and Evolution* 12(2): 189-197.
- Raymond, J., J. L. Siefert, C. R. Staples y R.E. Blankenship. 2004. The natural history of nitrogen fixation. *Molecular Biology and Evolution* 21(3): 541-554.
- Raz, Y. y E. Tannenbaum. 2010. The influence of horizontal gene transfer on the mean fitness of unicellular populations in static environments. *Genetics* 185(1): 327-337.
- Rodríguez-Martínez, J. M. (2004). Secuenciación de genomas. *Revista Arbor* CLXXVII 698: 285-310.
- Romero, D. y S. Brom. 2004. The symbiotic plasmids of the Rhizobiaceae. 271-290 p. En: G. Phillips y B. Funnell (Eds.), *Plasmid Biology*. ASM Press. Washington DC., Estados Unidos.
- Ronaghi, M., M. Uhlen y P. Nyren. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281(5375): 363-365.

- Sambrook, J., E. F. Fritsch y T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press. Nueva York, Estados Unidos.
- Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, C. A. Fiddes, et al. 1977. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*: 265(5596): 687-695.
- Schäfer, A., A. Tauch, W. Jäger, J. Kalinowski, G. Thierbach y A. Pühler. 1994. Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* 145(1): 69-73.
- Slater, S. C., B. S. Goldman, B. Goodner, J. C. Setubal, S. K. Farrand, E. W. Nester, et al. 2009. Genome sequences of three agrobacterium biovars help elucidate the evolution of multichromosome genomes in bacteria. *Journal of Bacteriology* 191(8): 2501-2511.
- Sprent, J. I. 2008. 60Ma of legume nodulation. What's new? What's changing? *Journal of Experimental Botany* 59(5): 1081-1084.
- Sullivan, J. T., H. N. Patrick, W. L. Lowther, D. B. Scott y C. W. Ronson. 1995. Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proceedings of the National Academy of Sciences* 92(19): 8985-8989.
- Sullivan, J. T. y C. W. Ronson. 1998. Evolution of *Rhizobia* by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proceedings of the National Academy of Sciences* 95: 5145-5149.
- Taylor, R. G. 1993. *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Research* 21: 1677-1678
- Torres Tejerizo, G., L. Lozano, V. González, P. Bustos, D. Romero y S. Brom. 2012. Draft Genome Sequence of the Bean-Nodulating *Sinorhizobium fredii* Strain GR64. *Journal of Bacteriology* 194(24): 6978.
- Tsai, I. J., T.D. Otto y M. Berriman. 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biology* 11(4): 41.
- Wheatcroft, R. y R. J. Watson. 1988. Distribution of insertion sequence ISRm1 in *Rhizobium meliloti* and other gram-negative bacteria. *Journal of General Microbiology* 134(1): 113-121.
- Woese, C. R. y G. E. Fox. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences* 74(11): 5088-5090.