UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

DOCTORADO EN CIENCIAS BIOMÉDICAS

CENTRO DE CIENCIAS GENÓMICAS

ESTUDIO METAGENÓMICO DE MICROORGANISMOS EXTREMÓFILOS DEL CAMPO GEOTÉRMICO DE LOS AZUFRES Y DIVERSIDAD GENÓMICA DE SIMBIONTES DE *PHASEOLUS* E INSECTOS NATIVOS DE MÉXICO

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

LUIS EDUARDO SERVÍN GARCIDUEÑAS

TUTORA PRINCIPAL DE TESIS:
DRA. MARÍA ESPERANZA MARTÍNEZ ROMERO
CENTRO DE CIENCIAS GENÓMICAS

COMITÉ TUTOR:
DRA. ELDA GUADALUPE ESPÍN OCAMPO
INSTITUTO DE BIOTECNOLOGÍA
DR. FEDERICO ESTEBAN SÁNCHEZ RODRÍGUEZ
INSTITUTO DE BIOTECNOLOGÍA

CUERNAVACA, MORELOS SEPTIEMBRE, 2015

Agradecimientos

A la Universidad Nacional Autónoma de México y al Programa de Doctorado en Ciencias Biomédicas, UNAM por darme las herramientas para desarrollarme académicamente asistiendo a cursos, congresos y estancias de investigación.

Al Consejo Nacional de Ciencia y Tecnología por la beca otorgada para realizar mis estudios de posgrado. Al Programa de Apoyo a los Estudios de Posgrado de la UNAM y al CONACYT por los apoyos de viaje y de Beca Mixta para realizar una estancia de investigación en Dinamarca.

Al CONACYT por aprobar y financiar el proyecto de Ciencia Básica con número de referencia 154453. Al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica de la Dirección General Asuntos del Personal Académico de la UNAM por apoyar los proyectos PAPIIT UNAM IN205412 y PAPIIT UNAM IN207615.

Al Subsistema Nacional de Recursos Genéticos Microbianos de la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación por el financiamiento recibido.

A mi tutora la Dra. Esperanza Martínez Romero por sus pláticas maravillosas. Sus enseñanzas y consejos han contribuido a mi formación académica y personal. Estoy agradecido por su esfuerzo, apoyo incondicional, motivación, paciencia y confianza. Los momentos en el laboratorio, los seminarios y todas sus clases fueron experiencias fascinantes.

A la Dra. Lupita Espín y al Dr. Federico Sánchez por formar parte de mi comité tutor y por brindarme consejos y apoyo. Al Dr. Jesús Campos por permitirme utilizar su laboratorio en la Universidad Michoacana de San Nicolás de Hidalgo. A las Doctoras Luisa Falcón, Katy Juárez y Esperanza Martínez y a los Doctores Alfredo Martínez y Diego Cortez por formar parte de mi comité sinodal, por sus recomendaciones y por su apoyo.

Agradecimientos a título personal

A la Dra. Susana Brom, al Dr. Pablo Vinuesa y al Dr. Otto Geiger por su apoyo como responsables del Programa de Doctorado en Ciencias Biomédicas durante diferentes periodos de mi formación académica. A la Lic. Gladys Avilés por todo su apoyo y orientación.

A los profesores Roger Garrett, Xu Peng y Susanne Erdmann por permitirme realizar una estancia agradable y de provecho en Dinamarca y por sus contribuciones.

A Ernesto Ormeño, Mónica Rosenblueth, Marco Rogel y Julio Martínez por su paciencia y ayuda en el laboratorio, y por todos sus consejos y charlas. A todos los integrantes del programa de Ecología Genómica que me ha tocado conocer a lo largo de estos años.

A mis compañeros en el laboratorio, Tania Rosas, Luis Bolaños, Diana Sahonero, Tabita Ramírez, Miguel Vences, Arturo Vera, Mariana Reyes y Berenice Jiménez.

A mis grandes amigas Alejandra Zayas, Alejandra Medina y Ared Mendoza por todos los momentos en Cuernavaca así como todo el apoyo para la realización de proyectos. A todos mis amigos y alumnos de la Licenciatura en Ciencias Genómicas por su motivación.

Dedicatoria

Este trabajo se lo dedico a mis padres, Cecilia Garcidueñas y José Luis Servín, y a mi amigo Jonathan López. Estoy muy agradecido por su apoyo, compañía y buenos deseos. Mi trayectoria académica no hubiera sido posible sin sus esfuerzos.

PREFACIO

Mis estudios de doctorado se realizaron en el Programa de Ecología Genómica del Centro de Ciencias Genómicas de la UNAM que se concentra en el estudio de poblaciones bacterianas, su diversidad y taxonomía, así como en la base molecular de las funciones bacterianas que participan en las interacciones de las bacterias con otros organismos.

Dentro del Programa, formé parte del Grupo de Microbiología Ambiental y Simbiótica a cargo de la Dra. Esperanza Martínez Romero donde tuve la oportunidad de proponer y colaborar en proyectos enfocados en estudiar a las bacterias fijadoras de nitrógeno en los nódulos de las leguminosas y a los simbiontes de artrópodos.

Por otra parte, el grupo de la Dra. Esperanza Martínez Romero tiene interés de analizar la diversidad microbiana en ambientes naturales de México. Dentro de esta área de investigación surge mi interés y curiosidad personal de explorar la diversidad microbiana del campo geotérmico de Los Azufres ubicado en Michoacán, México. Al ser originario de Michoacán es mi intención realizar investigaciones sobre microorganismos presentes en este Estado.

El proyecto de Los Azufres emerge por la fascinación de conocer microorganismos capaces de habitar en condiciones volcánicas así como analizar su potencial genético de resistencia a temperatura, acidez y niveles elevados de metales pesados. De esta manera, mi reto principal consistió en obtener ADN ambiental y generar secuencias metagenómicas para sobrepasar las limitantes de las técnicas de cultivo de microorganismos.

En el primer capítulo de esta tesis se presentan los resultados obtenidos a partir del análisis de un metagenoma de una manifestación termal de Los Azufres. En este primer estudio recuperamos los primeros genomas de arqueas y arqueovirus de ambientes termales dentro de territorio continental de México. Se presentan también avances obtenidos para comprender la diversidad microbiana presente en otras comunidades microbianas del mismo campo geotérmico.

Durante el transcurso de mi licenciatura y doctorado, mostré un especial interés en conocer el fenómeno de simbiosis entre plantas leguminosas y bacterias fijadoras de nitrógeno. En el segundo capítulo de la tesis se presentan resultados sobre la diversidad genómica de rizobios de frijoles nativos de México. De manera especial se profundizó en analizar los rizobios de una especie de frijol nativa de las regiones montañosas de Michoacán y Jalisco.

Al participar en los seminarios del grupo de investigación surgió la posibilidad de analizar la diversidad de bacterias de insectos nativos de México. Mi propuesta fue explorar la microbiota intestinal de la mariposa monarca ya que había permanecido sin estudiar mediante técnicas moleculares. Las mariposas monarca son insectos icónicos y en riesgo que realizan una fascinante migración a través de Norteamérica para arribar a una región de bosques de oyamel en Michoacán. En el tercer capítulo de la tesis se presentan los primeros resultados sobre la diversidad de la microbiota intestinal de las mariposas monarca y resultados adicionales que surgen de proyectos de colaboración sobre otros insectos y artrópodos de México.

Al final de esta tesis se incluyen resúmenes para cada uno de los capítulos.

Resumen

La metagenómica permite analizar la diversidad microbiana que habita en distintas condiciones ambientales tales como las presentes en aéreas volcánicas. El presente estudio analizó la diversidad microbiana que reside en el campo geotérmico de Los Azufres usando técnicas metagenómicas. La diversidad genómica de los microorganismos de Los Azufres no se había estudiado previamente. Se emplearon técnicas de secuenciación de ADN ambiental y análisis bioinformáticos para revelar la diversidad microbiana y para generar ensambles de genomas de microorganismos abundantes. En el metagenoma de una fuente termal se identificó una comunidad integrada por arqueas termoacidófilas. Del metagenoma de la fuente termal se obtuvieron genomas de arqueovirus y de arqueas *Sulfolobales* y *Thermoproteales.* En un metagenoma de sedimentos termales se identificó una comunidad dominada por una microalga de la familia *Cyanidiaceae*. A partir de este metagenoma se obtuvo el genoma de una nueva arquea del filo *Parvarchaeota*. También se identificó una comunidad compuesta principalmente por bacterias acidófilas y por una microalga de la clase *Trebouxiophyceae* en el metagenoma de una laguna ácida. Los genomas de los organelos de la microalga se ensamblaron a partir de las secuencias del metagenoma. Además se obtuvo de células en cultivo, el primer genoma de bacterias del género *Acidocella*. Las manifestaciones termales del campo geotérmico de Los Azufres contienen comunidades microbianas novedosas con diversidad limitada. Los microorganismos identificados poseen atributos funcionales potenciales que son consistentes con las condiciones geoquímicas. Por último, la tesis contiene capítulos adicionales con resultados sobre la diversidad genómica de bacterias fijadoras de nitrógeno de especies de *Phaseolus* y de bacterias simbiontes de insectos de México.

Abstract

Metagenomics allow the study of the microbial diversity that inhabit different environmental conditions such as the ones present in volcanic areas. This study analyzed the microbial diversity that thrives at the Los Azufres geothermal field by using metagenomic techniques. The genomic diversity of the microorganisms residing at Los Azufres had not been previously studied. Environmental DNA sequencing and bioinformatic analyses were used to reveal the microbial diversity and to obtain genomic assemblies of abundant microorganisms. A community of thermoacidiphilic archaeons was identified from a hot spring. From the hot spring metagenome, novel genomes were reconstructed for archaeovirus and *Sulfolobales* and *Thermoproteales* archaeons. A metagenome of thermal sediments contains a community dominated by a microalga belonging to the *Cyanidiaceae* family. The first genome of a novel archaeon from the *Parvarchaeota* phylum was obtained from this metagenome. Also, a community integrated mainly by acidophilic bacteria and by a microalga belonging to the *Trebouxiophyceae* class was identified from the metagenome of an acidic lagoon. The microalga organelle genomes were assembled by using metagenomic sequences. Moreover, the first bacterial genome of the *Acidocella* genus was obtained from cultured cells. To conclude, the thermal features at the Los Azufres geothermal field contain novel microbial communities with limited diversity. The identified microorganisms have potential functional attributes that are consistent with geochemical conditions. Finally, this thesis contains additional chapters with results about the genomic diversity of nitrogen fixing bacteria of *Phaseolus* species and symbiotic bacteria of insects from Mexico.

ÍNDICE

Capítulo I

Diversidad genómica de microorganismos extremófilos de Los Azufres

Capítulo II

Diversidad genómica de rizobios de *Phaseolus* nativos de México

Capítulo III

Resultados adicionales

Diversidad genómica de simbiontes de insectos y artrópodos nativos de México

Apéndices

Capítulo I

Diversidad genómica de microorganismos extremófilos de Los Azufres

Introducción y antecedentes

En algunos sitios volcánicos se encuentran manifestaciones termales con condiciones de acidez extrema, de temperaturas elevadas y de concentraciones tóxicas de metales (condiciones extremas). Los análisis de diversidad de comunidades microbianas de dichas manifestaciones termales han identificado microorganismos que pertenecen a los dominios Bacteria, Archaea y Eucaria y que son capaces de contender con las presiones ambientales (Huber et al., 2000). Estas comunidades microbianas pudieran ser aprovechadas para mejorar los procesos actuales de biorremediación, de biomineria y otros procesos industriales (van den Burg, 2003; Podar y Reysenbach, 2006; González-Pastor y Mirete, 2010).

En las manifestaciones termales tenemos la oportunidad de analizar comunidades conformadas por virus y microorganismos desconocidos o pobremente caracterizados. En México, el campo geotérmico de Los Azufres alberga una gran cantidad de manifestaciones geotérmicas de origen natural. El campo geotérmico de Los Azufres se localiza en el Estado de Michoacán, dentro de la sierra de San Andrés en el Cinturón Volcánico Transmexicano en el occidente de México. Las manifestaciones geotérmicas que se pueden encontrar en Los Azufres son principalmente manantiales termales, fumarolas y solfataras (pozas termales de lodos ácidos) que llegan a alcanzar los 90 ºC, con pH menores que 4.0 y altas concentraciones de sulfatos, silicatos y metales pesados (Tello-López y Suarez-Arriaga, 2000).

Los estudios microbiológicos que se han realizado en Los Azufres se han enfocado en determinar microorganismos de muestras de metales corroidos o de tuberias empleadas para la generación de energía geotérmica (Torres-Sanchez et al. 1996; Valdez-Salas et al., 2000; Alfaro-Cuevas-Villanueva et al., 2006; Castorena et al., 2006). Estos estudios han detectato la presencia de bacterias pertenecientes a los géneros *Desulfotomaculum*, *Desulforomonas*, *Desulfobacter*, *Desulfovibrio*, *Burkholderia y Thermodesulfobacterium*. El estudio de Navarrette-Bedolla et al., 1999 fue el primero en reportar arqueas del género *Thermoproteus* en tuberías de las plantas geotérmicas. La identificación de los microorganismos en estos estudios se realizó analizando imágenes de microscopía electrónica o pruebas bioquímicas.

La gran mayoría de la diversidad microbiana de las manifestaciones termales de Los Azufres se desconoce por completo. En el año 2014 se publicó un primer trabajo que utilizó secuencias y patrónes de digestión enzimática de genes ribosomales 16S rRNA de muestras de lodo y agua de una área pequeña que es utilizada como un spa natural para cientos de turistas (Brito et al., 2014). En este estudio se identificaron y se lograron aislar bacterias pertenecientes a los géneros *Rhodobacter*, *Acidithiobacillus*, *Lyzobacter*, *Thermodesulfobium*, *Desulfurella*, *Thiomonas* y *Thermodesulfobium*.

Los estudios microbiológicos realizados en Los Azufres han analizado la diversidad microbiana presente en manifestaciones termales que han sufrido impacto de actividades humanas, principalmente debido a las actividades recreativas y de generación de energía geotérmica. Es común encontrar manantiales termales que han sido entubados y sellados con concreto para bombear las aguas termales con el propósito de llenar albercas para turistas. En otros casos, se extraen los lodos y sedimentos ácidos con propósitos de exfoliar la piel o se pavimentan los caminos por donde se pueden encontrar pequeñas pozas termales. Además es frecuente encontrar ganado vacuno que genera desechos orgánicos que modifican la composición de nutrientes presentes en suelos y manantiales. Finalmente, la generación de energía geotérmica también tiene un impacto ecologíco al bombear agua geotérmica, generar salmuera y provocar la acumulación de metales pesados en la superficie.

La diversidad genómica de los microorganismos que residen en las manifestaciones termales de Los Azufres había permanecido sin explorar. En el laboratorio de la Dra. Esperanza Martínez se desarrolla una línea de investigación en genómica ambiental en la que propuse analizar la diversidad de comunidades microbianas de Los Azufres mediante técnicas metagenómicas. Este proyecto representa una oportunidad para analizar comunidades microbianas que de otra forma permanecerían desconocidas.

Los mecanismos microbianos de resistencia a condiciones ambientales son conocidos en su mayoria a partir del estudio de microorganismos cultivados (van den Burg, 2003). Una enorme cantidad de microorganismos permanecen como entidades desconocidas para la ciencia si se considera que alrededor del 99% de la diversidad microbiana que reside en un ambiente en particular no se puede cultivar con técnicas tradicionales en el laboratorio (Amann et al., 1995; Hugenholtz et al., 1998; Rappé y Giovannoni, 2003; DeLong y Pace, 2001; Pace, 2007; Epstein, 2009; Chaffron, et al., 2010; Pham y Kim, 2012).

La metagenómica es una aproximación que trata de revertir el desconocimiento que tenemos sobre los microorganismos que conforman las comunidades microbianas tal como se encuentran en la naturaleza (Podar y Reysenbach, 2006). La metagenómica contempla una serie de métodos independientes del cultivo microbiano que permite, en teoría, tener acceso al ADN de todos los microorganismos cultivables y no cultivables de un ambiente específico y con ello, es posible determinar la composición, la estructura y el metabolismo potencial de tales microorganismos (Handelsman et al., 2002; Riesenfeld et al., 2004).

La secuenciación de metagenomas es efectiva para caracterizar comunidades microbianas de complejidad limitada ya que poseen un número reducido de diferentes microorganismos. Algunos ejemplos incluyen las comunidades microbianas de biofilmes de una mina ácida de California que presentan altas concentraciones de metales pesados (Tyson et al., 2004), las aguas profundas que drenan una mina de oro en África a 2.8 km de profundidad, con nutrientes limitados y en un ambiente aislado de la luz solar (Chivian et al., 2008) y las comunidades microbianas de las pozas termales de Yellowstone que presentan altas temperaturas y acidez extrema (Inskeep et al., 2010). Recientemente se han obtenido cerca de 800 genomas casi completos de bacterias diminutas (de incluso menos de 400 nm) que no se han podido cultivar y que representan más de 35 filos nuevos. Los genomas de estas bacterias se obtuvieron al realizar análisis metagenómicos de muestras de aguas subterraneas de un sitio de remediacion en Colorado. Se ha estimado que estos filos nuevos representan más del 15% de los grupos conocidos de bacterias (Brown et al., 2015). Los hallagos de los estudios metagenómicos han permitido descubrir una enorme diversidad microbiana y han remodelado el árbol de la vida.

De igual forma se han publicado estudios metagenómicos fascinantes que han identificado nuevas ramas dentro del dominio Archaea incluyendo filos de arqueas hiperhalófilas (Narasingarao et al. 2012) y de arqueas del fondo marino (Spang et al., 2015). Además las técnicas metagenómicas en conjunto con análisis de células únicas han permitido recuperar una gran cantidad de genomas de microorganismos de ambientes extremos (Rinke et al., 2013).

En los ejemplos antes expuestos, los análisis metagenómicos recuperaron los genomas consenso de las poblaciones de microorganismos más abundantes. Los estudios metagenómicos ayudan a detectar microorganismos totalmente nuevos para la ciencia y abren la posibilidad de analizar sus potenciales genéticos sobrepasando las limitantes de su cultivo.

Planteamiento del problema

La diversidad de las comunidades microbianas que residen en las manifestaciones termales del campo geotérmico de Los Azufres no se había estudiado mediante técnicas metagenómicas. A la fecha no se habían realizado análisis metagenómicos de sitios geotérmicos que presentan condiciones extremas de temperatura y de acidez extrema del Eje Volcánico de México.

Hipótesis

Las condiciones geoquímicas y térmicas imponen una presión de selección significativa en la diversidad de las comunidades microbianas. Bajo las restricciones de los nichos ecológicos de Los Azufres, las comunidades microbianas de este campo geotérmico presentan baja complejidad y poseen atributos funcionales potenciales de resistencia a condiciones ambientales extremas.

Objetivos

Realizar censos de diversidad microbiana de nichos representativos y con condiciones ambientales extremas del campo geotérmico de Los Azufres (Figura 1) usando análisis filogenéticos de genes ribosomales.

Seleccionar comunidades microbianas de baja complejidad y con microoganismos filogenéticamente novedosos en base a los resultados de los análisis de diversidad. Posteriormente, realizar la secuenciación metagenómica de las comunidades seleccionadas.

Evaluar la diversidad de las comunidades microbianas seleccionadas mediante el análisis de secuencias metagenómicas para evitar los sesgos inherentes a las técnicas de amplificación de genes ribosomales.

Determinar el metabolismo potencial contenido en las secuencias metagenómicas utilizando análisis comparativos de identidad de secuencia con genes ya conocidos. Realizar análisis de presencia de dominios funcionales en secuencias de genes desconocidos que pudieran dar alguna evidencia de su función posible.

Identificar genes codificantes que sirvan como marcadores clave para determinar los mecanismos de resistencia a las condiciones extremas tales como los ya conocidos para la detoxificación de metales pesados, fijación de nitrógeno y carbono y estabilidad del ADN.

Resultados y discusión

Muestreo de los sitios termales de Los Azufres

Se realizaron muestreos en seis sitios termales de Los Azufres para obtener ADN ambiental. Los sitios muestreados fueron una solfatara ácida (Figura 1A), un manantial hidrotermal conocido como manantial de Marítaro (Figura 1B), sedimentos fotosintéticos expuestos a una fumarola (Figura 1C), una laguna ácida (Figura 1D), un pozo de enfriamiento de agua geotérmica (Figura 1E) y además se analizó agua condensada de una tubería que transporta vapor desde el reservorio geotérmico hasta la superficie.

Los sitios termales se eligieron en base a su fácil acceso y por el bajo nivel de impacto de actividades humanas. Los sitios se localizaron en viajes de exploración y por recomendación del personal del campo geotérmico. En estos sitios se esperaba encontrar microorganismos y consorcios microbianos no descritos debido al aislamiento geográfico de Los Azufres. También se esperaba encontrar microorganismos capaces de contender con las condiciones ambientales extremas particulares a cada sitio termal.

Figura 1. Sitios termales de Los Azufres de donde se tomaron muestras para analizar la diversidad microbiana.

Uno de los objetivos primarios fue realizar los censos de diversidad microbiana de los sitios elegidos (Figura 1) usando análisis filogenéticos de genes ribosomales. Por ello se realizaron muestreos múltiples de agua y sedimentos que variaron dependiendo de cada sitio. En algunos casos solo fue posible realizar una única toma de muestras debido a que los consorcios microbianos se encontraban en abundancia limitada como fue el caso de los sedimentos termales (Figura 1C). Otros sitios fueron muestreados en múltiples ocasiones debido a la extensión de las manifestaciones o para verificar la estabilidad de las comunidades microbianas cuando la cantidad de material biológico no era limitante, como por ejemplo las muestras de agua de la laguna verde (Figura 1D) y de la solfata ácida (Figura 1A).

Cada sitio termal se analizó siguiendo una dinámica particular sin perder de vista el objetivo común de explorar su diversidad microbiana con fines a seleccionar las comunidades microbianas que pudieran ser las mejores candidatas a ser analizadas mediante enfoques metagenómicos. Se buscó encontrar comunidades microbianas de baja complejidad o que poseyeran consorcios microbianos poco comunes o con miembros no reportados. También se deseaba que las comunidades microbianas fueran estables y que se encontraran en sitios termales donde la toma de muestras fuera accesible y abundante, esto con miras a seguir estudiando las comunidades en su ambiente natural en proyectos futuros.

Las muestras de todos los sitios fueron de un litro por triplicado y las muestras se conservaron en frascos estériles a temperatura ambiente. En el caso de las muestras de sedimento la colecta se realizó directamente con cajas de Petri estériles.

La laguna verde (Figura 1D) y la solfatara ácida (Figura 1A) recibieron interés especial debido a que su acceso es sencillo, a que presentan condiciones geoquímicas y térmicas contrastantes y a que permiten la toma de muestras en abundancia. Los muestreos de agua de la laguna ácida y de la solfatara ácida se realizaron durante el mes de marzo (estación de primavera) de tres años consecutivos (2008, 2009 y 2010) para determinar la estabilidad de las estructuras de las comunidades microbianas.

Dos muestreos adicionales se realizaron durante julio (sequia) y noviembre (posterior a la época de lluvias) de 2009 para determinar la estabilidad de las estructuras de las comunidades durante las estaciones de un mismo año. Estos muestreos se realizaron en los mismos lugares de colecta. El lugar de colecta de la laguna ácida estaba seco durante julio de 2009 debido a la sequía por tanto se colectó agua cerca del centro de la laguna a cuatro metros de una manifestación geotérmica que normalmente se encuentra en el fondo.

La colecta de agua del pozo de enfriamiento (Figura 1E) y del manantial Marítaro (Figura 1B) se realizó durante marzo de 2009. La colecta de agua geotérmica de las tuberías se realizó durante mayo de 2010. Finalmente la colecta de sedimentos termales y fotosintéticos se realizó durante julio de 2009 en el área de Maritaro (Figura 1C) y durante abril de 2013.

Análisis geoquímicos

Los análisis de la composición química del agua de la laguna ácida y de la solfatara ácida se presentan en la Tabla 1. Los análisis se llevaron a cabo en el Instituto Mexicano de Tecnología del Agua (IMTA). Las muestras empleadas para los análisis químicos fueron las mismas que se utilizaron para los análisis de secuenciación masiva.

Tabla 1. Composición química de las muestras de agua colectadas durante 2010 de la laguna ácida y de la solfatara ácida y que se utilizaron para purificar y secuenciar ADN metagenómico.

Element/Compound (mg/L)	Acidic lake	Acidic solfatar
Carbonates	0	0
Chlorides	69.5	<1.06
Phosphates	<1.13	<1.13
Nitrites	2.49	0.263
Sulphates	287	314
TOC	5.67	9.70
As	0.1440	1.0714
Fe	16.511	11.314
Mn	0.104	0.460
Al	7.905	12.454
Cd	/	<0.02
Cu	/	<0.05
Cr	/	<0.05
Si	/	160.311
Zn	/	0.115
Pb	/	<0.10

La temperatura y el pH se determinaron durante los muestreos utilizando termómetros estándar y tiras indicadoras de pH (Fermont). Adicionalmente, el pH de las muestras se determinó en el laboratorio utilizando un potenciómetro (accumet AB15). En general los pHs se mantuvieron estables en la laguna verde y en la solfatara ácida y solo presentaron variaciones pequeñas en las épocas de lluvia. Esto se explicará más adelante en el capítulo.

Extracción de ADN

Las muestras colectadas de cada sitio se procesaron el mismo día de colecta o al día siguiente para evitar que las poblaciones microbianas se alteraran. Las muestras de ADN de la laguna ácida, de los sedimentos, del pozo de enfriamiento y de la muestra de agua geotérmica se purificaron utilizando el *Ultra-Clean microbial DNA Isolation Kit* (MoBio Laboratories, Inc., Carlsbad, CA) siguiendo las instrucciones del fabricante. Las muestras de ADN del manantial de Marítaro y de la solfatara ácida se purificaron utilizando el *Ultra-Clean mega soil DNA Kit* (MoBio Laboratories, Inc., Solana Beach, CA) siguiendo las instrucciones del fabricante.

Análisis de diversidad

Los genes ribosomales 16S rRNA y 18S rRNA se amplificaron mediante reacciones de PCR para determinar la diversidad de las comunidades microbianas (Apéndice IV). Se utilizaron distintas parejas de oligonucleótidos para evitar el sesgo inherente a las técnicas de PCR (Tabla 2). Los productos de PCR se amplificaron y se clonaron utilizando el *TOPO-TA Cloning Kit for Sequencing* (pCR4- TOPO) de Invitrogen y se siguieron las instrucciones del fabricante. Los productos de PCR se purificaron utilizando el *High Pure PCR Product Purification Kit* de Roche. Los productos de PCR purificados se secuenciaron en Macrogen Inc. en Seúl, Corea del Sur. La identidad de secuencia de los genes ribosomales se determinó mediante búsquedas de BLASTN (las identidades de secuencia se presentan en la sección correspondiente a cada sitio termal). Se obtuvieron secuencias de GenBank para realizar análisis filogenéticos. Las secuencias se alinearon utilizando MUSCLE v.3.8.31. Las reconstrucciones filogenéticas se infirieron con MrBayes v.3.1.2 (Ronquist y Huelsenbeck, 2003) y PhyML v.3.0 (Guindon et al., 2010). Los árboles filogenéticos se editaron en TreeDyn v.198.3 (Chevenet et al., 2006). Los datos para realizar las curvas de rarefacción se calcularon en mothur v.1.7.2 (Schloss et al., 2010).

Tabla 2. Oligonucleótidos utilizados para los análisis de diversidad microbiana.

Oligonucleótido	Tamaño	Sentido	Reverso	Referencia
16S rRNA Bacteria 1	1500 pb	FD1	RD1	Weisburg et al., 1991
16S rRNA Bacteria 2	700 pb	799f	1492r	Chelius y Triplett, 2001
16S rRNA Arquea 1	600 pb	Arch0333aS15	Arch0958aA19	Lepp et al., 1999
16S rRNA Arquea 2	660/1300 pb	UA751F	UA1406R	Baker et al., 2004
16S rRNA Arquea 3	1000/1700 pb	Arch0333aS15	UA1406R	Lepp et al., 1999, Baker et al., 2004
16S rRNA Arquea 4	1460/2000 pb	A2F	U1510R	Hugenholtz et al., 1998
18S rRNA Eucariota	1700 pb	90F	B	Yuasa et al., 2004
arsenito oxidasa *aoxB*	1100 pb	aoxBM1-2F	aoxBM3-2R	Quéméneur et al., 2008

Diversidad de comunidades de arqueas de manantiales termales de Los Azufres

El manantial de Marítaro (Figura 1B) mostró una temperatura de 89 °C y un pH de 3.7 mientras que la solfatara ácida (Figura 1A) tiene pH y temperaturas constantes en todos los muestreos (~pH =3, ~T=65 °C). Las comunidades de la solfatara ácida y del manantial de Marítaro presentaron diversidad limitada y están conformadas por poblaciones de arqueas. Las reacciones de PCR no amplificaron genes ribosomales de bacterias ni de eucariotas. Se identificaron dos secuencias distintas de genes ribosomales 16S rRNA de arqueas en las librerías.

Las secuencias de la arquea predominante representa más de tres cuartas partes de la librería de la solfatara ácida (40 de 50 secuencias) y casi la totalidad de la librería del manantial de Marítaro (28 de 30 secuencias). La secuencia del gen ribosomal 16S rRNA de la arquea predominante es similar a secuencias de arqueas del orden *Sulfolobales* y las cepas más cercanas corresponden al género *Acidianus*. La cepa tipo más cercana corresponde a *Acidianus infernus* So4a^T que presenta una identidad de 92.8% en 1334 nucleótidos.

El género *Acidianus* no es monofilético (Figura 2). El género *Acidianus sensu stricto* está definido por la cepa tipo de *Acidianus infernus* So4a^T. Un análisis filogenético basado en genes ribosomales 16S rRNA agrupa las secuencias de la arquea predominante de la solfatara ácida de Los Azufres en un grupo independiente las cepas del grupo de *Acidianus sensu stricto* (Figura 2). La arquea predominante se designó como arquea *Sulfolobales* AZ1. Las cepas de "*Acidianus brierleyi*" DSM 1651 y "*Acidianus tengchongensis*" S5 poseen una identidad de 92% en 1334 nucleótidos pero su clasificación taxonómica no se ha resuelto.

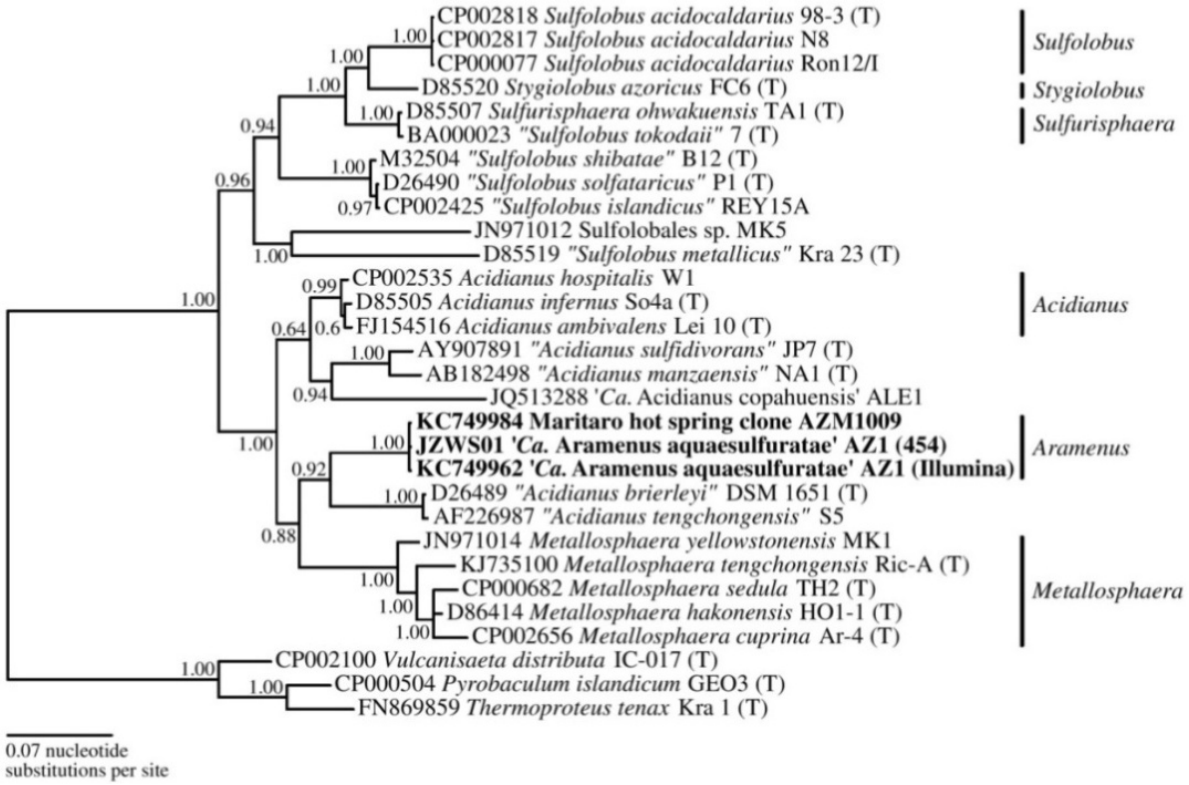


Figura 2. Reconstrucción filogenética de arqueas del orden *Sulfolobales*. En negritas se indica la posición de secuencias correspondientes a la arquea *Sulfolobales* AZ1. El alineamiento de secuencias contiene 1426 caracteres. La reconstrucción filogenética bayesiana se realizó en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Se muestran los números de acceso de GenBank de las secuencias de referencia. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. La barra de escala representa el número promedio de sustitución de nucleótidos por sitio.

La arquea *Sulfolobales* AZ1 cumple los requisitos para considerarse como la representante de un nuevo género candidato si consideramos los criterios de clasificación taxonómica actuales (Hugenholtz et al, 1998; Tindall et al., 2010; Yarza et al., 2008; Yarza et al., 2010; Yarza et al., 2014). Los criterios incluyen que la identidad de secuencia del gen ribosomal 16S rRNA sea menor a 94.9±0.4%, que la secuencia sea mayor a 1000 nucleótidos, que la secuencia se recupere tres o más veces en productos de PCR independientes y que las reconstrucciones filogenéticas den soporte a clados independientes.

Las secuencias del gen ribosomal 16S rRNA de la arquea minoritaria presentan una identidad de entre 97 y 98% en 1,031 nucleótidos con secuencias de arqueas *Thermoproteales* del género *Thermoproteus*. La arquea minoritaria se designó como *Thermoproteus* sp. AZ2. En un análisis filogenético basado en genes ribosomales 16S rRNA se encontró que la posición filogenética de *Thermoproteus* sp. AZ2 es externa dentro del clado *Thermoproteus,* lo que provee evidencia de que pudiera corresponder a una especie nueva (Figura 3).

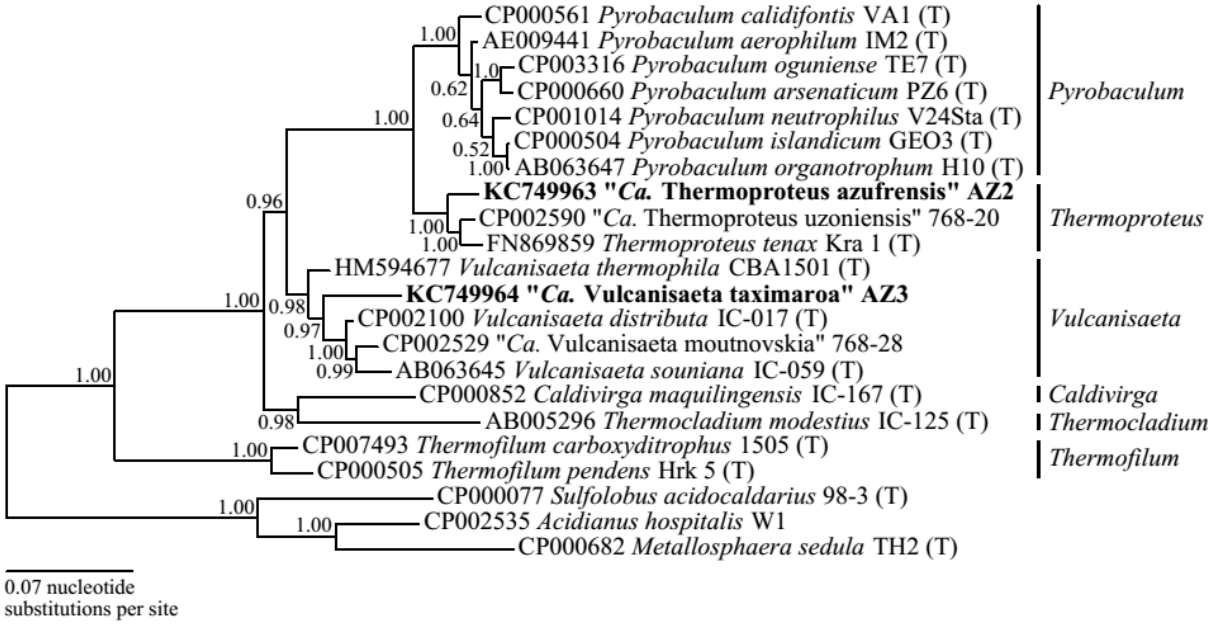


Figura 3. Reconstrucción filogenética de arqueas del orden *Thermoproteales*. En negritas se indica la posición de secuencias correspondientes a *Thermoproteus* sp. AZ2 ("*Ca*. Thermoproteus azufrensis"). También se presenta la posición de la arquea Vulcanisaeta sp. AZ3 ("*Ca*. Vulcanisaeta taximaroa" AZ3). La arqueas *Thermoproteales* y sus nombres candidatos serán presentados más adelante en el capítulo I. El alineamiento de secuencias contiene 1445 caracteres. La reconstrucción filogenética bayesiana se realizó en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Se muestran los números de acceso de GenBank de las secuencias de referencia. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. La barra de escala representa el número promedio de sustitución de nucleótidos por sitio.

En conjunto, estos resultados nos indican que las comunidades microbianas son de complejidad limitada en la solfatara ácida y en el manantial de Marítaro, que las comunidades están integradas predominantemente por arqueas y que la población de arqueas mayoritaria pertenece a un grupo filogenético novedoso del orden *Sulfolobales*.

Se realizaron estos análisis de diversidad con la finalidad de determinar si las comunidades microbianas de la solfatara ácida son estables a través del tiempo. Los análisis de diversidad de la solfatara ácida realizados durante el mes de marzo de tres años consecutivos (2008, 2009 y 2010) y de julio y noviembre de 2009 identifican a los mismos grupos filogenéticos. Varias manifestaciones termales alrededor del mundo presentan comunidades microbianas estables (Wilson et al., 2008; Satoh et al., 2013; Wemheuer et al., 2013). En general se esperaban pocos cambios en las poblaciones de arqueas de la solfatara ácida debido a la baja complejidad y a la estabilidad de las condiciones geoquímicas y de temperatura.

Análisis de secuencias metagenómicas de la solfatara ácida

La solfatara ácida se eligió como uno de los sitios para realizar la secuenciación de un metagenoma debido a que sus comunidades microbianas son estables a través del tiempo, son de diversidad limitada y están integradas por una población abundante de arqueas *Sulfolobales* de un nuevo género candidato. La solfatara ácida se encuentra en un sitio aislado que no muestra impacto humano y presenta condiciones extremas de pH y de temperatura. Además no existían genomas de referencia de los grupos filogenéticos identificados.

La zona donde se encuentra el manantial de Marítaro se seca casi por completo durante la estación de sequía y solo permanecen algunos pequeños riachuelos. A partir de finales del año 2009 la zona del manantial de Maritaro se convirtió en una zona de recreación humana. El manantial de Maritaro quedo inaccesible debido a que fue sellado con varillas y concreto y sus aguas termales son bombeadas a través de tuberías a piscinas para turistas. También se ha permitido el ingreso de ganado a la zona. Debido a estas razones ya no se continuó analizando el manantial de Marítaro.

El primer metagenoma de la solfatara se secuenció en la plataforma de Illumina-Solexa como uno de los proyectos piloto de la Unidad Universitaria de Secuenciación Masiva de la UNAM. Las secuencias metagenómicas comprenden 12.5 millones de lecturas pareadas de 35 bases que integran 436 Mb. El ensamble de las lecturas se realizó durante 2009 usando Velvet v.1.0.13 (Zerbino y Birney, 2008). Posteriormente se continuaron realizando ensambles adicionales con los programas Edena (Hernandez et al., 2008), SOAPdenovo (Luo et al., 2012) y Metavelvet (Namiki et al., 2012) pero no fue posible mejorar los ensambles.

Las secuencias metagenómicas revelaron la composición filogenética de la comunidad de la solfatara ácida y permitieron la reconstrucción del genoma parcial de la arquea *Sulfolobales* AZ1 y de arqueovirus novedosos.

Artículo:

Servín-Garcidueñas LE, Martínez-Romero E. 2014. Draft genome sequence of the *Sulfolobales* archaeon AZ1, obtained through metagenomic analysis of a Mexican hot spring. Genome Announc. 2: e00164-14.

A partir del ensamble del primer metagenoma de la solfatara ácida de Los Azufres logramos identificar solo un gen ribosomal 16S rRNA que corresponde a la arquea *Sulfolobales* AZ1. El primer metagenoma se obtuvo de una muestra de agua de la superficie de la solfatara ácida. Es decir, la muestra de ADN ambiental que se secuenció estaba naturalmente enriquecida en arqueas aeróbicas o aeróbicas facultativas, que en este caso correspondieron a la población de la arquea *Sulfolobales* AZ1. En este primer metagenoma no se obtuvieron secuencias genómicas de las poblaciones de arqueas *Thermoproteales* ya que son anaeróbicas.

En base a la comparación de secuencias de genes ribosomales y de secuencias genómicas se encontró que la arquea *Sulfolobales* AZ1 representa una especie candidata de un género nuevo dentro del orden *Sulfolobales* y se ha propuesto el nombre '*Candidatus* Aramenus sulfurataquae'. En este artículo se reporta el ensamble del genoma de la arquea *Sulfolobales* AZ1 a partir de las secuencias metagenómicas de la solfatara ácida. Este ensamble representa el primer genoma obtenido de una arquea que habita en el Eje Volcánico de México.

Se encontró que el contenido de G+C del genoma de la arquea *Sulfolobales* AZ1 (47%) es de los más altos reportados para arqueas *Sulfolobales*. Además el contenido de G+C promedio de sus 47 genes de ARN de transferencia predichos es incluso más alto, de 65%, lo que reflejaría las adaptaciones a las condiciones termales extremas. Las principales características del genoma de la arquea *Sulfolobales* AZ1 se presentan en el artículo.

genomeA
Journals.ASM.org

Draft Genome Sequence of the *Sulfolobales* Archaeon AZ1, Obtained through Metagenomic Analysis of a Mexican Hot Spring

Luis E. Servín-Garcidueñas, Esperanza Martínez-Romero

Center for Genomic Sciences, Department of Ecological Genomics, National University of Mexico, Cuernavaca, Morelos, Mexico

The *Sulfolobales* archaea have been found inhabiting acidic hot springs all over the world. Here, we report the 1.798-Mbp draft genome sequence of the thermoacidophilic *Sulfolobales* archaeon AZ1, reconstructed from the metagenome of a Mexican hot spring. Sequence-based comparisons revealed that the *Sulfolobales* archaeon AZ1 represents a novel candidate genus.

Received 10 February 2014 **Accepted** 13 February 2014 **Published** 6 March 2014

Citation Servín-Garcidueñas LE, Martínez-Romero E. 2014. Draft genome sequence of the *Sulfolobales* archaeon AZ1, obtained through metagenomic analysis of a Mexican hot spring. Genome Announc. 2(2):e00164-14. doi:10.1128/genomeA.00164-14.

Address correspondence to Luis E. Servín-Garcidueñas, luis.e.servin@gmail.com.

The order *Sulfolobales* is placed within the phylum *Crenarchaeota* (1), and some of its species are considered model organisms (2). The order *Sulfolobales* comprises the genera *Sulfolobus*, *Acidianus*, *Metallosphaera*, *Stygiolobus*, and *Sulfurisphaera* (3–7). *Sulfolobus* contains the highest number of sequenced strains (8–15), and until now, only one complete *Acidianus* genome sequence was available (16). *Metallosphaera* contains three sequenced species (17, 18, 19), including "*Metallosphaera yellowstonensis*," which was first described through metagenomic efforts (20). A novel *Sulfolobales* archaeon has also been discovered from a metagenomic study (21). We report the draft metagenomic sequence of the *Sulfolobales* archaeon AZ1, the first member of the "*Candidatus* Aramenus" genus.

Samples were collected from a hot spring (pH 3.6 and 65°C) located at Los Azufres National Park, Mexico, during March 2009. Environmental DNA was purified using the Ultra-Clean microbial DNA and the Ultra-Clean mega soil DNA kits (MoBio Laboratories, Inc., Carlsbad and Solana Beach, CA). Sequencing was performed with an Illumina-GAIIx platform, producing 36-bp paired-end reads with 300-bp inserts representing 216 Mbp. Reads were assembled *de novo* using Velvet version 1.2.10 (22). A total of 163 contigs were verified by BLASTN searches to be of archaeal origin. All other contigs were assembled into the *Sulfolobales* Mexican rudivirus 1 (SMR1) (23) and the *Sulfolobales* Mexican fusellovirus 1 (SMF1) (24) sequences. All reads were mapped to the archaeal contigs using Maq 0.7.1 (25). The mapping reads were reassembled to eliminate gaps from the archaeal contigs by sequence extensions. Genome annotation was performed with the NCBI Prokaryotic Genomes Automatic Annotation Pipeline version 2.0 (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/). DNA-DNA hybridization (DDH) values were computed using the Genome-to-Genome Distance Calculator (26, 27) version 2.0 (28).

The metagenome assembly was 1,798,894 bp, and only one type of 16S rRNA gene was detected. We retrieved a consensus genome of a population consisting of a dominant *Sulfolobales* archaeon that was designated AZ1. The consensus genome contains 46 contigs with a coverage of 71.9× and an N_{50} value of 223,688 bp. A total of 2,002 genes were predicted, including 1,975 protein-coding genes. The consensus genome had a G+C content of 47%, higher than the 34.1% of the *Acidianus hospitalis* W1 genome and the 32.8 to 36.7% of the *Sulfolobus* genomes. The G+C content more closely resembles the 42 to 47.7% G+C contents of *Metallosphaera* genomes.

The 16S rRNA gene from the archaeon AZ1 shares 93% sequence identity with the corresponding gene from *A. hospitalis* W1. A 95% sequence identity has been proposed as a reasonable value to limit different genera (29, 30). Genome sequence comparisons between the archaeon AZ1 and *A. hospitalis* W1 revealed a DDH estimate of 16.10%, well below the 70% proposed for species definition (30, 31). The archaeon AZ1 would then correspond to a novel *Sulfolobales* genus. The name "*Candidatus* Aramenus sulfurataquae," meaning "the guardian of the sulfurated water," is proposed. The word "Arameni," from the Mexican Purepecha language, means "guardian/custodian of the water." Further characterization of the "*Candidatus* Aramenus" genus is in progress.

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. ASRH00000000.

ACKNOWLEDGMENTS

This research was supported by PAPIIT IN205412 from DGAPA, UNAM, and SUBNARGEM, SAGARPA. L.E.S.-G. received a Ph.D. scholarship from CONACYT (Mexico).

We thank Jean P. Euzéby (nomenclature reviewer of the *International Journal of Systematic and Evolutionary Microbiology*) for reviewing scientific names. We also thank the Programa de Doctorado en Ciencias Biomédicas from UNAM. We thank the UUSM from UNAM for sample sequencing. The Comisión Federal de Electricidad personnel provided a permit for sampling. We thank Jesús Campos García from the UMSNH for providing laboratory facilities. Samplings were carried out with the efforts of Jonathan Lopez, José Luis Servín, and Cecilia Garcidueñas.

REFERENCES

1. **Stetter KO, López-Casillas F, Bai DH, Luo X, Pape ME.** 1989. Order III. *Sulfolobales* ord. nov. family *Sulfolobaceae* fam. nov., p 2250–2251. *In* Staley JT, Bryant MP, Pfennig N, Holt JG, López-Casillas F, Bai DH, Luo X, Pape ME (ed), Bergey's manual of systematic bacteriology, vol 3, 1st ed. Williams & Wilkins, Baltimore, MD.
2. **Leigh JA, Albers SV, Atomi H, Allers T.** 2011. Model organisms for genetics in the domain *Archaea*: methanogens, halophiles, *Thermococcales* and *Sulfolobales*. FEMS Microbiol. Rev. **35:**577–608. http://dx.doi.org/10.1111/j.1574-6976.2011.00265.x.
3. **Brock TD, Brock KM, Belly RT, Weiss RL.** 1972. *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch. Mikrobiol. **84:**54–68. http://dx.doi.org/10.1007/BF00408082.
4. **Segerer A, Neuner AM, Kristjansson JK, Stetter KO.** 1986. *Acidianus infernus* gen. nov., sp. nov., and *Acidianus brierleyi* comb. nov.: facultatively aerobic, extremely acidophilic thermophilic sulfur-metabolizing archaebacteria. Int. J. Syst. Bacteriol. **36:**559–564. http://dx.doi.org/10.1099/00207713-36-4-559.
5. **Huber G, Spinnler C, Gambacorta A, Stetter KO.** 1989. *Metallosphaera sedula* gen. nov. and sp. nov. represents a new genus of aerobic, metal-mobilizing, thermoacidophilic archaebacteria. Syst. Appl. Microbiol. **12:**38–47. http://dx.doi.org/10.1016/S0723-2020(89)80038-4.
6. **Segerer AH, Trincone A, Gahrtz M, Stetter KO.** 1991. *Stygiolobus azoricus* gen. nov., sp. nov. represents a novel genus of anaerobic, extremely thermoacidophilic archaebacteria of the order *Sulfolobales*. Int. J. Syst. Bacteriol. **41:**495–501. http://dx.doi.org/10.1099/00207713-41-4-495.
7. **Kurosawa N, Itoh YH, Iwai T, Sugai A, Uda I, Kimura N, Horiuchi T, Itoh T.** 1998. *Sulfurisphaera ohwakuensis* gen. nov., sp. nov., a novel extremely thermophilic acidophile of the order *Sulfolobales*. Int. J. Syst. Bacteriol. **48**(Pt 2):451–456. http://dx.doi.org/10.1099/00207713-48-2-451.
8. **She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, Van der Oost J.** 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. Proc. Natl. Acad. Sci. U. S. A. **98:**7835–7840. http://dx.doi.org/10.1073/pnas.141222098.
9. **Kawarabayasi Y, Hino Y, Horikawa H, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A, Kosugi H, Hosoyama A, Fukui S, Nagai Y, Nishijima K, Otsuka R, Nakazawa H, Takamiya M, Kato Y, Yoshizawa T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Masuda S, Yanagii M, Nishimura M, Yamagishi A, Oshima T, Kikuchi H.** 2001. Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. DNA Res. **8:**123–140. http://dx.doi.org/10.1093/dnares/8.4.123.
10. **Chen L, Brügger K, Skovgaard M, Redder P, She Q, Torarinsson E, Greve B, Awayez M, Zibat A, Klenk HP, Garrett RA.** 2005. The genome of *Sulfolobus acidocaldarius*, a model organism of the *Crenarchaeota*. J. Bacteriol. **187:**4992–4999. http://dx.doi.org/10.1128/JB.187.14.4992-4999.2005.
11. **Reno ML, Held NL, Fields CJ, Burke PV, Whitaker RJ.** 2009. Biogeography of the *Sulfolobus islandicus* pan-genome. Proc. Natl. Acad. Sci. U. S. A. **106:**8605–8610. http://dx.doi.org/10.1073/pnas.0808945106.
12. **Guo L, Brügger K, Liu C, Shah SA, Zheng H, Zhu Y, Wang S, Lillestøl RK, Chen L, Frank J, Prangishvili D, Paulin L, She Q, Huang L, Garrett RA.** 2011. Genome analyses of Icelandic strains of *Sulfolobus islandicus*, model organisms for genetic and virus-host interaction studies. J. Bacteriol. **193:**1672–1680. http://dx.doi.org/10.1128/JB.01487-10.
13. **Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, Krause DJ, Whitaker RJ.** 2012. Patterns of gene flow define species of thermophilic *Archaea*. PLoS Biol. **10:**e1001265. http://dx.doi.org/10.1371/journal.pbio.1001265.
14. **Mao D, Grogan D.** 2012. Genomic evidence of rapid, global-scale gene flow in a *Sulfolobus* species. ISME J. **6:**1613–1616. http://dx.doi.org/10.1038/ismej.2012.20.
15. **Jaubert C, Danioux C, Oberto J, Cortez D, Bize A, Krupovic M, She Q, Forterre P, Prangishvili D, Sezonov G.** 2013. Genomics and genetics of *Sulfolobus islandicus* LAL14/1, a model hyperthermophilic archaeon. Open Biol. **3:**130010. http://dx.doi.org/10.1098/rsob.130010.
16. **You XY, Liu C, Wang SY, Jiang CY, Shah SA, Prangishvili D, She Q, Liu SJ, Garrett RA.** 2011. Genomic analysis of *Acidianus hospitalis* W1 a host for studying crenarchaeal virus and plasmid life cycles. Extremophiles **15:**487–497. http://dx.doi.org/10.1007/s00792-011-0379-y.
17. **Auernik KS, Maezato Y, Blum PH, Kelly RM.** 2008. The genome sequence of the metal-mobilizing, extremely thermoacidophilic archaeon *Metallosphaera sedula* provides insights into bioleaching-associated metabolism. Appl. Environ. Microbiol. **74:**682–692. http://dx.doi.org/10.1128/AEM.02019-07.
18. **Liu LJ, You XY, Zheng H, Wang S, Jiang CY, Liu SJ.** 2011. Complete genome sequence of *Metallosphaera cuprina*, a metal sulfide-oxidizing archaeon from a hot spring. J. Bacteriol. **193:**3387–3388. http://dx.doi.org/10.1128/JB.05038-11.
19. **Kozubal MA, Dlakic M, Macur RE, Inskeep WP.** 2011. Terminal oxidase diversity and function in "*Metallosphaera yellowstonensis*": gene expression and protein modeling suggest mechanisms of Fe(II) oxidation in the *Sulfolobales*. Appl. Environ. Microbiol. **77:**1844–1853. http://dx.doi.org/10.1128/AEM.01646-10.
20. **Inskeep WP, Rusch DB, Jay ZJ, Herrgard MJ, Kozubal MA, Richardson TH, Macur RE, Hamamura N, Jennings Rd, Fouke BW, Reysenbach AL, Roberto F, Young M, Schwartz A, Boyd ES, Badger JH, Mathur EJ, Ortmann AC, Bateson M, Geesey G, Frazier M.** 2010. Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. PLoS One **5:**e9773. http://dx.doi.org/10.1371/journal.pone.0009773.
21. **Podar M, Makarova KS, Graham DE, Wolf YI, Koonin EV, Reysenbach AL.** 2013. Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. Biol. Direct **8:**9. http://dx.doi.org/10.1186/1745-6150-8-9.
22. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
23. **Servín-Garcidueñas LE, Peng X, Garrett RA, Martínez-Romero E.** 2013. Genome sequence of a novel archaeal rudivirus recovered from a Mexican hot spring. Genome Announc. **1**(1):e00040-12. http://dx.doi.org/10.1128/genomeA.00040-12.
24. **Servín-Garcidueñas LE, Peng X, Garrett RA, Martínez-Romero E.** 2013. Genome sequence of a novel archaeal fusellovirus assembled from the metagenome of a Mexican hot spring. Genome Announc. **1**(2):e00164-13. http://dx.doi.org/10.1128/genomeA.00164-13.
25. **Li H, Ruan J, Durbin R.** 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18:**1851–1858. http://dx.doi.org/10.1101/gr.078212.108.
26. **Auch AF, Klenk HP, Göker M.** 2010. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand. Genomic Sci. **2:**142–148. http://dx.doi.org/10.4056/sigs.541628.
27. **Auch AF, von Jan M, Klenk HP, Göker M.** 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand. Genomic Sci. **2:**117–134. http://dx.doi.org/10.4056/sigs.531120.
28. **Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M.** 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics **14:**60. http://dx.doi.org/10.1186/1471-2105-14-60.
29. **Yarza P, Richter M, Peplies J, Euzeby J, Amann R, Schleifer KH, Ludwig W, Glöckner FO, Rosselló-Móra R.** 2008. The all-species living tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. Syst. Appl. Microbiol. **31:**241–250. http://dx.doi.org/10.1016/j.syapm.2008.07.001.
30. **Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W, Kämpfer P.** 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. Int. J. Syst. Evol. Microbiol. **60:**249–266. http://dx.doi.org/10.1099/ijs.0.016949-0.
31. **Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, Moore LH, Moore WEC, Murray RGE, Stackebrandt E, Starr MP, Truper HG.** 1987. Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. **37:**463–464. http://dx.doi.org/10.1099/00207713-37-4-463.

Diversidad filogenética de las arqueas Sulfolobales

Al realizar análisis filogenéticos de la arquea *Sulfolobales* AZ1 nos dimos cuenta de que existían inconsistencias en la clasificación de las arqueas *Sulfolobales*. El orden *Sulfolobales* contiene varios grupos de arqueas que no han sido descritas y que pueden representar géneros diferentes. Además varias especies descritas formalmente no tienen una clasificación adecuada.

Para estudiar la diversidad de las arqueas *Sulfolobales* realizamos un análisis de agrupamiento de 94 secuencias de genes ribosomales 16S rRNA de cepas tipo, de aislados y de secuencias ambientales. Este análisis reveló 15 grupos de arqueas *Sulfolobales* cuando se aplicó un valor de corte de identidad de secuencia de 94.9%, que es el valor del límite máximo para definir géneros en base a genes ribosomales 16S rRNA. Solo cinco de los 15 grupos contienen cepas tipo de los géneros *Acidianus*, *Metallosphaera, Stygiolobus*, *Sulfolobus* y *Sulfurisphaera*.

Una reconstrucción filogenética del mismo conjunto de secuencias de genes ribosomales reveló que 14 de los 15 grupos identificados en el análisis de agrupamiento son monofiléticos (Figura 4). La única excepción fue la arquea "*Sulfolobus vallisabyssus*" F que no ha sido descrita y que es un grupo externo dentro del grupo de *Sulfurisphaera ohwakuensis* TA1^{T}. También comprobamos que los géneros *Acidianus* y *Sulfolobus* no son monofiléticos.

Las inconsistencias en la clasificación de las arqueas *Sulfolobales* han sido documentadas extensivamente (Stetter et al., 1989; Goebel et al., 2000; Fuchs et al., 1996; Trevisanato et al., 1996; Burggraf et al., 1997; Kurosawa et al., 1998; Huber y Prangishvili, 2006).

A la fecha no se ha adoptado una reclasificación formal de los géneros del orden *Sulfolobales*. Las principales limitantes han sido tradicionalmente la falta de morfologías, de pruebas bioquímicas y de metabolismos diferenciales, además de que muchas cepas reportadas en la literatura no han sido depositadas en centro de cultivo internacionales.

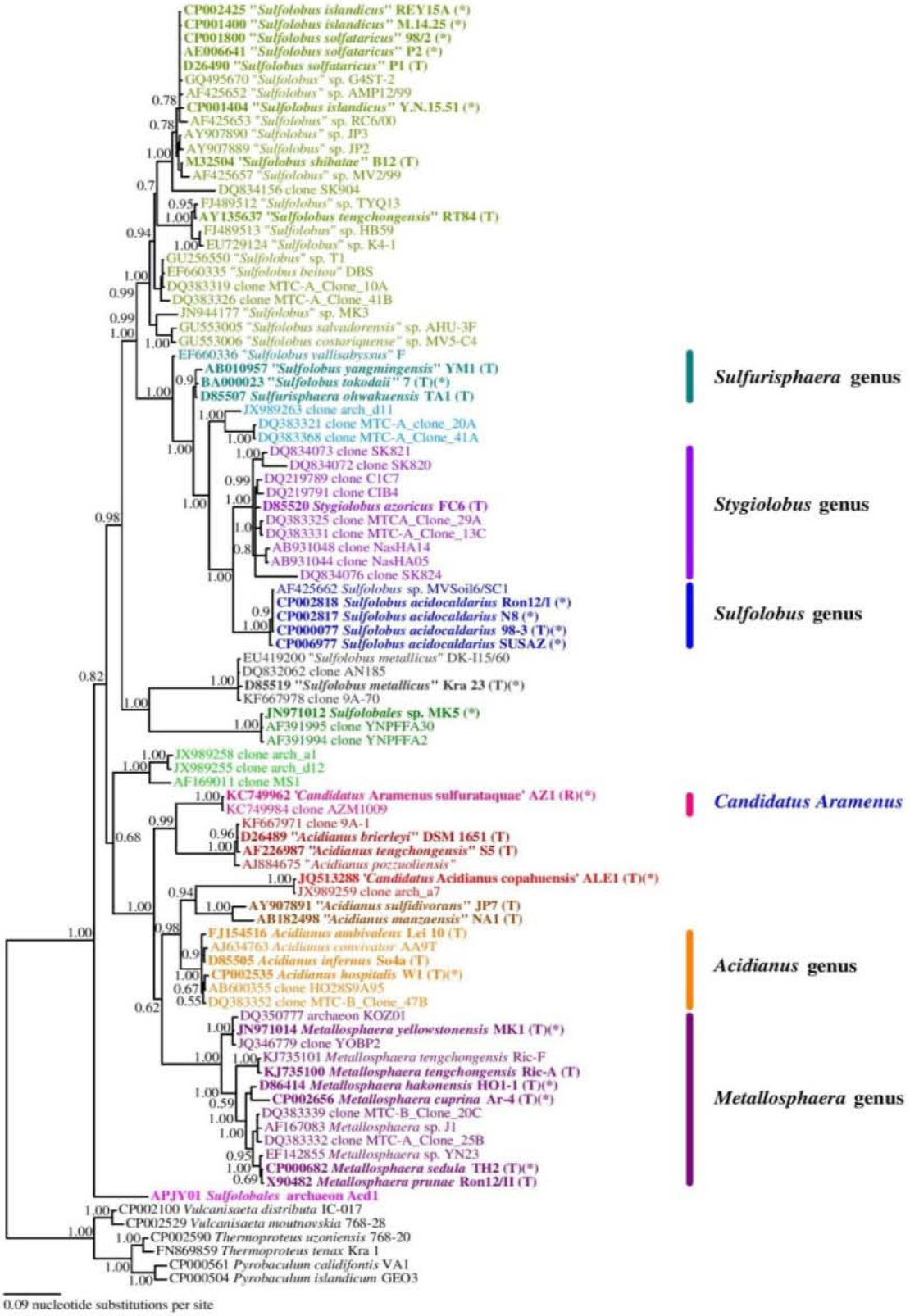


Figura 4. Reconstrucción filogenética de arqueas del orden *Sulfolobales* basada en genes ribosomales 16S rRNA (ver siguiente página para explicación de la reconstrucción filogenética).

(Continuación Figura 4). Se aplicó un valor de corte de identidad de secuencia de 94.9%, que es el valor del límite máximo para definir géneros en base a genes ribosomales 16S rRNA. Los colores indican a los grupos definidos con base en el punto de corte. El alineamiento de secuencias contiene 1335 caracteres. La reconstrucción filogenética bayesiana se realizó en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Se muestran los números de acceso de GenBank de las secuencias utilizadas. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. La barra de escala representa el número promedio de sustitución de nucleótidos por sitio.

Posición filogenómica de la arquea Sulfolobales AZ1

El orden *Sulfolobales* representa un linaje del filo Crenarchaeota dentro del dominio Archaea. El filo Crenarchaeota a la vez es parte del supergrupo 'TACK' junto con los filos Thaumarchaeota, Aigarchaeota y Korarchaeota. El filo Crenarchaeota está integrado por las arqueas *Sulfolobales*, *Acidilobales*, *Desulfurococcales*, *Fervidicoccales* y *Thermoproteales*. Una reconstrucción filogenómica ubica a la arquea *Sulfolobales* AZ1 en una rama externa dentro del grupo que contiene a *Acidianus hospitalis* W1 y '*Candidatus* Acidianus copahuensis' ALE1 (Figura 5). Sin embargo, es necesario contar con genomas adicionales de otros grupos filogenéticos para poder sustentar la posición filogenómica de la arquea *Sulfolobales* AZ1. La posición de la arquea *Sulfolobales* AZ1 depende del número y tipo de secuencias incluidas en las reconstrucciones filogenéticas (Fig. 3, Fig. 4, Fig. 5).

Para realizar la reconstrucción filogenómica se usaron proteomas de arqueas depositados en GenBank. Se utilizó AMPHORA2 para identificar un conjunto de 104 proteínas conservadas en arqueas y que incluyen proteínas ribosomales, proteínas involucradas en transcripción y replicación del ADN así como otras proteínas del metabolismo celular (Wu y Scott, 2012). Las proteínas conservadas además se presentan en copia única en los genomas de arqueas y no presentan eventos de transferencia horizontal. Las secuencias de proteínas se concatenaron y se alinearon usando MUSCLE v.3.8.31 (Edgar, 2004). El alineamiento se editó utilizando Gblocks (Castresana, 2000) para eliminar regiones divergentes y pobremente alineadas.

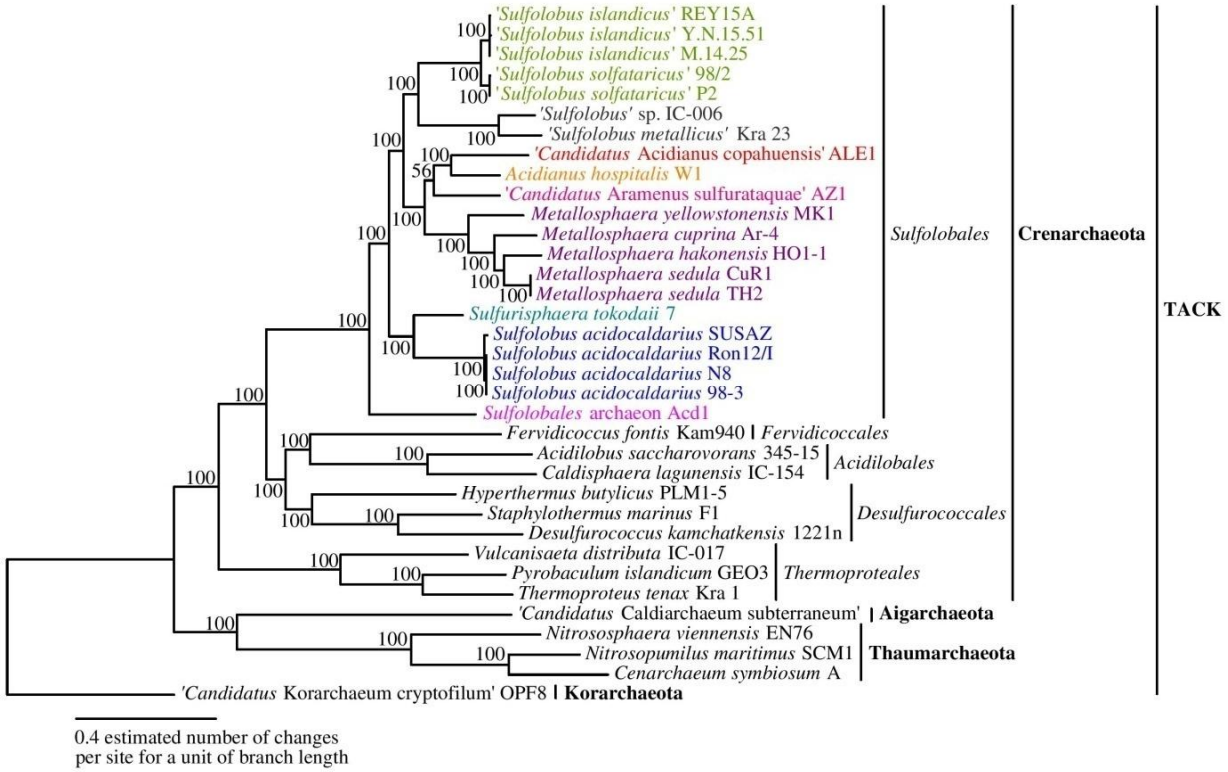


Figura 5. Reconstrucción filogenómica de arqueas *Sulfolobales* y de otras arqueas representativas del supergrupo TACK. El alineamiento editado contiene 24,569 posiciones de amino ácidos. El análisis filogenético corresponde a un análisis de máxima verosimilitud realizado en PhyML (Guindon et al., 2010) bajo el modelo de sustitución WAG. Los valores de soporte de las ramas corresponden a 100 réplicas de Bootstrap debido al alto tiempo de cómputo requerido y por la longitud del alineamiento de secuencias de proteínas. La barra de escala representa el número promedio de sustitución de aminoácidos por sitio.

Metabolismo potencial de la arquea Sulfolobales AZ1

El metabolismo potencial de la arquea *Sulfolobales* AZ1 es consistente con las condiciones geoquímicas de la solfatara ácida (Tabla 1). A partir de la anotación del genoma de la arquea *Sulfolobales* AZ1 se predijeron genes codificantes que potencialmente participan en la fijación de carbono, en la detoxificación de metales pesados y en los procesos de resistencia a temperatura y acidez. Primeramente se encontraron los genes de la ruta completa de biosíntesis de lípidos isoprenoides, que son fundamentales para las membranas de las arqueas.

También se encontraron genes comunes de arqueas termófilas tales como una girasa reversa (superenrollamiento del ADN), la proteína de unión a ADN Sac7d (estabilidad al ADN) y las subunidades del termosoma (chaperonas de proteínas). También se encontraron genes relacionados con la capa S lo que podría indicar la presencia de algún tipo de arquitectura celular.

Las arqueas *Sulfolobales* son capaces de crecer en dióxido de carbono como única fuente de carbono. La identificación de una deshidrogenasa de monóxido de carbono (EWG06769.1) que pudiera permitir a las células hacer uso de monóxido de carbono como fuente de energía y utilizar dióxido de carbono como fuente de carbono indicarían la posibilidad de un metabolismo autotrófico. Esta enzima sería esencial dada la ausencia de carbonatos en el agua de la solfatara ácida (Tabla 1). También se identificaron genes clave del ciclo 3-hidroxipropanoato/4-hidroxibutanato y de la vía de reducción de acetil coenzima A para la fijación de carbono.

Las arqueas *Sulfolobales* presentan metabolismos diversos pero dependen de compuestos azufrados para su crecimiento autotrófico o heterotrófico. Varias arqueas *Sulfolobales* pueden oxidar el azufre elemental a ácido sulfúrico. El potencial genético de la arquea *Sulfolobales* AZ1 incluye enzimas para el metabolismo de compuestos azufrados que se encuentran en altas concentraciones en al agua de la solfatara ácida (Tabla 1). El sulfuro de hidrógeno sería oxidado con la acción de una oxidoreductasa de sulfido-quinona (EWG08227.1) y una sulfuro deshidrogenasa (EWG07347.1). El azufre elemental sería oxidado a través de una oxigenasa-reductasa de azufre (EWG07991.1) para generar sulfitos, tiosulfatos y sulfuro de hidrógeno. Posteriormente una oxidasa de sulfitos (EWG07614.1) estaría oxidando los sulfitos a sulfatos y los sulfatos serían excluidos de la célula con la acción de una permeasa de transporte de sulfatos (EWG07143.1). Además se detectaron genes codificantes para otras oxidoreductasas de sulfuros (EWG06962.1, EWG07174.1, EWG07347.1, EWG07348.1, EWG07964.1, EWG08187.1).

Los análisis químicos del agua de la solfatara ácida indican que los fosfatos son limitantes (Tabla 1). En el genoma de la arquea *Sulfolobales* AZ1 se detectaron genes involucrados en la asimilación y la regulación del transporte de fósforo incluyendo un transportador de fosfato (EWG07892.1), reguladores del transporte de fosfatos (EWG07893.1, EWG08278.1, EWG08059.1), una pirofosfatasa (EWG06940.1) y una exopolifosfatasa (EWG06476.1).

Respecto a resistencia a metales pesados, encontramos que el genoma de la arquea *Sulfolobales* AZ1 codifica para una bomba de exclusión de arsénico (KJR79114.1), una reductasa de mercurio (EWG07641.1), una proteína periplásmica de tolerancia a cationes divalentes (EWG07078.1), una proteína de resistencia a cobalto/zinc/cadmio (EWG07495.1), un transportador de níquel (EWG06504.1) y una ATPasa de transporte de metales pesados (EWG06513.1). La afinidad de estos transportadores debe ser explorada pero es probable que estos transportadores sean esenciales para la detoxificación de varios metales pesados que se encuentran presentes en el agua de la solfatara ácida (Tabla 1). La solfatara ácida también posee concentraciones elevadas de hierro (Tabla 1) y en relación, se identificaron dos transportadores de hierro (EWG08191.1, EWG07908.1) en el genoma de la arquea *Sulfolobales* AZ1.

No se detectaron genes para fijación de nitrógeno como ocurre en otras arqueas *Sulfolobales*. Las búsquedas genómicas sugieren que el amonio es la principal fuente de nitrógeno. Se detectaron genes que codifican para un transportador de amonio (EWG07125.1), una carbamoil fosfato sintetasa (EWG08080.1, EWG08081.1), una glutamina sintetasa (EWG07704.1, EWG07765.1, EWG06893.1), y una glutamato deshidrogenasa (EWG08210.1). La presencia de estos genes sugiere que el amonio pudiera ser asimilado mediante la formación de carbamoil fosfato, glutamina y glutamato. Los análisis químicos indican la presencia de nitratos en el agua de la solfatara ácida (Tabla 1) y en relación el genoma de la arquea *Sulfolobales* AZ1 codifica un transportador ABC para el transporte de nitratos (EWG07885.1).

Algunas otras arqueas *Sulfolobales* son heterótrofas facultativas y pueden utilizar compuestos orgánicos complejos. El genoma de la arquea *Sulfolobales* AZ1 codifica varios transportadores ABC para el transporte de azúcares (EWG06676.1, EWG06677.1, EWG06678.1, EWG06681.1, EWG07766.1, EWG07886.1). También se identificaron permeasas y transportadores involucrados en el transporte de amino ácidos (EWG06646.1, EWG06703.1, EWG06792.1, EWG06878.1, EWG07452.1, EWG07707.1, EWG07937.1, EWG07972.1, EWG06473.1, EWG06484.1, EWG07668.1, EWG07683.1). Además se identificaron tres clústeres de genes que codifican para transportadores ABC de oligopéptidos y dipéptidos (EWG07878.1-EWG07881.1, EWG06691.1-EWG06694.1, EWG06414.1-EWG06418.1).

Finalmente, se identificaron genes que codifican para enzimas proteolíticas tales como una carboxipeptidasa (EWG07607.1), endopeptidasas (EWG07258.1, EWG08229.1, EWG08286.1, EWG08287.1), pepsinas (EWG07940.1, EWG07555.1), aminopeptidasas (EWG06756.1, EWG07436.1, EWG07574.1, EWG08162.1) y una asparaginasa (EWG07860.1). La presencia de estos genes en el genoma de la arquea *Sulfolobales* AZ1 era esperada ya que la gran mayoría de las arqueas *Sulfolobales* son heterotróficas facultativas. Los genes predichos indican por tanto que la arquea *Sulfolobales* AZ1 puede crecer heterotróficamente usando compuestos orgánicos tales como extracto de levadura, peptona y triptona.

Artículo:

Servín-Garcidueñas LE, Peng X, Garrett RA, Martínez-Romero E. 2013. Genome sequence of a novel archaeal rudivirus recovered from a Mexican hot spring. Genome Announc. 1: e00040-12.

La secuencias metagenómicas de la solfatara ácida permitieron identificar la presencia y la reconstrucción de genomas de arqueovirus novedosos. Los arqueovirus son abundantes en manifestaciones termales en todo el planeta, sin embargo no se habían realizado estudios sobre los virus que habitan en sitios termales de México.

En este trabajo reportamos el genoma de un arqueovirus que representa un linaje novedoso dentro de la familia *Rudiviridae* y que se nombró como SMR1 (*Sulfolobales Mexican rudivirus* 1). Se logró obtener un genoma de ADN lineal con una cobertura de 240X debido a la abundancia de secuencias metagenómicas del arqueovirus SMR1.

El análisis del contenido génico del genoma del arqueovirus SMR1 ayudó en la delimitación de los genes conservados en otros genomas de virus de la familia *Rudiviridae*. El genoma del rudivirus SMR1 es el quinto disponible para la familia *Rudiviridae* y el primero recuperado de un sitio termal del continente americano. A la fecha no se cuenta con rudivirus representativos y secuenciados completamente provenientes del Parque Nacional de Yellowstone.

Se encontró que el genoma del arqueovirus SMR1 exhibe el contenido de G+C (46.6%) más alto de todos los rudivirus secuenciados a la fecha. El contenido de G+C del arqueovirus SMR1 se relaciona con el de la arquea *Sulfolobales* AZ1 (47%), quien es su hospedero probable. En la Figura 6 se presenta una representación gráfica del genoma de SMR1. Las características del genoma y su potencial génico se describen en detalle en el artículo.

genomeA
Journals.ASM.org

Genome Sequence of a Novel Archaeal Rudivirus Recovered from a Mexican Hot Spring

Luis E. Servín-Garcidueñas,[a] Xu Peng,[b] Roger A. Garrett,[b] Esperanza Martínez-Romero[a]

Department of Ecological Genomics, Center for Genomic Sciences, National University of Mexico, Cuernavaca, Morelos, Mexico[a]; Department of Biology, Archaea Centre, University of Copenhagen,Copenhagen, Denmark[b]

We report the consensus genome sequence of a novel GC-rich rudivirus, designated SMR1 (*Sulfolobales* Mexican rudivirus 1), assembled from a high-throughput sequenced environmental sample from a hot spring in Los Azufres National Park in western Mexico.

Received 12 October 2012 **Accepted** 29 October 2012 **Published** 15 January 2013

Citation Servín-Garcidueñas LE, Peng X, Garrett RA, Martínez-Romero E. 2013. Genome sequence of a novel archaeal rudivirus recovered from a Mexican hot spring. Genome Announc. 1(1):e00040-12. doi:10.1128/genomeA.00040-12.

Address correspondence to Luis E. Servín-Garcidueñas, luis.e.servin@gmail.com.

Thermophilic archaeal viruses have been isolated from thermal terrestrial sites, revealing an incredibly large viral diversity (1–3). Mexico contains active volcanoes and geothermal areas extending along the Trans-Mexican Volcanic Belt (TMVB) (4). However, the thermophilic viral diversity present within the TMVB hot springs remains unexplored. Here, we present the consensus genome sequence of a novel rudivirus recovered by iterative *de novo* read mapping and assembly from the metagenome of a hot spring located along the northern edge of the TMVB.

Aqueous sediment samples were collected from an acidic hot spring (pH 3.6 and 65°C) located at Los Azufres, Mexico, in March 2009. The DNA was purified using the UltraClean microbial DNA and UltraClean Mega soil DNA kits (MoBio Laboratories, Inc., Carlsbad and Solana Beach, CA). The metagenomic DNA was sequenced with an Illumina GAIIx platform producing 36-bp paired-end reads with 300-bp inserts representing 216 Mbp. The reads were assembled *de novo* using Velvet 1.2.07 (5). The contigs with overrepresented coverage were verified by BLASTX searches to be of viral origins. The reads were mapped to the viral contigs using Maq 0.7.1 (6), and the mapping reads were reassembled to eliminate gaps. The coding sequences were predicted using GeneMark.hmm 2.0 (7) and were manually verified using Artemis (8).

The sequence coverage of the 27,431-bp double-stranded DNA genome was 240-fold. The presence of an inverted terminal repeat (1,240 bp), characteristic of the linear rudiviral genomic termini, indicated that the genome was complete or almost complete. The G+C content of 46.6% was higher than the 25 to 39% content of the four rudiviral genomes characterized previously (9–11). The host is likely to be a member of the order *Sulfolobales*, the sequences of which dominated the metagenome. Moreover, the G+C content of *Sulfolobales* Mexican rudivirus 1 (SMR1) was similar to those of the *Metallosphaera* genomes (~45%). Thirty-seven open reading frames (ORFs) were identified, 19 of which have putative homologs in the other characterized rudiviruses; this strongly supports SMR1 being a member of the *Rudiviridae* family. Common annotated gene products include the major coat protein, three minor structural proteins, two glycosyl transferases, a clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas4-like protein, a putative replication protein, a Holliday junction helicase, a Holliday junction resolvase, an *S*-adenosylmethionine-dependent methyltransferase, and a putative transcriptional regulator. The seven other shared rudiviral proteins were not assigned functions.

Three additional ORFs show sequence similarities to archaeal ORFs, including one carrying a zinc finger SWIM domain and a predicted CopG domain. Four other ORFs contain domains related to the thioredoxin-like superfamily and the GTP-binding proteins, as well as a ribbon-helix-helix protein and a nop25 domain-containing protein. Eight additional ORFs showed no significant matches. Interestingly, three ORFs were related to viral ORFs of the *Lipothrixviridae* family. Of the 6,000,792 environmental reads, 183,365 (3.05%) mapped to the consensus viral genome and 115 candidate single nucleotide polymorphisms (SNPs) were detected by Maq.

In conclusion, despite the large geographical distance from the locations of other sequenced rudiviruses, SMR1 retained a core set of conserved rudiviral genes that were inferred to be important for the viral life cycle.

Nucleotide sequence accession number. The genome sequence was deposited in GenBank under the accession no. JX944686.

ACKNOWLEDGMENTS

The UUSM from UNAM is thanked for sample sequencing. The Comisión Federal de Electricidad personnel provided a permit for samplings. We thank professor Jesús Campos Garcia from the UMSNH for providing laboratory facilities. The samplings were carried out with the efforts of Hans Román, Jonathan Lopez, José Luis Servín, and Cecilia Garcidueñas.

This research was supported by PAPIIT IN205412 from DGAPA, UNAM and SUBNARGEM, SAGARPA. L.E.S.G. received a PhD scholarship from CONACYT (Mexico) under the Programa de Doctorado en Ciencias Biomédicas from UNAM. L.E.S.G. had a "Beca Mixta" from CONACYT and a fellowship from PAEP during a research internship at the Danish Archaea Center.

REFERENCES

1. **Lawrence CM, Menon S, Eilers BJ, Bothner B, Khayat R, Douglas T, Young MJ.** 2009. Structural and functional studies of archaeal viruses. J. Biol. Chem. **284**:12599–12603.
2. **Pina M, Bize A, Forterre P, Prangishvili D.** 2011. The archeoviruses. FEMS Microbiol. Rev. **35**:1035–1054.
3. **Prangishvili D, Forterre P, Garrett RA.** 2006. Viruses of the Archaea: a unifying view. Nat. Rev. Microbiol. **4**:837–848.
4. **Ferrari L, Orozco-Esquivel MT, Manea V, Manea M.** 2011. The dynamic history of the trans-Mexican volcanic belt and the Mexico subduction zone. Tectonophysics **522–523**:122–149.
5. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.
6. **Li H, Ruan J, Durbin R.** 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18**: 1851–1858.
7. **Besemer J, Borodovsky M.** 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Res. **27**:3911–3920.
8. **Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J.** 2005. ACT: the Artemis comparison Tool. Bioinformatics **21**: 3422–3423.
9. **Peng X, Blum H, She Q, Mallok S, Brügger K, Garrett RA, Zillig W, Prangishvili D.** 2001. Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. Virology **291**:226–234.
10. **Vestergaard G, Häring M, Peng X, Rachel R, Garrett RA, Prangishvili D.** 2005. A novel rudivirus, ARV1, of the hyperthermophilic archaeal genus *Acidianus*. Virology **336**:83–92.
11. **Vestergaard G, Shah SA, Bize A, Reitberger W, Reuter M, Phan H, Briegel A, Rachel R, Garrett RA, Prangishvili D.** 2008. *Stygiolobus* rod-shaped virus and the interplay of crenarchaeal rudiviruses with the CRISPR antiviral system. J. Bacteriol. **190**:6837–6845.

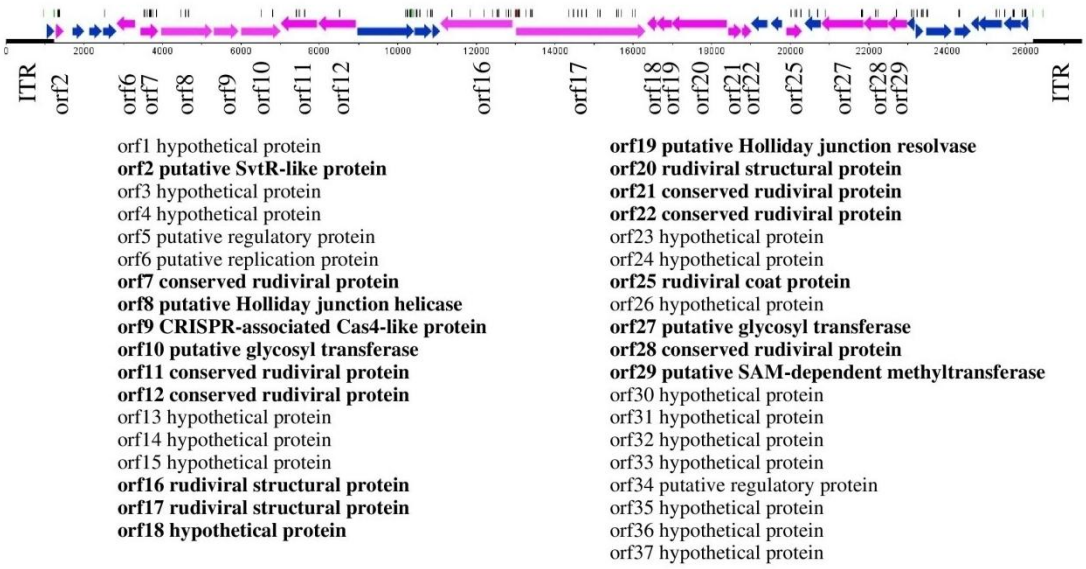


Figura 6. Representación gráfica del genoma del arqueovirus SMR1 (*Sulfolobales Mexican rudivirus* 1). Los ORFs en magenta están conservados en genomas de rudivirus y sus funciones predichas se muestran en las descripciones en negritas. Los ORFs en azul no tienen homólogos en otros genomas de rudivirus. El genoma de SMR1 es lineal y en cada extremo presenta secuencias repetidas terminales invertidas, *ITRs* (*Inverted Terminal Repeats*). Las líneas en negro indican polimorfismos de un solo nucleótido, *SNPs* (*Single Nucleotide Polymorphisms*) localizados en regiones codificantes. Las líneas en rojo indican *SNPs* localizados en regiones no codificantes. Las líneas verdes indican la presencia de indeles.

Artículo:

Servín-Garcidueñas LE, Peng X, Garrett RA, Martínez-Romero E. 2013. Genome sequence of a novel archaeal fusellovirus assembled from the metagenome of a Mexican hot spring. Genome Announc. 1: e0016413.

Por último, el metagenoma de la solfatara ácida contenía secuencias de un segundo arqueovirus perteneciente a la familia *Fuselloviridae***.** El ensamblaje del metagenoma permitió recuperar el genoma de ADN circular de un nuevo fusellovirus que se nombró como SMF1 (*Sulfolobales Mexican fusellovirus 1*). El fusellovirus SMF1 es el más abundante en el metagenoma de la solfatara ya que su genoma se logró obtener con una cobertura de 1,257X.

El genoma de SMF1 presenta un alto sesgo en sus cadenas codificantes ya que de los 24 genes predichos solo uno se encuentra codificado en una de las cadenas. Esto no se había observado en otros genomas de fusellovirus. Además 22 de los 24 genes predichos se encuentran organizados dentro de operones.

El genoma SMF1 codifica para el menor número de genes conservados entre genomas de fusellovirus y exhibe el contenido de G+C más alto (45.43%). El contenido de G+C del arqueovirus SMF1 se relaciona con el de la arquea *Sulfolobales* AZ1 (47%), quien es su hospedero probable. En la Figura 7 se presenta una representación gráfica del genoma de SMF1. Las características del genoma y su potencial génico se describen en detalle en el artículo.

genomeA
Journals.ASM.org

Genome Sequence of a Novel Archaeal Fusellovirus Assembled from the Metagenome of a Mexican Hot Spring

Luis E. Servín-Garcidueñas,[a] Xu Peng,[b] Roger A. Garrett,[b] Esperanza Martínez-Romero[a]

Department of Ecological Genomics, Center for Genomic Sciences, National University of Mexico, Cuernavaca, Morelos, Mexico[a]; Department of Biology, Archaea Centre, University of Copenhagen, Copenhagen, Denmark[b]

The consensus genome sequence of a new member of the family *Fuselloviridae* designated as SMF1 (*Sulfolobales* Mexican fusello-exhibits an exceptional coding strand bias and a reduced set of fuselloviral core genes.

Received 5 March 2013 **Accepted** 15 March 2013 **Published** 25 April 2013

Citation Servín-Garcidueñas LE, Peng X, Garrett RA, Martínez-Romero E. 2013. Genome sequence of a novel archaeal fusellovirus assembled from the metagenome of a Mexican hot spring. Genome Announc. 1(2):e00164-13. doi:10.1128/genomeA.00164-13.

Address correspondence to Luis E. Servín-Garcidueñas, luis.e.servin@gmail.com.

Members of the *Fuselloviridae* family from the crenarchaeal order *Sulfolobales* have been characterized, and they are abundant in extreme geothermal environments (1, 2). They carry circular double-stranded DNA (dsDNA) genomes and exhibit spindle-shaped morphologies. Here, we report the consensus genome sequence of a novel fusellovirus recovered from aqueous sediments from Los Azufres, Mexico.

Samples were collected from a hot spring with a pH of 3.6 and a temperature of 65°C. DNA was purified using the UltraClean microbial and the UltraClean Mega soil DNA kits (MoBio Laboratories, Inc., Carlsbad, CA). Sequencing was performed on an Illumina GAIIx platform, producing 36-bp paired-end reads with 300-bp inserts representing 216 Mb. Reads were assembled using Velvet 1.2.07 (3). A set of contigs were predicted by BLASTX searches to be of fuselloviral origin. Gaps were closed iteratively by mapping and reassembling reads to these contigs using Maq 0.7.1 (4) and Velvet. Open reading frames (ORFs) were predicted using GeneMark.hmm2.0 (5) and were manually verified using Artemis (6).

The average sequence coverage of the 14,847-bp circular dsDNA genome was 1,257-fold. We detected 57 candidate single nucleotide polymorphisms by Maq. The G+C content was 45.43%, higher than the 37.5 to 39.7% content of other fuselloviral genomes (1, 2, 7–9).

The genome has a strong coding-strand bias, not previously seen for fuselloviruses, with only the ubiquitous fuselloviral integrase encoded on one strand. The gene organization is also exceptional for fuselloviruses, with a high incidence of genes arranged in operons, which are also likely to encode cofunctional proteins.

Twenty-four genes were predicted, 22 of which are arranged in five operons. Fourteen genes have putative fuselloviral homologs, consistent with SMF1 being a member of the *Fuselloviridae* family. Most gene products show 30 to 70% amino acid sequence similarity to the best fuselloviral matches. Previous studies identified thirteen genes conserved in all fusellovirus genomes (2), and nine of these were localized in a "core" genomic region of SMF1. The core genes encode a DnaA-like protein, the integrase, one VP1-like structural protein, a putative helix-turn-helix (HTH) transcriptional regulator, and five proteins with unknown functions.

Five additional putative gene products shared with other fuselloviruses include a second VP1-like protein, a VP2-like structural protein, a putative end-filament protein, a regulatory protein, and a hypothetical protein. Three further nonconserved ORF products showed sequence similarities to putative regulatory proteins.

The host of SMF1 is likely to be a member of the order *Sulfolobales*. Fuselloviruses can replicate in both *Sulfolobus* and *Acidianus* species of the order *Sulfolobales* (2), and they are predicted to have an extended host range that may include as-yet-uncultured species (10).

In conclusion, the SMF1 genome was recovered from a site widely separated geographically from the locations of other sequenced fuselloviruses. The SMF1 genome shows exceptional properties, including a coding-strand bias and a high incidence of genes organized in operon structures, but nevertheless, it retains a large set of conserved fusellovirus genes, which lends further support to the exchange of genetic material over intercontinental distances (2, 10).

Nucleotide sequence accession number. The genome sequence was deposited in GenBank under the accession no. KC618393.

ACKNOWLEDGMENTS

This research was supported by PAPIIT IN205412 from DGAPA, UNAM, and by SUBNARGEM, SAGARPA. L.E.S.G. received a PhD scholarship from CONACYT (Mexico). L.E.S.-G. had a "Beca Mixta" from CONACYT and a fellowship from PAEP during a research internship at the Danish Archaea Center.

We thank the Programa de Doctorado en Ciencias Biomédicas from UNAM, the UUSM from UNAM for sample sequencing, and Jesús Campos García from the UMSNH for providing laboratory facilities. The CFE personnel provided a permit for samplings.

REFERENCES

1. **Wiedenheft B, Stedman K, Roberto F, Willits D, Gleske AK, Zoeller L, Snyder J, Douglas T, Young M.** 2004. Comparative genomic analysis of hyperthermophilic archaeal *Fuselloviridae* viruses. J. Virol. **78**:1954–1961.
2. **Redder P, Peng X, Brügger K, Shah SA, Roesch F, Greve B, She Q, Schleper C, Forterre P, Garrett RA, Prangishvili D.** 2009. Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. Environ. Microbiol. **11**:2849–2862.
3. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.
4. **Li H, Ruan J, Durbin R.** 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18**: 1851–1858.
5. **Besemer J, Borodovsky M.** 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Res. **27**:3911–3920.
6. **Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J.** 2005. ACT: the Artemis comparison tool. Bioinformatics **21**: 3422–3423.
7. **Palm P, Schleper C, Grampp B, Yeats S, McWilliam P, Reiter WD, Zillig W.** 1991. Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. Virology **185**:242–250.
8. **Stedman KM, She Q, Phan H, Arnold HP, Holz I, Garrett RA, Zillig W.** 2003. Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2. Res. Microbiol. **154**: 295–302.
9. **Peng X.** 2008. Evidence for the horizontal transfer of an integrase gene from a fusellovirus to a pRN-like plasmid within a single strain of *Sulfolobus* and the implications for plasmid survival. Microbiology **154**:383–391.
10. **Snyder JC, Wiedenheft B, Lavin M, Roberto FF, Spuhler J, Ortmann AC, Douglas T, Young M.** 2007. Virus movement maintains local virus population diversity. Proc. Natl. Acad. Sci. U. S. A. **104**:9102–19107.

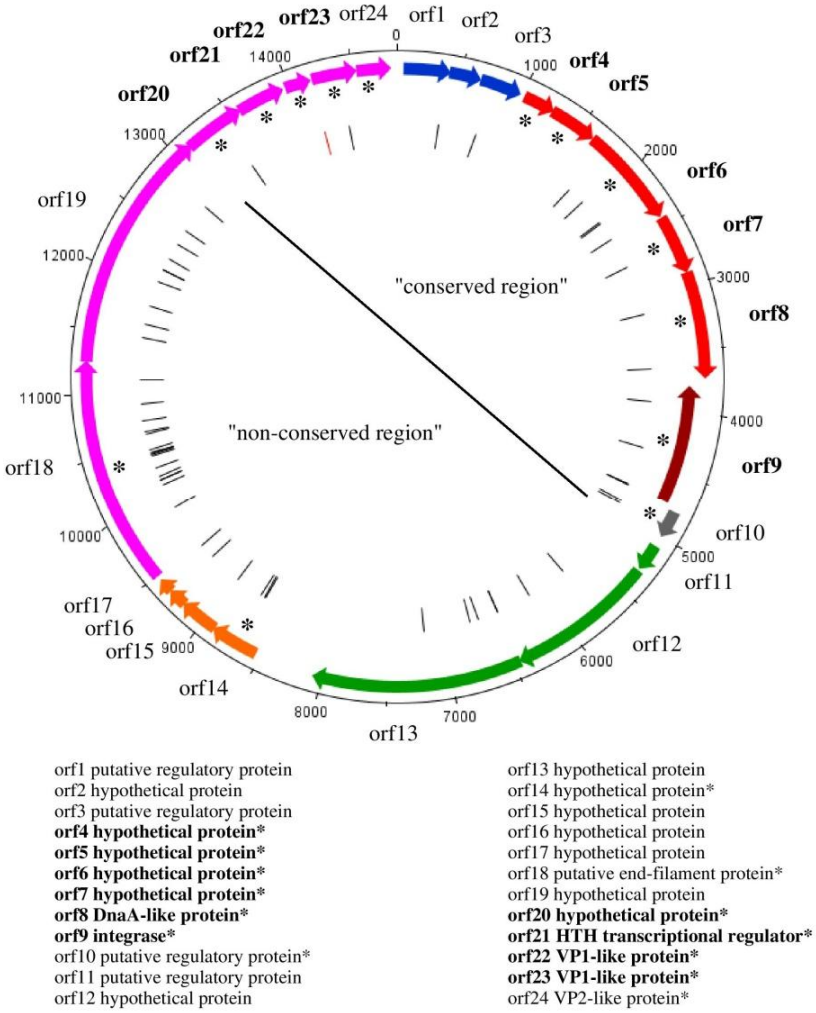


Figura 7. Representación gráfica del genoma del arqueovirus SMF1 (*Sulfolobales Mexican fusellovirus* 1). Los ORFs en colores distintos corresponden a diferentes operones. Los ORFs conservados en genomas de fusellovirus se indican con asteriscos. Los ORFs en negritas corresponden a los genes *core*. Las funciones predichas de los ORFs se indican en las descripciones. Las líneas en negro indican *SNPs* localizados en regiones codificantes. Las líneas rojas indican *SNPs* localizados en regiones no codificantes.

Diversidad genómica de un metagenoma adicional de la solfatara ácida

Como se describió anteriormente, a partir del análisis del primer metagenoma de la solfatara ácida se pudieron recuperar secuencias genómicas de la arquea *Sulfolobales* AZ1 (ensamble Illumina) y de los arqueovirus SMR1 y SMF1. Sin embargo, las librerías de clonas de genes ribosomales 16S rRNA lograron identificar la presencia adicional de arqueas del género *Thermoproteus* de muestras de agua con sedimentos del fondo de la solfatara.

Se realizó la secuenciación de un segundo metagenoma a partir de ADN purificado de muestras colectadas durante el año 2010 del fondo de la solfatara ácida con la finalidad de recuperar secuencias genómicas de otras arqueas. El segundo metagenoma se secuenció en un equipo 454 GS-FLX Titanium en el Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO). El metagenoma está conformado por 706,719 lecturas con una longitud promedio de 425 bases y comprenden 292 megabases. El ensamble de secuencias se realizó usando Newbler v.2.3 (454 *Life Sciences*) y la agrupación de los *contigs* en categorías taxonómicas (*binning*) se realizó utilizando el programa MaxBin que considera la cobertura de lecturas, la frecuencia de tetranucleótidos y la presencia de genes marcadores (Wu et al., 2014). La afiliación taxonómica de los *contigs* fue verificada manualmente realizando búsquedas de BLASTN.

Se logró obtener un genoma adicional de la arquea *Sulfolobales* AZ1 (ensamble 454) a partir del ensamble del segundo metagenoma (Tabla 3). Un alineamiento global de Mauve (Darling et al., 2010) muestra que los dos ensambles de la arquea *Sulfolobales* AZ1 están altamente conservados (Figura 8). Estos resultados proveen evidencia sobre la presencia estable a través del tiempo de la arquea *Sulfolobales* AZ1 en la solfatara ácida.

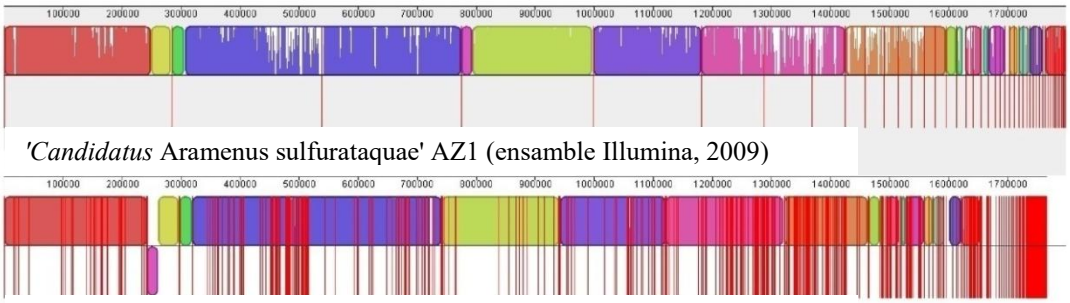


Figura 8. Alineamiento global de los ensambles genómicos de la arquea *Sulfolobales* AZ1 recuperados de los metagenomas de la solfatara ácida de dos años consecutivos.

El ensamble del segundo metagenoma además permitió recuperar los genomas parciales de arqueas anaeróbicas del orden *Thermoproteales* que corresponden a *Thermoproteus* sp. AZ2 y *Vulcanisaeta* sp. AZ3 (Tabla 3). Las librerías de clonas habían permitido identificar la presencia de *Thermoproteus* pero no se pudieron recuperar secuencias genómicas del primer metagenoma que se hizo de la superficie aeróbica de la solfatara. La presencia de *Vulcanisaeta* no se había detectado ni en las librerías de clonas ni en el primer metagenoma. Estos resultados nos indican que la solfatara ácida presenta variación en la diversidad de arqueas dependiendo del gradiente de oxígeno. Basados en estas observaciones, la arquea *Sulfolobales* AZ1 correspondería a un organismo que es aeróbico facultativo mientras que las arqueas *Thermoproteus* sp. AZ2 y *Vulcanisaeta* sp. AZ3 corresponderían a organismos anaeróbicos.

Tabla 3. Características de los genomas obtenidos del segundo metagenoma de la solfatara ácida y de genomas relacionados de cepas tipo.

Arquea secuenciada	Número de acceso	Número de contigs	Tamaño (pb)	G+C
'*Candidatus* Aramenus sulfurataquae' AZ1 (ensamble Illumina)	ASRH00000000	46	1,798,894	47
'*Candidatus* Aramenus sulfurataquae' AZ1 (ensamble 454)	JZWS00000000	374	1,765,812	46.5
Thermoproteus tenax Kra 1^{T}	FN869859	1	1,841,542	55.1
Thermoproteus uzoniensis 768-20	CP002590	1	1,936,063	59.7
Thermoproteus sp. AZ2	JZWT00000000	207	1,691,324	58.0
Vulcanisaeta distributa IC-017^{T}	CP002100	1	2,374,137	45.4
Vulcanisaeta souniana IC-059^{T}	DRX015960	81	2,438,900	45.5
Vulcanisaeta moutnovskia 768-28	CP002529	1	2,298,983	42.4
Vulcanisaeta sp. AZ3	JZWU00000000	478	1,446,340	42.4

Secuencia genómica de Sulfolobus acidocaldarius SUSAZ

Sulfolobus acidocaldarius es una arquea termoacidófila que habita frecuentemente en fuentes y sedimentos termales en todo el mundo. A la fecha, las comparaciones genómicas han mostrado que los genomas de *S. acidocaldarius* están altamente conservados a pesar de los orígenes geográficos distintos de las cepas actualmente secuenciadas (Mao y Grogan, 2012). Las cepas secuenciadas corresponden a la cepa tipo 98-3 aislada del Parque Nacional de Yellowstone, a la cepa Ron 12/I aislada de una área minera de Alemania y a la cepa N8 aislada de un campo termal en la isla japonesa de Hokkaido. Las cepas en promedio solo exhiben cerca de 26 diferentes polimorfismos de una sola base (Mao y Grogan, 2012). La comunidad científica trabaja principalmente con estas tres cepas de *S. acidocaldarius*.

Es necesario contar con más genomas de cepas de *S. acidocaldarius* de distintas regiones geográficas con la finalidad de revelar la diversidad genómica de la especie. En este estudio, analizamos la secuencia genómica de una cepa de *S. acidocaldarius* aislada de Los Azufres.

La cepa SUSAZ se aisló de una muestra de sedimentos termales y ácidos (pH 2.6 y 66.8 °C) que rodean a la solfatara ácida de Los Azufres (Fig. 1A). Un gramo de muestra se añadió a 50 ml de medio TYS suplementado con 0.2% de triptona, 0.1% de extracto de levadura y 0.2% de sacarosa (Zillig et al., 1994). El medio de cultivo general se incubó aeróbicamente por 5 días a 67 °C. Después se establecieron dos litros del cultivo enriquecido en medio TYS a 67 °C. Posteriormente, un mililitro del cultivo enriquecido se diluyó en 250 ml de medio TYS y se incubó la mezcla por 3 días adicionales. Para el aislamiento de colonias, la muestra diluida se inoculó en medio sólido TYS y se incubó a 67 °C por una semana. Las colonias únicas se purificaron y se extrajo su ADN para verificar su afiliación taxonómica amplificando y secuenciado sus genes ribosomales 16S rRNA.

Todas las colonias purificadas correspondieron a la especie *Sulfolobus acidocaldarius*. No fue posible recuperar colonias de la arquea *Sulfolobales* AZ1 ni de *Thermoproteus* ni *Vulcanisaeta* que fueron identificadas en los metagenomas de la solfatara ácida.

La cepa SUSAZ se cultivó en medio líquido TYS para su almacenamiento y para generar cultivos para extraer ADN genómico para la secuenciación de su genoma. El ADN se purificó utilizando el kit *DNeasy Blood & Tissue Kit* (Qiagen, Hilden, Germany) y se secuenció en Macrogen, Corea del Sur.

Las secuencias de los genes ribosomales 16S rRNA de cepas de *S. acidocaldarius* están altamente conservadas. Las cepas 98-3, Ron12/I y N8 comparten 100% de identidad de secuencia en ese gen mientras que la cepa SUSAZ muestra 99% de identidad de secuencia con las otras tres cepas. Un análisis filogenético basado en el gen ribosomal 16S rRNA localiza a la cepa SUSAZ como una rama externa al grupo que contiene a las cepas 98-3, Ron12/I y N8 (Figura 9).

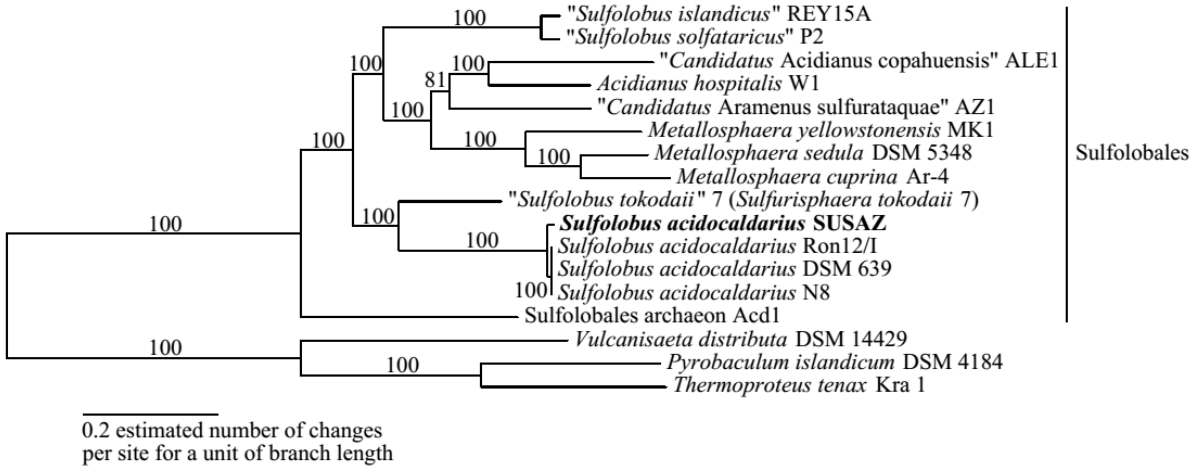


Figura 9. Reconstrucción filogenética de arqueas del orden *Sulfolobales*. El alineamiento de secuencias contiene 1424 caracteres. El análisis filogenético corresponde a un análisis de máxima verosimilitud realizado en PhyML (Guindon et al., 2010) utilizando el modelo de sustitución GTR+I+G. La posición de la cepa SUSAZ recuperada de Los Azufres se resalta en negritas. Los valores de soporte de las ramas corresponden a 1000 réplicas de Bootstrap. La barra de escala representa el número promedio de sustitución de nucleótidos por sitio.

El genoma de la cepa SUSAZ se secuenció en la plataforma de Illumina MiSeq con una cobertura de 1,568X. Las lecturas pareadas de MiSeq se generaron para sobrelaparse y generar lectura mixtas más largas. Las lecturas pareadas se unieron utilizando FLASH v. 1.2.11 (Magoc y Salzberg, 2011) y después se ensamblaron utilizando Newbler v. 2.3 (454 *Life Science*s).

El genoma completo de la cepa SUSAZ es circular, presenta un contenido de G+C de 36.3%, consiste de 2,061,920 pares de bases y codifica para 2,146 genes predichos (Figura 10). La anotación del genoma se realizó en GenBank utilizando el sistema del *NCBI Prokaryotic Genome Annotation Pipeline*. El genoma de la cepa SUSAZ ya se encuentra depositado en GenBank con el número de acceso CP006977. La secuencia del genoma también se cargó y se registró en el sistema IMG (*Integrated Microbial Genomes*) para obtener estadísticos sobre su genoma y para asignar los genes predichos dentro de categorías funcionales (Tabla 4).

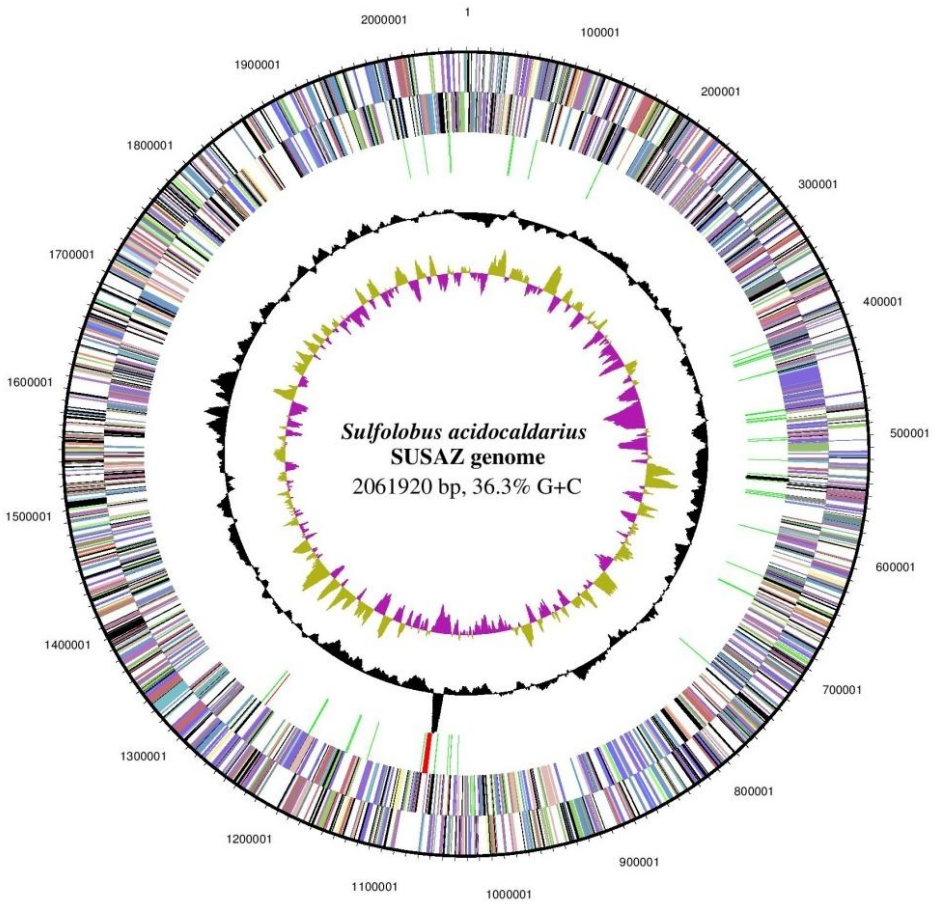


Figura 10. Representación gráfica del genoma de *S. acidocaldarius* SUSAZ. Los anillos de afuera hacia el centro muestran genes de la cadena sentido, genes de la cadena reversa, genes de ARN (tARNs en verde, rARNs en rojo y otros ARNs en negro), contenido de G+C y sesgo de G+C.

Tabla 4. Genes asociados a categorías funcionales *COGs* (*Clusters of Orthologous Groups*).

Código	Valor	% del total	Descripción
J	167	9.59	*Translation, ribosomal structure and biogenesis*
A	2	0.11	*RNA processing and modification*
K	102	5.86	*Transcription*
L	80	4.59	*Replication, recombination and repair*
B	3	0.17	*Chromatin structure and dynamics*
D	8	0.46	*Cell cycle control, cell division, chromosome partitioning*
V	16	0.92	*Defense mechanisms*
T	23	1.32	*Signal transduction mechanisms*
M	47	2.70	*Cell wall/membrane/envelope biogenesis*
N	7	0.40	*Cell motility*
U	12	0.69	*Intracellular trafficking and secretion*
O	70	4.02	*Posttranslational modification, protein turnover, chaperones*
C	152	8.73	*Energy production and conversion*
G	96	5.51	*Carbohydrate transport and metabolism*
E	147	8.44	*Amino acid transport and metabolism*
F	62	3.56	*Nucleotide transport and metabolism*
H	109	6.26	*Coenzyme transport and metabolism*
I	95	5.45	*Lipid transport and metabolism*
P	67	3.85	*Inorganic ion transport and metabolism*
Q	54	3.10	*Secondary metabolites biosynthesis, transport and catabolism*
R	279	16.02	*General function prediction only*
S	144	8.27	*Function unknown*
-	629	28.23	*Not in COGs*

Se encontró que el genoma de la cepa SUSAZ es de menor longitud en comparación con los otros genomas secuenciados de *S. acidocaldarius.* El genoma de la cepa 98-3 contiene 2,225,959 pares de bases, la cepa Ron12/I contiene 2,223,983 pares de bases y la cepa N8 contiene 2,176,362 pares de bases. En cambio la cepa SUSAZ contiene 2,061,920 pares de bases. El genoma de la cepa SUSAZ muestra regiones pequeñas de ganancia de genes pero principalmente ausencia de elementos integrados (Figura 11). Se encontró que los elementos integrados en los genomas de *S. acidocaldarius* no están presentes en el genoma de la cepa SUSAZ. Otra diferencia importante es que solo se encontraron dos arreglos de CRISPRs en el genoma de la cepa SUSAZ en lugar de los cuatro comunes en otras cepas de *S. acidocaldarius* (Figura 12). Además las secuencias espaciadoras de los CRISPRs son divergentes respecto a las secuencias presentes en otros genomas de *Sulfolobus*. Esto indica que la cepa SUSAZ ha estado expuesta a otros virus o plásmidos invasores en su ambiente natural en Los Azufres.

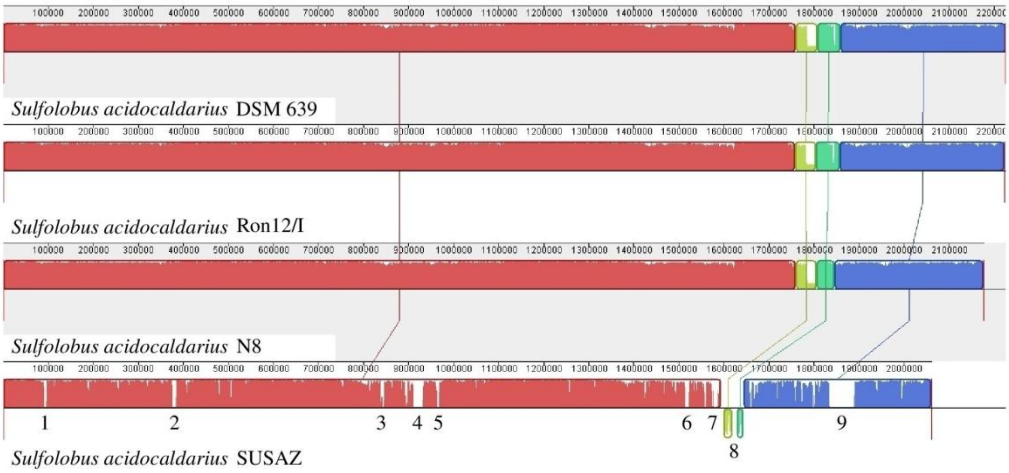


Figura 11. Alineamiento de genomas globales de las cepas secuenciadas de *S. acidocaldarius*. El alineamiento global se realizó en Mauve (Darling et al., 2010). El genoma de la cepa SUSAZ no presenta los elementos integrados comunes de otros genomas de *Sulfolobus*. Las regiones variables del genoma de la cepa SUSAZ se muestran con números. Las regiones 1 a 7 corresponden a elementos integrados comunes en los genomas de *S. acidocaldarius* que están ausentes en el genoma de la cepa SUSAZ. Las regiones 8 a 9 corresponden a las regiones donde se encuentran localizados los arreglos de CRISPRs.

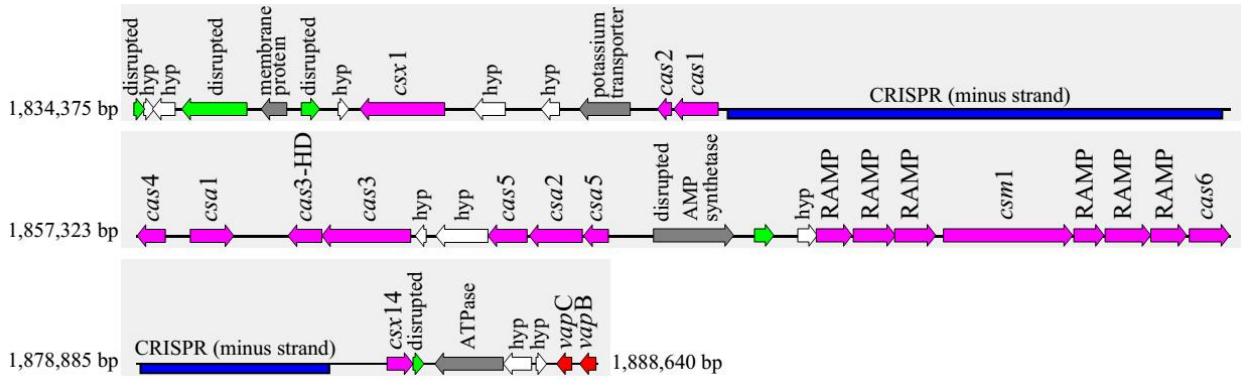


Figura 12. Arreglos de CRISPRs y proteínas asociadas del genoma de *S. acidocaldarius* SUSAZ.

Actualmente se están analizando los potenciales genéticos de las cepas de *S. acidocaldarius* y sus genes ortólogos. Hemos encontrado que la cepa SUSAZ es la que presenta el mayor números de genes únicos (95 genes) y que aproximadamente 1,863 genes (86.81% de los genes de la cepa SUSAZ) están conservados en todas las cepas secuenciadas (Figura 13). El genoma de la cepa SUSAZ es el primer genoma completamente secuenciado de una arquea aislada del Eje Volcánico de México y ayudará en el entendimiento de las relaciones evolutivas y de la dinámica genómica de la especie *S. acidocaldarius*.

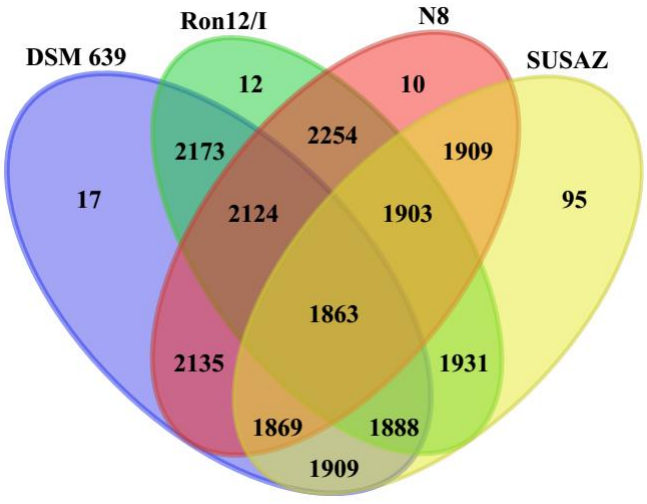


Figura 13. Diagrama de Venn que muestra el número de genes ortólogos que están compartidos entre las cepas secuenciadas de *Sulfolobus acidocaldarius*.

Microscopía electrónica de arqueas y arqueovirus de sedimentos termales

Se obtuvieron imágenes de microscopía electrónica con la finalidad de conocer la morfología de arqueas y arqueovirus identificados en los estudios metagenómicos y genómicos de la solfatara ácida. La visualización de partículas virales fue importante ya que los ensambles de los genomas de los arqueovirus SMR1 y SMF1 podrían solo estar integrados en los genomas de arqueas. Se lograron identificar partículas virales con morfologías características de fusellovirus (Figura 14A-C, con forma de limón alargado) y de rudivirus (Figura 14D-F, con forma de bastones largos). Además se observaron partículas virales desconocidas, lo que implica la presencia de otros virus de los que no se han obtenido secuencias genómicas (Figura 14G-I).

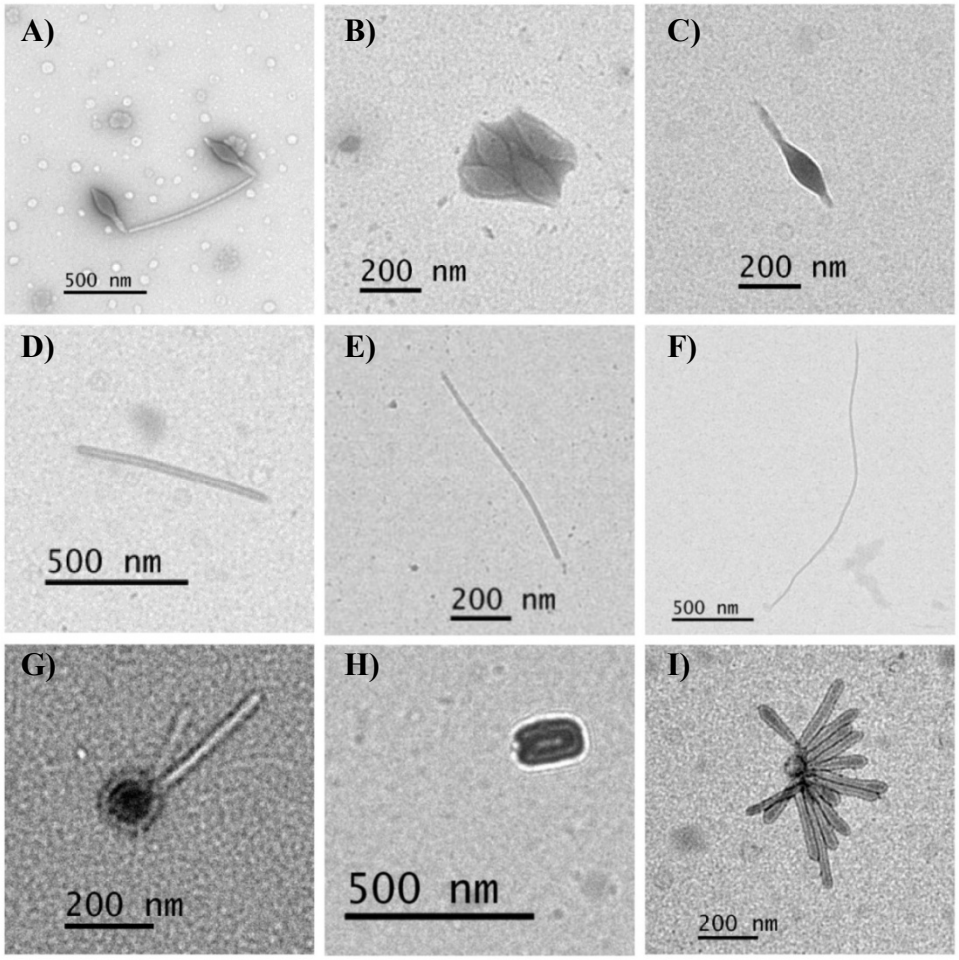


Figura 14. Imágenes de microscopía electrónica de partículas virales de muestras enriquecidas de sedimentos termales de la solfatara ácida de Los Azufres.

La detección de partículas virales adicionales mediante técnicas de microscopía se debe en parte al hecho de que se realizaron enriquecimientos mediante centrifugaciones múltiples para poder concentrar las muestras. Se utilizaron sedimentos termales de la solfatara ya que las partículas virales son más abundantes en sedimentos que en la columna de agua. En cambio, las muestras que se utilizaron para secuenciación masiva fueron tomadas de la columna de agua y no fueron enriquecidas por lo que solo se pudieron obtener secuencias metagenómicas de los virus más abundantes.

No ha sido posible obtener imágenes de microscopía electrónica de la arquea *Sulfolobales* AZ1 ni de *Thermoproteus* ni *Vulcanisaeta* ya que las muestras de agua de la solfatara ácida contienen una gran cantidad de partículas de sedimentos que han impedido su visualización. Los ensayos de enriquecimiento se han hecho de muestras de sedimento de las orillas de la solfatara ácida (Figura 15A). A la fecha solo se han obtenidos colonias de *Sulfolobus acidocaldarius*. Fue posible obtener imágenes de microscopía electrónica para la cepa *S. acidocaldarius* SUSAZ. La cepa SUSAZ presenta la morfología característica de arqueas *Sulfolobales* con una forma lobular irregular y con presencia de flagelos (Figura 15B-C).

Para obtener las imágenes de microscopía se utilizaron muestras de cultivos enriquecidos siguiendo los procedimientos descritos por Erdmann y Garrett, 2015. Las imágenes fueron obtenidas en el Centro Danés de Arqueas usando un microscopio electrónico de transmisión Tecnai G2 (FEI, Eindhoven, Holanda).

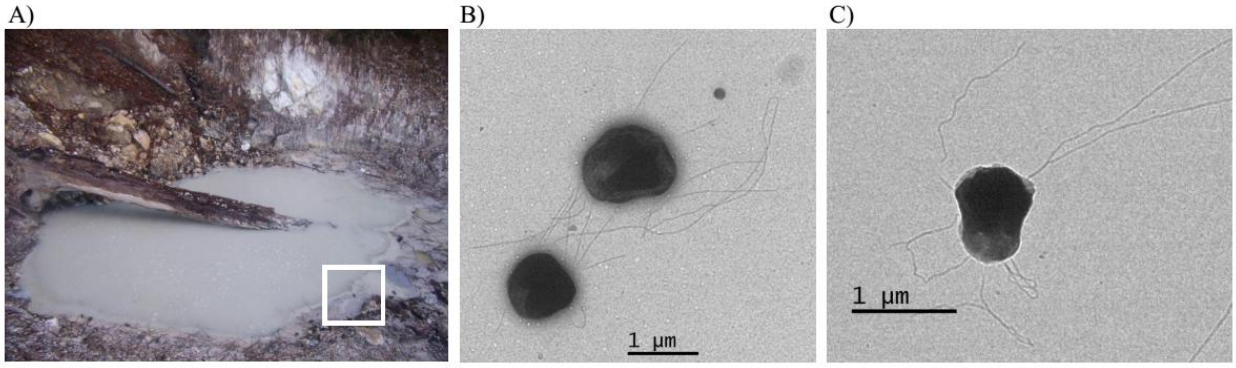


Figura 15. Imágenes de microscopía electrónica de *Sulfolobus acidocaldaricus* SUSAZ. A) zona de donde se obtuvieron las muestras de sedimentos termales de la solfatara ácida.

Diversidad de comunidades microbianas de sedimentos fotosintéticos de Los Azufres

Los estudios de diversidad microbiana de sedimentos termales de sitios volcánicos han revelado una gran diversidad de microorganismos novedosos. Algunos de estos estudios se han realizado a partir de comunidades microbianas de sedimentos expuestos a altas temperaturas alrededor de fumarolas de los Parques Nacionales de Yellowstone y de Lassen, de depósitos volcánicos de Hawaii y de Nueva Zelanda e incluso de la Antártica (Norris et al. 2002; Stott et al. 2008; Soo et al. 2009; Benson et al. 2011).

A la fecha no se han realizado estudios de diversidad microbiana de sedimentos termales del campo geotérmico de Los Azufres. En este estudio realizamos un primer censo de diversidad microbiana mediante la creación de una librería de genes ribosomales 16S rRNA y de análisis filogenéticos. Posteriormente, se purifico ADN ambiental para secuenciar y analizar un metagenoma de sedimentos termales de Los Azufres.

Se colectaron sedimentos termales para purificar y secuenciar ADN ambiental en los años 2009 y 2013 (Fig. 1C, Fig. 16). En ambos años se muestrearon comunidades microbianas de sedimentos de color verde oscuro depositadas sobre rocas en salidas de vapor de fumarolas (Figura 16). Los sedimentos se colectaron frotando cajas Petri sobre las rocas para obtener la mayor cantidad de material posible ya que los sedimentos son de menos un milímetro de grosor.

Se obtuvieron 60 secuencias de genes ribosomales 16S rRNA de bacterias de una librería de clonas generada a partir del ADN purificado de sedimentos colectados durante el año 2009. Las PCRs específicas para arqueas y eucariontes no amplificaron los correspondientes genes ribosomales. El sedimento utilizado estaba expuesto a 65 °C y presentó un pH de 3.5.

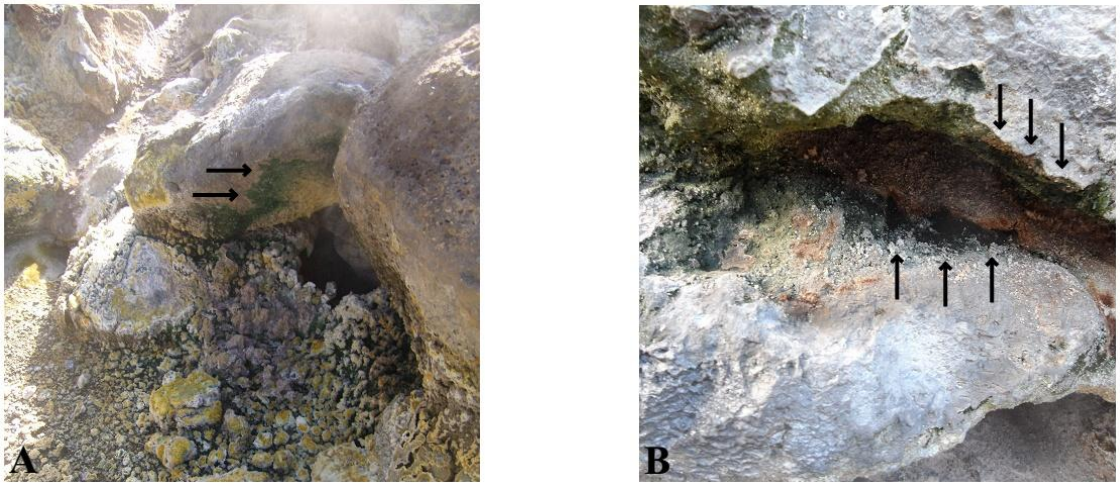


Figura 16. Comunidades microbianas depositados sobre sedimentos termales en fumarolas de Los Azufres. A) Sedimentos colectados durante Julio de 2009 en el área de Maritaro. B) Sedimentos colectados durante Abril de 2013 en un área rocosa de Los Azufres.

La curva de rarefacción de las secuencias de genes ribosomales 16S rRNA de las bacterias del sedimento termal no supera 10 OTUs y ha alcanzado la estabilidad (Figura 17). A partir de las secuencias obtenidas de la librería de clonas se identificaron a los fila bacterianos *Proteobacteria*, *Nitrospirae*, *Thermotogae*, *Firmicutes* y *Chloroflexi* (Tabla 5).

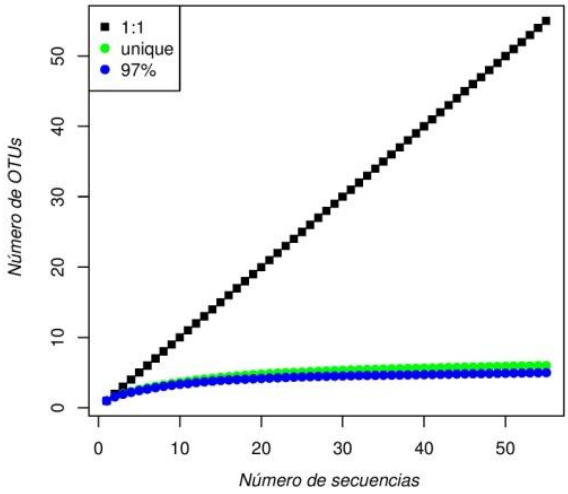


Figura 17. Curva de rarefacción de las secuencias de los genes ribosomales 16S rRNA del censo de diversidad de los sedimentos termales de Los Azufres. La curva de rarefacción incluye las 60 secuencias generadas (longitud= ~1500 nucleótidos). Los datos se obtuvieron utilizando mothur (Oakley et al., 2009) y la gráfica se reconstruyó utilizando R. Se muestran las curvas correspondientes a distintos puntos de corte de acuerdo a la identidad de secuencia.

Las secuencias más abundantes correspondieron a bacterias sin describir de la clase *Ktedonobacteria* del filo *Chloroflexi*. Estás secuencias representan el 58% (35 secuencias) de las obtenidas de la librería de clonas. Las secuencias presentan una identidad de entre 91 y 93% con secuencias de bacterias ambientales (Tabla 5). La identidad de secuencia fue de entre 86 a 88% con las cepas tipo de *Thermosporothrix hazakensis* SK20-1^T, *Ktedonobacter racemifer* DSM 44963^T, *Thermogemmatispora onikobensis* ONI-1^T y *Thermogemmatispora foliorum* ONI-5^T.

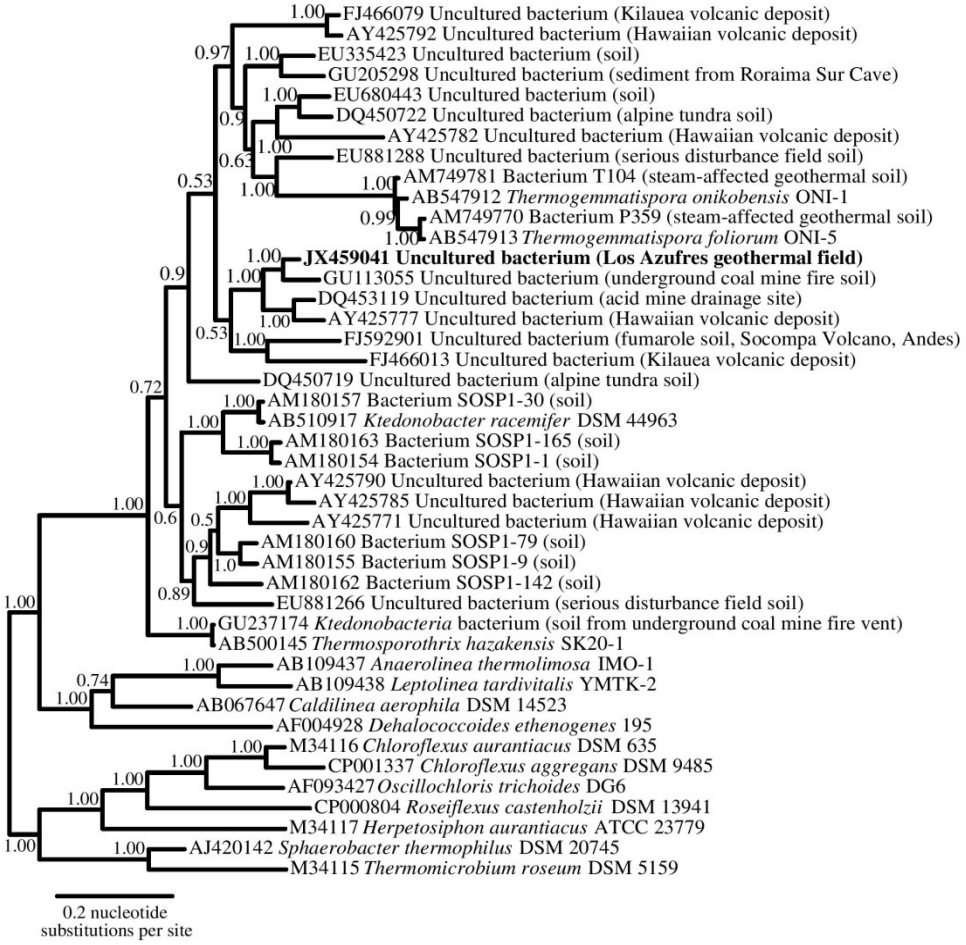


Figura 18. Análisis filogenético basado en genes ribosomales 16S rRNA de bacterias del filo *Chloroflexi*. El alineamiento de secuencias contiene 1382 caracteres. La reconstrucción filogenética bayesiana se realizó en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Se muestran los números de acceso de GenBank de las secuencias utilizadas. En negritas se muestra una secuencia representativa de la librería de clonas del sedimento termal de Los Azufres. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. La barra de escala representa el número promedio de sustitución de nucleótidos por sitio.

Las reconstrucciones filogenéticas posicionan a las secuencias en un grupo filogenético que incluye secuencias de bacterias de minas y de depósitos volcánicos (Figura 18). En este caso, no existe una relación geográfica clara entre las bacterias más cercanas filogenéticamente. En cambio se reconocen condiciones ambientales similares, en este caso termales y volcánicas.

También se obtuvieron siete secuencias idénticas que tienen una identidad de entre 96 y 98% con genes ribosomales 16S rRNA de bacterias del filo *Termotogae*. La identidad de secuencia es de 82% con la cepa tipo más cercana que corresponde a *Kosmotoga olearia* TBF 19.5.1^{T}. Las reconstrucciones filogenéticas agrupan la secuencia de ésta bacteria en un grupo filogenético que incluye secuencias de bacterias ambientales recuperadas de diferentes sitios geotérmicos a nivel mundial. El grupo filogenético que incluye la secuencia de Los Azufres está compuesto únicamente por secuencias de bacterias obtenidas de sitios térmicos continentales tales como Vulcano en Italia, del Parque Nacional de Yellowstone, y de la isla volcánica de Montserrat en el Caribe (Figura 19). A la fecha, solo se encuentran disponibles secuencias de genes ribosomales 16S rRNA para las bacterias de este grupo filogenético.

Otras secuencias presentan 99% de identidad con genes ribosomales 16S rRNA de bacterias relacionadas a *Leptospirillum ferrodiazotrophum* (7 secuencias) y *Acidithiobacillus caldus* (4 secuencias). Los géneros *Leptospirillum* y *Acidithiobacillus* son comunes en las áreas termales y en los drenajes ácidos de minas. También se identificaron secuencias que muestran entre 96 y 97% de identidad con genes ribosomales 16S rRNA de bacterias del filo *Firmicutes* relacionadas con cepas de la especie *Sulfobacillus acidophilus* (3 secuencias). Finalmente, se identificaron secuencias correspondientes a una α-proteobacteria relacionada con bacterias no descritas de ambientes extremos y pertenecientes a la familia *Acetobacteriaceae* (4 secuencias). Las secuencias de la α-proteobacteria presentan 95% de identidad con los genes ribosomales de las cepas tipo *Rhodovastum atsumiense* G2-11^{T} y 94% con *Acidisphaera rubrifaciens* JCM10600^{T} y *Rhodopila globiformis* DSM 161^{T}. Las posiciones filogenéticas de todas estas bacterias están estrechamente relacionadas con bacterias de otros sitios termales (Figura 20). Las secuencias de genes ribosomales 16S rRNA obtenidas de la librería de clonas se encuentran depositadas en GenBank con los números de acceso JX459035 a JX459094.

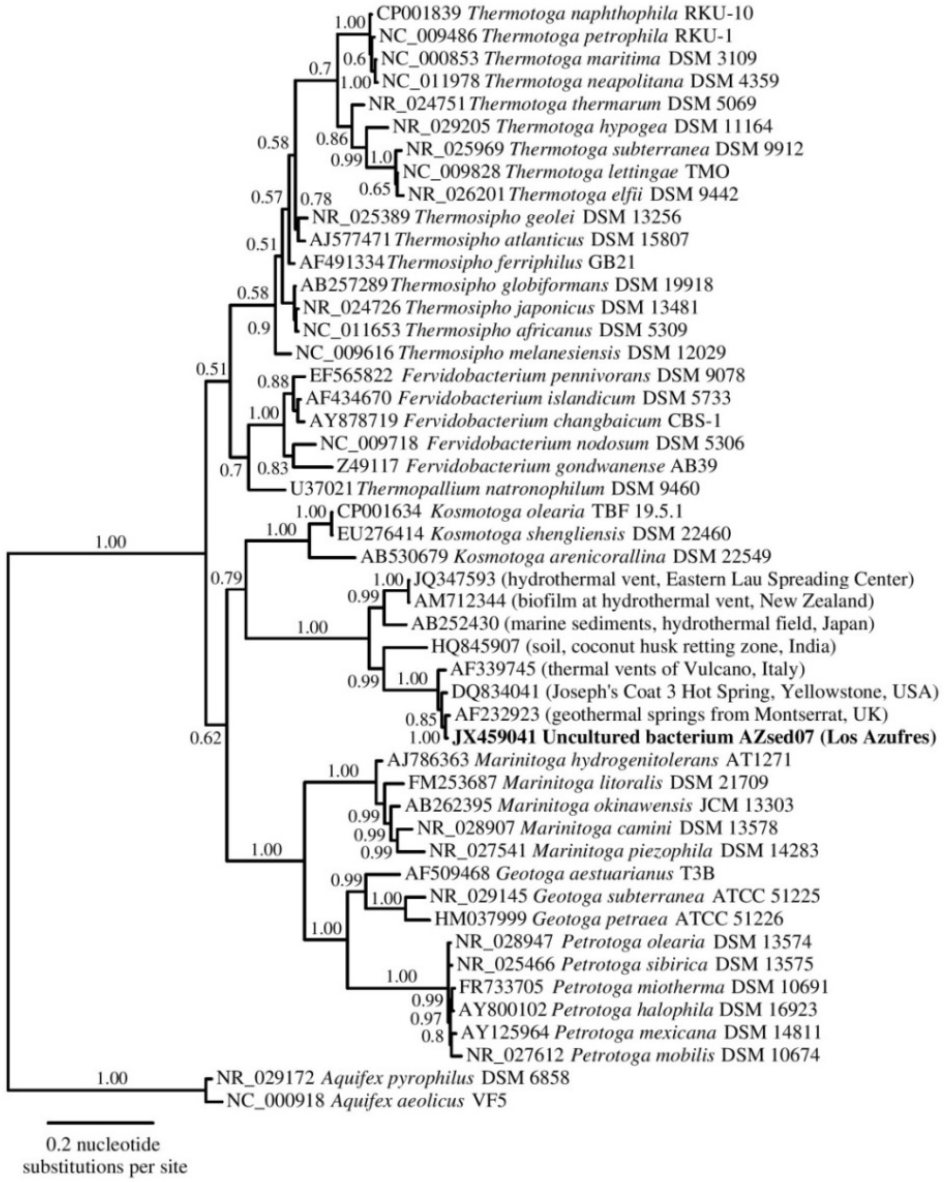


Figura 19. Análisis filogenético basado en genes ribosomales 16S rRNA de bacterias del filo *Thermotogae*. El alineamiento de secuencias contiene 1210 caracteres. La reconstrucción filogenética bayesiana se realizó en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Se muestran los números de acceso de GenBank de las secuencias utilizadas. En negritas se muestra una secuencia representativa de la librería de clonas del sedimento termal de Los Azufres. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. La barra de escala representa el número promedio de sustitución de nucleótidos por sitio.

Tabla 5. Filotipos identificados en los sedimentos termales de Los Azufres colectados en 2009.

Orden	Familia	Género	% de identidad	número de acceso GenBank
Ktedonobacterales	/	/	93%	GU113055
Thermotogales	*Thermotogaceae*	/	99%	EU541481
Nitrospirales	*Nitrospiraceae*	*Leptospirillum*	99%	EF065178
Acidithiobacillales	*Acidithiobacillaceae*	*Acidithiobacillus*	99%	EU499920
Clostridiales	*Clostridiales*	*Sulfobacillus*	97%	GU168017
Rhodospirillales	*Acetobacteraceae*	/	99%	DQ464129

Metagenoma de un sedimento termal de Los Azufres

Los sedimentos termales de Los Azufres se eligieron para purificar ADN metagenómico y hacer su secuenciación debido a que sus comunidades microbianas presentan diversidad limitada y contienen microorganismos filogenéticamente novedosos. Los sedimentos termales están expuestos a condiciones extremas de pH y de temperatura debido a su cercanía a fumarolas y además no existen genomas para algunos de los grupos filogenéticos predominantes.

La muestra de ADN secuenciada se purificó a partir de sedimentos colectados en 2013 y que estaban expuestos a 67 °C y pH 3 (Figura 16). La muestra de ADN de los sedimentos colectados en el año 2009 se utilizó para secuenciación. Sin embargo, la secuenciación no se realizó debido a problemas técnicos de los equipos de secuenciación del LANGEBIO. No se encontraron sedimentos termales parecidos en años subsecuentes hasta el año 2013.

El metagenoma se secuenció en 2014 en la plataforma de Illumina MiSeq en Macrogen, Corea del Sur. Las lecturas pareadas de MiSeq se generaron para sobrelaparse y generar lectura más largas. Las lecturas pareadas se unieron utilizando FLASH v. 1.2.11 (Magoc y Salzberg, 2011) y después se ensamblaron utilizando Newbler v. 2.3 (454 Life Science).

El metagenoma del sedimento termal primeramente se analizó para explorar su diversidad microbiana mediante la identificación de secuencias de genes ribosomales y sus correspondientes análisis filogenéticos.

La identificación de genes ribosomales en el ensamble metagenómico se realizó usando los servidores RNAmmer y WebMGA (Wu et al., 2011; Lagesen et al., 2007). La identidad de las secuencias ribosomales obtenidas se verificó mediante búsquedas de BLASTN en GenBank. No se detectaron secuencias quiméricas usando Bellerophon (Huber et al., 2004) en la base de datos GreenGenes (DeSantis et al., 2006). Las secuencias de genes ribosomales se depositaron en GenBank con los números de acceso KJ907763 a KJ907775 (bacterias), KJ907754 a KJ907762 (arqueas), KJ907776 a KJ907782 (microalgas) y KJ907783 a KJ907785 (otros eucariotas).

En las siguientes tablas y análisis filogenéticos se presentan resultados obtenidos sobre la diversidad de bacterias, arqueas y eucariotas identificados en el metagenoma de los sedimentos. Las secuencias metagenómicas revelaron la composición filogenética de la comunidad de los sedimentos termales y permitieron identificar a los microorganismos fotosintéticos. Además se logró la reconstrucción de los genomas completos de los organelos de una microalga no descrita y la obtención casi completa del primer genoma de una arquea del filo *Parvarchaeota*.

Tabla 6. Secuencias de genes ribosomales 16S rRNA de bacterias identificadas en el metagenoma de sedimentos termales de Los Azufres colectados en 2013.

Bacteria	Longitud (pb)	Número de lecturas	Filo	Clase	GenBank hit	Número de acceso	Identidad
Uncultured bacterium	1424	42	*Proteobacteria*	*α-protebacteria*	Uncultured bacterium clone SX1-107	DQ469200	99%
Uncultured *Acidithiobacillus* sp.	1526	607	*Proteobacteria*	*γ-proteobacteria*	*Acidithiobacillus caldus* ATCC 51756	CP005986	100%
Uncultured gamma proteobacterium	1530	384	*Proteobacteria*	*γ-proteobacteria*	Uncultured bacterium clone YTW-24-06	EF409834	97%
Uncultured *Sinobacteraceae* bacterium	1386	42	*Proteobacteria*	*γ-proteobacteria*	Uncultured bacterium clone CEM_Cya_d1	EU370295	99%
Uncultured *Acidimicrobiaceae* bacterium	1514	1643	*Actinobacteria*	*Actinobacteria*	Uncultured bacterium clone W2bXIIb78	EU419189	99%
Uncultured *Aciditerrimonas* sp.	1514	415	*Actinobacteria*	*Actinobacteria*	*Aciditerrimonas ferrireducens* IC-180	NR_112972	99%
Uncultured *Ferrimicrobium* sp.	1512	306	*Actinobacteria*	*Actinobacteria*	*Ferrimicrobium* sp. Mc9Kl-1-4	HM769774	99%
Uncultured *Mycobacterium* sp.	1510	289	*Actinobacteria*	*Actinobacteria*	*Mycobacterium holsaticum* 1406	NR_028945	98%
Uncultured actinobacterium	1493	263	*Actinobacteria*	*Actinobacteria*	Uncultured bacterium clone WB27	JX133618	95%
Uncultured *Conexibacteraceae* bacterium	1537	107	*Actinobacteria*	*Actinobacteria*	Uncultured bacterium clone bac2nit73	EU861962	98%
Uncultured *Leptospirillum* sp.	1518	29	*Nitrospirae*	*Nitrospira*	*Leptospirillum ferriphilum* BN	JQ820324	99%
Uncultured *Alicyclobacillus* sp.	1517	74	*Firmicutes*	*Bacilli*	*A.disulfidooxidans* DSM 12064	NR_040944	99%
Uncultured *Sulfobacillus* sp.	1312	24	*Firmicutes*	*Clostridia*	Uncultured bacterium clone K17bMu17	EU419137	97%

Tabla 7. Secuencias de genes ribosomales 16S rRNA de arqueas identificadas en el metagenoma de sedimentos termales de Los Azufres colectados en 2013.

Arquea	Longitud (pb)	Número de lecturas	Filo	Clase	GenBank hit	Número de acceso	Identidad
Uncultured *Thermoplasmatales* archaeon (G-plasma)	1468	2100	*Euryarchaeota*	*Thermoplasmata*	Uncultured *Thermoplasmatales* archaeon G-plasma	JX997948	100%
Uncultured *Thermoplasmatales* archaeon (C-plasma)	1470	1764	*Euryarchaeota*	*Thermoplasmata*	Uncultured *Thermoplasmatales* archaeon C-plasma	JX997949	99%
Uncultured *Thermoplasmatales* archaeon (A-plasma)	1469	1419	*Euryarchaeota*	*Thermoplasmata*	Uncultured *Thermoplasmatales* archaeon A-plasma	JX997946	100%
Uncultured *Thermoplasmatales* archaeon (D-plasma)	1470	1213	*Euryarchaeota*	*Thermoplasmata*	Uncultured archaeon clone 1000m_arch_e9	HM745447	99%
Uncultured *Thermoplasmatales* archaeon (E-plasma)	1468	447	*Euryarchaeota*	*Thermoplasmata*	Uncultured *Thermoplasmatales* archaeon E-plasma	JX997947	100%
Uncultured *Thermoplasmatales* archaeon (B-plasma)	1470	154	*Euryarchaeota*	*Thermoplasmata*	Uncultured archaeon clone ant g10	DQ303253	99%
Uncultured *Thermoplasmatales* archaeon (I-plasma)	1474	108	*Euryarchaeota*	*Thermoplasmata*	Uncultured *Thermoplasmatales* archaeon I-plasma	JX997945	100%
Uncultured *Ferroplasma* sp.	1470	807	*Euryarchaeota*	*Thermoplasmata*	*Ferroplasma acidarmanus* fer1	NR_103941	99%
Uncultured *Micrarchaeum* sp.	2003	1151	*Parvarchaeota*	/	ARMAN-1	AY652726	99%

Tabla 8. Secuencias de genes ribosomales de eucariotas identificados en el metagenoma de sedimentos termales de Los Azufres colectados en 2013.

Eucariota	Longitud (pb)	Número de lecturas	Filo	Clase	GenBank hit	Número de acceso	Identidad
Uncultured *Cyanidiaceae* sp. (plastid)	1423	4381	*Rhodophyta*	*Bangiophyceae*	Uncultured *Cyanidiales* clone R0435B54	KJ569775	100%
Uncultured *Cyanidiaceae* sp. (plastid)	1480	493	*Rhodophyta*	*Bangiophyceae*	*Galdieria sulphuraria* 14-1-1	X52985	99%
Uncultured *Cyanidiaceae* sp. (plastid)	1480	473	*Rhodophyta*	*Bangiophyceae*	*Galdieria sulphuraria* 14-1-1	X52985	99%
Uncultured *Cyanidiaceae* sp. (mitochondrion)	1567	3102	*Rhodophyta*	*Bangiophyceae*	Uncultured *Cyanidiales* clone R0435B41	GQ141766	99%
Uncultured *Cyanidiaceae* sp. (nuclear)	1794	898	*Rhodophyta*	*Bangiophyceae*	*Cyanidium caldarium* 55B	AB091232	99%
Uncultured *Cyanidiaceae* sp. (nuclear)	1836	107	*Rhodophyta*	*Bangiophyceae*	*Galdieria sulphuraria* SAG107.79	AB091229	98%
Uncultured *Cyanidiaceae* sp. (nuclear)	1842	144	*Rhodophyta*	*Bangiophyceae*	*Galdieria sulphuraria* SAG21.92	AB091230	89%
Uncultured *Bodonidae* sp. (nuclear)	2137	58	*Euglenozoa*	*Kinetoplastida order*	Uncultured eukaryote clone L13.10	AY753963	93%
Uncultured *Acidomyces* sp. (nuclear)	1799	365	*Ascomycota*	*Dothideomycetes*	*A.acidothermus* NBRC 106060	AB537895	100%

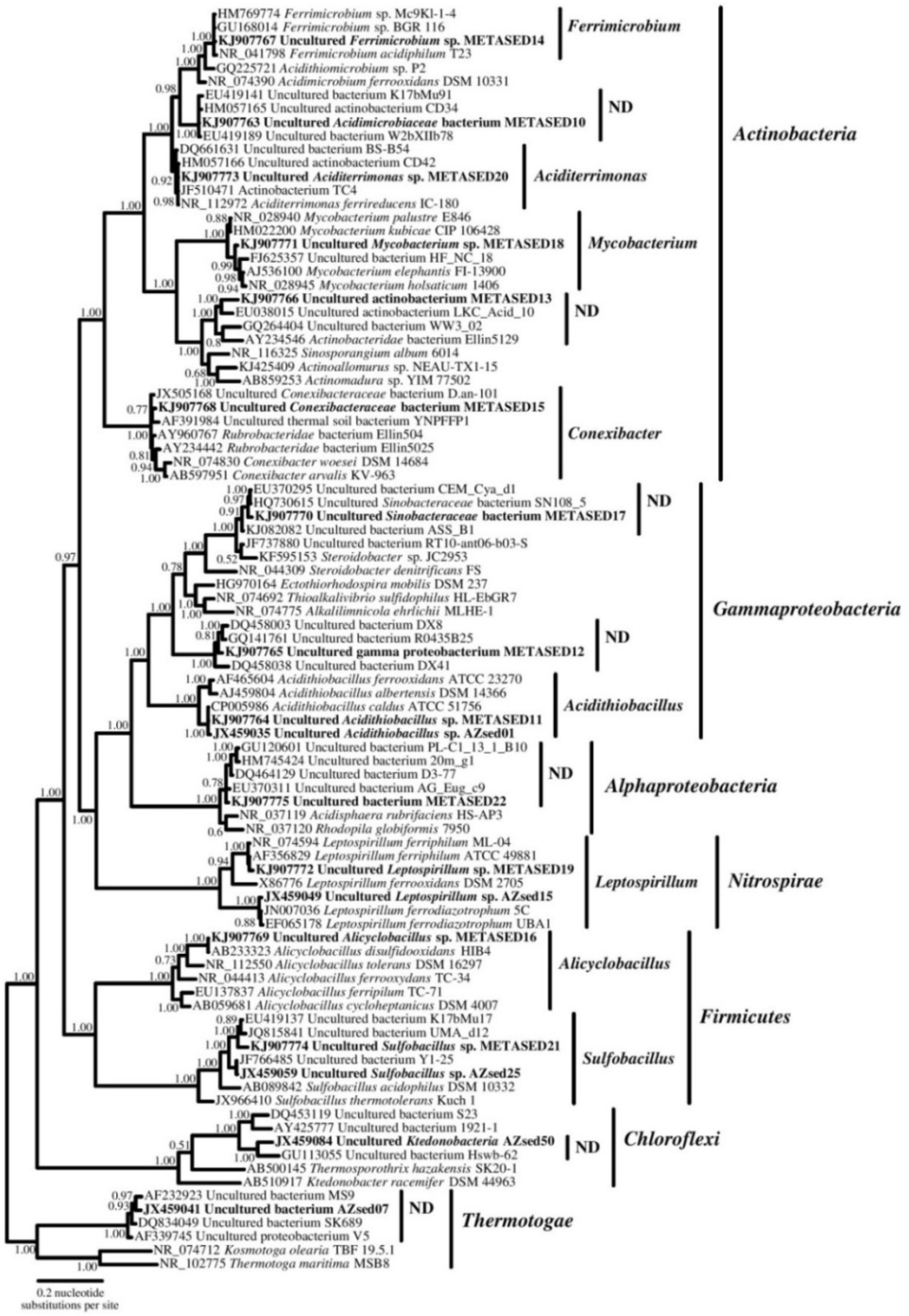


Figura 20. Análisis filogenético basado en secuencias de genes ribosomales 16S rRNA de bacterias. Las secuencias obtenidas de muestras de sedimentos de Los Azufres se muestran en negritas. Los identificadores METASED y AZsed indican a las secuencias metagenómicas y a las obtenidas de la librería de clonas, respectivamente. Los grupos sin géneros descritos se identifican con ND (no determinados). El alineamiento de secuencias contiene 1194 caracteres. La reconstrucción filogenética se realizó en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. La barra de escala representa el número promedio de sustitución de nucleótidos por sitio.

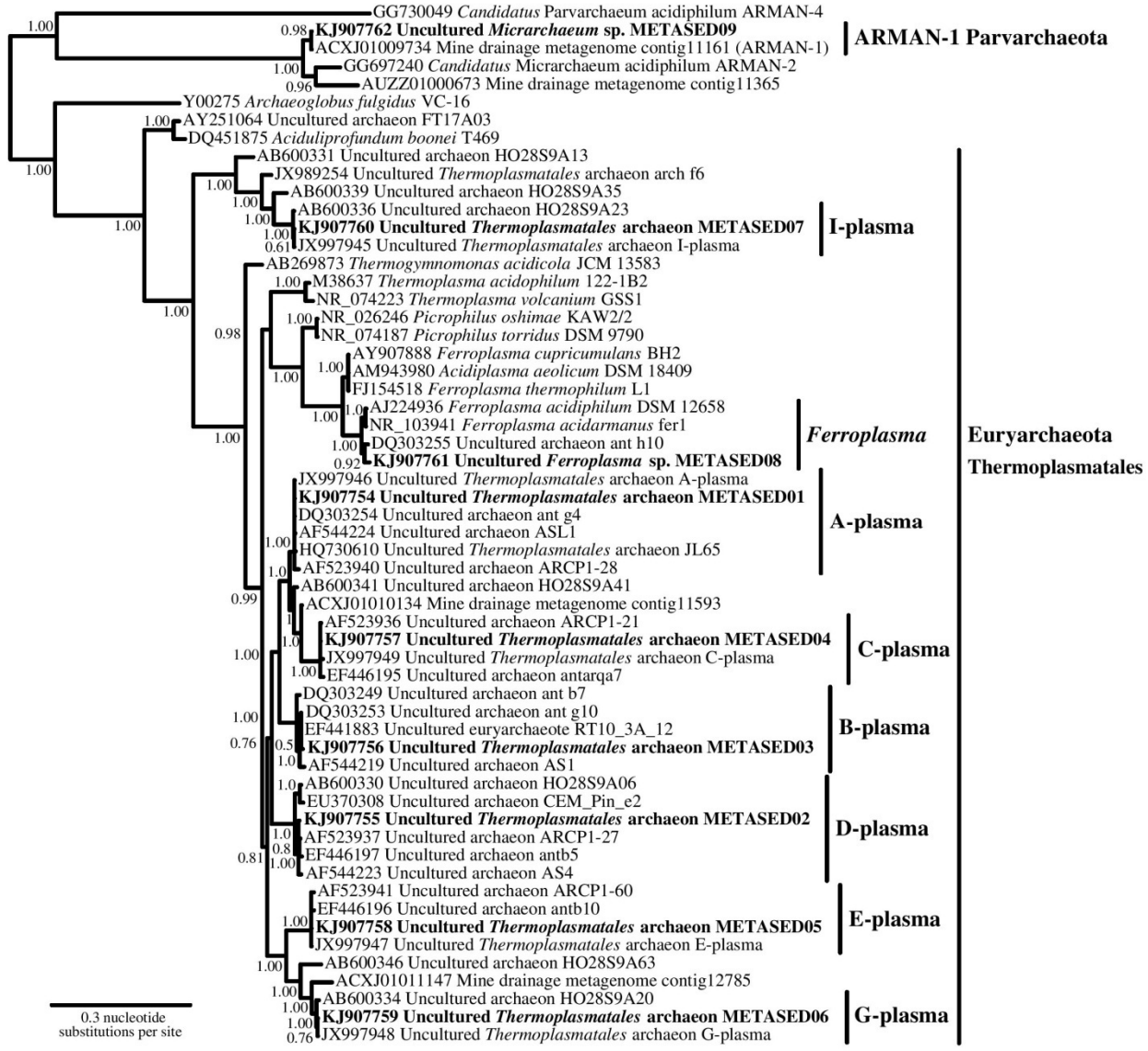


Figura 21. Análisis filogenético basado en secuencias de genes ribosomales 16S rRNA de arqueas. Las secuencias recuperadas del metagenoma del sedimento fotosintético de Los Azufres se muestran en negritas. El alineamiento de secuencias contiene 1385 caracteres. La reconstrucción filogenética se realizó en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Se muestran los identificadores de GenBank de las secuencias utilizadas. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. La barra de escala representa el número promedio de sustitución de nucleótidos por sitio.

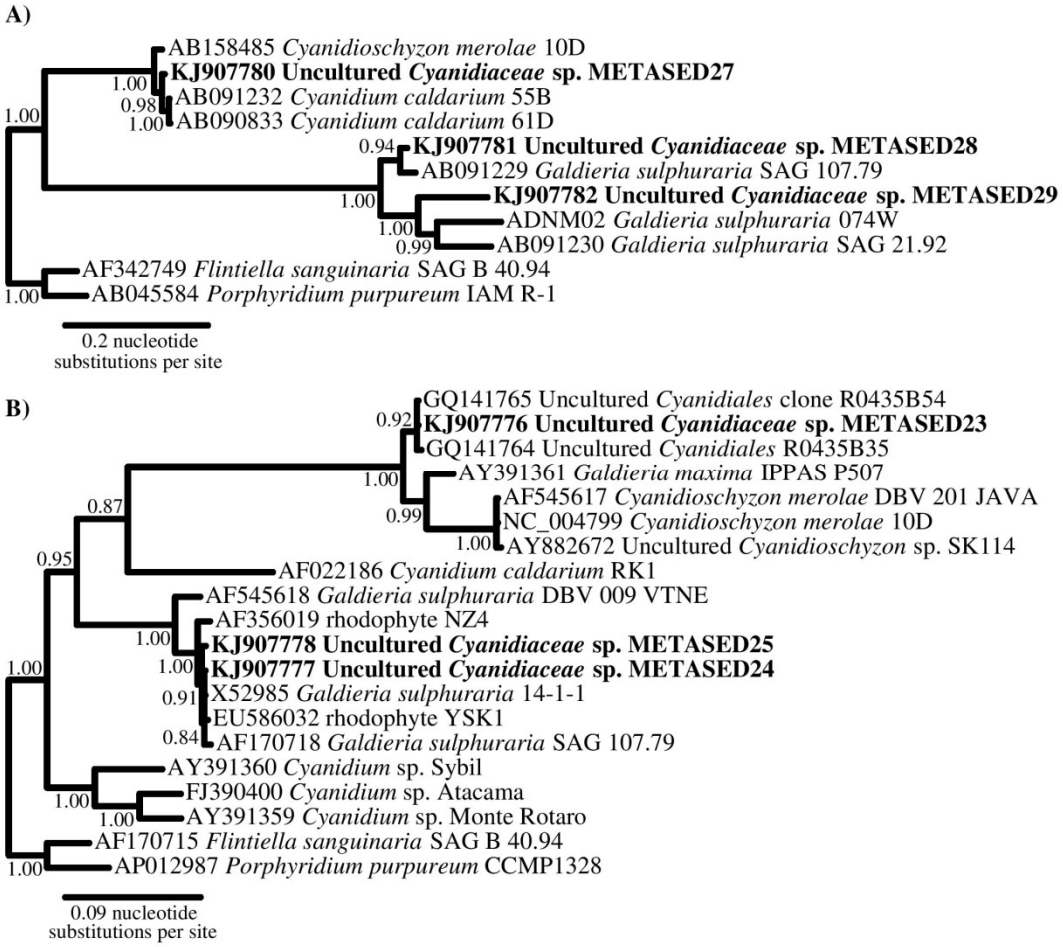


Figura 22. Análisis filogenéticos basados en secuencias de genes ribosomales de microalgas perteneciente a la familia *Cyanidiaceae*. A) Análisis filogenético basado en genes ribosomales 16S rRNA de cloroplastos. El alineamiento de secuencias contiene 1250 caracteres. B) Análisis filogenético basado en genes ribosomales 18S rRNA de microalgas. El alineamiento de secuencias contiene 1362 caracteres. Las secuencias recuperadas del metagenoma del sedimento fotosintético de Los Azufres se muestran en negritas. Las reconstrucciones filogenéticas se realizaron en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Se muestran los identificadores de GenBank de las secuencias utilizadas. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. Las barras de escala representa el número promedio de sustitución de nucleótidos por sitio.

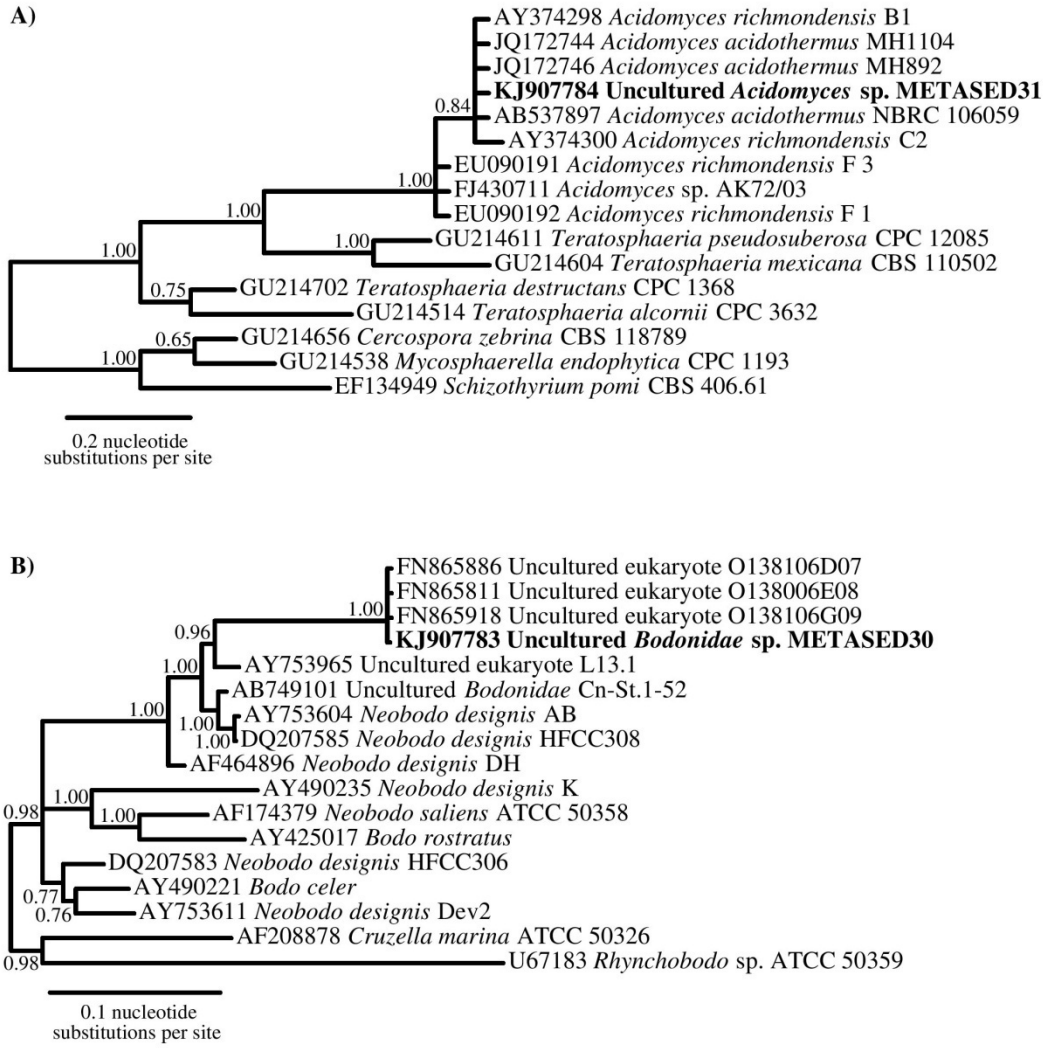


Figura 23. Análisis filogenéticos basados en secuencias de genes ribosomales de microeucariotas. A) Análisis filogenético basado en genes ribosomales 18S rRNA de hongos del género *Acidomyces*. El alineamiento de secuencias contiene 1592 caracteres. B) Análisis filogenético basado en genes ribosomales 18S rRNA relacionados al género *Neobodo*. El alineamiento de secuencias contiene 1408 caracteres. Las secuencias identificadas en el metagenoma del sedimento fotosintético de Los Azufres se muestran en negritas. Las reconstrucciones filogenéticas se realizaron en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Se muestran los identificadores de GenBank de las secuencias utilizadas. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. Las barras de escala representa el número promedio de sustitución de nucleótidos por sitio.

Genomas completos de los organelos de una nueva microalga

Las secuencias metagenómicas lograron revelar a los microorganismos fotosintéticos de los sedimentos termales de Los Azufres. Se encontraron secuencias de microalgas de la familia *Cyanidiaceae* que corresponden a genes ribosomales 18S rRNA y también 16S rRNA de los genomas de sus cloroplastos. Las secuencias obtenidas a partir de la librería de clonas no lograron identificar a las microalgas probablemente debido al sesgos inherentes a las técnicas basadas en reacciones de PCR.

Una de las microalgas es preponderante en la comunidad y es probable que su genoma se encuentre casi completamente representado en el metagenoma, se ha nombrado temporalmente a esta microalga como *Cyanidiaceae* sp. MX-AZ01. La microalga *Cyanidiaceae* sp. MX-AZ01 es cercana filogenéticamente a cepas de las especies descritas *Cyanidium caldarium* y *Cyanidioschyzon merolae* mientras que las poblaciones minoritarias de microalgas están relacionadas con la especie *Galdieria sulphuraria* (Figura 22).

Las microalgas de la familia *Cyanidiaceae* son capaces de habitar en ambientes termales extremos y los genomas de algunas cepas han sido secuenciados y publicados en revistas prestigiosas. Se ha encontrado que sus genomas contienen genes de bacterias y arqueas que se han incorporado al genoma nuclear mediante eventos de transferencia horizontal (Matsuzaki et al., 2004; Schönknecht et al., 2013; Schönknecht et al., 2014).

Las secuencias metagenómicas permitieron reconstruir los genomas completos de los genomas del cloroplasto y la mitocondria de la microalga *Cyanidiaceae* sp. MX-AZ01. Se encontró que el genoma de la mitocondria se encontraba completo en un único *contig* y solo se requirió eliminar la redundancia de secuencias en los extremos del *contig* para hacer su circularización. También se identificaron cinco *contigs* que correspondían al genoma del cloroplasto y fue necesario ensamblar sus extremos sobrelapantes para hacer su circularización.

Actualmente se están analizando las características de los genomas de los organelos. La anotación se realizó utilizando GeneMark.hmm 2.0 (Besemer y Borodovsky, 2005). Los genes rRNA y tRNA se predijeron usando RNAmmer 1.2 (Lagesen et al., 2007) y tRNAscan-SE 1.21 (Lowe et al., 1997). Las predicciones génicas se anotaron y se revisaron manualmente usando Artemis (Rutherford et al., 2000). Se utilizó el sistema de OrganellarGenomeDRAW (Lohse et al., 2013) para generar las representaciones de los genomas (Figura 24). Los genomas de los organelos han sido depositados en GenBank con los números de acceso KJ569775 y KJ569774.

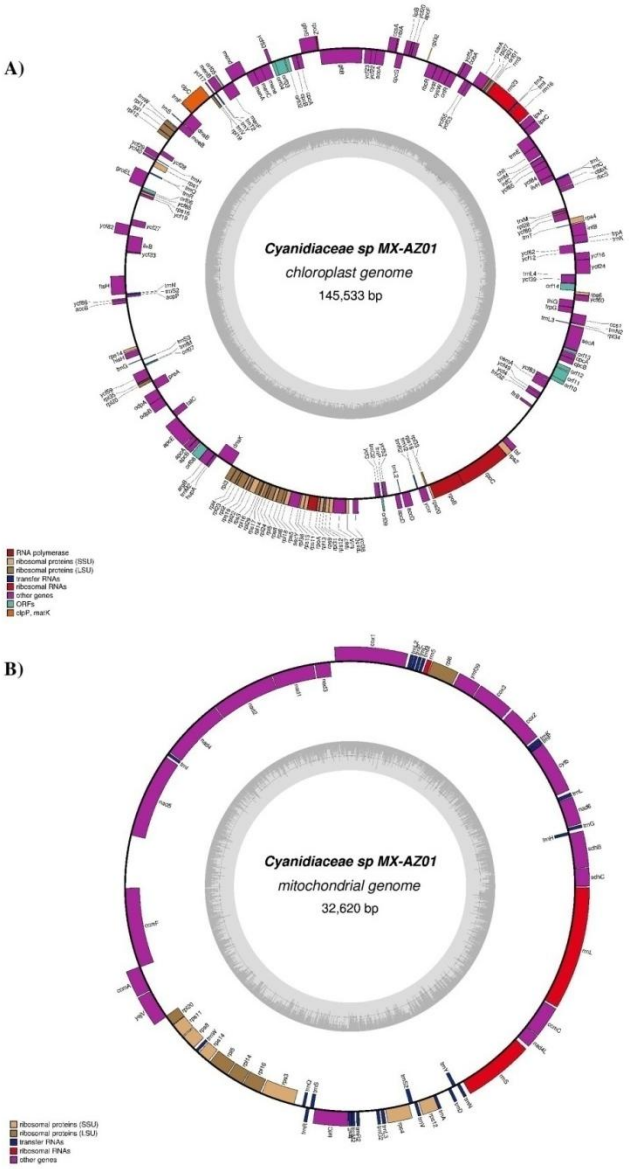


Figura 24. Representación de los genomas completos de los organelos de la microalga *Cyanidiaceae* sp. MX-AZ01 recuperados del metagenoma del sedimento fotosintético de Los Azufres. Los colores de los genes predichos corresponden a distintas categorías funcionales.

Genoma parcial de una nueva arquea del filo Parvarchaeota

Las arqueas de los linajes ARMAN (*Archaeal Richmond Mine Acidophilic Nanoorganisms*) fueron identificadas por primera vez en los biofilmes que flotan en arroyos de una mina de California conocida como *Iron Mountain* y reportadas en la revista Science (Baker et al., 2006). Las arqueas ARMAN se detectaron al analizar secuencias de genes ribosomales contenidas en los metagenomas de biofilmes. Estudios de diversidad previos no habían podido amplificar sus genes ribosomales 16S rRNA mediante reacciones de PCR debido a que sus secuencias son muy divergentes. Se ha encontrado que las arqueas ARMAN en conjunto representan poblaciones minoritarias dentro de las comunidades de los biofilmes (5-25%). Las arqueas ARMAN no se han podido cultivar. Mediante imágenes de microscopía electrónica se ha encontrado que sus células son las más pequeñas conocidas para formas de vida libre ya que solo miden cerca de ~300 nm de diámetro y además se ha visto que establecen interacciones desconocidas con arqueas del orden *Thermoplasmatales* y con virus que también son enigmáticos (Baker et al, 2006; Baker et al., 2010; Comolli y Banfield, 2014).

Se han podido reconstruir genomas parciales para los linajes ARMAN-2 (*Candidatus* Micrarchaeum acidiphilum), ARMAN-4 (*Candidatus* Parvarchaeum acidiphilum) y ARMAN-5 (*Candidatus* Parvarchaeum acidophilus) a partir de los metagenomas de los biofilmes de la mina de California (Baker et al., 2010) y se ha propuesto que corresponden a un filo nuevo nombrado *Parvarchaeota* (Rinke et al., 2013). Recientemente se obtuvieron genomas parciales adicionales para los linajes ARMAN-2, ARMAN-4 y ARMAN-5 a partir del metagenoma del drenaje ácido de una mina del sur de China (Hua et al., 2015). A la fecha no se han podido obtener genomas del linaje ARMAN-1 debido a que son miembros aún menos abundantes en las comunidades.

En el metagenoma del sedimento termal de Los Azufres identificamos una secuencia de un gen ribosomal 16S rRNA de una arquea correspondiente al linaje ARMAN-1. No se pudieron encontrar secuencias de genes ribosomales de otros linajes de arqueas ARMAN. Mediante búsquedas de BLASTP encontramos *contigs* que corresponderían a la arquea del linaje ARMAN-1 usando las secuencias de proteínas conocidas del genoma del linaje ARMAN-2 (el linaje filogenéticamente más cercano al linaje ARMAN-1).

Posteriormente hicimos un análisis de agrupación de los *contigs* en categorías taxonómicas (*binning*) usando el programa MaxBin que considera la cobertura de lecturas, la frecuencia de tetranucleótidos y la presencia de genes marcadores (Wu et al., 2014). Del análisis de agrupamiento, encontramos un único *bin* que contenía los mismos *contigs* que el análisis de búsquedas por BLAST además de *contigs* adicionales que no habían podido ser detectados mediante identidad de secuencia.

El ensamble actual del genoma representativo del linaje ARMAN-1 de Los Azufres presenta una longitud de 980,507 pares de bases, una alta cobertura de secuencia de 146X y está contenido en cinco *contigs*. El genoma está depositado en GenBank pero no se ha hecho público. Actualmente se está analizando el potencial genético de nuestro ensamble del linaje ARMAN-1 y las diferencias existentes con los genomas de las arqueas ARMAN. Un alineamiento global de los genomas de los linajes ARMAN-2 y ARMAN-1 muestra que los genomas presentan una gran cantidad de rearreglos genómicos pero tienen pequeños bloques conservados (Figura 25). Los genomas podrían presentar tantos rearreglos debido a la actividad de transposasas. En base a la anotación actual, solo se han identificado dos transposasas de la familia IS605 OrfB que se han identificado también en el linaje ARMAN-4 y en arqueas de los órdenes *Methanosarcinales* y *Thermoplasmatales*. Las secuencias de arqueas ARMAN solo se han podido identificar en drenajes ácidos de minas. Nuestro estudio las ha podido identificar en sedimentos fotosintéticos en sitios termales. Al realizar análisis filogenéticos nos hemos percatado que existen secuencias obtenidas de otros sitios termales que corresponden al linaje ARMAN-1 (Figura 26).

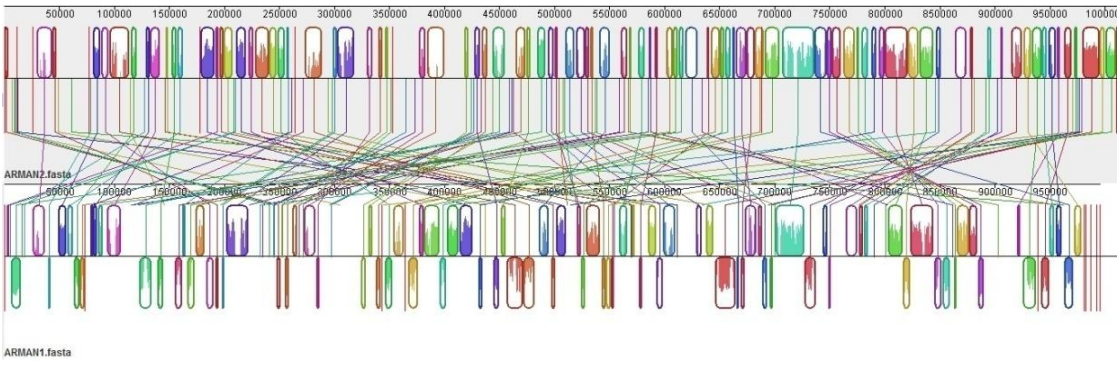


Figura 25. Alineamiento global de genomas de arqueas del filo *Parvarchaeota* de los linajes ARMAN-1 y ARMAN-2. A) genoma de ARMAN-2 recuperado de metagenomas de biofilmes de la Mina de California. B) genoma recuperado del metagenoma del sedimento fotosintético de Los Azufres y representante del linaje ARMAN-1.

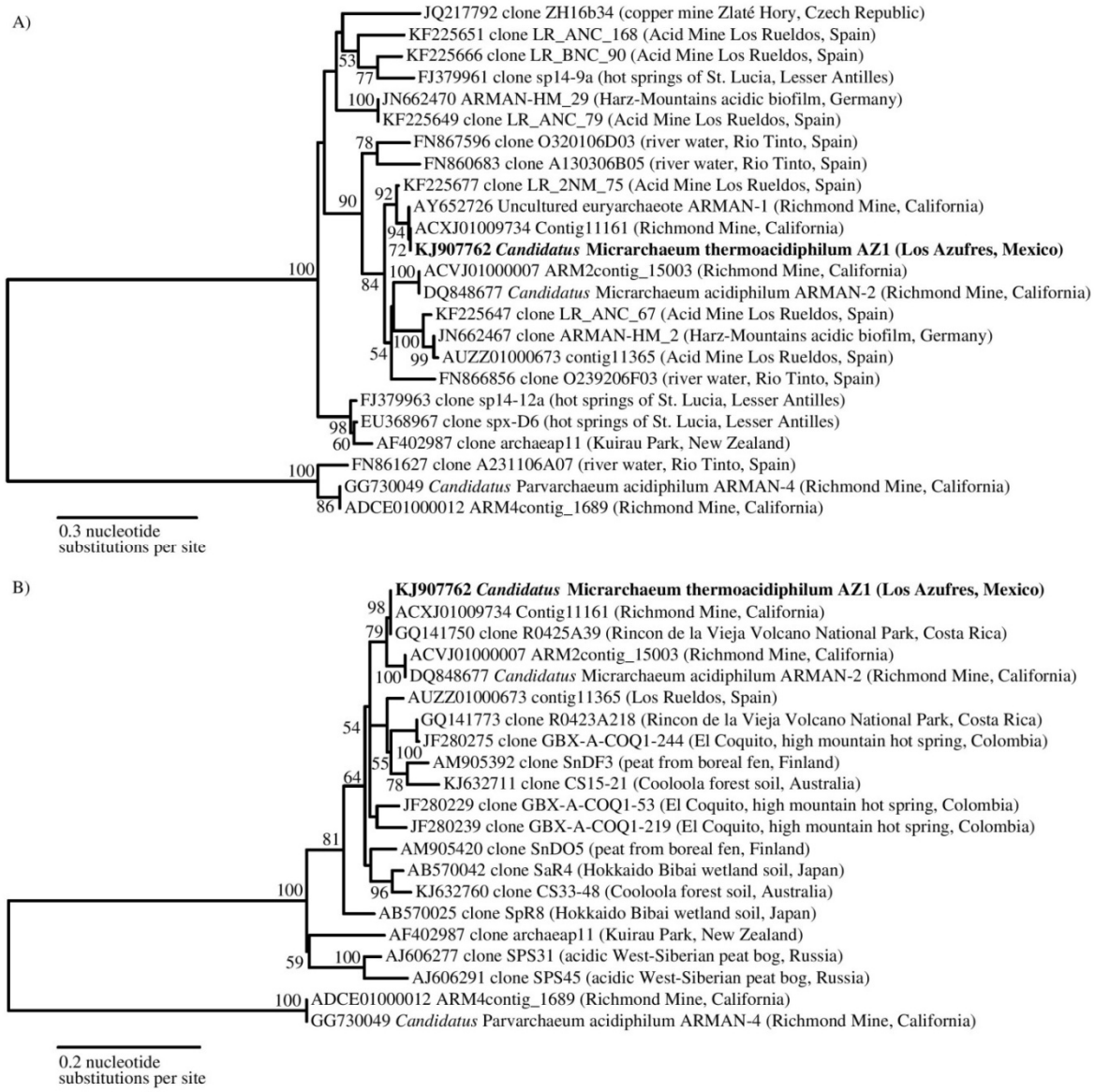


Figura 26. Análisis filogenéticos basados en secuencias de genes ribosomales de arqueas del filo *Parvarchaeota* relacionadas con el linaje ARMAN-1. Las figuras A) y B) muestran análisis filogenéticos basados en distintos fragmentos de la secuencia del gen ribosomal 16S rRNA. Los alineamientos de secuencias contienen 449 y 403 caracteres respectivamente para las figuras 26A y 26B. Las secuencias recuperadas del metagenoma del sedimento de Los Azufres se muestran en negritas. Las reconstrucciones filogenéticas se realizaron en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. Las barras de escala representan el número promedio de sustitución de nucleótidos por sitio.

Cultivo de microorganismos de sedimentos fotosintéticos de Los Azufres

Se lograron cultivar algunos microorganismos de la misma muestra de sedimentos termales que se utilizó para la purificación de ADN ambiental (Figura 27). Se pudieron obtener colonias de microalgas y hongos a temperaturas no extremas (25 °C) por lo que estos microorganismos pueden ser cultivados en rangos grandes de temperaturas. También crecieron colonias que presentan halos donde no crecen las microalgas ni los hongos, lo que provee evidencia de que existen interacciones microbianas antagonistas dentro de la comunidad. El aislamiento y caracterización molecular quedan pendientes de realizarse.

Figura 27. Caja de cultivo de microorganismos de sedimentos termales de Los Azufres. En verde se pueden apreciar colonias de microalgas. Las colonias en color café oscuro corresponden a un hongo que probablemente corresponde al género *Acidomyces*. Las flechas indican colonias que forman halos pequeños en donde no crecen microalgas. El medio de cultivo utilizado fue el DSM 35a (http://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium35a.pdf).

En conjunto, estos resultados nos indican que las comunidades microbianas de los sedimentos termales analizados de Los Azufres presentan diversidad limitada y contienen comunidades integradas por microalgas de la familia *Cyanidiaceae* y por algunas poblaciones de bacterias y arqueas de grupos filogenéticos novedosos.

Diversidad de comunidades microbianas de una laguna ácida de Los Azufres

Los lagos ácidos ocurren naturalmente en sitios volcánicos y en sitios con alta concentración de hierro y compuestos azufrados. Los lagos ácidos también pueden crearse como resultado de la explotación minera. Los pH ácidos característicos de este tipo de lagos se deben a la disolución y oxidación de minerales ricos en azufre y hierro en columnas de agua.

Las comunidades microbianas de los lagos ácidos están compuestas principalmente por microorganismos autótrofos tales como microalgas y por bacterias relacionadas con α-protebacterias, β-proteobacterias y γ-proteobacterias así como acidobacterias y actinobacterias. Las arqueas son en general poco abundantes en estas comunidades (González-Toril et al., 2003; Falagán et al., 2014; Stankovic et al., 2014). Actualmente no se han reportado estudios metagenómicos de este tipo de lagos ácidos que permitan evaluar la diversidad de sus comunidades con mayor resolución.

El campo geotérmico de Los Azufres contiene una laguna de origen natural (Figura 28) que presenta un pH estable de 2.3 y temperaturas de entre 12 y 18 °C. Se colectaron muestras de agua de la superficie para purificar ADN ambiental en los años 2008, 2009 y 2010 con el fin de realizar censos de diversidad en base a análisis de genes ribosomales 16S rRNA. El ADN purificado durante de 2010 también se utilizó para la secuenciación de un metagenoma.

Figura 28. Laguna ácida de Los Azufres de donde se colectaron muestras de agua para realizar análisis de diversidad microbiana.

Se obtuvieron 63 secuencias de genes ribosomales 16S rRNA de bacterias de las librerías de clonas generadas a partir del ADN purificado de la laguna ácida. Las comunidades bacterianas de laguna ácida son de complejidad limitada. La curva de rarefacción de las secuencias de genes ribosomales 16S rRNA de las bacterias de la laguna ácida no supera los 20 OTUs cuando ha alcanzado la estabilidad (Figura 29). Dadas las dimensiones de la laguna (largo=500 metros, ancho=200 metros) es notable identificar tan pocos ribotipos distintos.

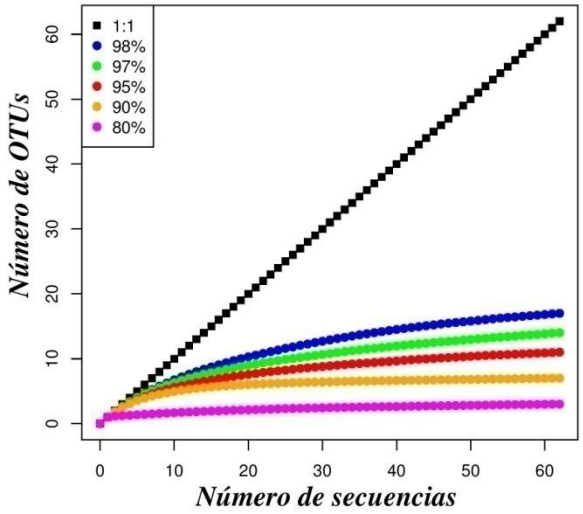


Figura 29. Curva de rarefacción de las secuencias de los genes ribosomales 16S rRNA del censo de bacterias de la laguna ácida de Los Azufres. La curva de rarefacción incluye 63 secuencias redundantes (longitud= ~1500 nucleótidos) de 500 clonas analizadas mediante patrones de restricción en cinco muestreos. Los datos se obtuvieron utilizando mothur (Oakley et al., 2009) y la gráfica se reconstruyó utilizando R. Se muestran las curvas correspondientes a distintos puntos de corte de acuerdo a la identidad de secuencia.

Las bacterias más abundantes de la laguna ácida pertenecen a los géneros *Acidiphilium* (α-protebacteria), *Acidithiobacillus* (γ-proteobacteria), *Acidobacterium (Acidobacteria)*, *Thiomonas* (β-proteobacteria) y también se identificaron bacterias del orden *Xanthomonadales* (γ-proteobacterias) y del filo *Thermotogae* (Tabla 9). Todas las bacterias identificadas están relacionadas filogenéticamente con grupos bacterianos que son comunes en sitios termales y ácidos alrededor del mundo (Figura 30).

Tabla 9. Filotipos identificados en muestras de agua de la laguna ácida de Los Azufres.

Orden	Familia	Género	% de identidad	número de acceso GenBank
Rhodospirillales	*Acetobacteraceae*	*Acidiphilium*	99%	FJ915145
Burkholderiales	/	*Thiomonas*	99%	CP002021
Acidithiobacillales	*Acidithiobacillaceae*	*Acidithiobacillus*	99%	NR_028982
Acidithiobacillales	*Acidithiobacillaceae*	*Acidithiobacillus*	96%	AF362022
Xanthomonadales	*Xanthomonadaceae*	/	93%	AY987369
Nitrospirales	*Nitrospiraceae*	*Leptospirillum*	99%	AF356839
Acidobacteriales	*Acidobacteriaceae*	*Acidobacterium*	97%	CP001472
Acidobacteriales	*Acidobacteriaceae*	*Acidobacterium*	95%	CP001472
Thermotogales	*Thermotogaceae*	/	81%	CP001634
Trebouxiophyceae	*Coccomyxaceae*	/	90% (16S rRNA)	AM292034
			97% (18S rRNA)	FJ946891

Las comunidades de eucariotas de la laguna ácida están dominadas por un grupo filogenético novedoso. Todas las secuencias de las librerías de clonas de genes ribosomales 18S rRNA son idénticas. La secuencia del gen ribosomal 18S rRNA presenta una identidad del 97% en 1,706 nucleótidos con genes de algas de la clase *Trebouxiophyceae* (reino *Viridiplantae,* filo *Chlorophyta*). Las librerías de clonas de la laguna ácida también contienen una secuencia que presentó una identidad del 90% en 1445 nucleótidos con genes ribosomales 16S rRNA de cloroplastos de algas trebouxiofitas (Tabla 9).

Un análisis filogenético indica que la secuencia del gen ribosomal 16S rRNA del cloroplasto forma un grupo independiente (Figura 31). La presencia de un único eucariota dominante es intrigante ya que los análisis de otras comunidades de ambientes ácidos revelaron una diversidad muy amplia de eucariotas (Amaral et al., 2002; Aguilera et al.,2006).

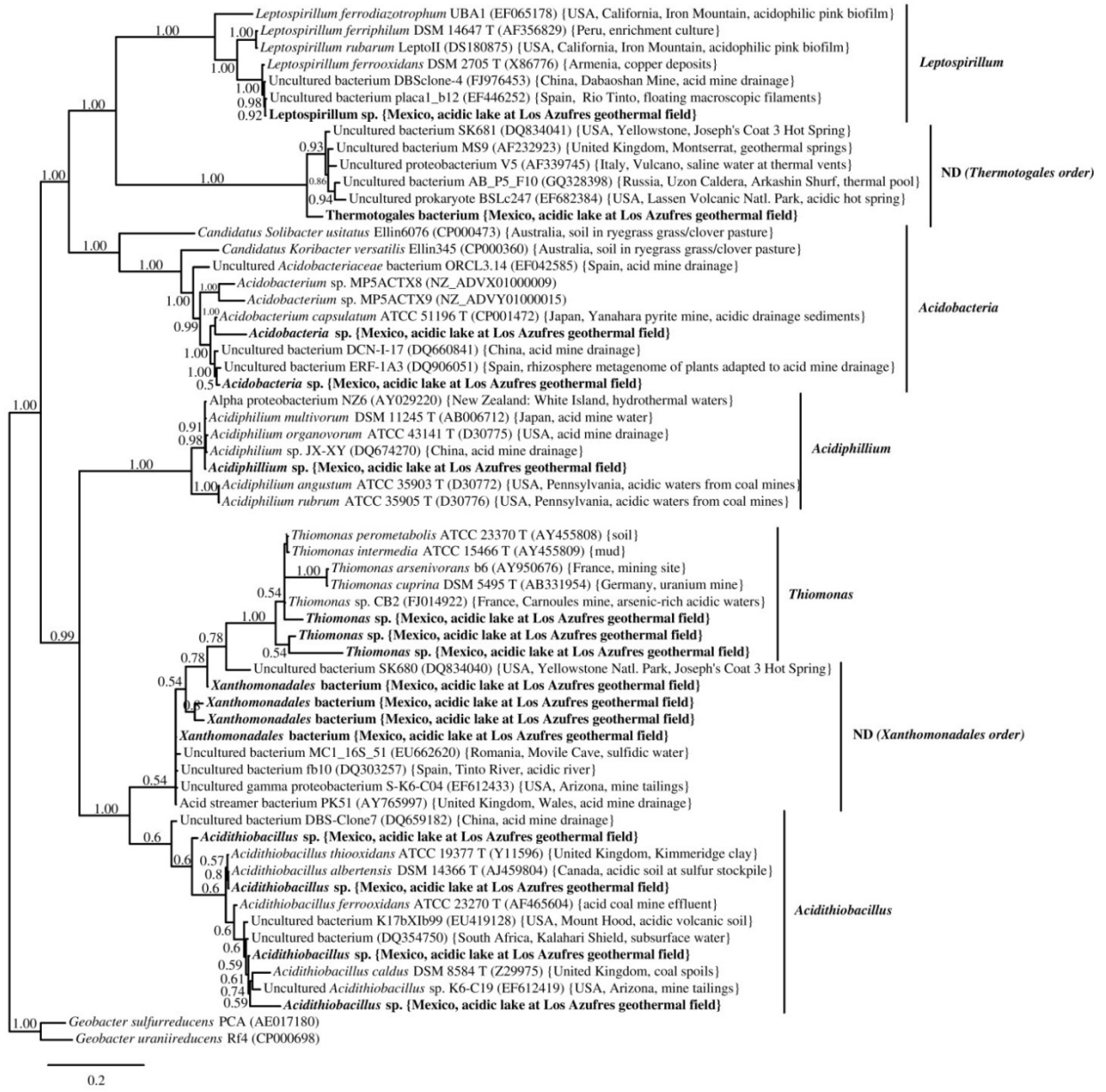


Figura 30. Análisis filogenético basado en secuencias de genes ribosomales 16S rRNA de bacterias. Las secuencias de las librerías de clonas de la laguna ácida de Los Azufres se muestran en negritas. Los grupos sin géneros descritos se identifican con ND (no determinados). El alineamiento de secuencias contiene 1381 caracteres. La reconstrucción filogenética bayesiana se realizó en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Se muestran los números de acceso de GenBank de las secuencias utilizadas. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. La barra de escala representa el número promedio de sustitución de nucleótidos por sitio.

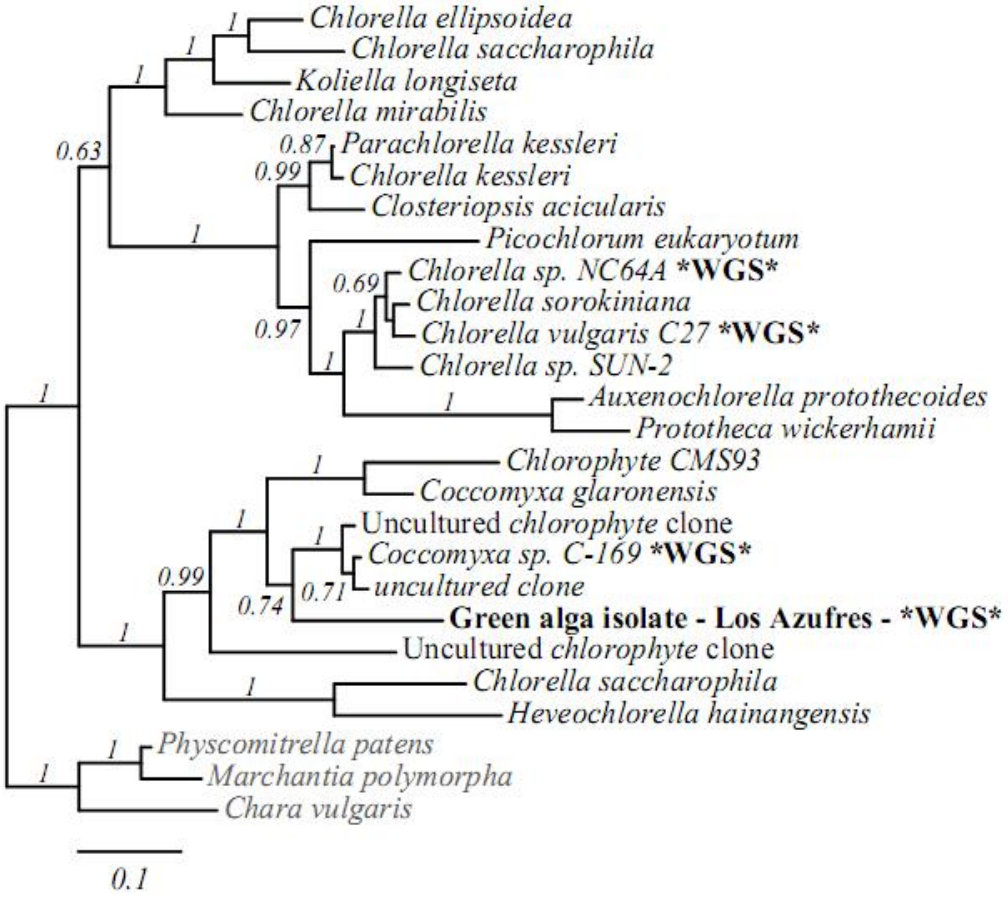


Figura 31. Análisis filogenético basado en secuencias de genes ribosomales 16S rRNA de cloroplastos de microalgas de la clase *Trebouxiophyceae*. Se muestra en negritas una secuencia representativa de un cloroplasto de la microalga de la laguna ácida de Los Azufres obtenida de la librería de clonas. Las microalgas con secuencias genómicas disponibles se identifican con las siglas WGS (*Whole Genome Shotgun*). El alineamiento de secuencias contiene 1614 caracteres. La reconstrucción filogenética bayesiana se realizó en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. La barra de escala representa el número promedio de sustitución de nucleótidos por sitio.

Las comunidades microbianas de la laguna ácida son estables a través del tiempo en base a los censos de diversidad de secuencias obtenidas por librerías de clonas de genes ribosomales. Los análisis de diversidad de cinco muestras de agua colectadas en fechas distintas indican que las comunidades microbianas de la laguna ácida modifican sus estructuras pero conservan los mismos ribotipos (Figura 32). Sin embargo, queda pendiente realizar análisis metagenómicos más finos en distintas épocas del año para corroborar la estabilidad de las comunidades microbianas de la laguna ácida de Los Azufres. Las secuencias metagenómicas adicionales ayudarían a eliminar posibles sesgos en la amplificación de genes ribosomales.

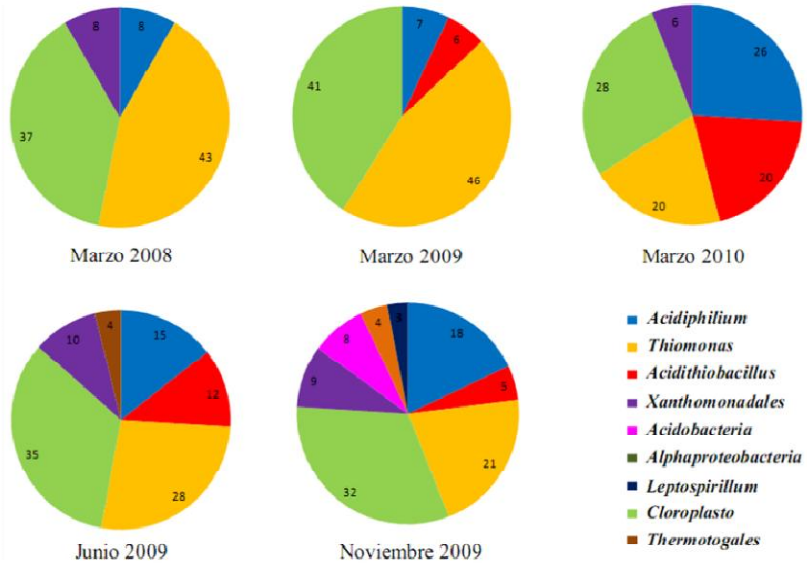


Figura 32. Proporción de ribotipos en los análisis de diversidad de la laguna ácida de Los Azufres. Los análisis están conformados en base a 100 clonas ambientales en cada mes.

Las diferencias en las estructuras de las comunidades quizás se deben a las variaciones ambientales. Por ejemplo, la comunidad con más ribotipos (noviembre 2009) corresponde al muestreo posterior a la época de lluvias. Durante la época de lluvias seguramente ingresa una mayor cantidad de nutrientes del suelo circundante a la laguna. La presencia extra de nutrientes o un cambio en el pH del agua podrían favorecer la proliferación de grupos filogenéticos minoritarios. Tal vez, grupos que se encuentran presentes en los sedimentos prosperan temporalmente antes de que las condiciones se estabilicen en la época posterior a las lluvias cuando quizá ocurra una selección de microorganismos.

Todas las muestras analizadas son de un solo sitio para reducir las variables. Sin embargo, debido a que el sitio de muestro se encontraba seco durante la época de sequía el análisis de diversidad de julio de 2009 se realizó a 4 metros de una manifestación termal que se encuentra en lo que normalmente sería el fondo de la laguna ácida. Solo en el análisis de diversidad de julio de 2009 es posible identificar microorganismos del filo *Thermotoga* (que solo incluye microorganismos de manifestaciones hidrotermales). Durante la época de sequía el agua de la laguna tiene una mayor interacción con los fluidos y con los sedimentos que emanan de las manifestaciones geotérmicas del fondo, que suponemos afectan la estructura de la comunidad.

Los análisis de diversidad hechos a lo largo de tres años corresponden a la estación de inicio de primavera (marzo) que es cuando las condiciones químicas parecen más estables (sin influencia directa ni de las lluvias ni de la sequía). No se analizó la diversidad durante la época de lluvias, tampoco se analizó la diversidad de los sedimentos y solo se cuenta con el análisis químico de una muestra de agua. Análisis futuros que consideren estas limitantes podrían identificar mejor las variables que afectan la estructura de la comunidad.

En general los grupos filogenéticos son los mismos en todas las muestras a pesar de las variaciones ambientales. Con el tiempo, la acidez constante del agua de la laguna ácida parece ser el factor principal que selecciona y mantiene a los microorganismos de la comunidad microbiana.

Análisis de secuencias metagenómicas de una laguna ácida

Los laguna ácida se eligió para purificar ADN metagenómico y hacer su secuenciación debido a que presenta comunidades estables y de complejidad limitada. La laguna ácida presenta condiciones extremas de pH y metales pesados. Además las comunidades contienen grupos bacterianos que no se habían identificado en el Eje Volcánico de México. Finalmente fue importante realizar la secuenciación de un metagenoma para confirmar la diversidad limitada de eucariotas y confirmar si las microalgas son los microorganismos dominantes.

Se realizó la secuenciación del metagenoma a partir de ADN purificado de muestras de agua colectadas durante el año 2010 de la superficie de la laguna ácida. El metagenoma se secuenció en un equipo 454 GS-FLX Titanium en el Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO). El metagenoma está conformado por 278 megabases de secuencias. El ensamble preliminar se realizó usando Newbler v.2.3 (454 Life Sciences) y generó 31,762 *contigs* que contienen 49 megabases. La agrupación preliminar de los *contigs* en categorías taxonómicas (*binning*) se realizó utilizando el servidor MG-RAST (Glass et al., 2010).

El metagenoma de la laguna ácida primeramente se analizó para explorar su diversidad microbiana usando las secuencias de genes ribosomales. La identificación de genes ribosomales se realizó usando el servidor WebMGA (Wu et al., 2011). La identidad de las secuencias ribosomales obtenidas se verificó mediante búsquedas de BLASTN en GenBank.

Se identificaron 71 lecturas de pirosecuenciación que presentan identidad con genes ribosomales 16S rRNA de bacterias (Figura 33). La mayoría de las lecturas corresponden a *Acidiphilium* (16 lecturas), a *Thiomonas* (14 lecturas), a *Acidithiobacillus* (10 lecturas), a la *Gammaproteobacteria* del orden *Xanthomonadales* (8 lecturas) y a *Acidobacterium* (7 lecturas). Las restantes 16 lecturas corresponden a bacterias de los géneros *Acidocella* y *Leptospirillum* y a bacterias de grupos filogenéticos identificados en el Rio Tinto de España. Con respecto a eucariotas, la mayoría de las lecturas corresponden a los genes ribosomales de la microalga de la clase *Trebouxiophyceae* (Figura 31).

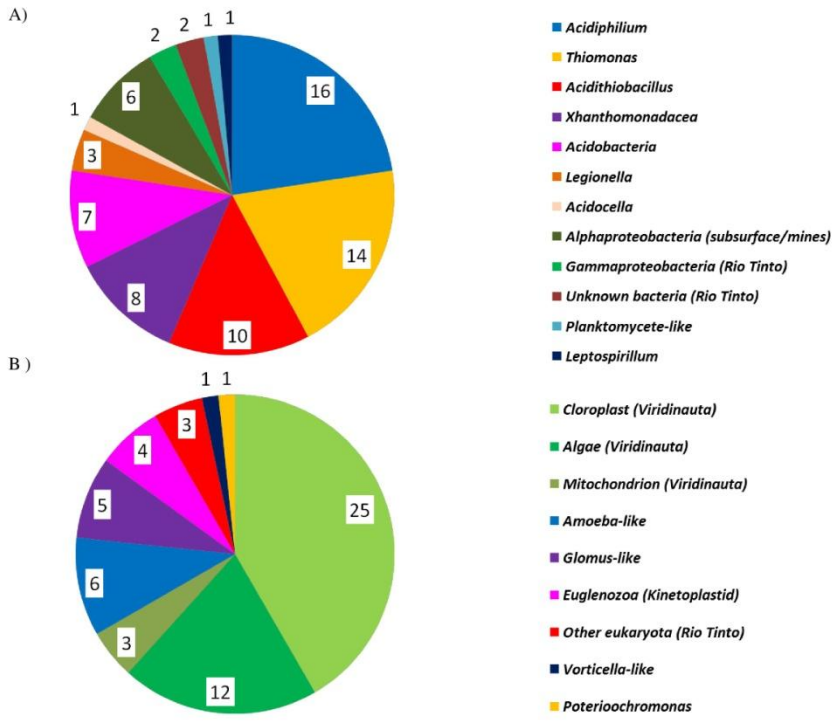


Figura 33. Número de lecturas del metagenoma de la laguna ácida que presentan identidad con genes ribosomales. A) Lecturas que presentan identidad con genes ribosomales 16S rRNA de bacterias. B) Lecturas que presentan identidad con genes ribosomales de eucariotas.

Se encontró que 39.59% de los *contigs* (12,576 *contigs*) corresponden a un grupo filogenético conocido. 60.41% de los *contigs* (19,186 *contigs*) permanecen sin asignación filogenética (Figura 34). Los *contigs* clasificados tienen identidad con secuencias de bacterias (8,961 *contigs*) y con secuencias de eucariotas (3,532 *contigs*).

Casi la totalidad de los *contigs* con identidad a secuencias de bacterias pertenecen a los filos *Acidobacteria* (5.28%) y *Proteobacteria* (82.55%). Las secuencias metagenómicas más abundantes de bacterias corresponden a α-protebacterias, β-proteobacterias, γ-proteobacterias y acidobacterias. 72.45% de los *contigs* tienen identidad con secuencias de microalgas y 22% de los *contigs* corresponden al grupo Fungi/Metazoa (Figura 34). Los grupos filogenéticos identificados en el metagenoma son congruentes con los identificados en los análisis de diversidad basados en productos de PCR de los genes ribosomales.

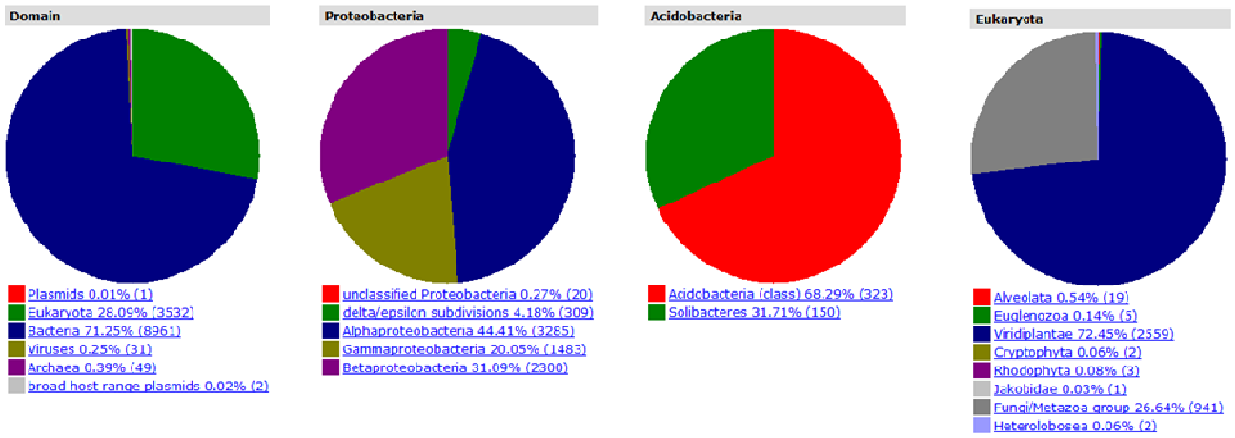


Figura 34. Composición filogenética del metagenoma de la laguna ácida. El número de *contigs* para cada grupo filogenético se muestra entre paréntesis. Los *contigs* están clasificados filogenéticamente en base a identidad de secuencia utilizando la base de datos SEED en el servidor MG-RAST (Glass et al., 2010). El valor de expectativa es de 1e−10.

Se logró obtener la secuencia completa de un plásmido que contiene genes de resistencia a metales pesados (Figura 35) y también se han identificaron genes relevantes en las secuencias metagenómicas que participan en la conversión de compuestos de arsénico y en la fijación de nitrógeno (Figura 36).

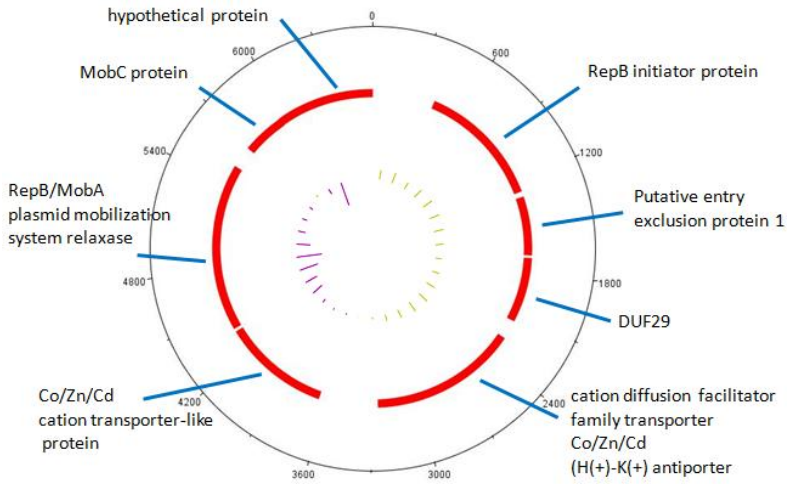
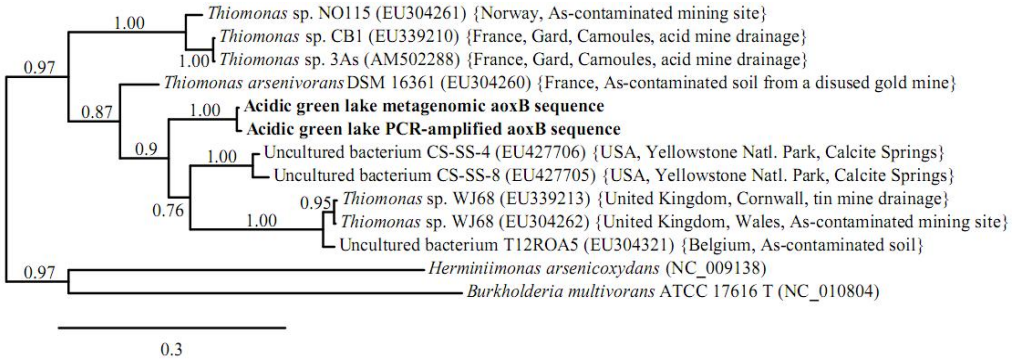
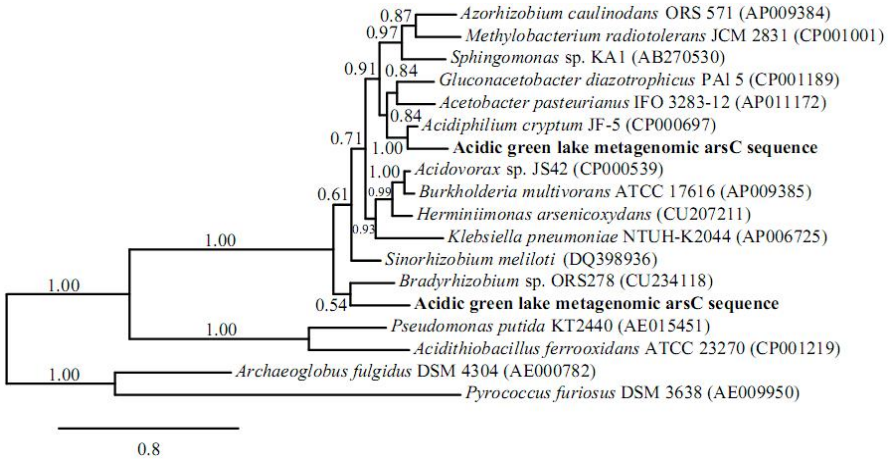


Figura 35. Secuencia completa de un plásmido obtenida del ensamble del metagenoma de la laguna ácida de Los Azufres. El plásmido contiene genes de replicación y genes que participan en la resistencia a metales pesados.

A) *aoxB* (conversión biológica de arsenatos a arsenitos)



B) *arsC* (conversión biológica de arsenitos a arsenatos)



C) *nifH* (fijación biológica de nitrógeno)

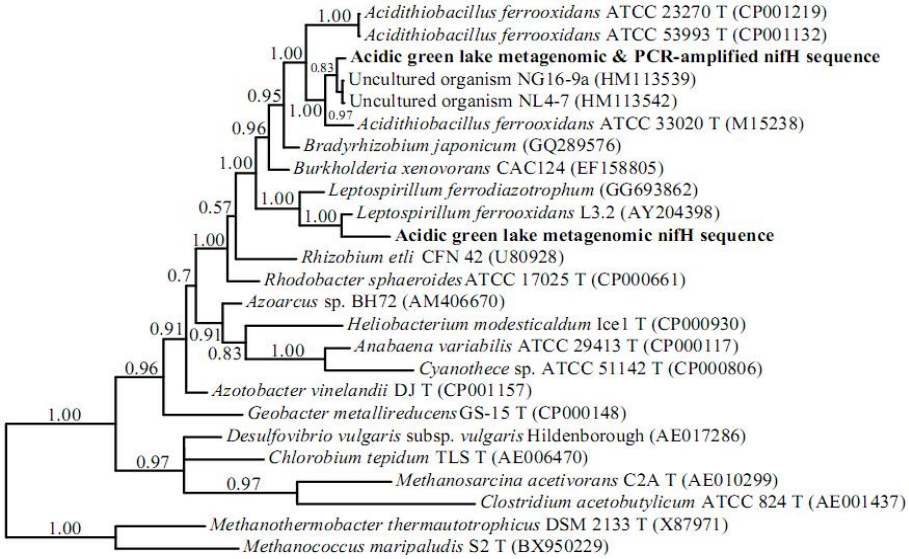


Figura 36. Análisis filogenéticos de genes identificados en secuencias metagenómicas de la laguna ácida participan en procesos de detoxificación biológica de compuestos de arsénico y en el proceso de fijación de nitrógeno.

Las secuencias metagenómicas revelaron la composición filogenética limitada de la comunidad de la laguna ácida y permitieron identificar a una microalga verde de la clase *Trebouxiophyceae* que suponemos es el productor primario de la comunidad. Además, las secuencias metagenómicas confirmaron que las microalgas son los microorganismos dominantes. Finalmente el metagenoma contiene secuencias de bacterias cosmopolitas de ambientes ácidos pero no contiene secuencias de arqueas, apoyando así a que las bacterias son más abundantes.

Cultivo de microorganismos de la laguna ácida de Los Azufres

Se lograron obtener colonias de microalgas y de bacterias de muestras de agua de la laguna ácida inoculadas en medio de cultivo DSMZ 35a a 25 °C. Algunas cajas de cultivo se dejaron expuestas a luz solar indirecta para tratar de cultivar a las microalgas de la laguna ácida (Figura 37). En cambio, otras cajas de cultivo se dejaron en oscuridad para favorecer el crecimiento de bacterias. La caracterización molecular en base a secuencias de genes ribosomales indican que las colonias microalgas corresponden a la microalga de la clase *Trebouxiophyceae* que se detectó en las librerías de clonas y en el metagenoma de la laguna ácida. Las colonias bacterianas correspondieron a cepas de los géneros *Acidocella* y *Thiomonas.*

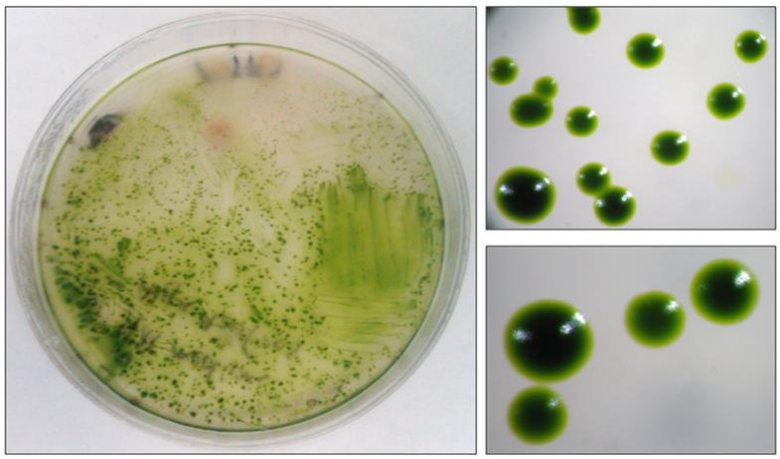


Figura 37. Microalgas cultivables de la clase *Trebouxiophyceae* de la laguna ácida de Los Azufres. Se muestra la primera caja de cultivo en donde se identificaron las colonias de microalgas y de la cual se obtuvo subsecuentemente a la cepa de la microalga *Trebouxiophyceae* sp. MX-AZ01. El medio de cultivo utilizado fue el DSM 35a.

Genoma de una microalga nueva de la clase Trebouxiophyceae

Se obtuvo la secuencia del genoma de la cepa de la microalga *Trebouxiophyceae* sp. MX-AZ01 que se logró cultivar y purificar en el laboratorio. Primero se realizó una ronda de secuenciación en la plataforma GAIIx de Illumina/Solexa. Se obtuvieron 79,838,082 lecturas pareadas de 32 pares de bases y con tamaños de inserto de 300 pares de bases que corresponden a 2,554 megabases de secuencia. Las lecturas se ensamblaron *de novo* utilizando Velvet v.1.2.07 (Zerbino y Birney, 2008). La primera versión del ensamble presenta una cobertura total de 64X y recuperó 320 *scaffolds* que contienen 46.41 megabases. Los 4,615 *contigs* sin *gaps* contienen 45.32 megabases. El genoma de la microalga *Trebouxiophyceae* sp. MX-AZ01 mide aproximadamente entre 45 y 46 megabases en base al primer ensamble.

Posteriormente se realizó una segunda secuenciación del genoma de la microalga utilizando un cuarto de placa de la plataforma de 454 Titanium para obtener lecturas más largas y con mayor tamaño de inserto para mejorar el ensamble de Illumina. Se obtuvieron 324,530 lecturas pareadas con tamaños de inserto de tres kilobases que corresponden a 206.75 megabases. Las secuencias se ensamblaron *de novo* utilizando Newbler (454 *Life Sciences*). El ensamble obtenido presenta una cobertura total de 4X y recuperó 1,889 *scaffolds* que contienen 30.43 megabases. Los 24,771 *contigs* sin *gaps* contienen 28.93 megabases. Actualmente se está trabajando en mejorar los ensambles usando las lecturas de Illumina y de 454 Titanium. La última versión del ensamble del genoma de la microalga contiene 1,537 *contigs* mayores a cinco kilobases, mide 43.2 megabases y presenta un contenido de G+C de 52.31%.

Los genomas parcialmente secuenciados de las microalgas filogenéticamente cercanas *Coccomyxa subellipsoidea* C-169 (38.8 megabases) y *Chlorella variabilis* NC64A (46 Mb) presentan tamaños de genomas parecidos. Nuestros ensambles proveen buenos estimados sobre el tamaño del genoma de la microalga *Trebouxiophyceae* sp. MX-AZ01.

Artículo:

Servín-Garcidueñas LE, Martínez-Romero E. 2012. Complete mitochondrial and plastid genomes of the green microalga *Trebouxiophyceae* sp. strain MX-AZ01 isolated from a highly acidic geothermal lake. Eukaryotic Cell. 11: 1417-1418.

Las secuencias genómicas del primer ensamble de lecturas de Illumina permitieron reconstruir los genomas completos de los genomas del cloroplasto y la mitocondria de la cepa de la microalga *Trebouxiophyceae* sp. MX-AZ01 (Figura 38).

En este artículo se reporta la secuenciación, ensamble, características y contenido genético los genomas de los organelos de la microalga *Trebouxiophyceae* sp. MX-AZ01. Los genomas de los organelos fueron los primeros obtenidos de microalgas trebouxiofitas acidófilas.

Se encontró que el genoma de la mitocondria es el más largo reportado para una microalga trebouxiofita debido a que presenta varias secuencias de intrones y endonucleases, presentes incluso en sus genes ribosomales. Además el genoma del cloroplasto presenta el contenido de G+C más alto descrito para microalgas trebouxiofitas.

EC Journals.ASM.org

GENOME ANNOUNCEMENT

Complete Mitochondrial and Plastid Genomes of the Green Microalga *Trebouxiophyceae* sp. Strain MX-AZ01 Isolated from a Highly Acidic Geothermal Lake

Luis E. Servín-Garcidueñas and Esperanza Martínez-Romero

Center for Genomic Sciences, UNAM, Cuernavaca, Morelos, Mexico

We report the complete organelle genome sequences of *Trebouxiophyceae* sp. strain MX-AZ01, an acidophilic green microalga isolated from a geothermal field in Mexico. This eukaryote has the remarkable ability to thrive in a particular shallow lake with emerging hot springs at the bottom, extremely low pH, and toxic heavy metal concentrations. *Trebouxiophyceae* sp. MX-AZ01 represents one of few described photosynthetic eukaryotes living in such a hostile environment. The organelle genomes of *Trebouxiophyceae* sp. MX-AZ01 are remarkable. The plastid genome sequence currently presents the highest G+C content for a trebouxiophyte. The mitochondrial genome sequence is the largest reported to date for the *Trebouxiophyceae* class of green algae. The analysis of the genome sequences presented here provides insight into the evolution of organelle genomes of trebouxiophytes and green algae.

The *Trebouxiophyceae* are a class of phylum *Chlorophyta*, which comprises algae from marine and fresh waters (5). *Trebouxiophyceae* sp. strain MX-AZ01 was recently identified and isolated from a lake from the Los Azufres geothermal field in western Mexico, where it seems to be endemic. This strain thrives in a lake with pH 2.3 and may represent a novel species. Here we present the complete and annotated organelle genomes of *Trebouxiophyceae* sp. MX-AZ01. These sequences represent the first organelle genomes reported for an acidophilic trebouxiophyte.

Trebouxiophyceae sp. MX-AZ01 was maintained at the Center for Genomic Sciences in the culture collection of the Ecological Genomics Department of UNAM. DNA was sequenced with the Illumina GAIIx platform. Reads were assembled *de novo* using Velvet 1.2.07 (9). Some contigs with overrepresented coverage corresponded to organelle sequences. Reads were mapped to gap-surrounding sequences by using Maq 0.7.1 (6). Mapping reads and PCR amplifications were used to eliminate gaps. Coding sequences were predicted using GeneMark.hmm 2.0 (1). rRNA and tRNA were predicted using RNAmmer 1.2 (4) and tRNAscan-SE 1.21 (7). Further, gene predictions were manually verified. The organelle genomes from *Coccomyxa subellipsoidea* C-169 (8) and *Trebouxiophyceae* sp. MX-AZ01 were aligned using Mauve 2.3.1 (3).

The mitochondrial genome sequence (74.4 kb; 197-fold coverage) is the largest currently deposited in GenBank for the *Trebouxiophyceae* and has a G+C content of 53.4%. The mitochondrial genome comprises 42 putative coding genes, 24 tRNAs, and 3 rRNAs. Seven putative endonucleases were detected as part of intronic regions, with six of them located in the large-subunit rRNA gene and one present in the cytochrome *c* oxidase subunit 1. Group II introns were detected among four tRNA-coding genes ($trnH_{gug}$, $trnS_{gcu}$, $trnS_{uga}$, and $trnW_{cca}$). The gene order between *Trebouxiophyceae* sp. MX-AZ01 and *C. subellipsoidea* mitochondrial genome is conserved.

The plastid genome (149.7 kb; 312-fold coverage) has the highest G+C content, 57.7%, exceeding the recently described 50.7% G+C content of the *C. subellipsoidea* plastid (8). The *Trebouxiophyceae* sp. MX-AZ01 plastid genome comprises 81 putative coding genes, 32 tRNAs, and 3 rRNAs. All three rRNA genes clustered together, different from the ribosomal gene organization of the plastid genome of *C. subellipsoidea*. Three putative endonucleases were detected as part of intronic regions, with two of them located in the large-subunit rRNA gene and another one found in *psbA*. The group I intron found in *psbB* in the plastid genome of *C. subellipsoidea* was absent (8). Comparisons of the plastid genomes of *C. subellipsoidea* and *Trebouxiophyceae* sp. MX-AZ01 suggested the occurrence of several genome rearrangements.

The organelle genomes were sequenced as part of a research initiative aimed at describing the diversity of extremophiles from the Los Azufres geothermal field. These are the first complete genome sequences obtained from a microorganism isolated from Los Azufres. The nuclear genome from *C. subellipsoidea* has been reported previously (2). Future comparative genomics between nuclear genomes will allow better taxonomic classification of *Trebouxiophyceae* sp. MX-AZ01.

Nucleotide sequence accession numbers. The genome sequences determined in this study have been deposited in DDBJ/EMBL/GenBank under accession numbers JX315601 (mitochondrion) and JX402620 (plastid).

ACKNOWLEDGMENTS

We are thankful with the University Unit for Massive Sequencing (UUSM) from the National University of Mexico (UNAM) for Illumina-Solexa sample sequencing. The Comisión Federal de Electricidad (CFE) personnel gave us field assistance and permits for samplings at Los Azufres. We are also thankful to Michael Dunn for reviewing the manuscript and Jesus Campos Garcia for allowing us to process some samples at the Chemical-Biology Research Institute (IIQB) from the Universidad Michoacana de San Nicolás de Hidalgo (UMSNH). Samplings were carried out with the efforts of Hans Román, Jonathan Lopez, Cecilia Garci-

Received 4 September 2012 **Accepted** 4 September 2012

Address correspondence to Luis E. Servín-Garcidueñas, luis.e.servin@gmail.com.

doi:10.1128/EC.00244-12

dueñas, and José Luis Servín. We are grateful to Paul Gaytán and Eugenio López for synthesis of oligonucleotides at the Synthesis Unit of the Biotechnology Institute of UNAM.

This work was supported by PAPIIT IN205412 from DGAPA, UNAM. L.E.S.-G. was supported with a scholarship from CONACyT as Ph.D. student at the Programa de Doctorado en Ciencias Biomédicas from UNAM.

REFERENCES

1. **Besemer J, Borodovsky M.** 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Res. **27**:3911–3920.
2. **Blanc G, et al.** 2012. The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. Genome Biol. **13**: R39.
3. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14**:1394–1403.
4. **Lagesen K, et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
5. **Leliaert F, et al.** 2012. Phylogeny and molecular evolution of the green algae. Crit. Rev. Plant Sci. **31**:1–46.
6. **Li H, Ruan J, Durbin R.** 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18**:1851–1858.
7. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**: 955–964.
8. **Smith DR, et al.** 2011. The GC-rich mitochondrial and plastid genomes of the green alga *Coccomyxa* give insight into the evolution of organelle DNA nucleotide landscape. PLoS One **6**:e23624. doi:10.1371/journal.pone.0023624.
9. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.

A)

B)

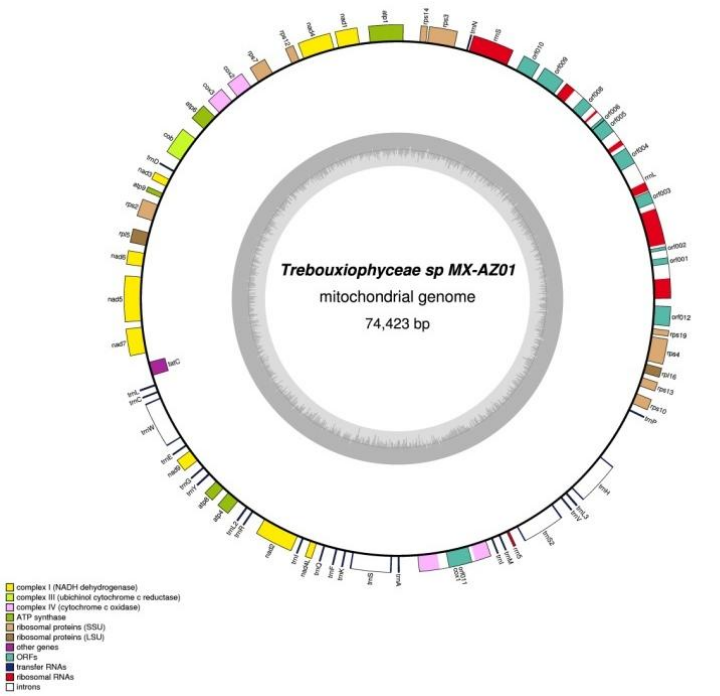
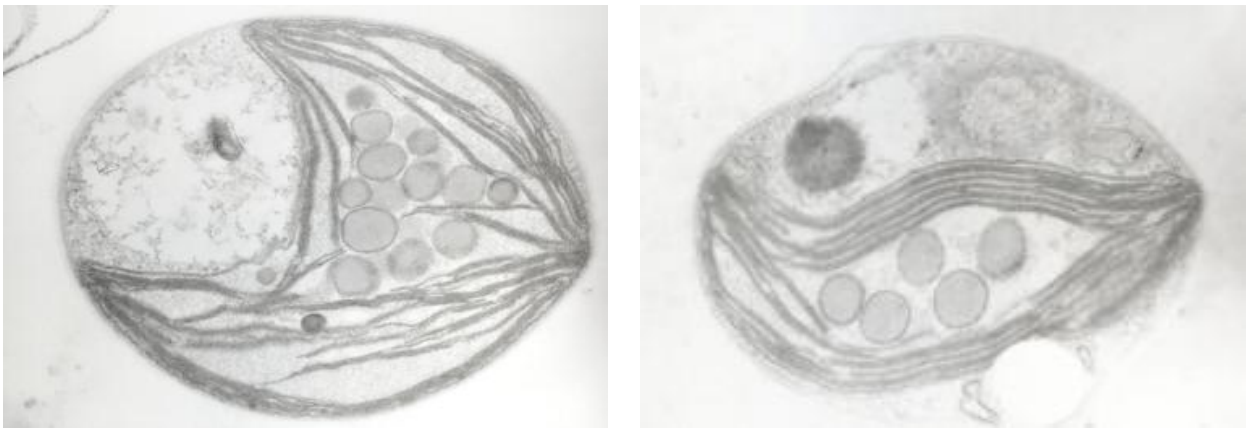


Figura 38. Representación de los genomas completos de los organelos de la microalga *Trebouxiophyceae* sp. MX-AZ01 recuperados del metagenoma del sedimento fotosintético de Los Azufres. Los colores de los genes predichos corresponden a distintas categorías funcionales.

Microscopía electrónica de microalgas de la laguna ácida de Los Azufres

Se obtuvieron imágenes de microscopía electrónica preliminares con la finalidad de conocer la morfología de las microalgas. A partir de muestras de agua de la laguna ácida se identificaron células con morfologías características de microalgas verdes de la clase *Trebouxiophyceae*. Las células vegetativas tienen forma globular o de cocos alargados y miden entre 3 nm de largo por 2 nm de ancho. Las células vegetativas contienen un cloroplasto grande que presenta varias membranas. También se observa la presencia de vesículas que pudieran estar almacenando almidón o ácidos grasos (Figura 39).

A) B)



3 nm x 2.25 nm 2.6 nm x 1.8 nm

Figura 39. Imágenes de microscopia electrónica de transmisión de células vegetativas de microalgas de la laguna ácida de Los Azufres.

Artículo:

Servín-Garcidueñas LE, Garrett RA, Amils R, Martínez-Romero E. 2013. Genome sequence of the acidophilic bacterium *Acidocella* sp. strain MX-AZ02. Genome Announc. 1: e00041-12.

Una de las bacterias de la laguna ácida que se logró cultivar en el laboratorio corresponde al género *Acidocella*. Las bacterias del género *Acidocella* son α-protebacterias que habitan comúnmente en ambientes con concentraciones elevadas de metales pesados y en condiciones de acidez extrema tales como las presentes en sitios termales y en de drenajes de minas ácidas. Las bacterias del género *Acidocella* son acidófilas, mesófilas y heterótrofas. En el metagenoma de la Laguna Verde se identificaron pocas lecturas de *Acidocella* sin embargo se pudieron identificar varias colonias usando medio de cultivo DSM 35a. El ADN de la cepa *Acidocella* sp. MX-AZ02 se purificó y se utilizó para obtener la secuencia de su genoma. El genoma de la cepa MX-AZ02 es el primero que se secuenció de una bacteria del género *Acidocella*. El genoma de la cepa MX-AZ02 reveló secuencias genómicas que permitirán conocer el potencial genético que posee para residir en las condiciones extremas de la laguna ácida. En este artículo se reporta la secuenciación, ensamble y principales características del genoma parcial de la cepa MX-AZ02.

El genoma de *Acidocella* sp. MX-AZ02 ha permitido identificar genes involucrados en el metabolismo energético que son basales en la historia evolutiva de las α-protebacterias. En base a análisis filogenéticos sugerimos que ocurrió una duplicación génica ancestral de la citocromo *bd* oxidasa. El evento de duplicación génica probablemente ocurrió en los antecesores de las α-proteobacterias actuales tales como las pertenecientes al orden Rhodospirillales, que incluyen a *Acidocella*. Sugerimos que las oxidasas del tipo *bd-I* ancestrales pudieron haberse transferido a otros linajes de proteobacterias. Encontramos también que las oxidasas del tipo CIO pudieron haberse diferenciado en varios subtipos como resultado de duplicaciones génicas posteriores. Los resultados completos de este análisis se presentan en el artículo: Degli Esposti M, Rosas-Pérez T, Servín-Garcidueñas LE, Bolaños LM, Rosenblueth M, Martínez-Romero E. 2015. *Molecular evolution of cytochrome bd oxidases across proteobacterial genomes*. Genome Biol. Evol. 7:801–820, que no se incluye como parte de esta tesis.

genomeA
Journals.ASM.org

Genome Sequence of the Acidophilic Bacterium *Acidocella* sp. Strain MX-AZ02

Luis E. Servín-Garcidueñas,[a] Roger A. Garrett,[b] Ricardo Amils,[c] Esperanza Martínez-Romero[a]

Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico[a]; Department of Biology, Archaea Centre, University of Copenhagen, Copenhagen, Denmark[b]; Centro de Biología Molecular Severo Ochoa, Universidad Autónoma de Madrid, Madrid, Spain[c]

Here, we report the draft genome sequence of *Acidocella* sp. strain MX-AZ02, an acidophilic and heterotrophic alphaproteobacterium isolated from a geothermal lake in western Mexico.

Received 12 October 2012 **Accepted** 26 October 2012 **Published** 15 January 2013

Citation Servín-Garcidueñas LE, Garrett RA, Amils R, Martínez-Romero E. 2013. Genome sequence of the acidophilic bacterium *Acidocella* sp. strain MX-AZ02. Genome Announc. 1(1):e00041-12. doi:10.1128/genomeA.00041-12.

Address correspondence to Luis E. Servín-Garcidueñas, luis.e.servin@gmail.com.

Acidocella sp. strain MX-AZ02 was isolated from a naturally acidic (pH 2.3) and heavy metal-containing shallow lake in the Los Azufres National Park in western Mexico. The *Acidocella* genus comprises aerobic, acidophilic, Gram-negative bacteria belonging to the class *Alphaproteobacteria* (1). *Acidocella* relatives have been identified both in natural and acid mine drainage environments exhibiting high heavy-metal levels (2–6). *Acidocella* has also been detected among *Sphagnum* moss microbiota growing under varying acidic conditions (7, 8). Currently, the genus contains three reference strains isolated from acidic environments (9–11).

DNA was isolated from *Acidocella* sp. MX-AZ02, which yields smooth, round, and translucent colonies on DSMZ medium 35a. the culture collection of the Ecological Genomics Department, National University of Mexico (UNAM). The sequencing was performed with the Roche 454 GS-FLX titanium technology generating 58.04 Mbp (~16-fold coverage) from a mate-paired library with 3-kb inserts. The reads were assembled *de novo* using Newbler assembler 2.3 (454 Life Sciences). The assembly produced 303 contigs of >500 bp each with an N_{50} size of 22.12 kb. Nine scaffolds were generated containing 250 contigs. Genome annotation was performed using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html).

The genome of *Acidocella* sp. MX-AZ02 was estimated to be 3.6 Mbp with a G+C content of 64.1% and it carried 3,553 open reading frames (ORFs). The 16S rRNA gene phylogeny indicated that the strain is closely related to type strains *Acidocella facilis* PW2, *Acidocella aluminiidurans* AL46, and *Acidocella aminolytica* 101, sharing 99.86%, 99.50%, and 97.93% sequence identities, respectively, over 1,407 bp.

Metal resistance determinants have been identified for *Acidocella* strains (12, 13, 14). The *Acidocella* sp. MX-AZ02 genome codes for arsenic, chromium, copper, and cobalt-zinc-cadmium transporters, as well as heavy-metal sensor signal transduction histidine kinases and chaperones. Carbonic anhydrases were also encoded, which may provide a means to cope with the low CO_2 levels in acidic waters.

One *Acidocella* strain was shown to metabolize fructose from medium containing cell-free algal exudates, but it was unable to metabolize mannitol or glucose (15). *Acidocella* sp. MX-AZ02 may use glucose in the isolation medium as a carbon source. An acidophilic and abundant unicellular green alga was recently characterized from the same lake from which *Acidocella* sp. MX-AZ02 was isolated (16). Possibly, *Acidocella* sp. MX-AZ02 utilizes organic compounds from the alga, as was proposed previously for acidophilic microalgae and acidophilic heterotrophic bacteria (15).

The draft genome of *Acidocella* sp. MX-AZ02 will facilitate the identification of metal resistance determinants and may help us understand bacterial–algal interactions. This is the first isolated bacterial genome for an *Acidocella* strain and is the first sequenced bacterial genome from the Los Azufres National Park.

Nucleotide sequence accession number. The draft of the genome sequence is deposited at DDBJ/EMBL/GenBank under the accession no. AMPS00000000.

ACKNOWLEDGMENTS

We thank the Next Generation Sequencing unit from Macrogen Inc. for pyrosequencing. The Comisión Federal de Electricidad (CFE) personnel provided a sampling permit.

This research was supported by PAPIIT IN205412 from DGAPA, UNAM. L.E.S.G. received a PhD scholarship from CONACYT (Mexico) under the Programa de Doctorado en Ciencias Biomédicas from UNAM. L.E.S.G. had a "Beca Mixta" from CONACYT and a fellowship from PAEP during a research internship at the Danish Archaea Centre.

REFERENCES

1. **Kishimoto N, Kosako Y, Wakao N, Tano T, Hiraishi A.** 1995. Transfer of *Acidiphilium facilis* and *Acidiphilium aminolytica* to the genus *Acidocella* gen. nov, and emendation of the genus *Acidiphilium*. Syst. Appl. Microbiol. **18**:85–91.
2. **Hallberg KB, Coupland K, Kimura S, Johnson DB.** 2006. Macroscopic streamer growths in acidic, metal-rich mine waters in north Wales consist of novel and remarkably simple bacterial communities. Appl. Environ. Microbiol. **72**:2022–2030.

3. **Hamamura N, Olson SH, Ward DM, Inskeep WP.** 2005. Diversity and functional analysis of bacterial communities associated with natural hydrocarbon seeps in acidic soils at rainbow springs, Yellowstone National Park. Appl. Environ. Microbiol. **71**:5943–5950.
4. **Johnson DB, Rolfe S, Hallberg KB, Iversen E.** 2001. Isolation and phylogenetic characterization of acidophilic microorganisms indigenous to acidic drainage waters at an abandoned Norwegian copper mine. Environ. Microbiol. **3**:630–637.
5. **Lear G, Niyogi D, Harding J, Dong Y, Lewis G.** 2009. Biofilm bacterial community structure in streams affected by acid mine drainage. Appl. Environ. Microbiol. **75**:3455–3460.
6. **Lu S, Gischkat S, Reiche M, Akob DM, Hallberg KB, Küsel K.** 2010. Ecophysiology of Fe-cycling bacteria in acidic sediments. Appl. Environ. Microbiol. **76**:8174–8183.
7. **Bragina A, Maier S, Berg C, Müller H, Chobot V, Hadacek F, Berg G.** 2011. Similar diversity of alphaproteobacteria and nitrogenase gene amplicons on two related *Sphagnum* mosses. Front. Microbiol. **2**:275.
8. **Opelt K, Berg G.** 2004. Diversity and antagonistic potential of bacteria associated with bryophytes from nutrient-poor habitats of the Baltic sea coast. Appl. Environ. Microbiol. **70**:6569–6579.
9. **Kimoto K, Aizawa T, Urai M, Ve NB, Suzuki K, Nakajima M, Sunairi M.** 2010. *Acidocella aluminiidurans* sp. nov., an aluminium-tolerant bacterium isolated from *Panicum repens* grown in a highly acidic swamp in actual acid sulfate soil area of Vietnam. Int. J. Syst. Evol. Microbiol. **60**:764–768.
10. **Kishimoto N, Kosako Y, Tano T.** 1993. *Acidiphilium aminolytica* sp. nov: an acidophilic chemoorganotrophic bacterium isolated from acidic mineral environment. Curr. Microbiol. **27**:131–136.
11. **Wichlacz PL, Unz RF, Langworthy TA.** 1986. *Acidiphilium angustum* sp. nov., *Acidiphilium facilis* sp. nov, and *Acidiphilium rubrum* sp. nov: acidophilic heterotrophic bacteria isolated from acidic coal mine drainage. Int. J. Syst. Bacteriol **36**:197–201.
12. **Ghosh S, Mahapatra NR, Banerjee PC.** 1997. Metal resistance in *Acidocella* strains and plasmid-mediated transfer of this characteristic to *Acidiphilium multivorum* and *Escherichia coli*. Appl. Environ. Microbiol. **63**:4523–4527.
13. **Ghosh S, Mahapatra NR, Ramamurthy T, Banerjee PC.** 2000. Plasmid curing from an acidophilic bacterium of the genus *Acidocella*. FEMS Microbiol. Lett. **183**:271–274.
14. **Ghosh S, Mahapatra NR, Nandi S, Banerjee PC.** 2005. Integration of metal-resistant determinants from the plasmid of an *Acidocella* strain into the chromosome of *Escherichia coli* DH5alpha. Curr. Microbiol. **50**:28–32.
15. **Nancucheo I, Barrie Johnson D.** 2012. Acidophilic algae isolated from mine-impacted environments and their roles in sustaining heterotrophic acidophiles. Front. Microbiol. **3**:325.
16. **Servín-Garcidueñas LE, Martínez-Romero E.** 2012. Complete mitochondrial and plastid genomes of the green microalga Trebouxiophyceae sp. strain MX-AZ01 isolated from a highly acidic geothermal lake. Eukaryot. Cell **11**:1417–1418.

Diversidad de otras comunidades bacterianas de Los Azufres

Diversidad bacteriana de agua colectada de un pozo de enfriamiento

El pozo de enfriamiento contiene la comunidad microbiana más diversa de todos los sitios analizados (Fig. 1E). Las secuencias de genes 16S rRNA obtenidas de la muestra del pozo de enfriamiento (55 secuencias) se agrupan filogenéticamente dentro de nueve géneros bacterianos (Tabla 10, Figura 40). Probablemente, la presencia de un número mayor de géneros bacterianos se debe al hecho de que las condiciones de la muestra del pozo de enfriamiento (pH=6.6, T=18°C) no son tan extremas como en los otros sitios.

La curva de rarefacción de las secuencias de genes 16S rRNA de las bacterias del agua del pozo de enfriamiento no supera los 20 OTUs y ha alcanzado la estabilidad (Figura 41). La curva de rarefacción incluye a las 55 secuencias generadas (longitud= ~1500 nucleótidos). La secuenciación del metagenoma de la muestra de agua del pozo no se realizó dado que el número de filotipos es mayor que el resto de los sitios, que los filotipos identificados son ya conocidos y también debido a que ya se cuenta con genomas completos para varios de los filotipos.

Tabla 10. Filotipos identificados en la muestra de agua de un pozo de enfriamiento de Los Azufres.

Orden	Familia	Género	% de identidad	número de acceso GenBank
Rhodobacterales	*Hyphomonadaceae*	*Hyphomonas*	97%	CP000158
Burkholderiales	*Burkholderiaceae*	*Limnobacter*	97%	GQ284439
Oceanospirillales	*Oceanospirillaceae*	*Marinomonas*	98%	EF382679
Rhizobiales	*Phyllobacteriaceae*	*Parvibaculum*	97%	CP000774
Sphingomonadales	*Erythrobacteraceae*	*Porphyrobacter*	97%	AB016518
Xanthomonadales	*Xanthomonadaceae*	*Rhodanobacter*	97%	FJ405366
Rhodobacterales	*Rhodobacteraceae*	*Rhodobacter*	97%	AM399030
Burkholderiales	/	*Thiomonas*	98%	CP002021
Chromatiales	*Halothiobacillaceae*	*Thiovirga*	81%	CP001634

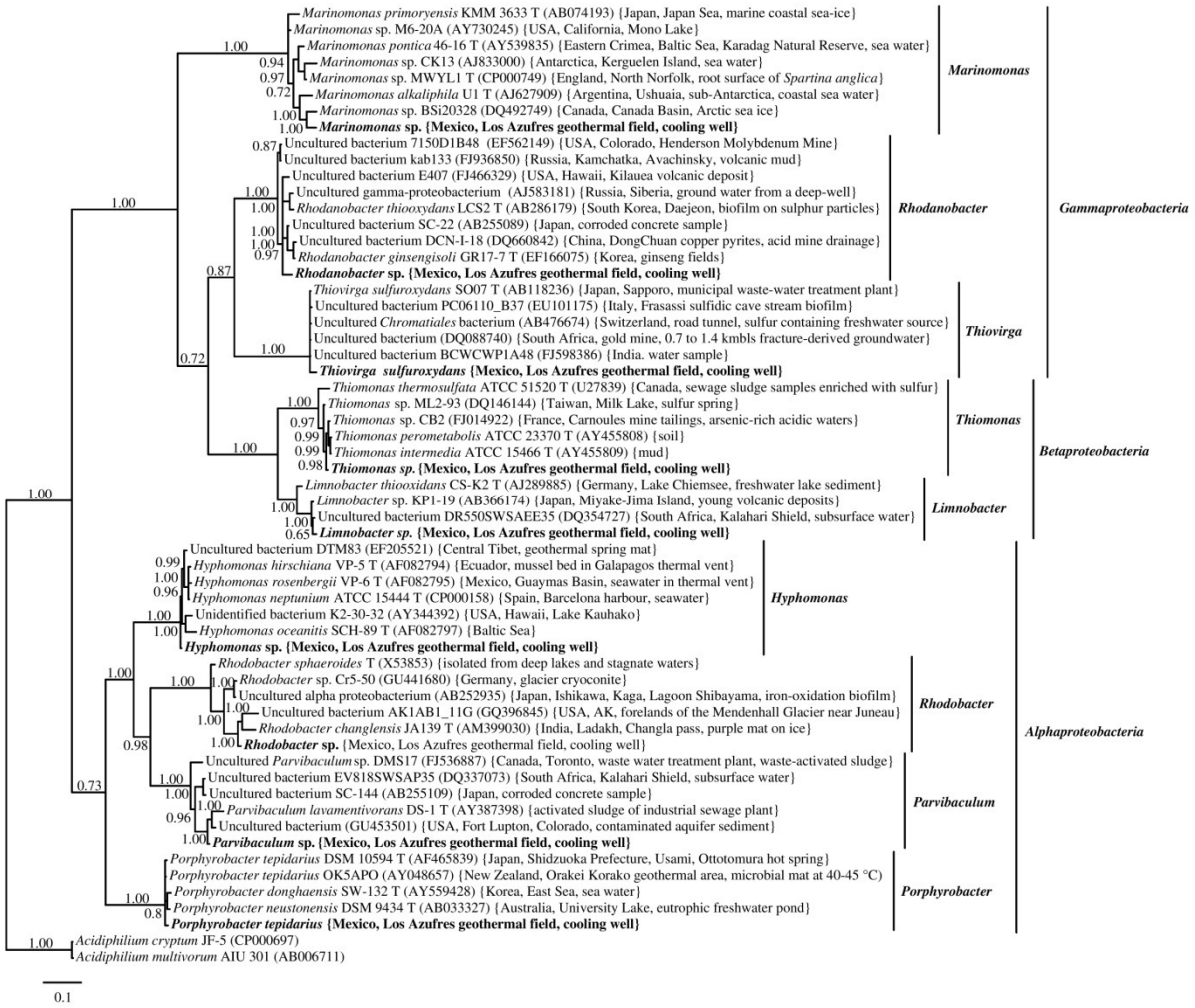


Figura 40. Posición filogenética de las secuencias de genes ribosomales 16S rRNA obtenidas de la muestra de ADN de agua del pozo de enfriamiento. El alineamiento de secuencias contiene 1376 caracteres. La reconstrucción filogenética bayesiana se realizó en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Los números de acceso de GenBank de las secuencias de referencia se indican en paréntesis. Las secuencias obtenidas en este estudio se muestran en negritas. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. La barra de escala representa el número promedio de sustitución de nucleótidos por sitio.

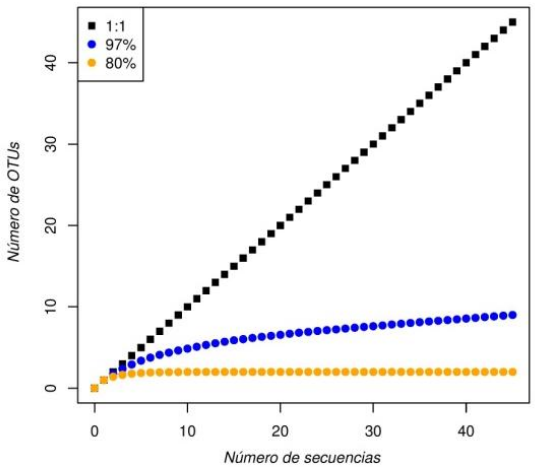


Figura 41. Curva de rarefacción de las secuencias de los genes ribosomales 16S rRNA de los análisis de diversidad de la muestra de agua del pozo de enfriamiento. Los datos se obtuvieron utilizando mothur (Oakley et al., 2009) y la gráfica se reconstruyó utilizando R. Se muestran las curvas correspondientes a distintos puntos de corte de acuerdo a la identidad de secuencia.

Diversidad bacteriana de agua geotérmica colectada de tuberías profundas

El análisis de diversidad de una muestra de agua geotérmica se realizó para determinar si los microorganismos identificados en las manifestaciones geotérmicas superficiales podrían estar presentes en el reservorio geotérmico profundo. Las secuencias de genes 16S rRNA de la muestra de agua son cercanas filogenéticamente a secuencias de bacterias de la clase *Betaproteobacteria* que han sido identificadas en manifestaciones geotérmicas y en muestras de agua de minas (Tabla 11, Figura 42).

El género *Thiomonas* se identificó tanto en la comunidad de la laguna ácida como en la comunidad del pozo de enfriamiento. En conjunto, estos resultados nos indican que la comunidad bacteriana de una muestra de agua del reservorio geotérmico profundo presenta una complejidad limitada y que la comunidad está conformada principalmente por *Betaproteobacterias*.

Tabla 11. Filotipos identificados en la muestra de agua geotérmica profunda de Los Azufres.

Orden	Familia	Género	% de identidad	número de acceso GenBank
Burkholderiales	*Comamonadaceae*	*Acidovorax*	99%	Y18617
Burkholderiales	/	*Thiomonas*	99%	DQ146144
Burkholderiales	*Comamonadaceae*	/	98%	EF562037

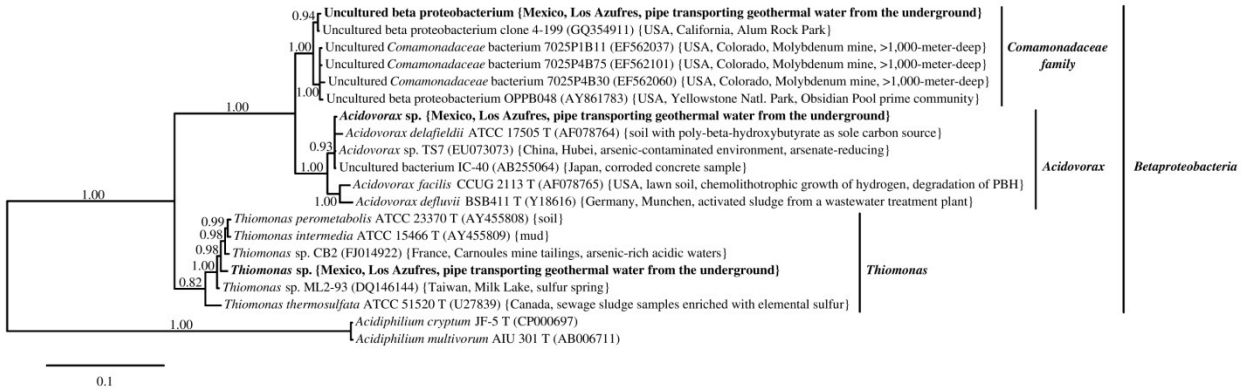


Figura 42. Posición filogenética de las secuencias de genes ribosomales 16S rRNA obtenidas de la muestra de agua geotérmica colectada de una tubería que conduce agua y vapor desde el reservorio geotérmico profundo. El alineamiento de secuencias contiene 1082 caracteres. La reconstrucción filogenética bayesiana se realizó en MrBayes3 (Ronquist & Huelsenbeck, 2003) utilizando el modelo de sustitución GTR+I+G. Los números de acceso de GenBank de las secuencias de referencia se indican en paréntesis. Las secuencias obtenidas en este estudio se muestran en negritas. Los valores de soporte de las ramas se muestran como probabilidades posteriores bayesianas. La barra de escala representa el número promedio de sustitución de nucleótidos por sitio.

Conclusiones

Las comunidades microbianas que poseen una complejidad limitada proveen la oportunidad única de utilizar las herramientas de secuenciación masiva para elucidar la diversidad filogenética y funcional con gran resolución. Nuestros análisis de diversidad mediante enfoques independientes de cultivo indican que de las manifestaciones geotérmicas de Los Azufres albergan comunidades microbianas de complejidad limitada.

En todas las comunidades analizadas se identificaron géneros conocidos de bacterias y de arqueas que se encuentran presentes a nivel mundial en drenajes ácidos de minas, en depósitos de aguas profundas, en cuevas y en ambientes de origen volcánico. De acuerdo a los análisis realizados, también se identificaron microorganismos filogenéticamente novedosos de los tres dominios de la vida y virus que permanecían desconocidos.

En general, los ensambles de las secuencias metagenómicas permitieron la reconstrucción de los genomas consenso de las poblaciones de los microorganismos mayoritarios lo que está permitiendo descubrir su potencial metabólico con gran resolución.

Los análisis metagenómicos de una solfatara ácida indican que las comunidades microbianas están compuestas por poblaciones de arqueas. Las poblaciones dominantes corresponden a un género nuevo del orden *Sulfolobales.* Las arqueas *Sulfolobales* se encuentran presentes tanto en la parte aeróbica como anaeróbica de la solfatara. También se descubrieron poblaciones minoritarias de arqueas anaeróbicas del orden *Thermoproteales* que corresponden a los géneros *Vulcanisaeta* y *Thermoproteus*.

Se obtuvieron genomas consenso de las poblaciones de arqueas *Sulfolobales* y *Thermoproteales* a partir del ensamble de metagenomas de la solfatara ácida. También se obtuvieron los genomas de los virus SMR1 (*Sulfolobales Mexican Rudivirus*) y SMF1 (*Sulfolobales Mexican Fusellovirus*) de la misma solfatara ácida. Se descubrió que los sedimentos circundantes a la solfatara ácida contienen una cepa nueva de *Sulfolobus acidocaldarius* que fue secuenciada completamente y cuyo genoma representa el más pequeño y diverso dentro del género *Sulfolobus.*

Se logró identificar una comunidad dominada por una microalga roja de la familia *Cyanidiaceae* y por arqueas conocidas como ABC-plasmas del orden *Thermoplasmatales* en un metagenoma de sedimentos fotosintéticos depositados en una fumarola ácida. El ensamble de secuencias permitió reconstruir los genomas de los organelos de la microalga. Además, el análisis reveló el primer genoma de una especie de arquea candidata del filo Parvarchaeota que contiene arqueas enigmáticas y ultra pequeñas. También se lograron identificar secuencias de genes ribosomales que corresponden a linajes nuevos de actinobacterias, *Chloroflexi* y *Thermotogales*.

La comunidad de la laguna ácida está compuesta en su mayoría por una población dominante de una microalga verde que corresponde a un linaje nuevo. También se identificó a un número limitado de bacterias acidófilas y de eucariotas desconocidos. El genoma parcial de la microalga verde pudo ser obtenido del metagenoma de la laguna ácida y permitió reconstruir los genomas completos de sus organelos.

Los análisis de diversidad de los sitios mencionados anteriormente y de muestras de agua de un pozo de enfriamiento y de agua geotérmica indican que los microorganismos de Los Azufres pueden transferirse entre el subsuelo y la superficie y se localizan tanto en sitios naturales como en sitios artificiales generados por el hombre para la generación de energía geotérmica.

Perspectivas

Solfatara ácida

Como primera meta se planean publicar los resultados del segundo metagenoma para reportar la identificación y los genomas consenso de las comunidades de arqueas *Thermoproteales*. Actualmente se está trabajando en el escrito.

Se desea realizar un tercer metagenoma de la solfatara ácida utilizando técnicas nuevas de secuenciación usando equipos tales como PacBio o HiSeq. De esta forma se podrán generar ensambles más completos para los genomas de las arqueas y arqueovirus.

Realizar ensayos adicionales para tratar de cultivar a las arqueas *Sulfolobales* y *Thermoproteales* identificadas mediante los metagenomas. El cultivo de arqueas permitiría además el enriquecimiento de arqueovirus nuevos.

Al realizar los análisis filogenéticos de la arquea *Sulfolobales* AZ01 nos percatamos que el orden *Sulfolobales* requiere esfuerzos para enmendar su taxonomía. Por lo tanto se están realizando ensayos filogenéticos para enviar artículos sobre las relaciones evolutivas de las arqueas que integran al orden *Sulfolobales*.

Sedimentos fotosintéticos

Como primer meta se planea realizar una publicación para reportar la diversidad revelada mediante técnicas metagenómicas.

Después se planean analizar los genomas ya obtenidos de los organelos de la microalga roja de la familia *Cyanidiaceae* y enviar una publicación.

Se desea identificar mediante técnicas moleculares a los microorganismos que se lograron cultivar en los primeros ensayos. Se podrán generar secuencias genómicas de microorganismos aislados si llegaran a ser filogenéticamente novedosos o si pertenecieran a grupos taxonómicos que carecen de genomas reportados.

Actualmente se están realizando análisis comparativos entre los genomas de las arqueas ARMAN obtenidos de los biofilmes de la mina de California (Baker et al., 2010) y el genoma recuperado del metagenoma de los sedimentos de Los Azufres. Los análisis comparativos permitirán identificar los genes y propiedades que tienen en común o que son exclusivas para cada una de ellas. Se planea enviar un artículo enfocado en reportar la obtención del genoma la arquea ARMAN de Los Azufres y los resultados obtenidos de los análisis comparativos.

Se tiene contemplado hacer un análisis para obtener otros genomas individuales del metagenoma. Los genomas de las arqueas ABC-plasmas son los primeros que se tienen considerados. Se harán comparaciones con los genomas recuperados para las arqueas ABC-plasmas obtenidas de los biofilmes de la mina de California (Yelton et al., 2013).

Se planean realizar búsquedas bioinformáticas para recuperar secuencias de virus del metagenoma del sedimento fotosintético debido a que las pocas imágenes de microscopía existentes de arqueas ARMAN revelan interacciones desconocidas con virus (Baker et al., 2010; Comolli y Banfield, 2014). A la fecha no se han reportado genomas completos de virus de arqueas ARMAN o de arqueas ABC-plasmas. Los genomas de los virus podrían ayudar a comprender a las interacciones mediante el análisis de su potencial genético.

Otra meta más ambiciosa consiste en analizar el genoma de la microalga roja que es dominante en los sedimentos ya que su genoma se encuentra casi completamente representado en el metagenoma. Se planean identificar eventos de transferencia horizontal de bacterias y arqueas al genoma de la microalga mediante análisis filogenéticos, anotación de genes y contenido diferencial de G+C. Se usaran como base los trabajos reportados sobre los genomas de otras microalgas rojas (Matsuzaki et al., 2004; Barbier et al., 2005; Schönknecht et al., 2013).

Laguna ácida

Se desea publicar un artículo para reportar la diversidad bacteriana de la laguna ácida recuperada mediante las técnicas de amplificación de genes ribosomales y de metagenómica.

También se planea realizar el ensamble del genoma de la microalga verde mediante el uso de secuencias del metagenoma y del genoma aislado obtenido de células en cultivo. Se tiene considerado realizar la secuenciación de transcriptomas de la microalga verde en condiciones de campo y de cultivo para ayudar en la anotación del genoma.

Actualmente se está secuenciando ADN del genoma de *Acidocella* sp. MXAZ02 con fines a generar su genoma completo. Ayudaré en ensamble del genoma cuando las secuencias se encuentren disponibles. También se tiene considerado generar los genomas de aislados de *Thiomonas* sp. y de *Acidiphilium* sp. para comprender mejor su potencial metabólico y para realizar análisis comparativos con otros genomas reportados.

Otros sitios

Se desean publicar los resultados obtenidos sobre la diversidad bacteriana del agua del pozo de enfriamiento y del agua geotérmica. Se pondrá énfasis en reportar a los grupos bacterianos compartidos entre ambas muestras de agua y en la movilidad de los microorganismos entre el subsuelo y la superficie.

Bibliografía

Aguilera A, Manrubia SC, Gómez F, Rodríguez N, Amils R. 2006. Eukaryotic community distribution and its relationship to water physicochemical parameters in an extreme acidic environment, Rio Tinto (southwestern Spain). Appl. Environ. Microbiol. 72: 5325–5330.

Alfaro-Cuevas-Villanueva R, Cortes-Martinez R, García-Díaz JJ, Galvan-Martinez R, Torres-Sanchez R. 2006. Microbiologically influenced corrosion of steels by thermophilic and mesophilic bacteria. Materials and Corrosion. 57: 543–548.

Amann RI, Ludwig W, Schleifer KH. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev.. 59: 143–169.

Amaral Zettler LA, Gómez F, Zettler E, Keenan BG, Amils R, Sogin ML. 2002. Microbiology: eukaryotic diversity in Spain's River of Fire. Nature. 417: 137.

Baker BJ, Tyson GW, Webb RI, Flanagan J, Hugenholtz P, Allen EE, Banfield JF. 2006. Lineages of acidophilic archaea revealed by community genomic analysis. Science. 314: 1933–1935.

Baker BJ, Comolli LR, Dick GJ, Hauser LJ, Hyatt D, Dill BD, Land ML, Verberkmoes NC, Hettich RL, Banfield JF. 2010. Enigmatic, ultrasmall, uncultivated Archaea. Proc. Natl. Acad. Sci. U. S. A. 107: 8806–8811.

Baker GC, Cowan DA. 2004. 16S rDNA primers and the unbiased assessment of thermophile diversity. Biochem. Soc. Trans. 32: 218–221.

Barbier G, Oesterhelt C, Larson MD, Halgren RG, Wilkerson C, Garavito RM, Benning C, Weber AP. 2005. Comparative genomics of two closely related unicellular thermo-acidophilic red algae, Galdieria sulphuraria and Cyanidioschyzon merolae, reveals the molecular basis of the metabolic flexibility of Galdieria sulphuraria and significant differences in carbohydrate metabolism of both algae. Plant Physiol. 137: 460–474.

Benson CA, Bizzoco RW, Lipson DA, Kelley ST. 2011. Microbial diversity in nonsulfur, sulfur and iron geothermal steam vents. FEMS Microbiol. Ecol. 76: 74–88.

Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Research. 33: W451–454.

Brito EM, Villegas-Negrete N, Sotelo-González IA, Caretta CA, Goñi-Urriza M, Gassie C, Hakil F, Colin Y, Duran R, Gutiérrez-Corona F, Piñón-Castillo HA, Cuevas-Rodríguez G,

Malm O, Torres JP, Fahy A, Reyna-López GE, Guyoneaud R. 2014. Microbial diversity in Los Azufres geothermal field (Michoacán, Mexico) and isolation of representative sulfate and sulfur reducers. Extremophiles. 18: 385–398.

Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, Wilkins MJ, Wrighton KC, Williams KH, Banfield JF. 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature. 523: 208–211..

Burggraf S, Huber H, Stetter KO. 1997. Reclassification of the crenarchael orders and families in accordance with 16S rRNA sequence data. Int. J. Syst. Bacteriol. 47: 657–660.

Castorena G, Mugica V, Le Borgne S, Acun ME, Bustos-Jaimes I, Aburto J. 2006. Carbazole biodegradation in gas oil/water biphasic media by a new isolated bacterium *Burkholderia* sp. strain IMP5GC. J. Appl. Microbiol. 100: 739–745.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17: 540–552.

Chaffron S, Rehrauer H, Pernthaler J, von Mering C. 2010. A global network of coexisting microbes from environmental and whole-genome sequence data. Genome Res. 20: 947–959.

Chelius MK, Triplett EW. 2001. The diversity of Archaea and Bacteria in association with the roots of *Zea mays* L. Microb. Ecol. 41: 252–263.

Chevenet F, Brun C, Bañuls AL, Jacq B, Christen R. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics. 7: 439.

Chivian D, Brodie EL, Alm EJ, Culley DE, Dehal PS, DeSantis TZ, Gihring TM, Lapidus A, Lin LH, Lowry SR, Moser DP, Richardson PM, Southam G, Wanger G, Pratt LM, Andersen GL, Hazen TC, Brockman FJ, Arkin AP, Onstott TC. 2008. Environmental genomic sreveals a single-species ecosystem deep within Earth. Science. 322: 275–278.

Comolli LR, Banfield JF. 2014. Inter-species interconnections in acid mine drainage microbial communities. Front. Microbiol. 5: 367.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 5: e11147.

DeLong EF, Pace NR. 2001. Environmental diversity of bacteria and archaea. Syst. Biol.50: 470–478.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72: 5069–5072.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797.

Epstein SS. 2009. General model of microbial uncultivability. Microbiology Monographs.10: 131–159.

Erdmann S, Garrett RA. 2015. Archaeal viruses of the *Sulfolobales*: isolation, infection, and CRISPR spacer acquisition. Methods Mol. Biol. 1311:223–232.

Falagán C, Sánchez-España J, Johnson DB. 2014. New insights into the biogeochemistry of extremely acidic environments revealed by a combined cultivation-based and culture-independent study of two stratified pit lakes. FEMS Microbiol. Ecol. 87: 231–243

Fuchs T, Huber H, Burggraf S, Stetter KO. 1996. 16S rDNA-based phylogeny of the archaeal order *Sulfolobales* and reclassification of *Desulfurococcus ambivalens* as *Acidianus ambivalens* comb. nov. Syst. Appl. Microbiol. 19: 56–60.

Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F. 2010. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. Cold Spring Harb. Protoc. 2010: pdb.prot5368.

Goebel BM, Norris PR, Burton NP. 2000. Acidophiles in biomining. In: Priest FG, Goodfellow M. (Eds.), Applied Microbial Systematics. First Edition, Kluwer Academic Publishers, Dordrecht, pp. 293–314.

González-Pastor JE, Mirete S. 2010. Novel metal resistance genes from microorganisms: a functional metagenomic approach. Methods Mol. Biol. 668: 273–285.

González-Toril E, Llobet-Brossa E, Casamayor EO, Amann R, Amils R. 2003. Microbial ecology of an extreme acidic environment, the Tinto River. Appl Environ. Microbiol. 69: 4853–4865.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59: 307–321.

Handelsman J, Liles M, Mann D, Riesenfeld C, Goodman RM. 2002. Cloning the metagenome: culture-independent access to the diversity and functions of uncultivated microbial world. Methods in Microbiology. 33: 241–255.

Hernandez D1, François P, Farinelli L, Osterås M, Schrenzel J. 2008. *De novo* bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Genome Res. 18: 802–809.

Hua ZS, Han YJ, Chen LX, Liu J, Hu M, Li SJ, Kuang JL, Chain PS, Huang LN, Shu WS. 2015. Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. ISME J. 9: 1280–1294.

Huber R, Huber H, Stetter KO. 2000. Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. FEMS Microbiol. Rev. 24: 615–623.

Huber T, Faulkner G, Hugenholtz P. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics. 20: 2317–2319.

Huber H, Prangishvili D. 2006. *Sulfolobales*. In: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K-H. Stackebrandt, E. (Eds.), The Prokaryotes. Third Edition, Springer, New York, pp. 23–51.

Hugenholtz P, Goebel BM, Pace NR. 1998. Impact of culture-independent studies on the emerging phylogenetic view of microbial diversity. J. Bacteriol. 180: 4765–4774.

Inskeep WP, Rusch DB, Jay ZJ, Herrgard MJ, Kozubal MA, Richardson TH, Macur RE, Hamamura N, Jennings Rd, Fouke BW, Reysenbach AL, Roberto F, Young M, Schwartz A, Boyd ES, Badger JH, Mathur EJ, Ortmann AC, Bateson M, Geesey G, Frazier M. 2010. Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. PLoS ONE. 5: e9773.

Kurosawa N, Itoh YH, Iwai T, Sugai A, Uda I, Kimura N, Horiuchi T, Itoh T. 1998. *Sulfurisphaera ohwakuensis* gen. nov., sp. nov., a novel extremely thermophilic acidophile of the order *Sulfolobales*. Int. J. Syst. Bacteriol. 4: 451–456.

Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35: 3100–3108.

Lepp PW, Brinig MM, Ouverney CC, Palm K, Armitage GC, Relman DA. 2004. Methanogenic Archaea and human periodontal disease. Proc. Natl. Acad. Sci. USA. 101: 6176–6181.

Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 41: W575–581.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25: 955–964.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 1: 18.

Magoc T, Salzberg S. 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 27: 2957–2963.

Mao D, Grogan D. 2012. Genomic evidence of rapid, global-scale gene flow in a *Sulfolobus* species. ISME J. 6: 1613–1616.

Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Nishida K, Yoshida Y, Nishimura Y, Nakao S, Kobayashi T, Momoyama Y, Higashiyama T, Minoda A, Sano M, Nomoto H, Oishi K, Hayashi H, Ohta F, Nishizaka S, Haga S, Miura S, Morishita T, Kabeya Y, Terasawa K, Suzuki Y, Ishii Y, Asakawa S, Takano H, Ohta N, Kuroiwa H, Tanaka K, Shimizu N, Sugano S, Sato N, Nozaki H, Ogasawara N, Kohara Y, Kuroiwa T. 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature. 428: 653–657.

Namiki T, Hachiya T, Tanaka H, Sakakibara Y. 2012. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Res. 40: e155.

Narasingarao P, Podell S, Ugalde JA, Brochier-Armanet C, Emerson JB, Brocks JJ, Heidelberg KB, Banfield JF, Allen EE. 2012. *De novo* metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. ISME J. 6: 81–93.

Navarrette-Bedolla M, Ballesteros-Almanza ML, Sanchez-Yanez JM, Valdez-Salas B, Hernandez-Duque G. 1999. Biocorrosion in a geothermal power plant. Materials Performance. 38: 52–57.

Norris TB, Wraith JM, Castenholz RW, McDermott, TR. 2002. Soil microbial community structure across a thermal gradient following a geothermal heating event. Appl. Environ. Microbiol. 68: 6300–6309.

Pace NR. 1997. A molecular view of microbial diversity and the biosphere. Science. 276: 734–740.

Pham VH, Kim J. 2012. Cultivation of unculturable soil bacteria. Trends Biotechnol. 30: 475–484.

Podar M, Reysenbach AL. 2006. New opportunities revealed by biotechnological explorations of extremophiles. Curr. Opin. Biotechnol. 17: 250–255.

Quijano-León JL, Gutiérrez-Negrín LCA. 2003. An unfinished journey: 30 years of geothermal electric generation in Mexico. GRC Bulletin. 198–203.

Quéméneur M, Heinrich-Salmeron A, Muller D, Lièvremont D, Jauzein M, Bertin PN, Garrido F, Joulian C. 2008. Diversity surveys and evolutionary relationships of aoxB genes in aerobic arsenite-oxidizing bacteria. Appl. Environ. Microbiol. 74: 4567–4573.

Rappé M, Giovannoni S. 2003. The uncultured microbial majority. Annu. Rev. Microbiol. 57: 369–394.

Riesenfeld CS, Schloss PD, Handelsman J. 2004. Metagenomics: genomic analysis of microbial communities. Annu. Rev. Genet. 38: 525–552.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu WT, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature. 499: 431–437.

Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19: 1572–1574.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. Bioinformatics. 16: 944–945.

Satoh T, Watanabe K, Yamamoto H, Yamamoto S, Kurosawa N. 2013. Archaeal community structures in the solfataric acidic hot springs with different temperatures and elemental compositions. Archaea. 2013: 723871.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75: 7537–7541.

Schönknecht G, Chen WH, Ternes CM, Barbier GG, Shrestha RP, Stanke M, Bräutigam A, Baker BJ, Banfield JF, Garavito RM, Carr K, Wilkerson C, Rensing SA, Gagneul D, Dickenson NE, Oesterhelt C, Lercher MJ, Weber AP. 2013. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. Science. 339: 1207–1210.

Schönknecht G, Weber AP, Lercher MJ. 2014. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. Bioessays. 36: 9–20.

Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJ. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature. 521: 173–179.

Soo RM, Wood SA, Grzymski JJ, McDonald IR, Cary SC. 2009. Microbial biodiversity of thermophilic communities in hot mineral soils of Tramway Ridge, Mount Erebus, Antarctica. Environ. Microbiol 11: 715–728.

Stanković S, Morić I, Pavić A, Vasiljević B, Johnson BD, Cvetković V. 2014. Journal of the Serbian Chemical Society. 79: 729–741.

Stetter KO. 1989. "Order III. *Sulfolobales* ord. nov. family *Sulfolobaceae* fam. nov.". In: Staley JT, Bryant MP, Pfennig N, Holt JG. (Eds.), Bergey's Manual of Systematic Bacteriology. First Edition, The Williams & Wilkins Co., Baltimore, pp. 2250–2251.

Stott MB, Crowe MA, Mountain BW, Smirnova AV, Hou S, Alam M, Dunfield PF. 2008. Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. Environ. Microbiol 10: 2030–2041.

Tello-López MR, Suarez-Arriaga MC. 2000. Geochemical evolution of the Los Azufres, Mexico, geothermal reservoir. Proceedings World Geothermal Congress. Kyushu-Tohoku, Japan. 2257–2262.

Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W, Kämpfer P. 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. J. Syst. Evol. Microbiol. 60: 249–266.

Torres-Sanchez R, Magaña-Vazquez A, Sanchez-Yañez JM, Martinez-Gomez L. 1996. High temperature microbial corrosion in the condenser of a geothermal electric power unit. CORROSION96 The NACE Int. Ann. Conf. Exp. 293: 1–14.

Torres-Sanchez R, García-Vargas J, Alfonso-Alonso A, Martínez-Gómez L. 2001. Corrosion of AISI 304 stainless steel induced by thermophilic sulfate reducing bacteria (SRB) from a geothermal power unit. Materials and Corrosion. 52: 614–618.

Trevisanato SI, Larsen N, Segerer AH, Stetter KO, Garrett RA. 1996. Phylogenetic analysis of the archaeal order of *Sulfolobales* based on sequences of 23S rRNA genes and 16S/23S rDNA spacers. Syst. Appl. Microbiol. 19: 61–65.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature. 4: 37–43.

Valdez-Salas B, Schorr-Wiener M, Rioseco de la Peña L, Navarrete-Bedolla M. 2000. Deterioration of materials in geothermal fields in Mexico. Materials and Corrosion. 51: 698–704.

van den Burg B. 2003. Extremophiles as a source for novel enzymes. Curr. Opin. Microbiol. 6: 213–218.

Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. 173: 697–703.

Wemheuer B, Taube R, Akyol P, Wemheuer F, Daniel R. 2013. Microbial diversity and biochemical potential encoded by thermal spring metagenomes derived from the Kamchatka Peninsula. Archaea. 2013: 136714.

Wilson MS, Siering PL, White CL, Hauser ME, Bartles AN. 2008. Novel archaea and bacteria dominate stable microbial communities in North America's Largest Hot Spring. Microb. Ecol. 56: 292–305.

Wu M, Scott AJ. 2012. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics. 28: 1033–1034.

Wu S, Zhu Z, Fu L, Niu B, Li W. 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics. 12: 444.

Wu YW, Tang YH, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome. 2: 26.

Yarza P, Richter M, Peplies J, Euzeby J, Amann R, Schleifer KH, Ludwig W, Glöckner FO, Rosselló-Móra R. 2008. The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. Syst. Appl. Microbiol. 31: 241–250.

Yarza P, Ludwig W, Euzéby J, Amann R, Schleifer KH, Glöckner FO, Rosselló-Móra R. 2010. Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. Syst. Appl. Microbiol. 33: 291–299.

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured Bacteria and Archaea using 16S rRNA gene sequences. Nat. Rev. Microbiol. 12: 635–645.

Yelton AP, Comolli LR, Justice NB, Castelle C, Denef VJ, Thomas BC, Banfield JF. 2013. Comparative genomics in acid mine drainage biofilm communities reveals metabolic and structural differentiation of co-occurring archaea. BMC Genomics. 14: 485.

Yuasa T, Takahashi O, Honda D, Mayama S. 2004. PCR primers for the amplification of the nuclear small subunit ribosomal DNA sequences from polycystine radiolarians. Japanese Journal of Protozoology. 37: 133–137.

Zerbino DR, Birney E.. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18: 821–829.

Zillig W, Kletzin A, Schleper C, Holz I, Janekovic D, Hain J, Lanzendörfer M, Kristjansson JK. 1993. Screening for *Sulfolobales*, their plasmids and their viruses in Icelandic solfataras. Syst. Appl. Microbiol. 16: 609–628.

Capítulo II

Diversidad genómica de rizobios de *Phaseolus* nativos de México

Introducción y antecedentes

Los rizobios son α-proteobacterias que son capaces de fijar nitrógeno atmosférico para sus plantas hospederas. El uso de rizobios en la agricultura permite ahorros de miles de millones de dólares ya que reduce el uso de fertilizantes químicos (Martínez-Romero, E., 2003). El proceso de fijación de nitrógeno ocurre en los nódulos de las plantas donde los rizobios reciben nutrientes y expresan a la enzima nitrogenasa que requiere de bajas concentraciones de oxígeno. Los nódulos son pequeñas estructuras especializadas en las raíces, y en muy pocos casos en los tallos, de plantas leguminosas como la alfalfa, la soya y el frijol.

La formación de los nódulos depende del apropiado reconocimiento molecular entre los rizobios y las plantas leguminosas. Los factores Nod son las señales moleculares que producen los rizobios en presencia de las plantas (Lerouge et al., 1990). Los factores Nod son lipooligosacáridos que contienen residuos de un azúcar, glucosamina, y diferentes sustituciones químicas (Denarie et al., 1996). El descubrimiento del factor Nod fue reportado en la revista Nature hace 25 años (Lerouge et al., 1990) y se designó como la molécula del año.

Estudios posteriores encontraron variantes del factor Nod producidas por diferentes rizobios. Las variantes en el factor Nod se deben a la presencia de sustituciones químicas tales como acetilaciones, fucosilaciones, sulfataciones, etc. Algunas variantes del factor Nod han podido correlacionarse en cierta medida con la especificidad por distintas plantas leguminosas (Denarié et al., 1996; Poupot et al., 1993; Poupot et al., 1995; Laeremans et al., 1999).

Los genes *nod* son necesarios para la producción de las enzimas que producen los factores Nod y están codificados en diferentes regiones de los genomas de los rizobios. En *Bradyrhizobium* los genes *nod* se encuentran codificados en los cromosomas principales dentro de las regiones conocidas como islas simbióticas. En cambio, en *Rhizobium* los genes *nod* residen en los llamados plásmidos simbióticos que pueden transferirse entre células y por tanto tienen una evolución genómica más dinámica.

Las vías de señalización que se desencadenan en las plantas después del reconocimiento del factor Nod fueron dilucidadas a lo largo de más de una década de investigación (Geurts y Bisseling, 2001; Madsen et al., 2003). Parte de las vías de señalización también son usadas por las plantas para percibir las señales moleculares de hongos micorrízicos que pueden llegar a colonizar cerca de un 90% de las plantas (Parniske, 2008). La molécula señal producida por los hongos micorrízicos es estructuralmente semejante al factor Nod (Maillet et al., 2011).

Phaseolus vulgaris, el frijol común, ha sido por muchos años la especie modelo para estudiar interacciones moleculares entre *Rhizobium* y *Phaseolus. Phaseolus vulgaris* es además la especie de frijol más utilizada para la alimentación humana en el mundo. En México, las principales especies de *Rhizobium* que nodulan y fijan nitrógeno con el frijol común son *R. phaseoli* y *R. etli* (Martínez-Romero et al., 1985; Segovia et al., 1991; Segovia et al., 1993; Souza et al., 1994; Caballero-Mellado y Martínez-Romero, 1999; Aguilar et al., 2004).

Rhizobium phaseoli y *R. etli* producen factores Nod que no están sulfatados pero sí fucosilados (Poupot et al., 1995; Cárdenas et al., 1995). Sin embargo, el frijol común es una planta promiscua que puede llegar a nodular con rizobios tales como *R. tropici* y *R. leucaenae* que presentan rangos amplios de nodulación y que producen factores Nod sulfatados (Poupot et al., 1993; Folch-Mallol et al., 1996; Torres-Tejerizo et al., 2011).

México cuenta con la mayor diversidad de especies de *Phaseolus* silvestres. El género *Phaseolus* se distribuye a lo largo del continente americano y contiene cerca de 70 especies (Freytag y Debouck, 2002; Delgado-Salinas et al., 1999; Delgado-Salinas et al., 2006). En base a estudios moleculares, el género *Phaseolus* comprende ocho clados filogenéticos nombramos en base a especies representativas (*Phaseolus filiformis*, *P. leptostachyus*, *P. lunatus*, *P. pauciflorus*, *P. pedicellatus*, *P. polystachius*, *P. tuerckheimii* y *P. vulgaris*) además de especies con relaciones filogenéticas que no se han podido resolver tales como *P. microcarpus*. El clado de *Phaseolus vulgaris* está conformado por las especies *P. vulgaris*, *P. coccineus*, *P. dumosus*, *P. costaricensis* y *P. albescens* (Delgado-Salinas et al., 1999; Delgado-Salinas et al., 2006).

Los simbiontes de *P. vulgaris* han sido estudiados detalladamente mientras que los simbiontes de otras especies han recibido menor atención. Se conoce que muchas plantas leguminosas son noduladas por diversas especies de *Bradyrhizobium* (Martínez-Romero, 2009), y esto pudiera ser cierto también para las distintas especies de *Phaseolus*. En contraste con *P. vulgaris*, variedades cultivadas de *P. acutifolius* y *P. lunatus* junto con plantas silvestres de *P. parvulus* y *P. pauciflorus* son noduladas por *Bradyrhizobium.* (Somasegaran et al., 1991; Parker, 2002; Ormeño-Orrillo et al., 2006).

Se ha reportado que variedades cultivadas de *P. lunatus* pueden presentar simbiosis con diferentes rizobios (incluso de género distinto) aunque en baja proporción (Ormeño-Orrillo et al., 2006). Por tanto, las plantas de *Phaseolus* pueden llegar a presentar cierto grado de laxitud durante la selección de sus simbiontes. En ausencia de los simbiontes principales, las plantas de *Phaseolus* llegan a seleccionar otros simbiontes presentes en el suelo.

En el laboratorio de la Dra. Esperanza Martínez se han analizado los rizobios de los nódulos de diversas leguminosas de México, con especial énfasis en *P. vulgaris.* Propuse analizar la diversidad genética de rizobios de otras especies de *Phaseolus* nativos de México. La propuesta tuvo su fundamento en la tesis de licenciatura que realicé en el laboratorio del Dr. Federico Sánchez del Instituto de Biotecnología de la UNAM donde analicé filogenias de *Phaseolus* y con quien se estableció una colaboración para la realización del proyecto.

Planteamiento del problema

La diversidad genética de rizobios no se ha analizado exhaustivamente para las diversas especies de *Phaseolus* nativos de México.

Hipótesis

La nodulación con *Bradyrhizobium* es común en *Phaseolus* y la nodulación preferencial con *Rhizobium* ocurre en un grupo de especies relacionadas filogenéticamente con *P. vulgaris*.

Objetivos

Identificar molecularmente a los rizobios que establecen simbiosis con especies representativas de los grupos filogenéticos del género *Phaseolus*.

Específicamente, identificar molecularmente a los rizobios que establecen simbiosis con todas las especies reportadas del grupo filogenético de *Phaseolus vulgaris*.

Obtener cultivos de rizobios aislados de plantas de campo y realizar la secuenciación de sus genomas para comprender mejor las bases genéticas de su potencial simbiótico.

Analizar la posición filogenética y filogenómica de las cepas de rizobios recuperados.

Revisar la taxonomía de *Rhizobium* y otros géneros de la familia *Rhizobiaceae* mediante análisis comparativos de secuencias genómicas.

Resultados y discusión

Artículo:

Servín-Garcidueñas LE, Zayas-Del Moral A, Ormeño-Orrillo E, Rogel MA, Delgado-Salinas A, Sánchez F, Martínez-Romero E. 2014. Symbiont shift towards *Rhizobium* nodulation in a group of phylogenetically related *Phaseolus* species. Mol. Phylogenet. Evol. 79: 1-11.

Encontramos que las especies de *Phaseolus* cercanas a *P. vulgaris* establecen simbiosis preferentemente con *Rhizobium* independientemente de si son variedades silvestres o cultivadas. También observamos que la mayoría de las especies de *Phaseolus* forman nódulos con *Bradyrhizobium* como ocurre en otras leguminosas de la tribu *Phaseoleae*. Se obtuvo la secuencia del genoma parcial de la cepa *Bradyrhizobium* sp. CCGE-LA001 que se aisló de nódulos de campo de *P. microcarpus.* El genoma de *Bradyrhizobium* sp. CCGE-LA01 es el primero que se secuencia para un *Bradyrhizobium* que establece simbiosis con *Phaseolus*. El genoma reveló genes simbióticos con secuencias novedosas y se encontró que codifica para un repertorio amplio de genes *nod* involucrados en la adición de distintas modificaciones químicas. En este artículo también se presenta una revisión de las relaciones filogenéticas de las especies pertenecientes al clado *P. vulgaris*. Los resultados completos se describen en el artículo.

Molecular Phylogenetics and Evolution 79 (2014) 1–11

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Symbiont shift towards *Rhizobium* nodulation in a group of phylogenetically related *Phaseolus* species

Luis E. Servín-Garcidueñas [a], Alejandra Zayas-Del Moral [b], Ernesto Ormeño-Orrillo [a], Marco A. Rogel [a], Alfonso Delgado-Salinas [c], Federico Sánchez [b], Esperanza Martínez-Romero [a,*]

[a] Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, Mexico
[b] Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, Mexico
[c] Instituto de Biología, Universidad Nacional Autónoma de México, México D.F. 04510, Mexico

ARTICLE INFO

Article history:
Received 21 January 2014
Revised 5 June 2014
Accepted 6 June 2014
Available online 18 June 2014

Keywords:
Phaseolus
Rhizobium
Bradyrhizobium
Symbiosis
Nodulation

ABSTRACT

Bean plants from the *Phaseolus* genus are widely consumed and represent a nitrogen source for human nutrition. They provide biological fertilization by establishing root nodule symbiosis with nitrogen-fixing bacteria. To establish a successful interaction, bean plants and their symbiotic bacteria need to synchronize a proper molecular crosstalk. Within the *Phaseolus* genus, *P. vulgaris* has been the prominent species to study nodulation with *Rhizobium* symbionts. However the *Phaseolus* genus comprises diverse species whose symbionts have not been analyzed. Here we identified and studied nodule bacteria from representative *Phaseolus* species not previously analyzed and from all the described wild species related to *P. vulgaris*. We found *Bradyrhizobium* in nodules from most species representing all *Phaseolus* clades except in five phylogenetically related species from the *P. vulgaris* clade. Therefore we propose that *Bradyrhizobium* nodulation is common in *Phaseolus* and that there was a symbiont preference shift to *Rhizobium* nodulation in few related species. This work sets the basis to further study the genetic basis of this symbiont substitution.

1. Introduction

The *Phaseolus* genus is estimated to contain around 70 species and the majority of them are distributed from Northern Mexico to Central America (Freytag and Debouck, 2002; Delgado-Salinas et al., 2006); most of the wild species diversity is found in this area suggesting that the origin for the whole *Phaseolus* genus may reside within it. In Mexico, wild *Phaseolus* species thrive from rain forests to deserts and from lands close to the ocean to forests surrounding high mountains and volcanoes. *Phaseolus* diversity should be conserved as "we live in a period of rapid loss of biodiversity" (Ley et al., 2008). It is unfortunate that some wild species like *Phaseolus albescens* and *P. rotundatus* are endangered (Ramírez-Delgadillo and Delgado-Salinas, 1999; Salcedo-Castaño et al., 2009). The diversity of the natural *Phaseolus* symbionts may be at risk if their hosts are threatened.

Based on molecular studies, the genus *Phaseolus* comprises eight phylogenetic clades named after representative species from each group (*Phaseolus filiformis*, *P. leptostachyus*, *P. lunatus*, *P. pauciflorus*, *P. pedicellatus*, *P. polystachius*, *P. tuerckheimii* and *P. vulgaris*) and species with unclear phylogenetic relationships like *P. microcarpus*. The *Phaseolus vulgaris* clade is composed of seven species. In general, *P. vulgaris*, *P. coccineus*, *P. dumosus*, *P. costaricensis* and *P. albescens* prefer more temperate and cooler conditions and are commonly found in mountain ranges, valleys and volcanic hills from Mesoamerica and Central America. *P. vulgaris* is the only species from this clade that is naturally distributed in South America and its northern range even extends to Canada and the United States of America. The two outlier species, *P. parvifolius* and *P. acutifolius* are resistant to drought and warm conditions that are commonly found where these species thrive (Delgado-Salinas et al., 1999; Delgado-Salinas et al., 2006). Besides *P. vulgaris*, only a few species (*P. lunatus*, *P. coccineus*, *P. dumosus* and *P. acutifolius*) were domesticated for human consumption.

Knowledge of the specific symbionts for different *Phaseolus* species may be important for farmers as nitrogen-fixing bacteria

* Corresponding author. Address: Centro de Ciencias Genómicas, UNAM. Av. Universidad SN, Chamilpa, CP 62210, Cuernavaca, Morelos, Mexico. Fax: +52 777 3175581.
E-mail addresses: lservin@lcg.unam.mx (L.E. Servín-Garcidueñas), alezayas@ibt.unam.mx (A. Zayas-Del Moral), eormeno@ccg.unam.mx (E. Ormeño-Orrillo), marogel@ccg.unam.mx (M.A. Rogel), adelgado@ibunam2.ibiologia.unam.mx (A. Delgado-Salinas), federico@ibt.unam.mx (F. Sánchez), emartine@ccg.unam.mx (E. Martínez-Romero).

http://dx.doi.org/10.1016/j.ympev.2014.06.006
1055-7903/

can be used as natural fertilizers to increase crop yields. Wild and domesticated *P. vulgaris* beans are nodulated by *Rhizobium etli* and related species such as *Rhizobium phaseoli* (López-Guerrero et al., 2012), both in its sites of origin (Martínez-Romero et al., 1985; Segovia et al., 1991; Segovia et al., 1993; Souza et al., 1994; Caballero-Mellado and Martinez-Romero, 1999; Aguilar et al., 2004) and in some regions where beans have been introduced (Herrera-Cervera et al., 1999; Grange and Hungria, 2004; Tamimi and Young, 2004; Aserse et al., 2012). This is an indication of *Phaseolus* symbiont selection irrespective of the soil origin. Furthermore, when *P. vulgaris* was used as a trap plant with Los Tuxtlas forest soil that is rich in *Bradyrhizobium*, it did not trap any bradyrhizobial strain (Ormeño-Orrillo et al., 2012). Domesticated *P. coccineus* plants from milpa plots in Mexico have been reported to be nodulated by *Rhizobium gallicum* (Silva et al., 2003), which is uncommonly found in *P. vulgaris* nodules. The genome of *Rhizobium* strain CCGE 510 obtained from field nodules of *P. albescens* was recently sequenced (Servín-Garcidueñas et al., 2012).

Even though incomplete information still exists, most legumes reported in the tribe Phaseoleae (Leguminosae) are nodulated by *Bradyrhizobium* (Martínez-Romero, 2009) and this could also be the case in *Phaseolus*. In contrast to *P. vulgaris*, *P. acutifolius* and *P. lunatus* cultivars together with wild *P. parvulus* and *P. pauciflorus* are nodulated by *Bradyrhizobium* species (Somasegaran et al., 1991; Parker, 2002; Ormeño-Orrillo et al., 2006). *Bradyrhizobium paxllaeri* and *Bradyrhizobium icense* were recently characterized as novel bradyrhizobial species isolated from nodules of *P. lunatus* from Peru (Durán et al., 2014).

It seems that "symbiosis with *Bradyrhizobium* represents the ancestral condition in the genus *Phaseolus* and that utilization of *Rhizobium* is a recent innovation that may be restricted to *P. vulgaris* and some of its close relatives, such as *P. coccineus*" (Parker, 2002). To support this assumption, other species should be analyzed as only a few *Phaseolus* species had been sampled for their nodule bacteria. We hypothesized that wild and domesticated *P. coccineus* and *P. dumosus* together with the wild species *P. costaricensis* and *P. albescens* may be naturally nodulated by *Rhizobium* strains due to their close phylogenetic affinities with *P. vulgaris*. The aim of this work was to study the symbiotic nodule bacteria from wild *Phaseolus* plants from representative species belonging to different clades and from all of the seven species that group within the *P. vulgaris* clade. We found that only species closely related to *P. vulgaris* were nodulated by *Rhizobium*.

Nod factors are produced by rhizobia during the early development of nodules upon perception of flavonoid molecules secreted by legume roots. The structure of Nod factors can vary between species, and chemical substitutions are commonly added that may affect legume specificity (D'Haeze and Holsters, 2002; Geurts and Bisseling, 2002). We compared potential *nod* genes encoded in the genomes of *Rhizobium* sp. CCGE 510 and *Bradyrhizobium* sp. CCGE-LA001 to predict Nod factors structures from two phylogenetically distinct strains isolated from field nodules of wild *Phaseolus* species.

2. Materials and methods

2.1. Wild Phaseolus nodule and soil samplings

P. vulgaris and *P. microcarpus* field nodules were collected from the town of Oaxtepec 1500 masl in the state of Morelos, Mexico. *P. vulgaris* and *P. leptostachyus* field-collected nodules were obtained in a pine and oak forest located in Cuernavaca City. *P. albescens* nodules were collected from the roots of a large vine (~12 m) growing in a mountain forest in the state of Jalisco, Municipio de Tecalitlán, Sierra del Halo. *P. coccineus* field collected nodules were recovered from two plants growing in the pine forest of Tetela del Bosque close to Cuernavaca City. Soils from sampling areas contained both rhizobial and bradyrhizobial symbionts and were mixed and used for subsequent trapping experiments.

2.2. Trapping experiments

Seeds were collected either from wild plants during field expeditions or provided by colleagues. Seeds were surface sterilized with serial washes of 70% ethanol, 1.5% sodium hypochlorite and several rinses with distilled water as described before (López-López et al., 2010). Surface sterilization was checked by rubbing the seeds on YEM medium (Vincent, 1970) in plates. After germination, seedlings were directly transferred to soils collected from our sampling areas. Plants were irrigated with sterile water every three days and were maintained in a temperature controlled room at 28 °C with a twelve hour light cycle. Nodules were collected from secondary roots after one month.

2.3. Phaseolus species identification

DNA was extracted from all the collected *Phaseolus* plants and seeds and the internal transcribed spacer marker (ITS) was amplified as described previously (Delgado-Salinas et al., 1999). DNA was extracted from leaves or nodules with the Genomic DNA Purification Kit (Fermentas) using a modified protocol. Briefly, leaves were chopped and mixed with 400 µl of lysis solution, incubated at 65 °C for five min and then 600 ml of chloroform was added. The mixed sample was then centrifuged at 11000*g* for 3 min. The supernatant was then purified using the High Pure PCR Product Purification Kit (Roche Applied Science). The amplified ITS regions were Sanger sequenced at Macrogen, Korea. *Phaseolus* ITS sequences of around 750 bp were searched against the GenBank database using BLASTN in order to validate their species identity and were used for phylogenetic analyses.

2.4. Bacterial isolation and identification

P. albescens field nodules were processed 4 days after collection and for all the other wild plants their nodules were processed within the same sampling day. In the laboratory, bacteria were recovered after nodule surface sterilization with 70% ethanol and then with 1.5% sodium hypochlorite followed by serial washes with sterile water as described previously (Ormeño-Orrillo et al., 2012). Surface sterilization was checked by rubbing the nodules on YEM medium (Vincent, 1970) in plates. Nodules were then crushed and their extracts were rubbed on YEM medium and on PY medium (Toledo et al., 2003). Plates were incubated at 28 °C for one week. One bacterial isolate was retained from each nodule. The same processing was performed for nodules retrieved from plant assays. DNA was extracted from each isolate and PCR amplifications were performed with 16S rRNA gene primers as described (Weisburg et al., 1991). PCR products were purified and sequenced. 16S rRNA gene sequences of 650–700 bp were searched against the GenBank database using BLASTN and their genus identities were verified.

2.5. Phylogenetic analyses

Sequences from related species were retrieved from GenBank. *rpoB*, *nodB* and *nifH* marker gene sequences were obtained from *R. phaseoli* strain ATCC 14482, *R. fabae* strain CCBAU 33202 and *R. pisi* strain DSM 30132 (Khamis et al., 2003; Silva et al., 2003; Zehr and McReynolds, 1989). Sequences were aligned with MUSCLE v3.8.31 (Edgar, 2004). Sequence alignments were inspected using MEGA5 software (Tamura et al., 2011). Phylogenetic

reconstructions were performed using MrBayes 3.2.1 (Ronquist et al., 2012). jModelTest (Posada, 2008) was used to find the model of evolution that best fit the data using the Akaike Information Criterion (Posada and Buckley, 2004). In all cases, the selected model was GTR + I + G. For Bayesian inferences, four independent runs were conducted for 5,000,000 generations. Each set was sampled every 100 generations with a burn-in of 25%. Standard deviations at the end of the runs were revised to have reached a value of 0.001.

2.6. Bacterial genome sequencing

The genome sequencing of *Rhizobium* sp. CCGE 510 was described previously (Servín-Garcidueñas et al., 2012). The plasmid content of *Rhizobium* sp. CCGE 510 was visualized by using the Eckhardt procedure (Eckhardt, 1978) to verify the genomic assembly of plasmids. Plasmids were observed in 0.7% agarose gels using plasmids of *R. etli* CFN 42 as molecular size references. The MAUVE genome alignment tool (Darling et al., 2010) was used to align symbiotic plasmid sequences from *Rhizobium* strains. DNA was extracted from *Bradyrhizobium* sp. CCGE-LA001 (maintained in the culture collection of the Ecological Genomics Department, CCG, UNAM) and sequenced with the Roche 454 GS-FLX Titanium technology from a mate-paired library with 3-kb inserts. Reads were assembled *de novo* using Newbler Assembler 2.3 (454 Life Science). Genome annotation was performed using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html).

2.7. Phylogenomic analyses

The genome assemblies of *Bradyrhizobium* sp. CCGE-LA001 and *Rhizobium* sp. CCGE 510 served as inputs to PhyloPhlAn to predict their evolutionary relationship with reference sequenced strains. Briefly, PhyloPhlAn is a new computational pipeline to assign microbial phylogeny and putative taxonomy by locating 400 conserved proteins in a given whole-genome sequence assembly (Segata et al., 2013). The pipeline performs individual protein alignments using MUSCLE v3.8.31 (Edgar, 2004) from the protein sets recovered from the input genomes. PhyloPhlAn then concatenates the most discriminative positions in each protein alignment into a single long sequence to reconstruct a phylogenetic tree using FastTree (Price et al., 2010). DNA–DNA hybridization (DDH) values were computed using the Genome-to-Genome Distance Calculator (Auch et al., 2010a,b) version 2.0 (Meier-Kolthoff et al., 2013). Average nucleotide identity (ANI) values were calculated as previously proposed (Goris et al., 2007) using the ANI calculator from the Kostas lab (http://enve-omics.ce.gatech.edu/ani/). *nod* gene sequences were recovered from genome assemblies by analyzing the *de novo* annotations and by performing searches against the GenBank database using BLASTN and BLASTP.

2.8. Plant nodulation

Rhizobium sp. CCGE 510 isolated from wild *P. albescens* nodules was tested in plant nodulation assays after colony purification. Nodulation tests were performed in triplicate using wild *P. vulgaris* and *P. albescens* seedlings in agar with Fahraeus nutrient solution (Fåhraeus, 1957) with no added nitrogen in flasks.

2.9. Light microscopy

P. vulgaris root nodules induced by *Rhizobium* sp. CCGE 510 and *R. etli* CE3 were quickly collected, fixed, embedded and polymerized in Epon Resin. Nodule sections were mounted on a slide with a coverslip for observation under a Zeiss Axiovert 200 M microscope. Images were processed using Adobe Photoshop 7.0 software (Adobe Systems Inc., Mountain View, CA, USA). Nodule sections (3 μm) were obtained and root nodule major features, including the presence of bacteroids, were determined by the Toluidine Blue O staining method described before (Chieco et al., 1993).

2.10. Electron microscopy

Nodule samples were cut in 1–2-mm sections, fixed, dehydrated (10–100% ethanol series), and pre-embedded in propylene oxide/epoxy resin mixtures (2:1, 1:1, and 1:2) (London Resin Company Limited, London) for several hours before embedding in 100% epoxy resin for 3 h. Nodules pieces in gelatin capsules (Electron Microscopy Sciences, Fort Washington, PA, USA) were ultrathin sectioned (70 nm) with a UCT-R ultramicrotome (Leica, Wetzlar, Germany) and mounted on nickel grids. They were then stained with 2% aqueous uranyl acetate and visualized under a Zeiss EM900 transmission electron microscope at 80 kV. Micrographs were taken with a digital CCD DualVision 300 W camera (Gatan, Inc., Pleasanton, CA, USA).

2.11. Data deposition

The *Phaseolus* ITS sequences reported in this paper were deposited in the GenBank database with accession numbers KF943718 to KF943753. *P. parvifolius* ITS sequences with accession numbers FJ853398 to FJ853403 and one *P. costaricensis* ITS sequence with accession number FJ853397 were previously generated by our group (unpublished). Bacterial 16S rRNA gene sequences were submitted to GenBank under accession numbers KF943754 to KF943817. Partial sequences from marker genes were from rhizobial species: *R. phaseoli* ATCC 14482 *nifH* (HQ670652) and *rpoB* (HQ670651); *R. fabae* CCBAU *33202 nifH* (HQ670650), *nodB* (HQ670649) and *rpoB* (HQ670648) and *R. pisi* DSM *30132 nodB* (HQ670647) and *rpoB* (HQ670646). The first version of the draft genome sequence from *Bradyrhizobium* sp. CCGE-LA001 has been deposited at DDBJ/EMBL/GenBank under the accession number AMCQ00000000.

3. Results

3.1. Phylogenetic analyses of bean species from the Phaseolus vulgaris clade

The *P. vulgaris* clade currently contains seven bean species. *P. acutifolius* and *P. parvifolius* are phylogenetically related and are the outlier species of the *P. vulgaris* clade (Fig. 1A). The *Phaseolus* molecular phylogeny has been largely based on ITS sequences but few *P. parvifolius* sequences were available. A phylogenetic analysis based on an increased set of ITS sequences clusters all *P. parvifolius* sequences in a different group apart from *P. acutifolius* (Fig. 1A). The other five species from the *P. vulgaris* clade, *P. albescens*, *P. coccineus*, *P. dumosus*, *P. costaricensis* and *P. vulgaris*, are phylogenetically closely related (Fig. 1A).

3.2. Native Phaseolus symbionts

To survey nodule bacteria from distinct *Phaseolus* species, field nodules were collected from a wild plant of *P. coccineus* growing in a temperate pine forest and from a large pine-climber vine of the endangered *P. albescens* species that was growing in a humid forest in Western Mexico. The field nodules from both species contained *Rhizobium* strains (Suppl. Table 1). *Bradyrhizobium* isolates were identified from field nodules of *P. microcarpus* and *P. leptostachyus* species that had not been studied before.

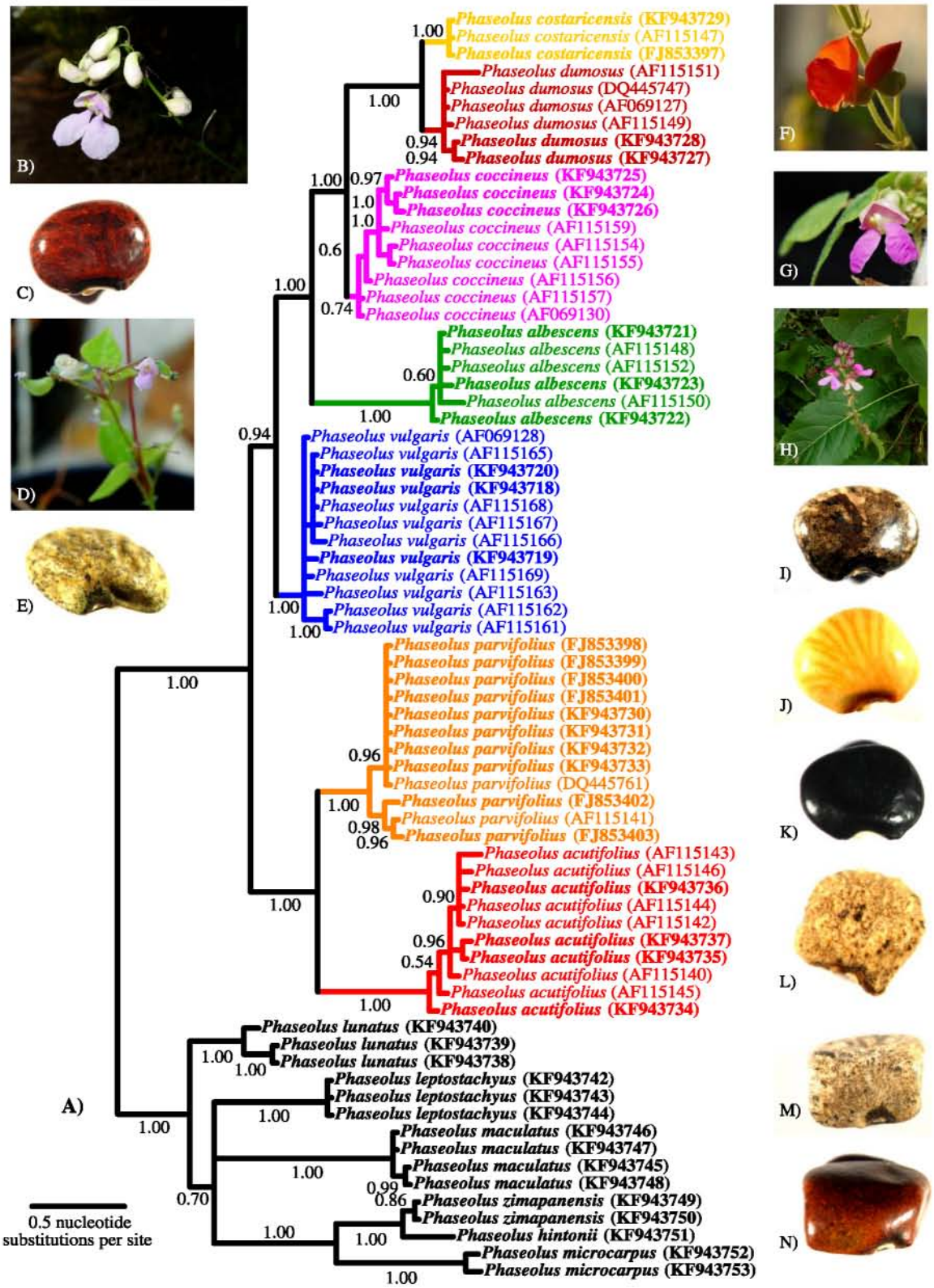


Fig. 1. (A) Phylogenetic relationships between the species from the *Phaseolus vulgaris* clade. Species from the *P. vulgaris* clade are shown with different colors. Sequences obtained by this study are shown in bold. GenBank sequence accession numbers are indicated in parenthesis. The Bayesian phylogenetic reconstruction is based on an alignment of partial ITS sequences and contains 746 characters. The support values for branches are shown as Bayesian posterior probabilities. The scale bar represents the average number of nucleotide replacements per site. (B–N) Representative *Phaseolus* species used in this study. Flowers and seed from the endangered *P. albescens* (B and C). Plant and seed from *P. microcarpus* (D and E). *P. coccineus* with characteristic red flowers (F). *P. vulgaris* pink flower (G). *P. leptostachyus* with a bunch of flowers (H). Seeds from wild beans are shown for representative *P. coccineus* (I), *P. lunatus* (J and K), *P. filiformis* (L), *P. acutifolius* (M) and *P. maculatus* (N).

3.3. Bradyrhizobia and Rhizobia from Phaseolus coexist in soil samples collected in the field

The *Bradyrhizobium* isolates from nodules of *P. leptostachyus* and *P. microcarpus* were recovered from natural fields where *P. vulgaris* was also present. At one location with deciduous forest we collected a plant of *P. microcarpus* whose nodule bacteria were *Bradyrhizobium* isolates and a few steps away we collected *P. vulgaris* plants and, as expected, their nodule bacteria were *Rhizobium* (Suppl. Table 1). At a second location with temperate pine forest we collected *P. leptostachyus* plants whose nodule bacteria were *Bradyrhizobium* isolates and we also found *P. vulgaris* plants a few meters away and their nodule bacteria were *Rhizobium* (Suppl. Table 1). The *P. vulgaris* plants with their associated rhizobial symbionts served as indicators for the presence of *Rhizobium* in the soil. Soil samples were collected from these locations and were used for trap plant assays as they contained both bradyrhizobia and rhizobia capable of nodulating *Phaseolus* species.

3.4. Insights into other Phaseolus symbionts

To further survey *Phaseolus* nodule bacterial species, isolates from representative species were recovered with trap plant assays using mainly wild seeds (Suppl. Table 1). Trapped nodule bacteria were not considered as natural symbionts but were used to assess the symbiont preference of some *Phaseolus* species. Bacterial isolates from *P. microcarpus* and *P. leptostachyus* trapping plants were identified as *Bradyrhizobium*, as determined previously from field nodules from both species. Two wild *P. lunatus* seed accessions native to Mexico had bradyrhizobia as nodule bacteria. From the *P. tuerckheimii* clade, the tiny and delicate *P. zimapanensis* and *P. hintonii* species were tested and *Bradyrhizobium* isolates were recovered from their nodules. From the *P. polystachius* clade we tested *P. maculatus* that is a species characterized for having robust seeds and is used for animal foraging. It is unclear if *P. maculatus* was domesticated by humans in ancient times. Four *P. maculatus* seed accessions were tested and their nodules were all occupied by *Bradyrhizobium*. Finally, nodules from the distinctive species *P. filiformis* that thrive in desert areas between rocks and sand dunes were occupied by *Bradyrhizobium*. A phylogenetic analysis based on partial 16S ribosomal rRNA sequences shows general phylogenetic affiliations of bradyrhizobial isolates with *Bradyrhizobium* reference strains (Suppl. Fig. 1).

In contrast, all trap plants from *P. vulgaris* close relatives (*P. albescens*, *P. coccineus*, *P. dumosus* and *P. costaricensis*) were nodulated by *Rhizobium* strains. Importantly, the outlier species from the *P. vulgaris* clade, *P. parvifolius* and *P. acutifolius*, were nodulated by *Bradyrhizobium* and thus helped define the boundaries of the *Rhizobium* preference shift (Fig. 2).

3.5. Bradyrhizobium sp. CCGE-LA001 genome-based phylogenetic analyses

We sequenced the genome of a bradyrhizobial strain from a wild *P. microcarpus* plant to provide a bradyrhizobial genome from a wild and poorly studied *Phaseolus* species. *P. microcarpus* represents an enigmatic species in terms of its phylogenetic position and its nodule bacteria may also carry particular symbiosis determinants.

The whole genome of *Bradyrhizobium* sp. CCGE-LA001 was sequenced with a light coverage (~6.9X) and was estimated to be 7.39 Mbp with a G + C content of 63.4% encoding 8,606 predicted open reading frames. The draft assembly produced 2807 contigs with a N_{50} size of 3.95 kb and 31 scaffolds containing 1929 contigs. The N_{50} size of the scaffolds was 505.47 kb. The produced draft assembly was useful for gene detection and genome comparisons. *Bradyrhizobium* sp. CCGE-LA001 16S rRNA gene had 99% sequence identity to the corresponding sequences of *B. japonicum* USDA 6 and *B. diazoefficiens* USDA110. A 16S ribosomal rRNA phylogenetic reconstruction shows that strain CCGE-LA001 is more closely related to *Bradyrhizobium daqingense* strain CCBAU 15774 (Suppl. Fig. 2). A concatenated phylogeny using *recA*, *glnII*, *gyrB* and *dnaK* genes provided increased resolution and confirmed the placement of strain CCGE-LA001 in the vicinity of *Bradyrhizobium daqingense* but seemingly representing a not yet described *Bradyrhizobium* species (Fig. 3A). Further, a phylogenomic approach clustered strain CCGE-LA001 within a group that includes the soybean isolates *B. japonicum* USDA 6 and *B. diazoefficiens* USDA 110. However strain CCGE-LA001 is resolved as a long branch without a close *Bradyrhizobium* sequenced strain (Fig. 3B). A comparison of strain CCGE-LA001 draft genome with *B. diazoefficiens* USDA 110 and *B. japonicum* USDA 6 complete genome sequences yielded average nucleotide identity (ANI) values of 89.30 ± 3.37% and 89.04 ± 3.34%, respectively. ANI values were under the 95–96% boundary for circumscribing prokaryotic species (Richter and Rosselló-Móra, 2009). *In silico* DNA–DNA hybridization (DDH) estimates were 35.70 ± 2.48% and 35.30 ± 2.48% between strain CCGE-LA001 and strains USDA 110 and USDA 6, also below the 70% proposed for species definition (Wayne et al., 1987; Tindall et al., 2010).

3.6. Predicted Nodulation factor structures from Rhizobium sp. CCGE 510 and Bradyrhizobium sp. CCGE-LA001 genomes

nodABC genes responsible for synthesis of the Nod factor core were located on the symbiotic plasmid of *Rhizobium* sp. CCGE 510. Gene products responsible for acetylation (NolL), glycosylation (NodZ, NolK), carbamoylation (NolO) and methylation (NodS) of the Nod factor structure were predicted. Fucose decorations are added on the reducing end by the NodZ fucosyltransferase. NodL adds an *O*-acetyl group on the non-reducing terminal *N*-acetylglucosamine residue. NolL also adds acetyl groups but only on fucose decorations attached to the reducing end. NolK is involved in the biosynthesis of GDP-fucose, a substrate for NodZ. The NolO protein is responsible for carbamoylation on the non-reducing end. NodS is a *N*-methyltransferase responsible for adding substitutions on the non-reducing *N*-acetyl-D-glucosamine of the Nod factor. Gene products responsible for sulfation decorations were not found. Based on the *nod* genes found, *Rhizobium* sp. CCGE 510 Nod factor structure was predicted as a chitin backbone of *N*-acetylglucosamine residues, *N*-acylated at the non-reducing end and fucosylated at the reducing end with additional acetylation on the fucose residue. Further decorations include acetylation, carbamoylation and *N*-methylation at the non-reducing end (Suppl. Fig. 3A).

The *nod* genes required for nodulation from *Bradyrhizobium* sp. CCGE-LA001 were remarkably divergent with around 70% sequence identity with the corresponding sequences of other *Bradyrhizobium* strains. The Nod factor structure of *Bradyrhizobium* sp. CCGE-LA001 was also predicted as a chitin backbone of N-acetylglucosamine residues acylated, carbamoylated and N-methylated at the non-reducing end and acetyl fucosylated at the reducing end as *nodABC*, *nolO*, *nodS*, *nodZ*, *nolK* and *nolL* genes were encoded in its genome. Nevertheless, the fucose residue can be further modified with sulfate and methyl groups by the action of *noeE* and *noeI* gene products, respectively. A further difference could be the presence of an additional carbamoyl group at the non-reducing end added by *nodU* in the same position where the structure of CCGE 510 possesses an acetyl group (Suppl. Fig. 3B).

3.7. Microscopic analysis of inefficient P. vulgaris symbiosis with P. albescens nodule bacteria

We determined if phylogenetically closely related *Rhizobium* symbionts could establish effective symbioses with related

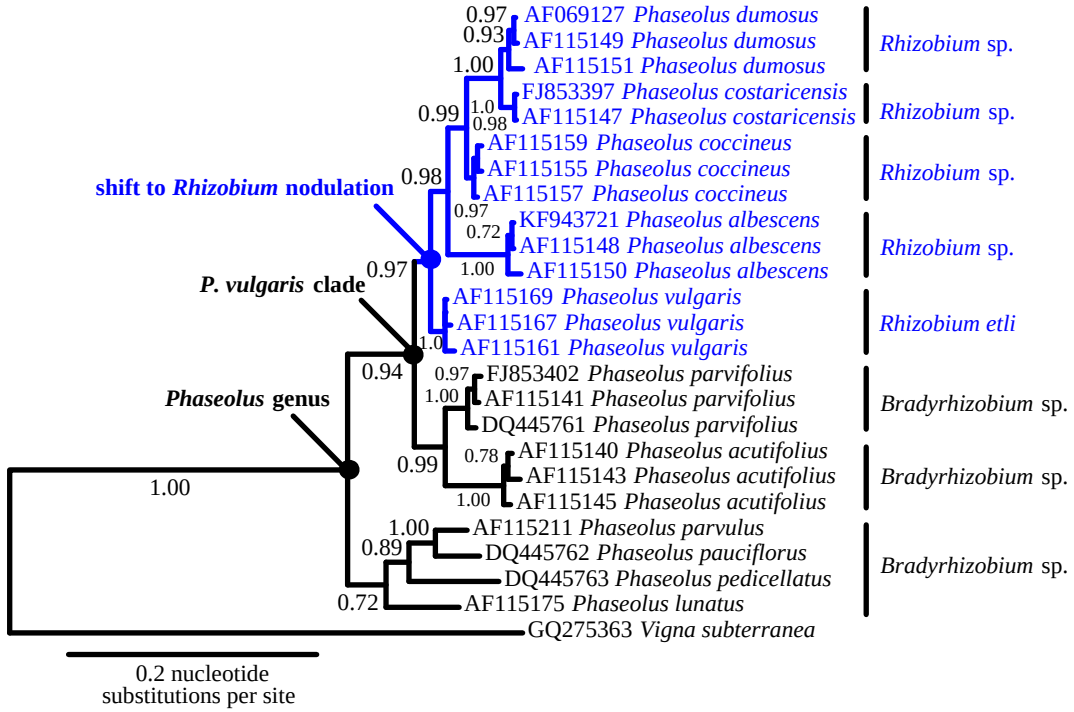


Fig. 2. Symbiont preference between the species from the *Phaseolus vulgaris* clade. The *Bradyrhizobium* to *Rhizobium* nodulation preference switch (shown in blue) occurred only in *P. vulgaris* phylogenetically related species, excluding *P. acutifolius* and *P. parvifolius*. All other wild species from different *Phaseolus* clades had been found to be nodulated by *Bradyrhizobium*. The Bayesian phylogenetic reconstruction is based on an alignment of partial ITS sequences and contains 735 characters. The support values for branches are shown as Bayesian posterior probabilities. The scale bar represents the average number of nucleotide replacements per site. Sequence accession numbers are indicated in parenthesis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Phaseolus species aside from their natural host. Thus, *P. vulgaris* nodulation and nitrogen fixation was tested with the representative and sequenced strain *Rhizobium* sp. CCGE 510 recovered from wild *P. albescens*. Nodules were mainly green and some red nodules turned to green rapidly. Nitrogen fixation was not detected and plants showed nitrogen deficiencies. Microscopically, there were nodule developmental alterations including an increased number of amyloplasts (Fig. 4F), bacteroids with different and aberrant morphologies (Fig. 4C and Fig. 4E), abnormal cell wall thickening in irregular infected cells and the presence of bacteria outside of uninfected cells (Fig. 4D and Fig. 4F). Unexpectedly, *P. albescens* symbionts did not establish an effective symbiosis in the phylogenetically-related and promiscuous *P. vulgaris* host.

3.8. Rhizobium sp. CCGE 510 genome-based phylogenetic analyses

We performed phylogenetic analyses of *Rhizobium* sp. strain CCGE 510 in order to better reveal its position in relation to *P. vulgaris* and other legume rhizobia. A 16S ribosomal rRNA phylogeny located strain CCGE 510 tightly clustered within an unresolved group of *Rhizobium leguminosarum* related strains including *R. leguminosarum* 3841 and *R. laguerreae* FB206 (Suppl. Fig. 4). 16S rRNA gene from strain CCGE 510 had 99% to 100% sequence identity with the corresponding sequences from *R. leguminosarum* allied strains. A concatenated phylogeny using partial *atpD*, *recA* and *rpoB* genes provided increased resolution and confirmed the placement of strain CCGE 510 closer to *R. leguminosarum* related strains than to *R. etli* or *R. phaseoli* strains (Fig. 5A). Furthermore, a phylogenomic approach including the available genomes from several *Rhizobium* strains placed strain CCGE 510 within a group including *R. leguminosarum* 3841, *Rhizobium* sp. WSM2304 and *Rhizobium* sp. WSM1325 (Fig. 5B). Genome comparisons between CCGE 510 and 3841 and WSM2304 strains revealed ANI values of 90.30 ± 3.14% and 91.22 ± 2.94%, respectively. The ANI values were below the 95–96% boundary for circumscribing prokaryotic species (Richter and Rosselló-Móra, 2009). Accordingly, *in silico* DDH estimates were 37.10 ± 2.49% and 40.40 ± 2.51% between the respective strains. DDH estimates were also below the 70% proposed for species definition (Wayne et al., 1987; Tindall et al., 2010).

The *nifH* and *nodB* genes from strain CCGE 510 are phylogenetically related to those from symbiovar phaseoli (Suppl. Fig. 5A and B). A phylogenetic analysis of the *teuB* gene involved in the specific uptake of *Phaseolus* bean-exudate compounds (Rosenblueth et al., 1998) showed that the *teuB* gene from *Rhizobium* sp. CCGE 510 is closer to that from *R. tropici* CFN 299 than to the corresponding gene in *R. etli*. (Suppl. Fig. 5C). Strain CCGE 510 possesses four plasmids and the symbiotic genes are contained in symbiotic plasmid c (Suppl. Fig. 6). Genome alignments between the complete symbiotic plasmid sequences from *R. etli* CFN 42, *R. phaseoli* CIAT 652 and the draft symbiotic plasmid sequence from strain CCGE 510 revealed that only certain genomic blocks are conserved (Suppl. Fig. 7). Sequence comparisons between the symbiotic plasmid of the CCGE 510 strain with the corresponding symbiotic counterparts of CFN 42 and CIAT 652 strains revealed ANI values of 94.51 ± 2.08% and 94.38 ± 2.16%), respectively. DDH estimates were 46.70 ± 2.58% and 46.10 ± 2.57% between the same respective species.

4. Discussion

4.1. Bradyrhizobium as dominant Phaseolus nodule bacteria

Phaseolus nodule isolates from most species corresponding to all *Phaseolus* clades were bradyrhizobia. Our data suggest an ancestral *Phaseolus* preference for *Bradyrhizobium*.

Native bradyrhizobial strains were identified from field plants of *P. microcarpus* and *P. leptostachyus* that had not been studied before. The genome sequence from *P. microcarpus* strain

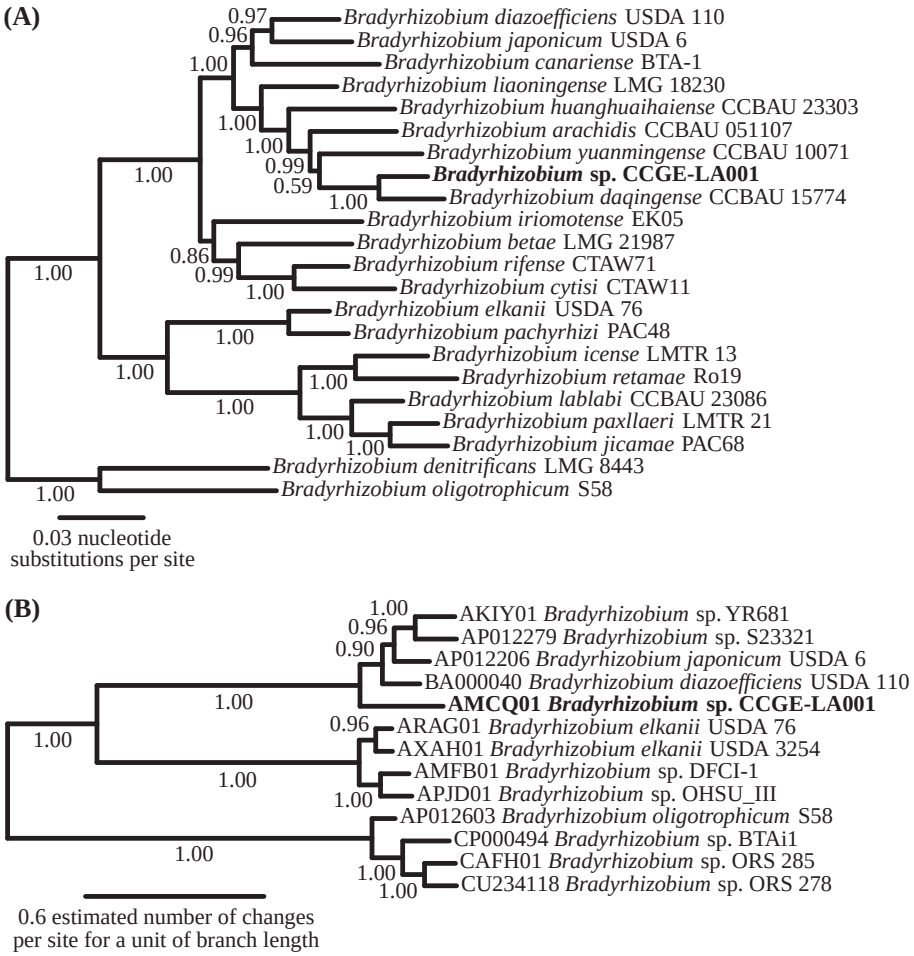


Fig. 3. (A) Topology of a multilocus sequence analysis of *Bradyrhizobium* reference strains using the nuclear marker genes *recA*, *gyrB*, *glnII* and *dnaK*. Partial *recA* (443 bp), *gyrB* (591 bp), *glnII* (524 bp) and *dnaK* (236 bp) gene sequences were used. *Bradyhizobium* sp. CCGE-LA001 is shown in bold. The Bayesian phylogenetic reconstruction is based on a concatenated alignment of the marker gene sequences and contains 1794 characters. The support values for branches are shown as Bayesian posterior probabilities. The scale bar represents the average number of nucleotide replacements per site. (B) Topology of a phylogenomic analysis showing the predicted evolutionary relationship between *Bradyhizobium* sp. CCGE-LA001 and sequenced *Bradyrhizobium* strains. The tree was reconstructed with PhyloPhlAn using a multisequence alignment of 375 conserved proteins. PhyloPhlAn performs individual alignments from each protein set recovered from the bradyrhizobial input genomes. PhyloPhlAn then concatenates the most discriminative positions in each protein alignment into a single long sequence to reconstruct a phylogenetic tree using FastTree. The position of *Bradyrhizobium* sp. CCGE-LA001 is shown in bold in the tree. Accession numbers are indicated for all sequenced genomes. Numbers at the branch points represent SH-like local support values (based on 1000 resamples). The scale bar represents the estimated number of amino acid changes per site for a unit of branch length.

CCGE-LA001 revealed that the symbiosis genes were different to the corresponding genes from other bradyrhizobia. Phylogenetic and phylogenomic approaches together with ANI values and DDH estimates indicate that strain CCGE-LA001 may represent a novel *Bradyrhizobium* species.

Bradyrhizobium had been reported as the main nodule bacteria in domesticated *P. lunatus* cultivars from Peru and Brazil (Ormeño-Orrillo et al., 2006; Santos et al., 2011) and some isolates have been described as novel bradyrhizobial species (Durán et al., 2014). Only recently, cultivated *P. lunatus* from milpa plots in Mexico were analyzed for their nodule bacteria and the bradyrhizobia encountered were described as native to the rain forest of Los Tuxtlas (López-López et al., 2013). No symbionts had been identified from wild *P. lunatus* in Mexico before. In a very low proportion, *Rhizobium* and sinorhizobial strains have been detected in nodules of domesticated *P. lunatus* cultivars (Ormeño-Orillo et al., 2007). In the absence of the preferred symbiont the plant may trap fortuitous or promiscuous strains.

We were unable to retrieve nodule bacteria from *Phaseolus* species corresponding to the *P. pedicellatus* and *P. pauciflorus* clades but it has been reported that some wild species from these clades are nodulated by *Bradyrhizobium* in Northern Mexico (Parker, 2002).

Wild *P. acutifolius* and *P. parvifolius* were nodulated by *Bradyrhizobium* and thus we propose that based on their symbiont preference and their phylogenetic relationships, both species should be defined as a separate group apart from the species closely related to *P. vulgaris*. This notion supports Freytag and Debouck morphological classification that consider *P. acutifolius* and *P. parvifolius* in the Acutifolii section (Freytag and Debouck, 2002). Additionally, our phylogenetic analyses based on ITS sequences support defining *P. parvifolius* as a different species apart from *P. acutifolius* (Fig. 1), as previously validated by AFLPs and microsatellite markers (Muñoz et al., 2006; Blair et al., 2012). The knowledge of *Phaseolus* symbionts will serve to guide inoculant choice for farmers in agricultural fields.

4.2. Nodulation by Rhizobium instead of Bradyrhizobium

Wild seed accessions from *P. vulgaris*, *P. albescens*, *P. coccineus*, *P. dumosus* and *P. costaricensis* were nodulated by *Rhizobium*. The natural rhizobial symbionts from wild *P. coccineus* nodules support previous results using domesticated plants (Silva et al., 2003). *P. albescens*, *P. dumosus* and *P. costaricensis* naturally grow in temperate-humid forest with similar environmental conditions within

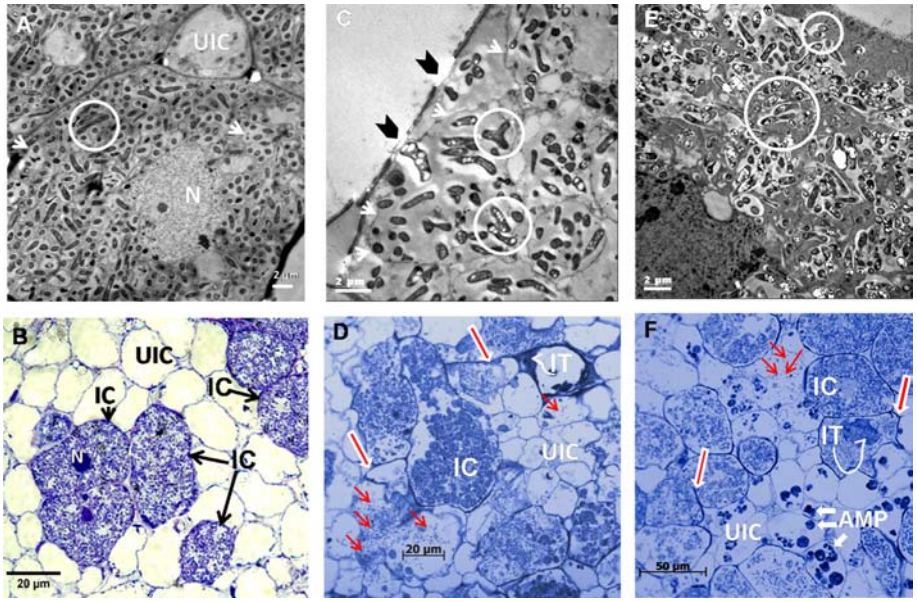


Fig. 4. Electron (A, C and E) and light (B, D, F) microscopy pictures depicting structural features from *Phaseolus vulgaris* root nodules inoculated with the *Rhizobium* sp. CCGE 510 strain originally isolated from *P. albescens* (C, D, E and F) and *R. etli* CE3 strain (A, B). Light micrographs were taken from 3 (μm) microtome slices stained with Toluidine Blue O. Infected cells (IC), uninfected cell (UIC), nucleus (N), infection thread (IT), amyloplast (AMP). White arrows show peribacteroid membranes, white circles show bacteroid different morphologies, red arrows show free bacteria present in uninfected cells, white arrows with red core show abnormal cell wall thickening in irregular infected cells.

restricted geographical areas. In contrast to these three species, *P. coccineus* distributes in Mesoamerica and Central America and *P. vulgaris* has an even wider distribution reaching South America. It is highly likely that the common ancestor to these species was subjected to an unknown pressure to shift their preference to establish symbiosis with *Rhizobium* strains.

Our sampling areas from distant Mesoamerican locations contained sympatric species from different *Phaseolus* clades and while *P. vulgaris* relatives were found to be nodulated by *Rhizobium*, the other bean species were found to establish symbiosis with bradyrhizobial strains. These data support the notion that different *Phaseolus* species "select" their symbionts from the coexisting rhizobia and bradyrhizobia in soil and that *P. vulgaris* related species are nodulated preferentially by *Rhizobium* strains at native sites (Parker 2002). Furthermore, we discarded the possibility that *Rhizobium* preference was determined by the domestication process as the symbiont shift was detected in wild species.

By testing nodulation and nitrogen fixation by a native *P. albescens* strain in *P. vulgaris* as a host we discovered a symbiotic misrecognition process that should be further analyzed. Unexpectedly, strain CCGE 510 formed ineffective green nodules on *P. vulgaris* that by being a promiscuous host (Hernández-Lucas et al., 1995; Martínez-Romero et al., 1985; Martínez-Romero 2003; Michiels et al., 1998) may form nitrogen fixing nodules with diverse rhizobia; therefore *Rhizobium* sp. CCGE 510 should correspond to a novel symbiovar. Green nodules formed by symbiotically defective strains have been described on *P. vulgaris* (Noel et al., 1984; Vandenbosch et al., 1985; Noel et al., 1986). Strain CCGE 510 is able to infect *P. vulgaris* and induce nodule formation, but later it is no longer found in infected cells and resembles saprophytic bacteria while the plant starts to accumulate starch vesicles. Similarly, *Bradyrhizobium* isolates from *P. pauciflorus* and *P. parvulus* produced inefficient nodules on *P. vulgaris* (Parker, 2002). Overall, results from phylogenetic and phylogenomic approaches together with ANI values and DDH estimates support strain CCGE 510 as a different *Rhizobium* species. The genome from strain CCGE 510 contains a chromosome related to *R. leguminosarum* strains and is capable of nodulating wild *P. albescens*, although inefficient in nodulating *P. vulgaris*. Strain CCGE 510 has recently been ascribed to the phaseoli-etli-leguminosarum lineage 6 (PEL6) based on multi-locus sequence analysis (Ribeiro et al., 2013).

4.3. The Phaseolus symbiont preference shift

The symbiont shift may have been facilitated by the transfer of Nod factor modification genes *nodZ* and *nolL* from *Bradyrhizobium* to *Rhizobium* (Ormeño-Orrillo et al., 2013). The genomes from *Rhizobium* sp. CCGE 510 and *Bradyrhizobium* sp. strain CCGE-LA001 potentially code *nod* genes involved in fucosylation (*nodZ* and *nolK*) and acetylation (*nolL*) of Nod factors (Suppl. Fig. 3). Fucosylated Nod factors have been found to be more efficient in their ability to induce bean nodules and were preferred by *P. vulgaris* plants (Laeremans et al., 1999).

Remarkably, the genome from *Bradyrhizobium* sp. strain CCGE-LA001 potentially codes for a large repertoire of *nod* genes involved in Nod factor modification. In contrast, *Rhizobium* sp. CCGE 510 has a more limited set of Nod factor decorating genes without genes required for sulfation in its genome. Nod factors produced by *R. etli* CFN 42 and *R. etli* CE3 are known to be non-sulfated (Poupot et al., 1995; Cárdenas et al., 1995). The addition of sulfate groups is performed by the sulfotransferases NodH and NoeE. We could not detect rhizobial *noeE* genes deposited in the GenBank database. We also could not find *nodH* genes in the genomes of *R. phaseoli* CIAT 652 and *R. phaseoli* CNPAF512. Other *P. vulgaris* nodulating strains like *R. tropici* CIAT 899, *R. leucaenae* CFN 299 and *Rhizobium* sp. LPU83 have broad-host ranges and produce sulfated Nod factors (Poupot et al., 1993; Folch-Mallol et al., 1996; Torres-Tejerizo et al., 2011).

The symbiont shift may have been driven by plant genetic determinants of unknown specificity in *Phaseolus*. Hybrids between *P. acutifolius* and *P. vulgaris* had been reported to be nodulated by *Rhizobium* strains, inheriting the symbiont preference of *P. vulgaris* cultivars (Somasegaran et al., 1991). In some soybean cultivars, belonging also to the Phaseoleae tribe, R genes have been discovered that dictate affinity to *Sinorhizobium* instead of *Bradyrhizobium* (Yang et al., 2010). It would be interesting to

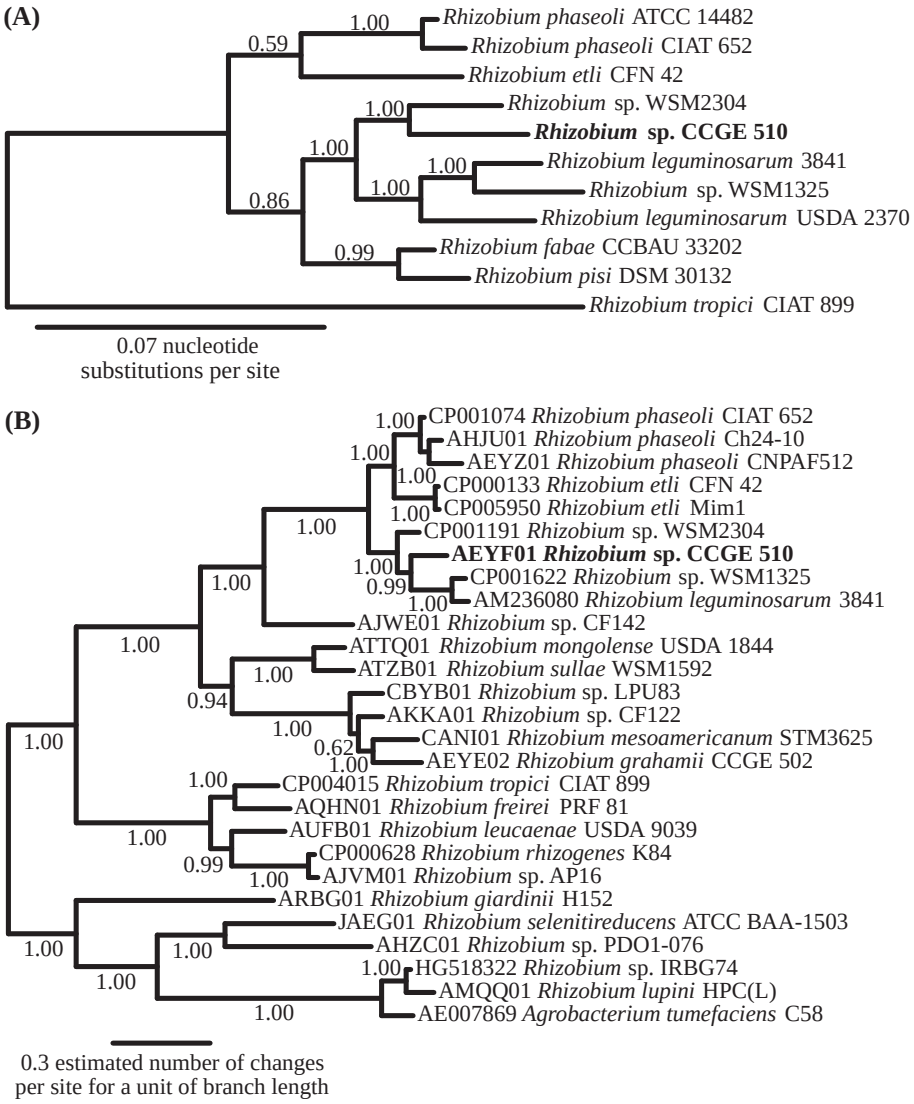


Fig. 5. (A) Topology of a multilocus sequence analysis of *Rhizobium* strains using the nuclear marker genes *atpD*, *recA* and *rpoB*. Partial *atpD* (432 bp), *recA* (453 bp) and *rpoB* (672 bp) gene sequences were used. *Rhizobium* sp. CCGE 510 is shown in bold. The Bayesian phylogenetic reconstruction is based on a concatenated alignment of the marker gene sequences and contains 1,557 characters. The support values for branches are shown as Bayesian posterior probabilities. The scale bar represents the average number of nucleotide replacements per site. (B) Topology of a phylogenomic analysis showing the predicted evolutionary relationship between *Rhizobium* sp. CCGE 510 and sequenced *Rhizobium* strains. The tree was reconstructed with PhyloPhlAn using a multisequence alignment of 377 conserved proteins. PhyloPhlAn performs individual alignments from each protein set recovered from the bradyrhizobial input genomes. PhyloPhlAn then concatenates the most discriminative positions in each protein alignment into a single long sequence to reconstruct a phylogenetic tree using FastTree. The position of *Rhizobium* sp. CCGE 510 is shown in bold in the tree. Accession numbers are indicated for all sequenced genomes. Numbers at the branch points represent SH-like local support values (based on 1000 resamples). The scale bar represents the estimated number of amino acid changes per site for a unit of branch length.

analyze symbiont specificity determined by R genes variants in *Phaseolus*. If this is the case, the R genes variants determining rhizobial specificity emerged in the ancestor of the *P. vulgaris*, *P. albescens*, *P. coccineus*, *P. dumosus* and *P. costaricensis* species (Fig. 2). It remains to be elucidated if this evolutionary replacement of nodule symbionts may confer novel ecological or metabolic capabilities to *Phaseolus* species. The common ancestor of *P. vulgaris* related species could have shifted its symbiont preference in response to a strong selective advantage that could have been pathogen or environment driven. It has been proposed that the microbiota can be considered as a genetic component that drives host speciation (Brucker and Bordenstein, 2012).

The fact that rhizobial symbiosis genes are located in plasmids instead of symbiosis islands on bradyrhizobial chromosomes may allow a larger plasticity of symbiotic determinants to keep pace with legume speciation events, due to frequent lateral transfer and recombination events. Field collected symbionts from *Phaseolus* species represent a rich source of novel symbiosis variants that are largely unexplored and that may provide clues on the molecular evolution of their symbiotic repertoires.

5. Conclusions

Here we proposed that a symbiont shift from *Bradyrhizobium* to *Rhizobium* occurred in an ancestor of a group of plants phylogenetically related to *P. vulgaris* but excluding *P. acutifolius* and *P. parvifolius* species. Studies should be carried in the future to study the genetic basis of this symbiont substitution.

Acknowledgments

This work was supported by PAPIIT IN 205412 from DGAPA, UNAM. We are thankful to the Programa de Doctorado en Ciencias Biomédicas from UNAM. L. E. Servín-Garcidueñas and A. Zayas-Del

Moral had CONACyT scholarship as graduate students. We thank X. Alvarado for technical help with tissue preparations and microscopy analyses. Samplings were carried out with the efforts of R. Ramírez, A. Frías, A. Mendoza, J. Olivares and S. Napsucialy. Some *Phaseolus* DNA samples were provided by M. Blair and some seeds were provided by J. Acosta, M. Rendón and F. Basurto. We thank M. Dunn for critically reading the paper. We appreciate the support of students and collaborators from the Programa Adopta un Talento (PAUTA).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2014.06.006.

References

Aguilar, O.M., Riva, O., Peltzer, E., 2004. Analysis of *Rhizobium etli* and of its symbiosis with wild *Phaseolus vulgaris* supports coevolution in centers of host diversification. Proc. Natl. Acad. Sci. U.S.A. 101, 13548–13553.

Auch, A.F., Klenk, H.P., Göker, M., 2010a. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand. Genomic Sci. 2, 142–148.

Auch, A.F., von Jan, M., Klenk, H.P., Göker, M., 2010b. Digital DNA–DNA hybridization for microbial species delineation by means of genome-to genome sequence comparison. Stand. Genomic Sci. 2, 117–134.

Aserse, A.A., Räsänen, L.A., Assefa, F., Hailemariam, A., Lindström, K., 2012. Phylogeny and genetic diversity of native rhizobia nodulating common bean (*Phaseolus vulgaris* L.) in Ethiopia. Syst. Appl. Microbiol. 35, 120–131.

Blair, M.W., Pantoja, W., Muñoz, L.C., 2012. First use of microsatellite markers in a large collection of cultivated and wild accessions of tepary bean (*Phaseolus acutifolius* A. Gray). Theor. Appl. Genet. 125, 1137–1147.

Brucker, R.M., Bordenstein, S.R., 2012. Speciation by symbiosis. Trends Ecol. Evol. 27, 443–451.

Caballero-Mellado, J., Martinez-Romero, E., 1999. Soil fertilization limits the genetic diversity of *Rhizobium* in bean nodules. Symbiosis 26, 111–121.

Cárdenas, L., Domínguez, J., Quinto, C., López-Lara, I.M., Lugtenberg, B.J., Spaink, H.P., Rademaker, G.J., Haverkamp, J., Thomas-Oates, J.E., 1995. Isolation, chemical structures and biological activity of the lipo-chitin oligosaccharide nodulation signals from *Rhizobium etli*. Plant Mol. Biol. 29, 453–464.

Chieco, P., Pagnoni, M., Romagnoli, E., Melchiorri, C., 1993. A rapid and simple staining method, using toluidine blue, for analysing mitotic figures in tissue sections. Histochem. J. 25, 569–577.

Darling, A.E., Mau, B., Perna, N.T., 2010. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE 5, e11147.

Delgado-Salinas, A., Turley, T., Richman, A., Lavin, M., 1999. Phylogenetic analysis of the cultivated and wild species of *Phaseolus* (Fabaceae). Syst. Bot. 24, 438–460.

Delgado-Salinas, A., Bibler, R., Lavin, M., 2006. Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. Syst. Bot. 31, 779–791.

D'Haeze, W., Holsters, M., 2002. Nod factor structures, responses, and perception during initiation of nodule development. Glycobiology 12, 79R–105R.

Durán, D., Rey, L., Mayo, J., Zúñiga-Dávila, D., Imperial, J., Ruiz-Argüeso, T., Martínez-Romero, E., Ormeño-Orrillo, E., 2014. *Bradyrhizobium paxllaeri* sp. nov. and *Bradyrhizobium icense* sp. nov., nitrogen-fixing rhizobial symbionts of Lima bean (*Phaseolus lunatus* L.) in Peru. Int. J. Syst. Evol. Microbiol. 64, 2072–2078.

Eckhardt, T., 1978. A rapid method for the identification of plasmid desoxyribonucleic acid in bacteria. Plasmid 1, 584–588.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Fåhraeus, G., 1957. The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique. J. Gen. Microbiol. 16, 374–381.

Folch-Mallol, J.L., Marroqui, S., Sousa, C., Manyani, H., López-Lara, I.M., van der Drift, K.M., Haverkamp, J., Quinto, C., Gil-Serrano, A., Thomas-Oates, J., Spaink, H.P., Megías, M., 1996. Characterization of *Rhizobium tropici* CIAT899 nodulation factors: the role of *nodH* and *nodPQ* genes in their sulfation. Mol. Plant-Microbe Interact. 9, 151–163.

Freytag, G.F., Debouck, D.G., 2002. Taxonomy, distribution, and ecology of the genus *Phaseolus* (Leguminosae–Papilionoideae) in North America, Mexico and Central America. Sida 23, 1–300.

Geurts, R., Bisseling, T., 2002. *Rhizobium* nod factor perception and signalling. Plant Cell. 14, S239–S249.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M., 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81–91.

Grange, L., Hungria, M., 2004. Genetic diversity of indigenous common bean (*Phaseolus vulgaris*) rhizobia in two Brazilian ecosystems. Soil Biol. Biochem. 36, 1389–1398.

Hernández-Lucas, I., Segovia, L., Martínez-Romero, E., Pueppke, S.G., 1995. Phylogenetic relationships and host range of *Rhizobium* spp. that nodulate *Phaseolus vulgaris* L.. Appl. Environ. Microbiol. 61, 2775–2779.

Herrera-Cervera, J.A., Caballero-Mellado, J., Laguerre, G., Tichy, H.-V., Requena, N., Amarger, N., Martínez-Romero, E., Olivares, J., Sanjuan, J., 1999. At least five rhizobial species nodulate *Phaseolus vulgaris* in a Spanish soil. FEMS Microbiol. Ecol. 30, 87–97.

Khamis, A., Colson, P., Raoult, D., Scola, B.L., 2003. Usefulness of *rpoB* gene sequencing for identification of *Afipia* and *Bosea* species, including a strategy for choosing discriminative partial sequences. Appl. Environ. Microbiol. 69, 6740–6749.

Laeremans, T., Snoeck, C., Mariën, J., Verreth, C., Martínez-Romero, E., Promé, J.C., Vanderleyden, J., 1999. *Phaseolus vulgaris* recognizes *Azorhizobium caulinodans* Nod factors with a variety of chemical substituents. Mol. Plant Microbe Interact. 12, 820–824.

Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R., Gordon, J.I., 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. Nat. Rev. Microbiol. 6, 776–788.

López-Guerrero, M.G., Ormeño-Orrillo, E., Velázquez, E., Rogel, M.A., Acosta, J.L., Gónzalez, V., Martínez, J., Martínez-Romero, E., 2012. *Rhizobium etli* taxonomy revised with novel genomic data and analyses. Syst. Appl. Microbiol. 35, 353–358.

López-López, A., Rogel, M.A., Ormeño-Orrillo, E., Martínez-Romero, J., Martínez-Romero, E., 2010. *Phaseolus vulgaris* seed-borne endophytic community with novel bacterial species such as *Rhizobium endophyticum* sp. nov. Syst. Appl. Microbiol. 33, 322–327.

López-López, A., Negrete-Yankelevich, S., Rogel, M.A., Ormeño-Orrillo, E., Martínez, J., Martínez-Romero, E., 2013. Native bradyrhizobia from Los Tuxtlas in Mexico are symbionts of *Phaseolus lunatus* (Lima bean). Syst. Appl. Microbiol. 36, 33–38.

Martínez-Romero, E., Pardo, M.A., Cevallos, M.A., Palacios, R., 1985. Reiteration of nitrogen fixation gene sequences and specificity of *Rhizobium* in nodulation and nitrogen fixation in *Phaseolus vulgaris*. J. Gen. Microbiol. 131, 1779–1786.

Martínez-Romero, E., 2003. Diversity of *Rhizobium-Phaseolus vulgaris* symbiosis: overview and perspectives. Plant Soil 252, 11–23.

Martínez-Romero, E., 2009. Coevolution in *Rhizobium*-legume symbiosis? DNA Cell Biol. 28, 361–370.

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.P., Göker, M., 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14, 60.

Michiels, J., Dombrecht, B., Vermeiren, N., Chuanwu, X., Luyten, E., Vanderleyden, J., 1998. *Phaseolus vulgaris* is a non-selective host for nodulation. FEMS Microbiol. Ecol. 26, 193–205.

Muñoz, L.C., Duque, M.C., Debouck, D.G., Blair, M.W., 2006. Taxonomy of tepary bean and wild relatives as determined by amplified fragment length polymorphism (AFLP) markers. Crop Sci. 46, 1744–1754.

Noel, K.D., Sanchez, A., Fernandez, L., Leemans, J., Cevallos, M.A., 1984. *Rhizobium phaseoli* symbiotic mutants with transposon Tn5 insertions. J. Bacteriol. 158, 148–155.

Noel, K.D., Vandenbosch, K.A., Kulpaca, B., 1986. Mutations in *Rhizobium phaseoli* that lead to arrested development of infection threads. J. Bacteriol. 168, 1392–1401.

Ormeño, E., Torres, R., Mayo, J., Rivas, R., Peix, A., Velázquez, E., Zúñiga, D., 2007. *Phaseolus lunatus* is nodulated by a phosphate solubilizing strain of *Sinorhizobium meliloti* in a Peruvian soil. In: Velázquez, E., Rodríguez-Barrueco, C. (Eds.), Developments in Plant and Soil Sciences. Springer-Verlag, The Netherlands, pp. 243–247.

Ormeño-Orrillo, E., Vinuesa, P., Zúñiga-Dávila, D., Martínez-Romero, E., 2006. Molecular diversity of native bradyrhizobia isolated from Lima bean (*Phaseolus lunatus* L.) in Peru. Syst. Appl. Microbiol. 29, 253–262.

Ormeño-Orrillo, E., Rogel-Hernández, M.A., Lloret, L., López-López, A., Martínez, J., Barois, I., Martínez-Romero, E., 2012. Change in land use alters the diversity and composition of *Bradyrhizobium* communities and led to the introduction of *Rhizobium etli* into the tropical rain forest of Los Tuxtlas (Mexico). Microb. Ecol. 63, 822–834.

Ormeño-Orrillo, E., Servín-Garcidueñas, L.E., Imperial, J., Rey, L., Ruiz-Argueso, T., Martínez-Romero, E., 2013. Phylogenetic evidence of the transfer of *nodZ* and *nolL* genes from *Bradyrhizobium* to other rhizobia. Mol. Phylogenet. Evol. 67, 626–630.

Parker, M.A., 2002. Bradyrhizobia from wild *Phaseolus*, *Desmodium*, and *Macroptilium* species in Northern Mexico. Appl. Environ. Microbiol. 68, 2044–2048.

Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.

Posada, D., 2008. JModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.

Poupot, R., Martinez-Romero, E., Promé, J.C., 1993. Nodulation factors from *Rhizobium tropici* are sulfated or nonsulfated chitopentasaccharides containing an N-methyl-N-acylglucosaminyl terminus. Biochemistry 32, 10430–10435.

Poupot, R., Martinez-Romero, E., Gautier, N., Promé, J.C., 1995. Wild type *Rhizobium etli*, a bean symbiont, produces acetyl-fucosylated, N-methylated, and carbamoylated nodulation factors. J. Biol. Chem. 270, 6050–6055.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE 5, e9490.

Ramírez-Delgadillo, R., Delgado-Salinas, A., 1999. A new species of *Phaseolus* (Fabaceae) from West-Central Mexico. Sida 18, 637–645.

Ribeiro, R.A., Ormeño-Orrillo, E., Dall'Agnol, R.F., Graham, P.H., Martinez-Romero, E., Hungria, M., 2013. Novel *Rhizobium* lineages isolated from root nodules of the common bean (*Phaseolus vulgaris* L.) in Andean and Mesoamerican areas. Res. Microbiol. 164, 740–748.

Richter, M., Rosselló-Móra, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. USA 106, 19126–19131.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Rosenblueth, M., Hynes, M.F., Martínez-Romero, E., 1998. *Rhizobium tropici teu* genes involved in specific uptake of *Phaseolus vulgaris* bean-exudate compounds. Mol. Gen. Genet. 258, 58–98.

Salcedo-Castaño, J.M., Lépiz-Ildefonso, R., Castañeda-Alvarez, N.P., Ocampo-Nahar, C.H., Debouck, D.G., 2009. Additional observations about *Phaseolus rotundatus* (Fabaceae), an endemic bean species from western Mexico. J. Bot. Res. Inst. Texas 3, 751–762.

Santos, J.O., Antunes, J.E., Araújo, A.S., Lyra, M.C., Gomes, R.L., Lopes, A.C., Figueiredo, M., 2011. Genetic diversity among native isolates of rhizobia from *Phaseolus lunatus*. Ann. Microbiol. 61, 437–444.

Segata, N., Börnigen, D., Morgan, X.C., Huttenhower, C., 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat. Commun. 4, 2304.

Segovia, L., Piñero, D., Palacios, R., Martínez-Romero, E., 1991. Genetic structure of a soil population of nonsymbiotic *Rhizobium leguminosarum*. Appl. Environ. Microbiol. 57, 426–433.

Segovia, L., Young, J.P.W., Martínez-Romero, E., 1993. Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* type I strains as *Rhizobium etli* sp. nov. Int. J. Syst. Bacteriol. 43, 374–377.

Servín-Garcidueñas, L.E., Rogel, M.A., Ormeño-Orrillo, E., Delgado-Salinas, A., Martínez-Romero, J., Sánchez, F., Martínez-Romero, E., 2012. Genome sequence of *Rhizobium* sp. strain CCGE 510, a symbiont isolated from nodules of the endangered wild bean *Phaseolus albescens*. J. Bacteriol. 194, 6310–6311.

Silva, C., Vinuesa, P., Eguiarte, L.E., Martínez-Romero, E., Souza, V., 2003. *Rhizobium etli* and *Rhizobium gallicum* nodulate common bean (*Phaseolus vulgaris*) in a traditionally managed milpa plot in Mexico: population genetics and biogeographic implications. Appl. Environ. Microbiol. 69, 884–893.

Somasegaran, P., Hoben, H.J., Lewinson, L., 1991. Symbiotic interactions of *Phaseolus acutifolius* and *P. acutifolius* × *P. vulgaris* hybrid progeny in symbiosis with *Bradyrhizobium* spp. and *Rhizobium leguminosarum* bv. *phaseoli*. Can. J. Microbiol. 37, 497–503.

Souza, V., Eguiarte, L., Avila, G., Cappello, R., Gallardo, C., Montoya, J., Piñero, D., 1994. Genetic structure of *Rhizobium etli* biovar *phaseoli* associated with wild and cultivated bean plants (*Phaseolus vulgaris* and *Phaseolus coccineus*) in Morelos, Mexico. Appl. Environ. Microbiol. 60, 1260–1268.

Tamimi, S.A., Young, J.P.W., 2004. *Rhizobium etli* is the dominant common bean nodulating rhizobia in cultivated soils from different locations in Jordan. Appl. Soil Ecol. 26, 193–200.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Tindall, B.J., Rosselló-Móra, R., Busse, H.J., Ludwig, W., Kämpfer, P., 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. Int. J. Syst. Evol. Microbiol. 60, 249–266.

Toledo, I., Lloret, L., Martínez-Romero, E., 2003. *Sinorhizobium americanus* sp. nov., a new *Sinorhizobium* species nodulating native *Acacia* spp. in Mexico. Syst. Appl. Microbiol. 26, 54–64.

Torres-Tejerizo, G., Del Papa, M.F., Soria-Diaz, M.E., Draghi, W., Lozano, M., Giusti, M.deL., Manyani, H., Megías, M., Gil Serrano, A., Pühler, A., Niehaus, K., Lagares, A., Pistorio, M., 2011. The nodulation of alfalfa by the acid-tolerant *Rhizobium* sp. strain LPU83 does not require sulfated forms of lipochitooligosaccharide nodulation signals. J. Bacteriol. 193, 30–39.

Vandenbosch, K.A., Noel, K.D., Kaneko, Y., Newcomb, E.H., 1985. Nodule initiation elicited by noninfective mutants of *Rhizobium phaseoli*. J. Bacteriol. 162, 950–959.

Vincent, J.M., 1970. A manual for the practical study of root-nodule bacteria. I.B.P. Handbook No. 15. Blackwell Scientific Publications, Oxford, 164 p.

Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., Moore, L.H., Moore, W.E.C., Murray, R.G.E., Stackebrandt, E., Starr, M.P., Truper, H.G., 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37, 463–464.

Weisburg, W.G., Barns, S.M., Pelletier, D.A., Lane, D.J., 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. 173, 697–703.

Yang, S., Tang, F., Gao, M., Krishnan, H.B., Zhu, H., 2010. R gene-controlled host specificity in the legume-rhizobia symbiosis. Proc. Natl. Acad. Sci. USA 107, 18735–18740.

Zehr, J.P., McReynolds, L.A., 1989. Use of degenerate oligonucleotides for amplification of the *nifH* gene from the marine cyanobacterium *Trichodesmium thiebautii*. Appl. Environ. Microbiol. 55, 2522–2526.

Artículo:

Servín-Garcidueñas LE, Rogel MA, Ormeño-Orrillo E, Delgado-Salinas A, Martínez-Romero J, Sánchez F, Martínez-Romero E. 2012. Genome sequence of *Rhizobium* sp. strain CCGE510, a symbiont isolated from nodules of the endangered wild bean *Phaseolus albescens*. J. Bacteriol. 194: 6310-6311.

Phaseolus albescens es una de las especies silvestres de frijol relacionadas con el grupo filogenético de *P. vulgaris.* Las plantas de *P. albescens* son endémicas de las regiones boscosas de la Sierra Madre Occidental localizadas en los Estados de Jalisco y Michoacán. A partir de los ensayos de cultivo de bacterias de nódulos de campo de *P. albescens* logramos recuperar a la cepa CCGE510 que representa a un nuevo linaje de rizobios. Las relaciones evolutivas de la cepa CCGE510 se clarificaron a partir de análisis filogenómicos usando marcadores genéticos recuperados de su genoma (Servín-Garcidueñas et al., 2014). En este artículo se reportan las principales características del genoma de la cepa CCGE510.

JB Journals.ASM.org

GENOME ANNOUNCEMENT

Genome Sequence of *Rhizobium* sp. Strain CCGE510, a Symbiont Isolated from Nodules of the Endangered Wild Bean *Phaseolus albescens*

Luis E. Servín-Garcidueñas,[a] Marco A. Rogel,[a] Ernesto Ormeño-Orrillo,[a] Alfonso Delgado-Salinas,[b] Julio Martínez-Romero,[a] Federico Sánchez,[c] and Esperanza Martínez-Romero[a]

Centro de Ciencias Genómicas, UNAM, Cuernavaca, Morelos, México[a]; Instituto de Biología, UNAM, México D.F., México[b]; and Instituto de Biotecnología, UNAM, Cuernavaca, Morelos, México[c]

We present the genome sequence of *Rhizobium* sp. strain CCGE510, a nitrogen fixing bacterium taxonomically affiliated with the *R. leguminosarum-R. etli* group, isolated from wild *Phaseolus albescens* nodules grown in native pine forests in western Mexico. *P. albescens* is an endangered bean species phylogenetically related to *P. vulgaris*. In spite of the close host relatedness, *Rhizobium* sp. CCGE510 does not establish an efficient symbiosis with *P. vulgaris*. This is the first genome of a *Rhizobium* symbiont from a *Phaseolus* species other than *P. vulgaris*, and it will provide valuable new insights about symbiont-host specificity.

Phaseolus albescens R. Ram. & A. Delgado is a nondomesticated species phylogenetically related to *Phaseolus vulgaris* (1, 2, 8), and its symbiotic bacteria have not been described. *P. albescens* is at risk (2, 8) because it is distributed in a restricted area and threatened by changing land use, and few seeds are safeguarded. We isolated novel rhizobia from field-collected *P. albescens* nodules, including a representative strain designated CCGE510. Strain CCGE510 established an effective symbiosis with *P. albescens*, as inoculated plants were green and nodules were pink and reduced acetylene. Interestingly, *P. vulgaris* plants inoculated with this strain were yellow and nodules turned green soon after appearance.

The genome of *Rhizobium* sp. strain CCGE510 was sequenced with the Illumina GAIIx platform, producing 6,329,550 36-bp reads (~32-fold coverage), and with the Roche 454 GS-FLX Titanium technology, generating 91.16 Mbp (~13-fold coverage) from a mate-paired library. Illumina paired reads were assembled *de novo* using Velvet 1.2.03 (11). Contigs were fragmented by using the EMBOSS splitter (9) and were assembled with 454 mate-reads using Newbler Assembler 2.3 (454 Life Science). Reads were mapped to gap-surrounding sequences using Maq 0.7.1 (5) and the Newbler runMapping option. Mapping contigs and PCR amplifications were used to eliminate gaps. The final assembly produced 142 contigs with an N_{50} size of 270.2 kb. Genome annotation was performed using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html). Plasmids were detected using a modified Eckhardt procedure (4), and their sizes were estimated using regression equations comparing them with *R. etli* CFN42 plasmids.

The genome (6.9 Mbp, 60.8% G+C content) consisted of a chromosome and four plasmids and coded for 6,642 predicted open reading frames. The chromosome (5.04 Mb) was distributed in one scaffold. Plasmid pRspCCGE510a (61.69 kb) seems to be unstable or recently acquired, as it is commonly absent in other *P. albescens* nodule strains. Plasmid pRspCCGE510b (270.27 kb), the symbiotic plasmid pRspCCGE510c (579.35 kb), and plasmid pRspCCGE510d (923.84 kb) were recovered as single scaffolds.

Small-subunit rRNA gene phylogeny indicated that *Rhizobium* sp. CCGE510 is related to *R. leguminosarum* strains; however, calculated average nucleotide identity (ANI) (10) separated it from that species. Symbiotic genes were related to those found in symbiovar (where "symbiovar" [sv.] means symbiotic variant) phaseoli, but overall symbiotic plasmid synteny was not as maintained as in phaseoli symbiotic plasmids (3). Some genes required for Nod factor synthesis were divergent from those of *R. etli* CFN42. Differences were also observed in secretion systems, including type III effector proteins and a transporter for the uptake of bean exudates. Strain CCGE510 may metabolize pine compounds, as *P. albescens* roots are intertwined with pine roots. *R. etli* sv. phaseoli strains are very competitive for *P. vulgaris* nodulation (6), and the evolution of the phaseoli plasmid could have been driven by host selective pressures (6, 7). Similarly, *P. albescens* bacteria seem to be better adapted to their host and not to *P. vulgaris*. *P. vulgaris* and *P. albescens* diverged about 1 to 2 million years ago (1); it is plausible that in the corresponding symbiotic plasmids (from *P. vulgaris* and *P. albescens* bacteria) differences have accumulated since host divergence.

Nucleotide sequence accession number. The genome sequence has been deposited in DDBJ/EMBL/GenBank under the accession number AEYF00000000.

ACKNOWLEDGMENTS

We are thankful to UUSM, UNAM, especially Ricardo Grande Cano and Verónica Jiménez Jacinto; to Paul Gaytán and Eugenio López for synthesis of oligonucleotides; and to Michael Dunn for critically reading the paper.

This work was supported by PAPIIT IN205412 from DGAPA, UNAM. L.E.S.-G. was supported by a CONACyT scholarship as a Ph.D.

Received 22 August 2012 **Accepted** 29 August 2012

Address correspondence to Esperanza Martínez-Romero, emartine@ccg.unam.mx.

This work is dedicated to the memory of Raymundo Ramírez Delgadillo; his enthusiastic work allowed the recovery of field nodules and conservation of *Phaseolus albescens* seeds.

doi:10.1128/JB.01536-12

student at the Programa de Doctorado en Ciencias Biomédicas from UNAM.

REFERENCES

1. **Delgado-Salinas A, Bibler R, Lavin M.** 2006. Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. Syst. Bot. **31**(4):779–791.
2. **Freytag GF, Debouck DG.** 2002. Sida botanical miscellany, vol 23. Taxonomy, distribution, and ecology of the genus *Phaseolus* (Leguminosae-Papilionoideae) in North America, Mexico and Central America. Botanical Research Institute of Texas, Fort Worth, TX.
3. **González V, et al.** 2010. Conserved symbiotic plasmid DNA sequences in the multireplicon pangenomic structure of *Rhizobium etli*. Appl. Environ. Microbiol. **76**:1604–1614.
4. **Hynes MF, McGregor NF.** 1990. Two plasmids other than the nodulation plasmid are necessary for formation of nitrogen-fixing nodules by *Rhizobium leguminosarum*. Mol. Microbiol. **4**:567–574.
5. **Li H, Ruan J, Durbin R.** 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18**:1851–1858.
6. **Martinez-Romero E, Rosenblueth M.** 1990. Increased bean (*Phaseolus vulgaris* L.) nodulation competitiveness of genetically modified *Rhizobium* strains. Appl. Environ. Microbiol. **56**:2384–2388.
7. **Martínez-Romero E.** 2009. Coevolution in *Rhizobium*-legume symbiosis? DNA Cell Biol. **28**:361–370.
8. **Ramírez-Delgadillo R, Delgado-Salinas A.** 1999. A new species of *Phaseolus* (Fabaceae) from West-Central México. Sida **18**(3):637–645.
9. **Rice P, Longden I, Bleasby A.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. **16**:276–277.
10. **Richter M, Rossello-Mora R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. **106**:19126–19131.
11. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.

Artículo:

Ormeño-Orrillo E, Servín-Garcidueñas LE, Imperial J, Rey L, Ruiz-Argueso T, Martinez-Romero E. 2013. Phylogenetic evidence of the transfer of *nodZ* and *nolL* genes from *Bradyrhizobium* to other rhizobia. Mol. Phylogenet. Evol. 67: 626-630.

En este trabajo se encontró que cepas de *Bradyrhizobium* presentan la mayor diversidad de genes *nodZ* y *nolL*, que están involucrados en la fucosilación y acetilación del factor Nod. Nuestros análisis filogenéticos revelan que han ocurrido eventos de transferencia horizontal de los genes *nodZ* y *nolL* entre linajes diversos de rizobios. También se encontró que la diversidad de estos genes en *Rhizobium* es más bien limitada. Así, sugerimos basados en los análisis filogenéticos la transferencia de los genes *nodZ* y *nolL* de *Bradyrhizobium* al plásmido simbiótico del simbiovar phaseoli de *Rhizobium etli* y otros simbiontes de *P. vulgaris*. Los resultados detallados se presentan en el artículo.

Molecular Phylogenetics and Evolution 67 (2013) 626–630

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

Phylogenetic evidence of the transfer of *nodZ* and *nolL* genes from *Bradyrhizobium* to other rhizobia

Ernesto Ormeño-Orrillo [a], Luis E. Servín-Garcidueñas [a], Juan Imperial [b], Luis Rey [b], Tomás Ruiz-Argueso [b], Esperanza Martinez-Romero [a,*]

[a] Centro de Ciencias Genómicas, UNAM, Cuernavaca, Mor., Mexico
[b] Departamento de Biotecnología (ETS de Ingenieros Agrónomos) and Centro de Biotecnología y Genómica de Plantas (CBGP), Universidad Politécnica de Madrid, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 29 January 2013
Revised 27 February 2013
Accepted 4 March 2013
Available online 14 March 2013

Keywords:
Nod genes
Legume symbiosis
Rhizobia
Lateral gene transfer

ABSTRACT

Nod factor modifications mediated by *nodZ* and *nolL* gene products (fucosylation and acetylation of fucose residues, respectively) were probably later acquisitions in the nodulation process. Novel phylogenetic analyses suggest that *nodZ* and *nolL* genes were transferred from *Bradyrhizobium* to other nodule bacteria. These bradyrhizobial genes are highly diverse while rhizobial, sinorhizobial and mesorhizobial *nodZ* and *nolL* genes are represented by few branches among those from bradyrhizobia. These genes in novel rhizobial backgrounds may have favored efficient nodulation in legume hosts commonly associated with *Bradyrhizobium* strains.

1. Introduction

Few legumes may form nodules with rhizobia that do not produce Nod factors (Giraud et al., 2007) but in most legumes nodulation is dependent on Nod factors that are key molecules in rhizobia-plant interactions (Dénarié et al., 1996). Nod factors are produced by enzymes coded by *nod*, *noe* or *nol* genes (collectively referred here as *nod* genes) that are inducible by plant exudates such as flavonoids or other unrelated molecules (Hassan and Mathesius, 2012).

Some of the modifications occurring in Nod factors are sulfation, methylation, acetylation, carbamoylation, fucosylation, arabinosylation (Debellé et al., 1986; Horvath et al., 1986; Carlson et al., 1994; Stepkowski et al., 2003, 2005, 2007; D'Haeze et al., 1999; Quinto et al., 1997). Some of the genes responsible for such modifications have been designated as host specifity genes and some of them are patchily distributed among *Sinorhizobium* (officially called *Ensifer*), *Rhizobium* (Martinez et al., 1995) or *Bradyrhizobium* (Steenkamp et al., 2008; Stepkowski et al., 2003, 2005, 2007) strains. Both fucosylation and sulfation occur alternatively at the same position of Nod factors and maybe fulfill the same function. Sulfate seems to be involved in protecting Nod factors from chitinase degradation (Staehelin et al., 1994). A relationship of *nod* genes or Nod factor structure and specificity is however not direct (Perret et al., 2000; López-Lara et al., 1996) probably because other bacterial systems such as type III secretion systems or different exopolysaccharides contribute to host specificity (Djordjevic et al., 1987; Marie et al., 2001; Jones, 2012).

In rhizobia, incongruent phylogenies of symbiotic and core genes have been commonly observed (reviewed in Rogel et al., 2011 and in Martínez-Romero, 2009) and are explained by the lateral transfer of symbiotic genes among rhizobia, in relation to adaptation to different hosts (Rogel et al., 2011; Lindström et al., 2010). Nodulation genes are found in symbiotic plasmids or islands that may be mobilized among rhizobia (Rogel et al., 2001; Sullivan and Ronson, 1998). The existence of beta-rhizobia is explained by an ancient transfer event of symbiosis genes from alpha-rhizobia (Chen et al., 2003; Bontemps et al., 2010). Similarly, it is speculated that *nod* genes were transferred to *Azorhizobium*, an epiphytic bacterium (Lee et al., 2008), or to *Methylobacterium* (Jourand et al., 2004), *Devosia* (Rivas et al., 2002), *Phyllobacterium* (Valverde et al., 2005), and from *Burkholderia* to *Cupriavidus* (Andam et al., 2007).

Azorhizobium caulinodans Nod factors are fucosylated (Mergaert et al., 1996). In *Bradyrhizobium japonicum* fucosylation of Nod factors by *nodZ* (Sanjuan et al., 1992) has been linked to nodulation of siratro (*Macroptilium atropurpureum*) and *Vigna umbellata* (Cohn et al., 1999) but *nodZ* mutants were still capable of nodulating soybean (Stacey et al., 1994). *Sinorhizobium fredii* mutants in *nodZ* genes have decreased competitiveness to nodulate soybean (Lamrabet et al., 1999). A NGR234 *nodZ* mutant does not nodulate *Pachyrhizus tuberosus* (Quesada-Vincens et al., 1997). In *Mesorhizobium*

* Corresponding author. Address: Centro de Ciencias Genómicas, UNAM Av. Universidad SN, Chamilpa, CP 62210, Cuernavaca, Mor., Mexico. Fax: +52 777 317 5581.
E-mail address: emartine@ccg.unam.mx (E. Martinez-Romero).

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.03.003

loti nodZ mutants did not form nodules in *Lotus filicaulis* (Rodpothong et al., 2009) not in *L. pedunculatus* (Scott et al., 1996), with conflicting results on nodulation of *L. corniculatus* (Scott et al., 1996; Rodpothong et al., 2009). An extended capacity to nodulate novel hosts such as siratro and cowpea or *Lotus* was promoted by transferring *nodZ* genes to *R. leguminosarum* sv. viciae (López-Lara et al., 1996; Pacios Bras et al., 2000). In *R. etli* and *R. phaseoli nodZ* genes are found in relation to *Phaseolus vulgaris* nodulation and fucosylation was a preferred Nod factor modification for the cultivar tested (Laeremans et al., 1999). However we found *nodZ* pseudogenes in some *R. phaseoli* strains such as CIAT 652, and in *R. popolucense* Pop5 (new taxa submitted).

2. Materials and methods

Bradyrhizobium strains previously isolated from *Phaseolus lunatus* (Ormeño-Orrillo et al., 2006), *Lupinus mariae-josephae* (Durán et al., 2013), *Pachyrhizus erosus* (Ramírez-Bahena et al., 2009) and *Lablab purpureus* (Chang et al., 2011) were used. Additionally, strains isolated in this study from *Phaseolus microcarpus* and *Phaseolus leptostachyus* were also analyzed (Supplementary Tables S1 and S2). Fragments of the *nodZ* and *nolL* genes were PCR amplified using primers shown in Supplementary Table S3 and Sanger sequenced using methods described by Moulin et al. (2004) and Stepkowski et al. (2005), respectively, or retrieved from whole genome sequences obtained by us (Durán et al., 2013; Servín et al. unpublished). All *nodZ* and *nolL* gene sequences available in the Genbank database were retrieved including those of a *Phaseolus albescens* symbiont recently reported by our group (Servín-Garcidueñas et al., 2012).

Sequences were aligned by using Muscle 3.8.31 (Edgar, 2004). Alignment lengths after gap removal were 426 and 655 characters for the *nodZ* and *nolL* gene sets, respectively. jModelTest (Posada, 2008) was used to find the model of evolution that best fit the data for subsequent phylogenetic analyses using the Akaike Information Criterion (Posada and Buckley, 2004). The models selected were TPM3uf + I + G and GTR + I + G for *nodZ* and *nolL* sets, respectively. All three codon positions were used as no substitution saturation was found in the third codon position of any gene ($Iss < Iss_c$, $P < 0.001$) (Xia et al., 2003). Maximum likelihood trees were generated with PhyML (Guindon et al., 2010) with tree node support evaluated by bootstrap analysis based on 1000 pseudoreplicate datasets. Phylogenetic relationships were also assessed by Bayesian inference using MrBayes 3.1 (Ronquist and Huelsenbeck, 2003). Analyses were initiated with random starting trees, run for 2,000,000 generations and three separate analyses were executed. Markov chains were sampled every 100 generations. We discarded 25% of trees as "burn in".

The *nodZ* and *nolL* gene sequences determined in this study were deposited in the GenBank database under accession numbers KC526990–KC527000 and KC527001–KC527013, respectively.

3. Results and discussion

nodZ bradyrhizobial genes are highly diverse while rhizobial, sinorhizobial and mesorhizobial *nodZ* genes are represented by a few branches related to bradyrhizobial clade IV and VII *nodZ* genes (Fig. 1A). Clade IV includes bradyrhizobial isolates LMTR13 and LMTR21 isolated from *P. lunatus* in Peru (Ormeño-Orrillo et al., 2006), *L. mariae-josephae* (Lmj) bradyrhizobia obtained from alka-

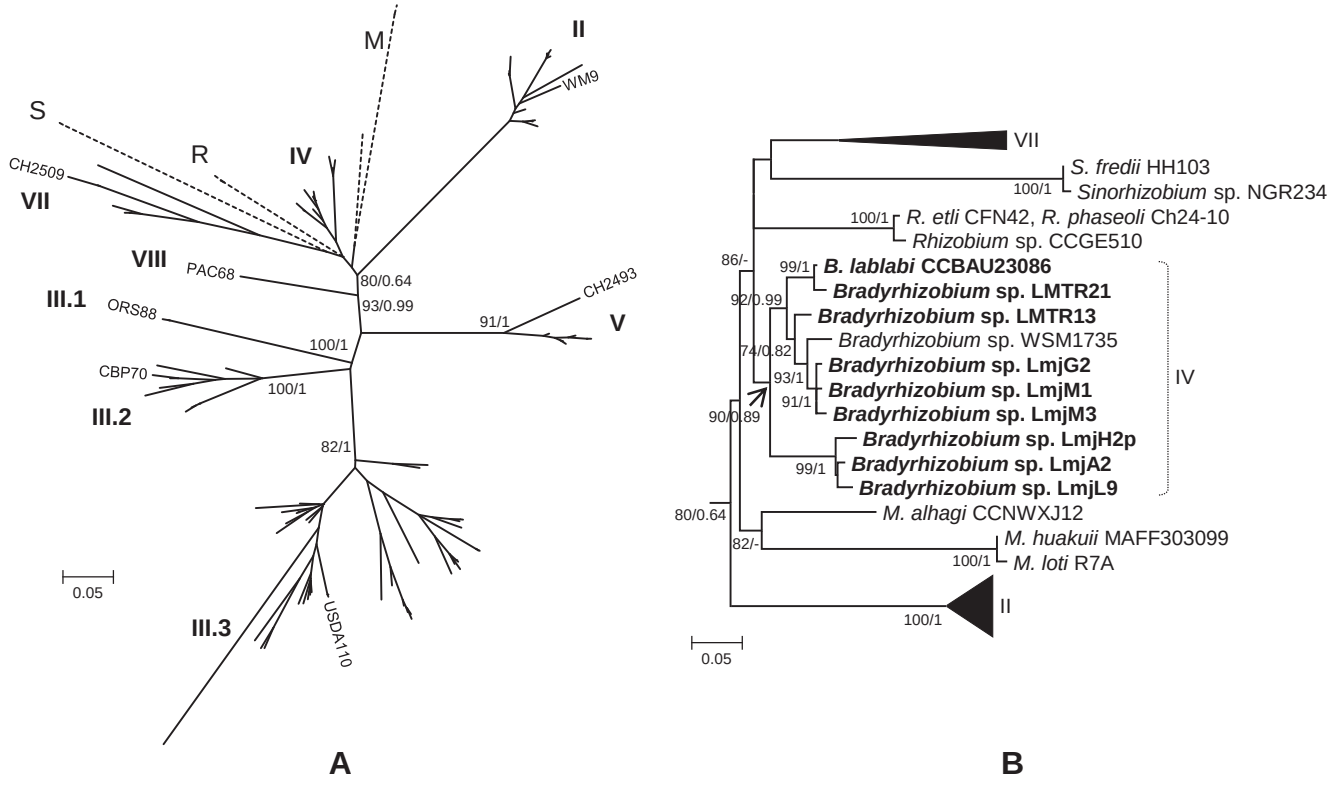


Fig. 1. (**A**) Maximum likelihood (ML) phylogeny of *nodZ* gene sequences. Roman numerals indicate bradyrhizobial *nodZ* clades as defined by Steenkamp et al. (2008) except clade VIII which comprises the gene of *B. jicamae* PAC68 obtained in this study. A representative strain is shown in each clade. R, *Rhizobium*; S, *Sinorhizobium*; M. *Mesorhizobium*. (**B**) Amplification of subtree including clades II, IV and VII. Sequences determined in this study are shown in bold. A similar tree was obtained by Bayesian inference (BI). Bootstrap supports values higher than 70% for the ML analysis as well as Bayesian posterior probabilities are indicated at tree nodes in the order ML/BI. Tree node support is indicated only for major clades in (A).

line soils in Spain (Durán et al., 2013), Chinese strain CCBAU 23086 from *Lablab purpureus* (Chang et al., 2011), and Australian isolate WSM1735 isolated from *Rhynchosia minima* (Stepkowski et al., 2005) (Fig. 1B). Clade VII comprises sequences from bradyrhizobia isolated from tropical legumes in America (Steenkamp et al., 2008; López-López et al., 2013). The *nodZ* phylogeny suggested that this gene was transferred from *Bradyrhizobium* to other rhizobia. In contrast, phylogeny of other *nod* genes like *nodA*, shows that *Rhizobium*, *Sinorhizobium* or *Mesorhizobium* genes are clearly separated from those of bradyrhizobia (Stepkowski et al., 2007; Martínez-Romero et al., 2010; Menna and Hungria, 2011). Interestingly, *nodZ* genes of symbiovar phaseoli *Rhizobium* strains CFN 42 and Ch24-10 isolated from *P. vulgaris* and of *Rhizobium* sp. CCGE510 isolated from *Phaseolus albescens* were related to clade IV *nodZ* that includes genes from *Phaseolus* bradyrhizobia (Fig. 1B). In most *Rhizobium* strains, *nodZ* genes are located in the symbiotic plasmids in close vicinity to a remnant of an IS21 transposase gene. Intriguingly, a remnant of an IS21 is also present close to *nodZ* in clade IV *P. lunatus* strain LMTR13 (our own unpublished data). Transfer of symbiotic genes was suggested to occur between bradyrhizobial strains based on phylogenetic analyses (Moulin et al., 2004) but there had not been any suggestion that there has been a transfer of *nodZ* genes from bradyrhizobia to *Rhizobium*. Previously it was suggested that *Bradyrhizobium* symbiotic genes were transferred to a Brazilian *Sinorhizobium* strain (Barcellos et al., 2007) but this looks like a recent event in view of sequence similarity. Remarkably, phylogenies of *ackA* and *pta* genes (Fournier and Gogarten, 2008) that served as the basis to recognize the lateral transfer of acetoclastic methanogenesis genes from clostridium to methanosarcinales resemble *nodZ* gene phylogenies (Fig. 1A) as methanosarcinales are a single branch among diverse clostridial sequences.

In *Sinorhizobium* NGR234 *nodZ* gene is part of *hsn* gene cluster 1 that includes *noeL* and *nolK* as part of the same operon neighbor to *noeK*, *noeJ* and *nodD1* (Freiberg et al., 1997). This gene organization, including *nodZ*, *noeL* and *nolK* is similar to that found in *Mesorhizobium loti* and both, sinorhizobia and mesorhizobia, are also related in *nodZ* gene phylogenies, such gene organization is not observed in the symbiovar phaseoli plasmids (not shown). *nodZ* gene evolution could have taken place around 60 million years ago maybe driven by fungal proliferation (Vajda and McLoughlin, 2004), if Nod factor modifications are related to chitinase defense (Staehelin et al., 1994; Stacey, 1995) that acts against fungi. Nod factor modifications could be an example of co-evolution of rhizo-

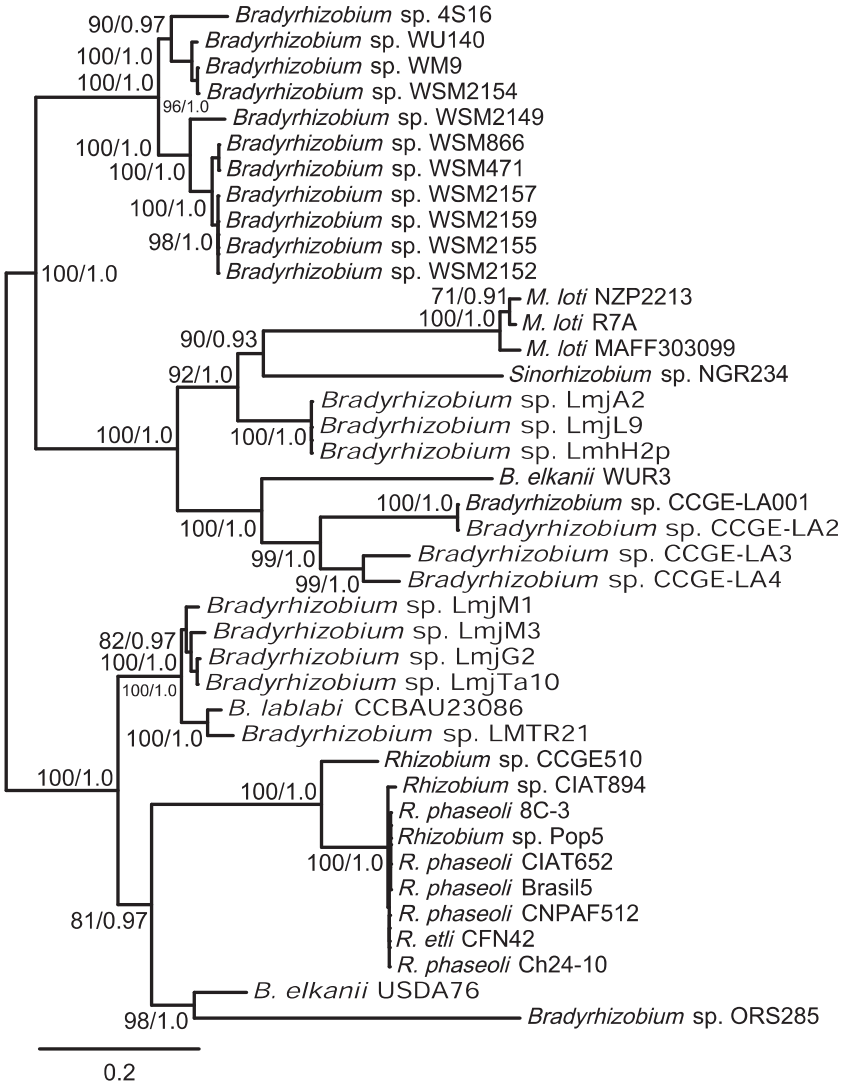


Fig. 2. Maximum likelihood (ML) phylogeny of *nolL* gene sequences. R, *Rhizobium*; S, *Sinorhizobium*; M. *Mesorhizobium*; B, *Bradyrhizobium*. Sequences determined in this study are shown in bold. A similar tree was obtained by Bayesian inference (BI). Bootstrap supports values higher than 70% for the ML analysis as well as Bayesian posterior probabilities are indicated at tree nodes in the order ML/BI. *nolL* pseudogene sequences of CIAT652 and Brasil5 were included in the analysis.

bia and their host-legume in response to fungal pathogens. *nodZ* gene transfer from bradyrhizobia could have occurred to an ancestor of *Rhizobium* and *Sinorhizobium* (around 50–45 million years ago) or alternatively, more recently to any rhizobial species and then transferred among bacteria followed by gene loss and rearrangements in some rhizobial lineages. An ancient transfer of *nod* genes has been considered to have occurred from alpha to beta-rhizobia (Chen et al., 2003; Bontemps et al., 2010) and later from *Burkholderia* to *Cupriavidus* (Andam et al., 2007). Gene rearrangements and losses account for the present organization of *nodZ* in different bacteria. *Sinorhizobium*, *Rhizobium* and *Agrobacterium* are closely related and the phylogenetic analysis of common *nod* genes showing intermixed *Rhizobium* and *Sinorhizobium nod* genes are suggestive of their lateral transfer between these genera (reviewed in Martínez-Romero, 2009).

A further Nod factor modification found in some rhizobia is mediated by *NolL* that acetylates the fucose residue (Berck et al., 1999). This Nod factor decoration seems to be related to nodulation efficiency as an *R. etli nolL* mutant formed a reduced number of nodules in comparison to the wild type on *P. vulgaris* and on *Vigna umbellata* (Corvera et al., 1999). Although relatively few sequences are available, *nolL* phylogeny also showed that *Rhizobium*, *Sinorhizobium* and *Mesorhizobium* sequences are found within those of *Bradyrhizobium* supporting the transfer of this gene from bradyrhizobia (Fig. 2 and Supplementary Fig. 1). All the symbiovar phaseoli *R. etli* and *R. phaseoli* sequences were found clustered, irrespective of the different geographical origin of the strains, and they were highly related to the *nolL* sequence of *Rhizobium* sp. CCGE510. Farther, but still related, are *nolL* sequence from *Bradyrhizobium* sp. LMTR21 isolated from cultivated *P. lunatus*. *Sinorhizobium* and *Mesorhizobium* sequences were related to those of *Bradyrhizobium* strains isolated from wild *P. leptostachyus* (CCGE-LA001, CCGE-LA2) and *P. microcarpus* (CCGE-LA3, CCGE-LA4), and of some *Lupinus mariae-josephae* (Lmj) bradyrhizobia (Fig. 2). Phylogenies of *nodZ* and *nolL* do not seem to be congruent (Supplementary Fig. S1), indicating that perhaps they were independent acquisition from bradyrhizobia. It is worth noting that *nolL* and *nodZ* genes are not contiguous in bradyrhizobia.

4. Conclusions

The two Nod factor modification genes analyzed here seem to have initially developed in the *Bradyrhizobium* genus and independently transferred to a few different fast-growing rhizobial genera. In comparison to the common *nodABC* genes, *nodZ* and *nolL* were probably later acquisitions in the nodulation process. In fast-growing rhizobia these genes may have favored efficient nodulation in legume hosts normally associated with *Bradyrhizobium* strains, like *Glycine* and *Pachyrhizus* for *nodZ*-bearing sinorhizobia. Analogously, transfer in the laboratory of *nodZ* genes to *R. leguminosarum* allowed the transconjugants to form nodules in soybean (López-Lara et al., 1996). Some *Lotus* species nodulating with mesorhizobia which possess *nodZ* genes also establish efficient symbiosis with *Bradyrhizobium* spp. (Vance et al., 1987). Finally, although *Rhizobium* strains are the preferred symbiont of *P. vulgaris*, it has been suggested that nodulation with *Bradyrhizobium* is the ancestral condition of the *Phaseolus* genus (Parker, 2002).

Acknowledgments

To PAPIIT Grant IN205412 from UNAM. To Fundación Bancomer FBBVA for financial support. To David Durán, Julio Martínez and Marco A. Rogel for their valuable help. To M. Dunn for critically reading the paper.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.03.003.

References

Andam, C.P., Mondo, S.J., Parker, M.A., 2007. Monophyly of *nodA* and *nifH* genes across Texan and Costa Rican populations of *Cupriavidus* nodule symbionts. Appl. Environ. Microbiol. 73, 4686–4690.

Barcellos, F.G., Menna, P., da Silva Batista, J.S., Hungria, M., 2007. Evidence of horizontal transfer of symbiotic genes from a *Bradyrhizobium japonicum* inoculant strain to indigenous diazotrophs *Sinorhizobium* (*Ensifer*) *fredii* and *Bradyrhizobium elkanii* in a Brazilian Savannah soil. Appl. Environ. Microbiol. 73, 2635–2643.

Berck, S., Perret, X., Quesada-Vincens, D., Promé, J., Broughton, W.J., Jabbouri, S., 1999. NolL of *Rhizobium* sp. strain NGR234 is required for O-acetyltransferase activity. J. Bacteriol. 181, 957–964.

Bontemps, C., Elliott, G.N., Simon, M.F., Dos Reis Jr., F.B., Gross, E., Lawton, R.C., Neto, N.E., de Fátima Loureiro, M., De Faria, S.M., Sprent, J.I., James, E.K., Young, J.P., 2010. *Burkholderia* species are ancient symbionts of legumes. Mol. Ecol. 19, 44–52.

Carlson, R.W., Price, N.P., Stacey, G., 1994. The biosynthesis of rhizobial lipo-oligosaccharide nodulation signal molecules. Mol. Plant-Microbe Interact. 7, 684–695.

Chang, Y.L., Wang, J.Y., Wang, E.T., Liu, H.C., Sui, X.H., Chen, W.X., 2011. *Bradyrhizobium lablabi* sp. nov., isolated from effective nodules of *Lablab purpureus* and *Arachis hypogaea*. Int. J. Syst. Evol. Microbiol. 61, 2496–2502.

Chen, W.M., Moulin, L., Bontemps, C., Vandamme, P., Béna, G., Boivin-Masson, C., 2003. Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in nature. J. Bacteriol. 185, 7266–7272.

Cohn, J., Stokkermans, T., Kolli, V.K., Day, R.B., Dunlap, J., Carlson, R., Hughes, D., Peters, N.K., Stacey, G., 1999. Aberrant nodulation response of *Vigna umbellata* to a *Bradyrhizobium japonicum* NodZ mutant and nodulation signals. Mol. Plant Microbe Interact. 12, 766–773.

Corvera, A., Promé, D., Promé, J.C., Martínez-Romero, E., Romero, D., 1999. The *nolL* gene from *Rhizobium etli* determines nodulation efficiency by mediating the acetylation of the fucosyl residue in the nodulation factor. Mol. Plant-Microbe Interact. 12, 236–246.

Debellé, F., Rosenberg, C., Vasse, J., Maillet, F., Martinez, E., Dénarié, J., Truchet, G., 1986. Assignment of symbiotic developmental phenotypes to common and specific nodulation (*nod*) genetic loci of *Rhizobium meliloti*. J. Bacteriol. 168, 1075–1086.

Dénarié, J., Debellé, F., Promé, J.C., 1996. *Rhizobium* lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. Ann. Rev. Biochem. 65, 503–535.

D'Haeze, W., Van Montagu, M., Promé, J.C., Holsters, M., 1999. Carbamoylation of azorhizobial Nod factors is mediated by NodU. Mol. Plant-Microbe Interact. 12, 68–73.

Djordjevic, S.P., Chen, H., Batley, M., Redmond, J.W., Rolfe, B.G., 1987. Nitrogen fixation ability of exopolysaccharide synthesis mutants of *Rhizobium* sp. strain NGR234 and *Rhizobium trifolii* is restored by the addition of homologous exopolysaccharides. J. Bacteriol. 169, 53–60.

Durán, D., Rey, L., Sánchez-Cañizares, C., Navarro, A., Imperial, J., Ruiz-Argueso, T., 2013. Genetic diversity of indigenous rhizobial symbionts of the *Lupinus mariae-josephae* endemism from alkaline-limed soils within its area of distribution in Eastern Spain. Syst. Appl. Microbiol. 36, 128–136.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Fournier, G.P., Gogarten, J.P., 2008. Evolution of acetoclastic methanogenesis in *Methanosarcina* via horizontal gene transfer from cellulolytic *Clostridia*. J. Bacteriol. 190, 1124–1127.

Freiberg, C., Fellay, R., Bairoch, A., Broughton, W.J., Rosenthal, A., Perret, X., 1997. Molecular basis of symbiosis between *Rhizobium* and legumes. Nature 387, 394–401.

Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J.C., Jaubert, M., Simon, D., Cartieaux, F., Prin, Y., Bena, G., Hannibal, L., Fardoux, J., Kojadinovic, M., Vuillet, L., Lajus, A., Cruveiller, S., Rouy, Z., Mangenot, S., Segurens, B., Dossat, C., Franck, W.L., Chang, W.S., Saunders, E., Bruce, D., Richardson, P., Normand, P., Dreyfus, B., Pignol, D., Stacey, G., Emerich, D., Verméglio, A., Médigue, C., Sadowsky, M., 2007. Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. Science 316, 1307–1312.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Hassan, S., Mathesius, U., 2012. The role of flavonoids in root-rhizosphere signalling: opportunities and challenges for improving plant-microbe interactions. J. Exp. Bot. 63, 3429–3444.

Horvath, B., Kondorosi, E., John, M., Schmidt, J., Török, I., Györgypal, Z., Barabas, I., Wieneke, U., Schell, J., Kondorosi, A., 1986. Organization, structure and symbiotic function of *Rhizobium meliloti* nodulation genes determining host specificity for alfalfa. Cell 46, 335–343.

Jones, K.M., 2012. Increased production of the exopolysaccharide succinoglycan enhances *Sinorhizobium meliloti* 1021 symbiosis with the host plant *Medicago truncatula*. J. Bacteriol. 194, 4322–4331.

Jourand, P., Giraud, E., Béna, G., Sy, A., Willems, A., Gillis, M., Dreyfus, B., de Lajudie, P., 2004. *Methylobacterium nodulans* sp. nov., for a group of aerobic, facultatively methylotrophic, legume root-nodule-forming and nitrogen-fixing bacteria. Int. J. Syst. Evol. Microbiol. 54, 2269–2273.

Laeremans, T., Snoeck, C., Mariën, J., Verreth, C., Martínez-Romero, E., Promé, J.C., Vanderleyden, J., 1999. *Phaseolus vulgaris* recognizes *Azorhizobium caulinodans* Nod factors with a variety of chemical substituents. Mol. Plant-Microbe Interact. 12, 820–824.

Lamrabet, Y., Bellogín, R.A., Cubo, T., Espuny, R., Gil, A., Krishnan, H.B., Megias, M., Ollero, F.J., Pueppke, S.G., Ruiz-Sainz, J.E., Spaink, H.P., Tejero-Mateo, P., Thomas-Oates, J., Vinardell, J.M., 1999. Mutation in GDP-fucose synthesis genes of *Sinorhizobium fredii* alters Nod factors and significantly decreases competitiveness to nodulate soybeans. Mol. Plant-Microbe Interact. 12, 207–217.

Lee, K.B., De Backer, P., Aono, T., Liu, C.T., Suzuki, S., Suzuki, T., Kaneko, T., Yamada, M., Tabata, S., Kupfer, D.M., Najar, F.Z., Wiley, G.B., Roe, B., Binnewies, T.T., Ussery, D.W., D'Haeze, W., Herder, J.D., Gevers, D., Vereecke, D., Holsters, M., Oyaizu, H., 2008. The genome of the versatile nitrogen fixer *Azorhizobium caulinodans* ORS571. BMC Genomics 9, 271.

Lindström, K., Murwira, M., Willems, A., Altier, N., 2010. The biodiversity of beneficial microbe-host mutualism: the case of rhizobia. Res. Microbiol. 161, 453–463.

López-Lara, I.M., Blok-Tip, L., Quinto, C., Garcia, M.L., Stacey, G., Bloemberg, G.V., Lamers, G.E., Lugtenberg, B.J., Thomas-Oates, J.E., Spaink, H.P., 1996. *NodZ* of *Bradyrhizobium* extends the nodulation host range of *Rhizobium* by adding a fucosyl residue to nodulation signals. Mol. Microbiol. 21, 397–408.

López-López, A., Negrete-Yankelevich, S., Rogel, M.A., Ormeño-Orrillo, E., Martínez, J., Martínez-Romero, E., 2013. Native bradyrhizobia from Los Tuxtlas in Mexico are symbionts of *Phaseolus lunatus* (Lima bean). Syst. Appl. Microbiol. 36, 33–38.

Marie, C., Broughton, W.J., Deakin, W.J., 2001. *Rhizobium* type III secretion systems: legume charmers or alarmers? Curr. Opin. Plant Biol. 4, 336–342.

Martinez, E., Laeremans, T., Poupot, R., Rogel, M.A., Lopez, L., Garcia, F., Vanderleyden, J., Prome, J.C., Lara, F., 1995. Nod metabolites and other compounds excreted by *Rhizobium* spp. In: Tikhonovich, I.A., Provorov, N.A., Romanov, V.I., Newton, W.E. (Eds.), Nitrogen fixation: fundamentals and applications. Proceedings of the 10th International Congress on Nitrogen Fixation, St. Petersburg, pp. 281–286.

Martínez-Romero, E., 2009. Coevolution in *Rhizobium*-legume symbiosis? DNA Cell Biol. 28, 361–370.

Martínez-Romero, J., Ormeño-Orrillo, E., Rogel, M.A., López-López, A., Martínez-Romero, E., 2010. Trends in rhizobial evolution and some taxonomic remarks. In: Pontarotti, P. (Ed.), Evolutionary Biology – Concepts, Molecular and Morphological Evolution. Springer, Berlin Heidelberg, pp. 301–315.

Menna, P., Hungria, M., 2011. Phylogeny of nodulation and nitrogen-fixation genes in *Bradyrhizobium*: supporting evidence for the theory of monophyletic origin, and spread and maintenance by both horizontal and vertical transfer. Int. J. Syst. Evol. Microbiol. 61, 3052–3067.

Mergaert, P., D'Haeze, W., Fernández-López, M., Geelen, D., Goethals, K., Promé, J.C., Van Montagu, M., Holsters, M., 1996. Fucosylation and arabinosylation of Nod factors in *Azorhizobium caulinodans*: involvement of *nolK*, *nodZ* as well as *noeC* and/or downstream genes. Mol. Microbiol. 21, 409–419.

Moulin, L., Béna, G., Boivin-Masson, C., Stepkowski, T., 2004. Phylogenetic analyses of symbiotic nodulation genes support vertical and lateral gene co-transfer within the *Bradyrhizobium* genus. Mol. Phylogenet. Evol. 30, 720–732.

Ormeño-Orrillo, E., Vinuesa, P., Zúñiga-Dávila, D., Martínez-Romero, E., 2006. Molecular diversity of native bradyrhizobia isolated from Lima bean (*Phaseolus lunatus* L.) in Peru. Syst. Appl. Microbiol. 29, 253–262.

Pacios Bras, C., Jordá, M.A., Wijfjes, A.H., Harteveld, M., Stuurman, N., Thomas-Oates, J.E., Spaink, H.P., 2000. A *Lotus japonicus* nodulation system based on heterologous expression of the fucosyl transferase NodZ and the acetyl transferase NolL in *Rhizobium leguminosarum*. Mol. Plant Microbe Interact. 13, 475–479.

Parker, M.A., 2002. Bradyrhizobia from wild *Phaseolus*, *Desmodium*, and *Macroptilium* species in Northern Mexico. Appl. Environ. Microbiol. 68, 2044–2048.

Perret, X., Staehelin, C., Broughton, W.J., 2000. Molecular basis of symbiotic promiscuity. Microbiol. Mol. Biol. Rev. 64, 180–201.

Posada, D., 2008. JModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.

Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.

Quesada-Vincens, D., Fellay, R., Nasim, T., Viprey, V., Burger, U., Prome, J.C., Broughton, W.J., Jabbouri, S., 1997. *Rhizobium* sp. strain NGR234 NodZ protein is a fucosyltransferase. J. Bacteriol. 179, 5087–5093.

Quinto, C., Wijfjes, A.H., Bloemberg, G.V., Blok-Tip, L., López-Lara, I.M., Lugtenberg, B.J., Thomas-Oates, J.E., Spaink, H.P., 1997. Bacterial nodulation protein NodZ is a chitin oligosaccharide fucosyltransferase which can also recognize related substrates of animal origin. Proc. Natl. Acad. Sci. USA 94, 4336–4341.

Ramírez-Bahena, M.H., Peix, A., Rivas, R., Camacho, M., Rodríguez-Navarro, D.N., Mateos, P.F., Martínez-Molina, E., Willems, A., Velázquez, E., 2009. *Bradyrhizobium pachyrhizi* sp. nov. and *Bradyrhizobium jicamae* sp. nov., isolated from effective nodules of *Pachyrhizus erosus*. Int. J. Syst. Evol. Microbiol. 59, 1929–1934.

Rivas, R., Velázquez, E., Willems, A., Vizcaíno, N., Subba-Rao, N.S., Mateos, P.F., Gillis, M., Dazzo, F.B., Martínez-Molina, E., 2002. A new species of *Devosia* that forms a unique nitrogen-fixing root-nodule symbiosis with the aquatic legume *Neptunia natans* (L.f.) druce. Appl. Environ. Microbiol. 68, 5217–5222.

Rodpothong, P., Sullivan, J.T., Songsrirote, K., Sumpton, D., Cheung, K.W., Thomas-Oates, J., Radutoiu, S., Stougaard, J., Ronson, C.W., 2009. Nodulation gene mutants of *Mesorhizobium loti* R7A-*nodZ* and *nolL* mutants have host-specific phenotypes on *Lotus* spp. Mol. Plant-Microbe Interact. 22, 1546–1554.

Rogel, M.A., Hernández-Lucas, I., Kuykendall, L.D., Balkwill, D.L., Martinez-Romero, E., 2001. Nitrogen-fixing nodules with *Ensifer adhaerens* harboring *Rhizobium tropici* symbiotic plasmids. Appl. Environ. Microbiol. 67, 3264–3268.

Rogel, M.A., Ormeño-Orrillo, E., Martinez-Romero, E., 2011. Symbiovars in rhizobia reflect bacterial adaptation to legumes. Syst. Appl. Microbiol. 34, 96–104.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Sanjuan, J., Carlson, R.W., Spaink, H.P., Bhat, U.R., Barbour, W.M., Glushka, J., Stacey, G., 1992. A 2-O-methylfucose moiety is present in the lipo-oligosaccharide nodulation signal of *Bradyrhizobium japonicum*. Proc. Natl. Acad. Sci. USA 89, 8789–8793.

Scott, D.B., Young, C.A., Collins-Emerson, J.M., Terzaghi, E.A., Rockman, E.S., Lewis, P.E., Pankhurst, C.E., 1996. Novel and complex chromosomal arrangement of *Rhizobium loti* nodulation genes. Mol. Plant Microbe Interact. 9, 187–197.

Servín-Garcidueñas, L.E., Rogel, M.A., Ormeño-Orrillo, E., Delgado-Salinas, A., Martínez, J., Sánchez, F., Martínez-Romero, E., 2012. Genome sequence of *Rhizobium* sp. strain CCGE510, a symbiont isolated from nodules of the endangered wild bean *Phaseolus albescens*. J. Bacteriol. 194, 6310.

Stacey, G., 1995. *Bradyrhizobium japonicum* nodulation genetics. FEMS Microbiol. Lett. 127, 1–9.

Stacey, G., Luka, S., Sanjuan, J., Banfalvi, Z., Nieuwkoop, A.J., Chun, J.Y., Forsberg, L.S., Carlson, R., 1994. *nodZ*, a unique host-specific nodulation gene, is involved in the fucosylation of the lipooligosaccharide nodulation signal of *Bradyrhizobium japonicum*. J. Bacteriol. 176, 620–633.

Staehelin, C., Schultze, M., Kondorosi, E., Mellor, R.B., Boiler, T., Kondorosi, A., 1994. Structural modifications in *Rhizobium meliloti* Nod factors influence their stability against hydrolysis by root chitinases. Plant J. 5, 319–330.

Steenkamp, E.T., Stepkowski, T., Przymusiak, A., Botha, W.J., Law, I.J., 2008. Cowpea and peanut in southern Africa are nodulated by diverse *Bradyrhizobium* strains harboring nodulation genes that belong to the large pantropical clade common in Africa. Mol. Phylogenet Evol. 48, 1131–1144.

Stepkowski, T., Swiderska, A., Miedzinska, K., Czaplińska, M., Swiderski, M., Biesiadka, J., Legocki, A.B., 2003. Low sequence similarity and gene content of symbiotic clusters of *Bradyrhizobium* sp. WM9 (*Lupinus*) indicate early divergence of "lupin" lineage in the genus *Bradyrhizobium*. Anton. Leeuw. 84, 115–124.

Stepkowski, T., Moulin, L., Krzyzańska, A., McInnes, A., Law, I.J., Howieson, J., 2005. European origin of *Bradyrhizobium* populations infecting lupins and serradella in soils of Western Australia and South Africa. Appl. Environ. Microbiol. 71, 7041–7052.

Stepkowski, T., Hughes, C.E., Law, I.J., Markiewicz, L., Gurda, D., Chlebicka, A., Moulin, L., 2007. Diversification of lupine *Bradyrhizobium* strains: evidence from nodulation gene trees. Appl. Environ. Microbiol. 73, 3254–3264.

Sullivan, J.T., Ronson, C.W., 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proc. Natl. Acad. Sci. USA 95, 5145–5149.

Vajda, V., McLoughlin, S., 2004. Fungal proliferation at the Cretaceous–Tertiary boundary. Science 303, 1489–1490.

Valverde, A., Velázquez, E., Fernández-Santos, F., Vizcaíno, N., Rivas, R., Mateos, P.F., Martínez-Molina, E., Igual, J.M., Willems, A., 2005. *Phyllobacterium trifolii* sp. nov., nodulating *Trifolium* and *Lupinus* in Spanish soils. Int. J. Syst. Evol. Microbiol. 55, 1985–1989.

Vance, C.P., Reibach, P.H., Pankhurst, C.E., 1987. Symbiotic properties of *Lotus pedunculatus* root nodules induced by *Rhizobium loti* and *Bradyrhizobium* sp. (*Lotus*). Physiol. Plant 69, 435–442.

Xia, X., Xie, Z., Salemi, M., Chen, L., Wang, Y., 2003. An index of substitution saturation and its application. Mol. Phylogenet. Evol. 26, 1–7.

Artículo:

Ormeño-Orrillo E, Servín-Garcidueñas LE, Rogel MA, González V, Peralta H, Mora J, Martínez-Romero J, Martínez-Romero E. 2014. Taxonomy of rhizobia and agrobacteria from the *Rhizobiaceae* family in light of genomics. Syst. Appl. Microbiol. 2014 Dec. 27. pii: S0723-2020(14)00190-8.

En este trabajo se presenta un análisis amplio sobre la conservación de secuencias entre genomas de diversas cepas y especies de rizobios. Presentamos una revisión de la taxonomía de la familia *Rhizobiaceae*, fundamentalmente en métricas y parámetros derivados de comparaciones de secuencias genómicas como ANI (*Average Nucleotide Identity*) e *in silico* DDH (*DNA-DNA hybridization*). Se pudieron identificar dos superclados dentro de la familia *Rhizobiaceae* que corresponden a los géneros *Rhizobium*/*Agrobacterium* y *Shinella*/*Ensifer*. Dentro del superclado *Rhizobium*/*Agrobacterium* logramos reconocer cuatro grupos principales que corresponden a géneros diferentes. Los resultados completos se presentan en la revisión.

Systematic and Applied Microbiology 38 (2015) 287–291

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Taxonomy of rhizobia and agrobacteria from the *Rhizobiaceae* family in light of genomics

Ernesto Ormeño-Orrillo, Luis E. Servín-Garcidueñas, Marco A. Rogel, Víctor González, Humberto Peralta, Jaime Mora, Julio Martínez-Romero, Esperanza Martínez-Romero*

Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

ARTICLE INFO

Keywords:
Rhizobia
Root nodule bacteria
Systematics
Phylogeny
Genome
Next generation sequencing

ABSTRACT

Phylogenomic analyses showed two major superclades within the family *Rhizobiaceae* that corresponded to the *Rhizobium*/*Agrobacterium* and *Shinella*/*Ensifer* groups. Within the *Rhizobium*/*Agrobacterium* group, four highly supported clades were evident that could correspond to distinct genera. The *Shinella*/*Ensifer* group encompassed not only the genera *Shinella* and *Ensifer* but also a separate clade containing the type strain of *Rhizobium giardinii*. *Ensifer adhaerens* (Casida A^T) was an outlier within its group, separated from the rest of the *Ensifer* strains. The phylogenomic analysis presented provided support for the revival of *Allorhizobium* as a *bona fide* genus within the *Rhizobiaceae*, the distinctiveness of *Agrobacterium* and the recently proposed *Neorhizobium* genus, and suggested that *R. giardinii* may be transferred to a novel genus. Genomics has provided data for defining bacterial-species limits from estimates of average nucleotide identity (ANI) and *in silico* DNA–DNA hybridization (DDH). ANI reference values are becoming the gold standard in rhizobial taxonomy and are being used to recognize novel rhizobial lineages and species that seem to be biologically coherent, as shown in this study.

Introduction

Rhizobia are soil and rhizospheric bacteria that may form nitrogen fixing symbioses in leguminous plants allowing their growth in poor nitrogen soils. Thus, rhizobia have been considered as bio-fertilizers and have been used as inoculants in agriculture for over 120 years. Rhizobial genetic diversity, as well as their plant–bacteria molecular interactions, has been well studied. In 1991, Graham et al. [18] published a set of recommendations for the description of novel rhizobial species on the "basis of both phylogenetic and phenotypic traits" using "genomic relationships to the greatest degree possible" and as "the culmination of considerable research". A large number of species have been reported since, based mainly on polyphasic analysis using a number of molecular-marker phylogenies, DNA–DNA hybridization (DDH) results and the description of different distinctive phenotypic features. Studies based on molecular marker sequences represented a significant advance in rhizobial taxonomy and have been included in most studies. Newly described species are periodically revised by the International Taxonomy Subcommittee on *Agrobacterium* and *Rhizobium* [24,25]. Several reviews on rhizobial taxonomy have been published [41,52,55] but none have been specifically oriented toward genomics.

The following genera within the family *Rhizobiaceae* include rhizobial members: *Rhizobium*, *Ensifer* (former *Sinorhizobium*), *Agrobacterium* and *Shinella* [5]. *Agrobacterium* includes tumor-forming bacteria as well as nitrogen-fixing nodule bacteria. An additional genus of the family *Rhizobiaceae*, *Carbophilus* [5], has not been described as containing nodule bacteria. A characteristic of rhizobia belonging to the family *Rhizobiaceae* and agrobacteria is their genome organization in multireplicons [17,19,21,26]. Furthermore, phenotypic distinctive characteristics in rhizobia may be encoded in extrachromosomal replicons (ERs) [33], a feature not normally recognized in novel species descriptions. In *Rhizobium*, *Ensifer* and *Agrobacterium*, almost half of the genome may be contained in ERs (reviewed in [26]), and some ERs even have roles in rhizobial growth rate and survival [4,16,19,20]. Two types of ERs have been recognized: plasmids and chromids [19]. ERs that carry "essential" genes with conserved gene sequences and sharing similar GC content and codon usage with the chromosome have been named chromids [19]. Chromids have been proposed as characteristic of a genus and contain many genus-specific genes [26]. Taxonomic phenotypic characteristics are encoded in chromids in *Rhizobium* [33]. Interestingly, chromids carry many genes

* Corresponding author. Tel.: +52 777 3131697.
E-mail address: emartine@ccg.unam.mx (E. Martínez-Romero).

http://dx.doi.org/10.1016/j.syapm.2014.12.002
0723-2020/

that are highly expressed by rhizobia on plant roots [33]. On the other hand, plasmids are highly variable and confer adaptive traits, such as nodulation and nitrogen fixation in legumes [6,8,19,26,29,38,43,49,51], or they may be transferred between bacteria [30,31,40].

Genomic impact on rhizobial taxonomy

Genomics has revolutionized microbiology and is having a significant impact on taxonomy. For many years, results from DDH experiments were the basis for circumscribing prokaryotic species [46]. However, alternatives for estimating DNA relatedness, such as whole-genome average nucleotide identity (ANI) [23] and *in silico* DDH [2], are currently much better than wet lab DDH which has been shown to produce highly variable results from lab to lab and from different DNA samples [27]. Additionally, the G+C content normally reported in novel bacteria descriptions may be calculated accurately from genomic data.

The novel quantitative genomic analyses are beginning to be used in rhizobial taxonomy. Species descriptions where ANI and/or *in silico* DDH were used to support or complement wet lab DDH values have been published [9,10,13,28]. Likewise, limits obtained from ANI and *in silico* DDH estimates have led to the discovery of novel rhizobial lineages [27,37]. Lack of genome sequences for most type strains of rhizobia is a limiting factor for the use of these novel approaches, although the reducing cost of whole genome sequencing will ease this restriction. As an example, two recent studies coupled genome sequencing of the rhizobia being characterized with that of all related type strains, which allowed complete replacement of wet lab DDH with ANI [14,15].

ANI values derived from only the conserved core genes of a group (referred to as ANIo) have also been proposed as a replacement for wet lab DDH, with approximately 96% ANIo corresponding to 70% DDH [22]. A minimum of three but a recommended number of six to eight genes can give a good estimate of ANIo [22]. Consequently, ANI values based on concatenated sequences of a few partial sequences of conserved core genes are being used to delineate putative rhizobial species [1,11,39]. Nevertheless, care must be exercised in not assuming that values obtained with partial sequences will correspond exactly to ANIo, instead, intra- and inter-species identity values must be evaluated in order to find a suitable cut-off value for species delineation that is appropriate for the set of genes being used. Recently, a set of three novel conserved genes has been proposed as a suitable tool for rhizobial taxonomy because the concatenated partial sequences produced ANI that were closely correlated with whole genome ANI [56].

A phylogenomic view of rhizobia and agrobacteria within the *Rhizobiaceae*

Besides providing quantitative values for species delineation, whole genomes allow the reconstruction of phylogenetic trees based on hundreds or thousands of genes that depict evolutionary relationships better than phylogenies based on a few markers including 16S rRNA genes. To date, 29 complete and 141 draft whole genome sequences (WGS) from members of the family *Rhizobiaceae* are available from the GenBank database (Supplementary Table S1). These genomes include 23 type strains, three of which are completely sequenced. Additionally, one complete and three draft WGS genomes sequenced at CCG-UNAM were included in the analysis (Supplementary Table S1). A total of 166 (66%) of the strains had genomes encoding *nodC*. Most strains lacking this gene are labeled as agrobacteria.

We checked the identity of all sequenced type strains by comparing their genomes against partial sequences of genes previously obtained for the same strains available from GenBank, and two anomalies were found. The *A. radiobacter* DSM30147^T genome (accession number ASXY01, Bioproject PRJNA212112) had identical sequences to several previously reported *A. radiobacter* DSM30147^T genes (*aptD*, *rpoB*, *mutS*, *gyrB*, *gltD*, *glnII*) but showed only 90–97% identity with others (*rpoD*, *chvA*, *hrcA*). The *R. gallicum* R602sp^T genome (accession number ARDC01, BioProject PRJNA169700) had divergent sequences in all the genes compared, which clustered within the *R. leguminosarum* clade (data not shown). Both genomes were excluded from further analyses, as they did not correspond to the designated type strains.

Except for one comparison, type strain genomes shared a maximum ANI value of 92%, thus supporting the proposed cut-off level of 95% as a species delineation threshold [23]. *R. gallicum* R602^T (newly sequenced at CCG) and *R. mongolense* USDA 1844^T shared an ANI value of 95.1% that validates their previously proposed synonymy [44]. Based on a 95% ANI threshold, the 172 sequenced strains would represent 77 genospecies (Supplementary Table S1). Given the scarcity of sequenced type strains, most of these genospecies could not be ascribed to described taxa solely by ANI and so were assigned arbitrary labels (GS1-G48) in Supplementary Table S1. To date, there are 27 genomes available from different *Ensifer meliloti* strains and eight genomes from distinct *E. fredii* strains, while the remaining geno(species) have from 1 to 6 sequenced strains. We used a sample of 113 genome sequences representing all possible (geno)species in order to construct a genome-based phylogeny with the aim of shedding light on uncertainties or controversies in the taxonomy of several clades within the family *Rhizobiaceae*. Due to the unreliable classification or naming of many sequenced strains (Supplementary Table S1) we chose to include a species designation only for type strains or strains that had been assigned to a known species on the basis of DNA-DNA hybridization or ANI analyses. As shown in Fig. 1, two major superclades were observed within the family *Rhizobiaceae*, which corresponded to the *Rhizobium*/*Agrobacterium* and *Shinella*/*Ensifer* groups.

Within the *Rhizobium*/*Agrobacterium* group, several highly supported clades were evident. One clade included *Agrobacterium* biovar 1 strains, as well as the type strains of *Agrobacterium rubi* and *Agrobacterium larrymoorei*. This clade, referred to here as *Agrobacterium sensu stricto*, had been previously revealed by *recA* sequence analysis [7] and includes strains whose genomes encode a protelomerase, which are characterized by possessing a linear replicon [36].

A second clade included strains of the recently proposed genus *Neorhizobium* [32]. This clade was previously known as the "*Rhizobium galegae* complex". The genome phylogeny supported the proposal of Mousavi et al. [32] for including *Rhizobium vignae* in *Neorhizobium*. However, ANI values between *R. vignae* and *N. galegae* strains were lower than 91%, indicating that *R. vignae* should not be included in the *N. galegae* species as suggested by Mousavi et al. [32] and must therefore be referred to as *Neorhizobium vignae*.

A third clade included the type strains of *Rhizobium undicola* (former *Allorhizobium undicola* [12,54]), as well as strain S4 of *Agrobacterium vitis*. The isolated position of this clade in relation to *Agrobacterium sensu stricto* and *Neorhizobium* could support the revival of *Allorhizobium* as a genus within the *Rhizobiaceae*, as has been recently suggested [36], and which includes the species *Allorhizobium vitis* (formerly *Agrobacterium vitis*) and *Allorhizobium taibaishanense* (former *Rhizobium taibaishanense*) [32].

A fourth clade included species closely related to *Rhizobium leguminosarum*, the type species of the genus *Rhizobium*, hence, we referred to this clade as *Rhizobium sensu stricto*. Within this clade, distinct groups of closely related species were observed. Hairy-root forming bacteria, originally described as *Agrobacterium rhizogenes* (biovar 2 agrobacteria) were found within the

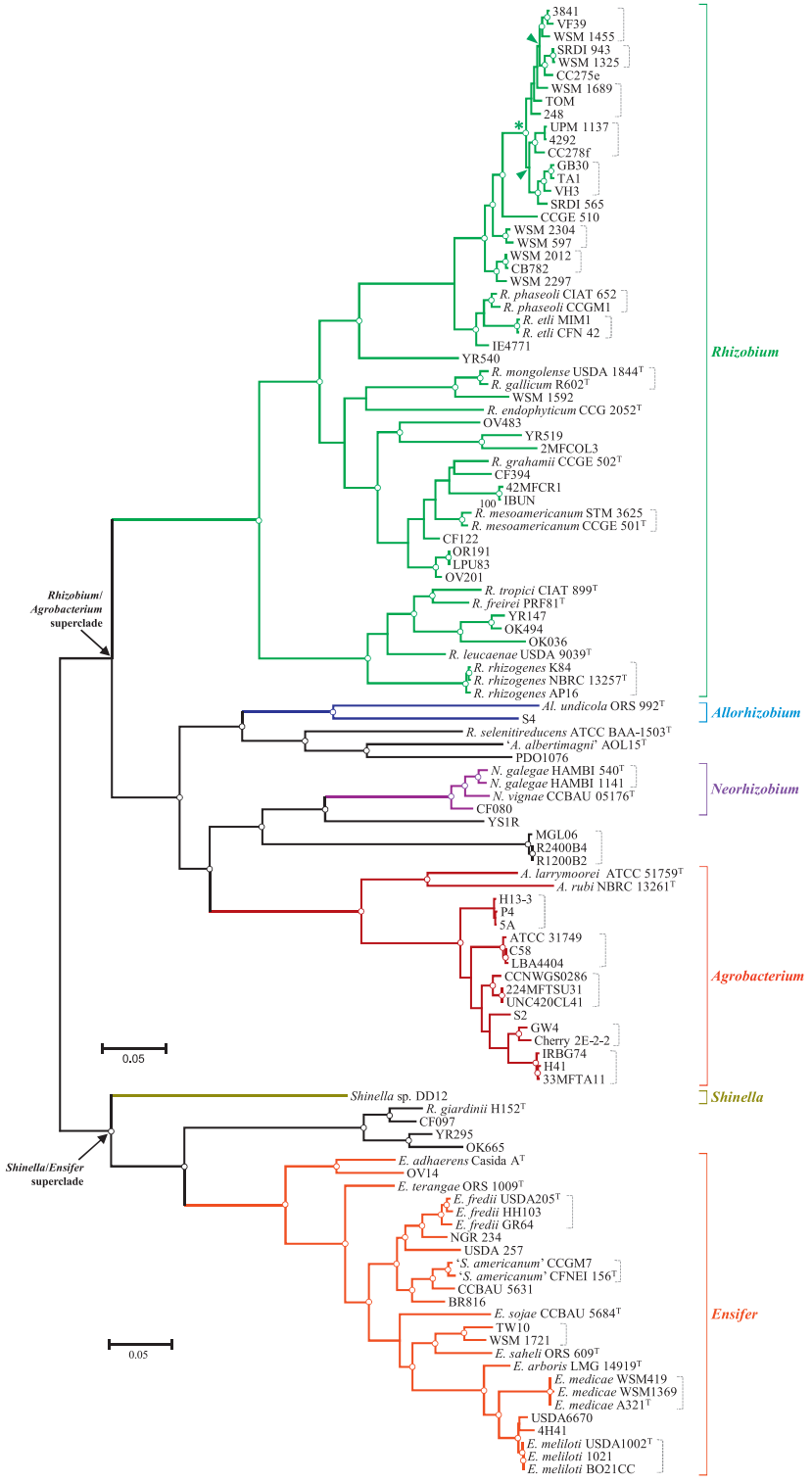


Figure 1. Phylogenomic analysis showing the evolutionary relationships between 113 sequenced strains from the family *Rhizobiaceae*. Putative genus-level clades are indicated with different colors. Strains forming species-level clades based on ANI, with a 95% cut-off level, are grouped with square brackets. Species designation was included only for type strains or strains assigned to a known species by DNA–DNA hybridization or ANI analyses. Type strains are indicated with a superscript T letter. Genus abbreviations: R, *Rhizobium*; A, *Agrobacterium*; Al, *Allorhizobium*; S, *Sinorhizobium*; E, *Ensifer*, N, *Neorhizobium*. The maximum likelihood phylogeny was reconstructed with FastTree 2 [35] using a concatenated alignment of the most discriminative amino acid positions of 384 proteins conserved in the chromosomes of all completely sequenced genomes identified with PhyloPhlAn [42]. Shimodaira-Hasegawa-like local support values for all nodes were ≥70% except for the two nodes marked with triangles. All genus-level clades had support values of 100%. Clustering of strains below nodes marked with circles was also observed in an independent phylogenomic analysis performed with AMPHORA [53] using 31 universally conserved and mostly chromosomally located proteins. Genome or replicon sequence accession numbers are indicated in Supplementary Table S1. The scale bar represents the estimated number of amino acid changes per site for a unit of branch length. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tropici group, thus their designation as *Rhizobium rhizogenes* is fully supported [50,54]. Inside the phaseoli-etli-leguminosarum (PEL) group, different genomic lineages were observed, including some that were previously identified and which probably correspond to distinct species [37]. A highly supported subgroup marked with an asterisk in Figure 1 included several closely related *R. leguminosarum* lineages showing ANI values that were borderline for species delineation (94–95%). This observation may be related to the findings of Tian et al. [48] that *R. leguminosarum* could be composed of sublineages that showed significant levels of genetic flow "preventing unlimited divergence" but "being uncommon enough to allow the creation and persistence of diversifying sublineages". Other strains designated as *R. leguminosarum*, such as WSM2304, seemed to represent distinct species.

In the *Shinella*/*Ensifer* group, three highly supported clades were observed. The sole sequenced strain of *Shinella* formed a clade of its own, and a second clade could be equated to the genus *Ensifer*. It is worth noting that the type species of this genus, *Ensifer adhaerens* (Casida A^T), was located in an outlier position separated from the rest of the *Ensifer* (formerly *Sinorhizobium*) strains (Fig. 1). Strains USDA 257 and NGR 234, sometimes ascribed to *Ensifer fredii*, clustered independently and showed low ANI values (<92%) with the *E. fredii* type strain and other tropical species, such as '*Sinorhizobium americanum*'. NGR 234 strain was proposed to correspond to a separate species on the basis of *nolR* polymorphisms [29].

Rhizobium giardinii $H152^T$ and four other rhizobial genomes formed an independent clade more related to *Ensifer* than to *Rhizobium*/*Agrobacterium*. This observation is in agreement with the phylogenetic placement of *R. giardinii* in a recently reported MLSA analysis [32] and supports the notion that this species represents a novel genus. This may also be the case for several strains presently ascribed to *Rhizobium*, as well as '*A. albertimagni*' $AOL15^T$, which occupy isolated positions suggesting that they may represent as yet undescribed genera within the family *Rhizobiaceae*.

Implications of genome-based taxonomy (genomotaxonomy)

We can speculate whether the new high resolution or high definition (HD) taxonomy that splits closely related groups into different species based on genomic data is biologically sound. These new species have 16S rRNA gene sequences that are almost indistinguishable from their closest relatives. Interestingly, the members of each of these newly described and tightly circumscribed species from HD taxonomy are highly coherent with very similar common characteristics, as observed in *R. phaseoli* [27] and in the novel species from the tropici group [9,10,34]. Similar situations have arisen in taxonomic studies of other rhizobia, such as *Bradyrhizobium diazoefficiens* (formerly considered as *B. japonicum*) [13,45], and other bacterial genera such as *Klebsiella* [3] and *Bacillus* [47].

Besides providing consistent measurements of DNA relatedness for defining bacterial-species limits, genomics would help provide a better knowledge of the phenotypic distinctiveness of species. The phenotypic characteristics used to date have been criticized as a requirement for proposing novel species and their substitution for a genomotaxonomy approach has been claimed where genomic data can be used to predict stable phenotypes [33]. The first steps have been taken in this direction for pathogenic bacteria such as vibrios [12]. In the case of rhizobia with partitioned genomes, chromosome- or chromid-based phenotypes should be considered valid in taxonomic descriptions but not those encoded in unstable plasmids.

Genomic data also provides an opportunity for exploring topics such as speciation that, although not being part of taxonomy itself, is inherently related to it and can provide novel criteria for use in taxonomy. The basis of speciation has not been reviewed in rhizobia, and key genes, mechanisms and processes driving speciation remain to be described. Is self-recognition occurring among members of a single species and not between members of closely related species? Is genomic architecture related to speciation? These and other questions should be addressed in the future.

After our article was peer-reviewed, we were aware that Mousavi et al (Syst. Appl. Microbiol., In Press, doi:10.1016/j.syapm.2014.12.003) proposed the novel genus *Pararhizobium* composed of *R. giardinii* and related species, which agree with our conclusion of the distinctiveness of *R. giardinii*.

Acknowledgements

This work was supported by PAPIIT IN 205412 from DGAPA, UNAM. We acknowledge M. Dunn for his critical review of the manuscript. L.E. Servín-Garcidueñas is thankful to the Programa de Doctorado en Ciencias Biomédicas, UNAM and a fellowship from CONACyT, Mexico.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.syapm.2014.12.002.

References

[1] Aserse, A.A., Rasanen, L.A., Assefa, F., Hailemariam, A., Lindström, K. (2012) Phylogeny and genetic diversity of native rhizobia nodulating common bean (*Phaseolus vulgaris* L.) in Ethiopia. Syst. Appl. Microbiol. 35, 120–131.
[2] Auch, A.F., von Jan, M., Klenk, H.P., Goker, M. (2010) Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand. Genomic Sci. 2, 117–134.
[3] Brisse, S., Passet, V., Grimont, P.A. (2014) Description of *Klebsiella quasipneumoniae* sp. nov., a novel species isolated from human infections, with two subspecies *Klebsiella quasipneumoniae* subsp. *quasipneumoniae* subsp. nov. and *Klebsiella quasipneumoniae* subsp. *similipneumoniae* subsp. nov., and demonstration that *K. singaporensis* is a later heterotypic synonym of *K. variicola*. Int. J. Syst. Evol. Microbiol. 64, 3146–3152.
[4] Brom, S., Girard, L., Garcia-de los Santos, A., Sanjuan-Pinilla, J.M., Olivares, J., Sanjuan, J. (2002) Conservation of plasmid-encoded traits among bean-nodulating *Rhizobium* species. Appl. Environ. Microbiol. 68, 2555–2561.
[5] Carrareto Alves, L.M., de Souza, J.A.M., Varani, A.d.M., Lemos, E.G.d.M. (2014) The family *Rhizobiaceae*. In: Rosenberg, E., DeLong, E., Lory, S., Stackebrandt, E., Thompson, F. (Eds.), The Prokaryotes, Springer, Berlin/Heidelberg, pp. 419–437.
[6] Cervantes, L., Bustos, P., Girard, L., Santamaria, R., Davila, G., Vinuesa, P., Romero, D., Brom, S. (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 Microbiol. 11, 149.
[7] Costechareyre, D., Rhouma, A., Lavire, C., Portier, P., Chapulliot, D., Bertolla, F., Boubaker, A., Dessaux, Y., Nesme, X. (2010) Rapid and efficient identification of *Agrobacterium* species by *recA* allele analysis. Microb. Ecol. 60, 862–872.
[8] Crossman, L.C., Castillo-Ramirez, S., McAnnula, C., Lozano, L., Vernikos, G.S., Acosta, J.L., Ghazoui, Z.F., Hernandez-Gonzalez, I., Meakin, G., Walker, A.W., Hynes, M.F., Young, J.P., Downie, J.A., Romero, D., Johnston, A.W., Davila, G., Parkhill, J., Gonzalez, V. (2008) A common genomic framework for a diverse assembly of plasmids in the symbiotic nitrogen fixing bacteria. PLoS ONE 3, e2567.
[9] Dall'Agnol, R.F., Ribeiro, R.A., Delamuta, J.R., Ormeño-Orrillo, E., Rogel, M.A., Andrade, D.S., Martínez-Romero, E., Hungria, M. (2014) *Rhizobium paranaense* sp. nov., an effective N_2-fixing symbiont of common bean (*Phaseolus vulgaris* L.) with broad geographical distribution in Brazil. Int. J. Syst. Evol. Microbiol. 64, 3222–3229.
[10] Dall'agnol, R.F., Ribeiro, R.A., Ormeño-Orrillo, E., Rogel, M.A., Delamuta, J.R., Andrade, D.S., Martínez-Romero, E., Hungria, M. (2013) *Rhizobium freirei*, a symbiont of *Phaseolus vulgaris* very effective in fixing nitrogen. Int. J. Syst. Evol. Microbiol. 63, 4167–4173.
[11] Degefu, T., Wolde-meskel, E., Frostegård, Å. (2012) Phylogenetic multilocus sequence analysis identifies seven novel *Ensifer* genospecies isolated from a less-well-explored biogeographical region in East Africa. Int. J. Syst. Evol. Microbiol. 62, 2286–2295.
[12] de Lajudie, P., Laurent-Fulele, E., Willems, A., Torck, U., Coopman, R., Collins, M.D., Kersters, K., Dreyfus, B., Gillis, M. (1998) *Allorhizobium undicola* gen. nov., sp. nov., nitrogen-fixing bacteria that efficiently nodulate *Neptunia natans* in Senegal. Int. J. Syst. Evol. Microbiol. 48, 1277–1290.

[13] Delamuta, J.R.M., Ribeiro, R.A., Ormeño-Orrillo, E., Melo, I.S., Martínez-Romero, E., Hungria, M. (2013) Polyphasic evidence supporting the reclassification of *Bradyrhizobium japonicum* group Ia strains as *Bradyrhizobium diazoefficiens* sp. nov. Int. J. Syst. Evol. Microbiol. 63, 3342–3351.
[14] Durán, D., Rey, L., Mayo, J., Zúñiga-Dávila, D., Imperial, J., Ruiz-Argüeso, T., Martínez-Romero, E., Ormeño-Orrillo, E. (2014) *Bradyrhizobium paxllaeri* sp. nov. and *Bradyrhizobium icense* sp. nov., nitrogen-fixing rhizobial symbionts of Lima bean (*Phaseolus lunatus* L.) in Peru. Int. J. Syst. Evol. Microbiol. 64, 2072–2078.
[15] Durán, D., Rey, L., Navarro, A., Busquets, A., Imperial, J., Ruiz-Argüeso, T. (2014) *Bradyrhizobium valentinum* sp. nov., isolated from effective nodules of *Lupinus mariae-josephae*, a lupine endemic of basic-lime soils in Eastern Spain. Syst. Appl. Microbiol. 37, 336–341.
[16] García-de los Santos, A., Brom, S., Romero, D. (1996) *Rhizobium* plasmids in bacteria–legume interactions. World J. Microbiol. Biotechnol. 12, 119–125.
[17] González, V., Santamaría, R.I., Bustos, P., Hernández-González, I., Medrano-Soto, A., Moreno-Hagelsieb, G., Janga, S.C., Ramírez, M.A., Jiménez-Jacinto, V., Collado-Vides, J., Dávila, G. (2006) The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. Proc. Natl. Acad. Sci. U. S. A. 103, 3834–3839.
[18] Graham, P.H., Sadowsky, M.J., Keyser, H.H., Barnet, Y.M., Bradley, R.S., Cooper, J.E., Deley, D.J., Jarvis, B.D.W., Roslycky, E.B., Strijdom, B.W., Young, J.P.W. (1991) Proposed minimal standards for the description of new genera and species of root-nodulating and stem-nodulating bacteria. Int. J. Syst. Evol. Microbiol. 41, 582–587.
[19] Harrison, P.W., Lower, R.P.J., Kim, N.K.D., Young, J.P.W. (2010) Introducing the bacterial 'chromid': not a chromosome, not a plasmid. Trends Microbiol. 18, 141–148.
[20] Hynes, M.F., Quandt, J., O'Connell, M.P., Pühler, A. (1989) Direct selection for curing and deletion of *Rhizobium* plasmids using transposons carrying the *Bacillus subtilis sacB* gene. Gene 78, 111–120.
[21] Jumas-Bilak, E., Michaux-Charachon, S., Bourg, G., Ramuz, M., Allardet-Servent, A. (1998) Unconventional genomic organization in the alpha subgroup of the *Proteobacteria*. J. Bacteriol. 180, 2749–2755.
[22] Konstantinidis, K.T., Ramette, A., Tiedje, J.M. (2006) Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. Appl. Environ. Microbiol. 72, 7286–7293.
[23] Konstantinidis, K.T., Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 102, 2567–2572.
[24] Lindström, K., Martínez-Romero, E. (2007) International Committee on Systematics of Prokaryotes. Subcommittee on the taxonomy of Agrobacterium and Rhizobium: Minutes of the meeting, 23–24 July 2006, Århus, Denmark. Int. J. Syst. Evol. Microbiol. 57, 1365–1366.
[25] Lindström, K., Young, J.P.W. (2011) International Committee on Systematics of Prokaryotes. Subcommittee on the taxonomy of Agrobacterium and Rhizobium: Minutes of the meeting, 7 September 2010, Geneva, Switzerland. Int. J. Syst. Evol. Microbiol. 61, 3089–3093.
[26] López-Guerrero, M.G., Ormeño-Orrillo, E., Acosta, J.L., Mendoza-Vargas, A., Rogel, M.A., Ramírez, M.A., Rosenblueth, M., Martínez-Romero, J., Martínez-Romero, E. (2012) Rhizobial extrachromosomal replicon variability, stability and expression in natural niches. Plasmid 68, 149–158.
[27] López-Guerrero, M.G., Ormeño-Orrillo, E., Velázquez, E., Rogel, M.A., Acosta, J.L., Gónzalez, V., Martínez, J., Martínez-Romero, E. (2012) *Rhizobium etli* taxonomy revised with novel genomic data and analyses. Syst. Appl. Microbiol. 35, 353–358.
[28] López-López, A., Rogel-Hernández, M.A., Barois, I., Ortiz Ceballos, A.I., Martínez, J., Ormeño-Orrillo, E., Martínez-Romero, E. (2012) *Rhizobium grahamii* sp. nov., from nodules of *Dalea leporina*, *Leucaena leucocephala* and *Clitoria ternatea*, and *Rhizobium mesoamericanum* sp. nov., from nodules of *Phaseolus vulgaris*, siratro, cowpea and *Mimosa pudica*. Int. J. Syst. Evol. Microbiol. 62, 2264–2271.
[29] Lloret, L., Ormeño-Orrillo, E., Rincón, R., Martínez-Romero, J., Rogel-Hernández, M.A., Martínez-Romero, E. (2007) *Ensifer mexicanus* sp. nov. a new species nodulating *Acacia angustissima* (Mill.) Kuntze in Mexico. Syst. Appl. Microbiol. 30, 280–290.
[30] Martínez, E., Palacios, R., Sanchez, F. (1987) Nitrogen-fixing nodules induced by *Agrobacterium tumefaciens* harboring *Rhizobium phaseoli* plasmids. J. Bacteriol. 169, 2828–2834.
[31] Martínez-Romero, E. (2009) Coevolution in *Rhizobium*-legume symbiosis? DNA Cell Biol. 28, 361–370.
[32] Mousavi, S.A., Osterman, J., Wahlberg, N., Nesme, X., Lavire, C., Vial, L., Paulin, L., de Lajudie, P., Lindstrom, K. (2014) Phylogeny of the *Rhizobium–Allorhizobium–Agrobacterium* clade supports the delineation of *Neorhizobium* gen. nov. Syst. Appl. Microbiol. 37, 208–215.
[33] Ormeño-Orrillo, E., Martínez-Romero, E. (2013) Phenotypic tests in *Rhizobium* species description: an opinion and (a sympatric speciation) hypothesis. Syst. Appl. Microbiol. 36, 145–147.
[34] Ormeño-Orrillo, E., Menna, P., Gonzaga Almeida, L., Ollero, F.J., Nicolas, M.F., Pains Rodrigues, E., Shigueyoshi Nakatani, A., Silva Batista, J.S., Oliveira Chueire, L.M., Souza, R.C., Ribeiro Vasconcelos, A., Megías, M., Hungria, M., Martínez-Romero, E. (2012) Genomic basis of broad host range and environmental adaptability of *Rhizobium tropici* CIAT 899 and *Rhizobium* sp. PRF 81 which are used in inoculants for common bean (*Phaseolus vulgaris* L.). BMC Genomics 13, 735.
[35] Price, M.N., Dehal, P.S., Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS ONE 5, e9490.
[36] Ramírez-Bahena, M.H., Vial, L., Lassalle, F., Diel, B., Chapulliot, D., Daubin, V., Nesme, X., Muller, D. (2014) Single acquisition of protelomerase gave rise to speciation of a large and diverse clade within the *Agrobacterium*/*Rhizobium* supercluster characterized by the presence of a linear chromid. Mol. Phylogenet. Evol. 73, 202–207.
[37] Ribeiro, R.A., Ormeño-Orrillo, E., Dall'Agnol, R.F., Graham, P.H., Martínez-Romero, E., Hungria, M. (2013) Novel *Rhizobium* lineages isolated from root nodules of the common bean (*Phaseolus vulgaris* L.) in Andean and Mesoamerican areas. Res. Microbiol. 164, 740–748.
[38] Rincón-Rosales, R., Lloret, L., Ponce, E., Martínez-Romero, E. (2009) Rhizobia with different symbiotic efficiencies nodulate *Acaciella angustissima* in Mexico, including *Sinorhizobium chiapanecum* sp. nov. which has common symbiotic genes with *Sinorhizobium mexicanum*. FEMS Microbiol. Ecol. 67, 103–117.
[39] Rincón-Rosales, R., Villalobos-Escobedo, J.M., Rogel, M.A., Martínez, J., Ormeño-Orrillo, E., Martínez-Romero, E. (2013) *Rhizobium calliandrae* sp. nov., *Rhizobium mayense* sp. nov. and *Rhizobium jaguaris* sp. nov., rhizobial species nodulating the medicinal legume *Calliandra grandiflora*. Int. J. Syst. Evol. Microbiol. 63, 3423–3429.
[40] Rogel, M.A., Ormeño-Orrillo, E., Martínez Romero, E. (2011) Symbiovars in rhizobia reflect bacterial adaptation to legumes. Syst. Appl. Microbiol. 34, 96–104.
[41] Sawada, H., Kuykendall, L.D., Young, J.M. (2003) Changing concepts in the systematics of bacterial nitrogen-fixing legume symbionts. J. Gen. Appl. Microbiol. 49, 155–179.
[42] Segata, N., Bornigen, D., Morgan, X.C., Huttenhower, C. (2013) PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat. Commun. 4, 2304.
[43] Silva, C., Vinuesa, P., Eguiarte, L.E., Martinez-Romero, E., Souza, V. (2003) *Rhizobium etli* and *Rhizobium gallicum* nodulate common bean (*Phaseolus vulgaris*) in a traditionally managed milpa plot in Mexico: population genetics and biogeographic implications. Appl. Environ. Microbiol. 69, 884–893.
[44] Silva, C., Vinuesa, P., Eguiarte, L.E., Souza, V., Martinez-Romero, E. (2005) Evolutionary genetics and biogeographic structure of *Rhizobium gallicum sensu lato*, a widely distributed bacterial symbiont of diverse legumes. Mol. Ecol. 14, 4033–4050.
[45] Siqueira, A., Ormeño-Orrillo, E., Souza, R., Rodrigues, E., Almeida, L.G., Barcellos, F., Batista, J.S., Nakatani, A., Martínez-Romero, E., Vasconcelos, A.T., Hungria, M. (2014) Comparative genomics of *Bradyrhizobium japonicum* CPAC 15 and *Bradyrhizobium diazoefficiens* CPAC 7: elite model strains for understanding symbiotic performance with soybean. BMC Genomics 15, 420.
[46] Stackebrandt, E., Goebel, B.M. (1994) Taxonomic note: a place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. 44, 846–849.
[47] Stropko, S.J., Pipes, S.E., Newman, J. (2014) Genome-based reclassification of *Bacillus cibi* as a latter heterotypic synonym of *Bacillus indicus* and emended description of *Bacillus indicus*. Int. J. Syst. Evol. Microbiol. 64, 3804–3809.
[48] Tian, C.F., Young, J.P., Wang, E.T., Tamimi, S.M., Chen, W.X. (2010) Population mixing of *Rhizobium leguminosarum* bv. viciae nodulating *Vicia faba*: the role of recombination and lateral gene transfer. FEMS Microbiol. Ecol. 73, 563–576.
[49] Toledo, I., Lloret, L., Martinez-Romero, E. (2003) *Sinorhizobium americanum* sp. nov., a new *Sinorhizobium* species nodulating native *Acacia* spp. in Mexico. Syst. Appl. Microbiol. 26, 54–64.
[50] Velázquez, E., Palomo, J.L., Rivas, R., Guerra, H., Peix, A., Trujillo, M.E., García-Benavides, P., Mateos, P.F., Wabiko, H., Martínez-Molina, E. (2010) Analysis of core genes supports the reclassification of strains *Agrobacterium radiobacter* K84 and *Agrobacterium tumefaciens* AKE10 into the species *Rhizobium rhizogenes*. Syst. Appl. Microbiol. 33, 247–251.
[51] Wang, E.T., Rogel, M.A., Garcia-de los Santos, A., Martinez-Romero, J., Cevallos, M.A., Martinez-Romero, E. (1999) *Rhizobium etli* bv. mimosae, a novel biovar isolated from *Mimosa affinis*. Int. J. Syst. Bacteriol. 49, 1479–1491.
[52] Willems, A. (2006) The taxonomy of rhizobia: an overview. Plant Soil 287, 3–14.
[53] Wu, M., Eisen, J.A. (2008) A simple, fast, and accurate method of phylogenomic inference. Genome Biol. 9, R151.
[54] Young, J.M., Kuykendall, L.D., Martínez-Romero, E., Kerr, A., Sawada, H. (2001) A revision of Rhizobium Frank 1889, with an emended description of the genus, and the inclusion of all species of Agrobacterium Conn 1942 and Allorhizobium undicola de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. Int. J. Syst. Evol. Microbiol. 51, 89–103.
[55] Zakhia, F., de Lajudie, P. (2001) Taxonomy of rhizobia. Agronomie 21, 569–576.
[56] Zhang, Y.M., Tian, C.F., Sui, X.H., Chen, W.F., Chen, W.X. (2012) Robust markers reflecting phylogeny and taxonomy of rhizobia. PLoS ONE 7, e44936.

Artículo:

Ramírez-Puebla ST, Servín-Garcidueñas LE, Jiménez-Marín B, Bolaños LM, Rosenblueth M, Martínez J, Rogel MA, Ormeño-Orrillo E, Martínez-Romero E. 2013. Gut and root microbiota commonalities. Appl. Environ. Microbiol. 79: 2-9.

Esta revisión surge a partir de los estudios realizados en el laboratorio de la Dra. Esperanza Martínez sobre comunidades bacterianas de raíces de plantas y de intestinos de animales. Logramos reconocer convergencias evolutivas entre microbiotas de plantas y animales. Las bacterias de raíces e intestinos participan en la degradación y modificación de nutrientes, regulan la expresión génica y amplían las capacidades metabólicas de sus hospederos, proveen nutrientes esenciales y ayudan en la protección contra patógenos. En la revisión se presenta una comparación global entre ambos tipos de microbiotas en base al análisis de literatura reciente y de resultados obtenidos en el laboratorio.

AEM Journals.ASM.org

MINIREVIEW

Gut and Root Microbiota Commonalities

Shamayim T. Ramírez-Puebla, Luis E. Servín-Garcidueñas, Berenice Jiménez-Marín, Luis M. Bolaños, Mónica Rosenblueth, Julio Martínez, Marco Antonio Rogel, Ernesto Ormeño-Orrillo, Esperanza Martínez-Romero

Genomic Sciences Center, UNAM, Cuernavaca, Morelos, Mexico

Animal guts and plant roots have absorption roles for nutrient uptake and converge in harboring large, complex, and dynamic groups of microbes that participate in degradation or modification of nutrients and other substances. Gut and root bacteria regulate host gene expression, provide metabolic capabilities, essential nutrients, and protection against pathogens, and seem to share evolutionary trends.

Guts and roots are inhabited by many different bacteria (1–5), archaea (6–12), and viruses (13–16), as well as by eukaryotes (17–20), with some of them containing bacteria of their own (21–24). Variations in gut microbiota respond to age (25–28), diet (29–31), or species (32). Most insects have dozens of microbial species in their guts, while mammalian guts may contain thousands. Herbivores exhibit the largest diversity (32, 33), including probably plant-associated bacteria, especially endophytes (34) that, by being inside plant tissues, may survive stomach digestion. Transiting diet-borne bacteria may contribute to gut metabolic capacities. Different soil types, moisture (35), plant genotypes (36), age (37), and root lysates, secretions, or exudates (38) are determinants of root microbiotas. Factors that determine root exudates, such as availability of inorganic nutrients, temperature, light intensity, O_2/CO_2 level, or root damage, may indirectly affect root microbiotas (39). The presence of pathogens induces changes in microbiota composition in roots and guts (40, 41).

Guts and roots have large surface areas, with microvilli and folds or root hairs in some parts. Both roots and guts are structured, nonhomogenous habitats with pH, nutrient, water, and oxygen differential levels or gradients. Gradients would favor colonization by distinct bacteria that are more successful in some root or gut regions. In consequence, the multiple microhabitats that exist in roots and guts contribute to high species richness (42, 43). Different conditions are found in the cecum and distal colon in humans, with cecal and colon microbiotas containing a larger proportion of facultative anaerobes (44). Colon mucosal folds exhibit particular bacteria adapted to colonic conditions and maybe to mucin degradation (45). Some insects have specialized structures in their gut, such as midgut sacs and tubular outgrowths called ceca or crypts, in which they harbor specific bacteria (46), and others with less-complex guts also have pH and oxygen gradients in their guts (47). A steep oxygen gradient including an anaerobic root environment in water-saturated roots parallels the gut oxygen gradient and anaerobic gut systems. Clostridia, and especially members of the family *Ruminococcaceae*, are more prevalent than other anaerobes and methanogens, a trend which is similar in the different gut systems (48). These communities take care of the degradation of the complex organic matter in the outer root layers. Some gut and root acid-tolerant bacteria can modify their environment by lowering the pH when producing diverse acids (49, 50). Along the roots, there are physiological differences, and their exudates are secreted differentially at the apical meristem, root cap, or root hairs (42), creating different microhabitats. A single *Burkholderia* strain colonizes only discrete root regions (51), and different burkholderias were found at different soil depths (37).

"*Arabidopsis thaliana* root microbiome might assemble by core ecological principles similar to those shaping the mammalian microbiome in which core phylum level enterotypes provide broad metabolic potential combined with modest levels of host genotype-dependent associations" (35). Metacommunity theory may be applied to root microbiotas, as has been used to explain the assembly of the gut microbial community (52). Metacommunity theory is based on the concept of discontinuous patches and interactions that can satisfactorily describe bacterial patchy colonization of roots. Future applications of these concepts will assert their usefulness.

Remarkably, there are individual-to-individual variations in bacterial composition of the gut (2, 53) and roots (54). Individual differences may be due to genetic differences and stochastic colonization processes (52). Limited patterns (enterotypes) in relation to stratified variation were distinguished in human and insect gut microbiotas (2, 55); however, it is controversial if there are only a few enterotypes in humans or gradients of diversity (28). In plants, similar bacterial genera are recurrently isolated from rhizospheres (soil surrounding roots affected by plants) or roots (34, 56). In roots, *Rhizobium* strain diversity with functional differentiation is high (57). Strain variability in vitamin production has been detected among gut bifidobacteria (reviewed in reference 58). Similarly, lactobacilli (reviewed in reference 59) are a heterogeneous group of bacteria with partly probiotic character which have considerable variation in terms of molecular characteristics and preferred natural habitats.

With few exceptions (see below), the gut microbiota is different from that of other host organs, and similarly, the root microbiota shares only some bacteria with those of other plant organs.

ENVIRONMENTAL AND MATERNAL ACQUISITION

Root and gut microorganisms are usually acquired from the environment. Roots are colonized by bulk soil microorganisms at-

Published ahead of print 26 October 2012

Address correspondence to Esperanza Martínez-Romero, emartine@ccg.unam.mx.

doi:10.1128/AEM.02553-12

The authors have paid a fee to allow immediate free access to this article.

tracted by chemotaxis and enriched by nutrients secreted by the roots in the rhizosphere. Animals also acquire their gut microbiota from their environment after they are born (60). In a few cases, microorganisms can be transferred vertically from mother to progenies. Endophytes present in plant seeds may subsequently colonize the roots and the rhizosphere. *Enterobacter asburiae*, found in maize kernels, is able to exit the roots and colonize the rhizosphere after the plant has established (61). Other seed bacteria do the same (54, 62). Animals can also acquire their gut microbiota from their mothers after being born, but there are cases of paternal transmission of symbionts, as in malaria vectors (63). Maternal transmission may occur before birth (64–66). When mammals are breast-fed, they acquire microorganisms that are present in the milk or on the mother's skin (67–69). Some stinkbug larvae acquire their mother's gut bacteria from contaminated eggs, by coprophagy, or by capsule-mediated transmission just after they have hatched (46). In view of the vertical and environmental transmission of root and gut microbes, gnotobiotic animals or plants are needed to clearly evaluate the effects of selected strains on hosts.

FUNCTIONAL REDUNDANCY AND ROLE OF MINORITIES

It seems that different microbiota composition may lead to the same and stable function. This may apply to gut and root bacteria and has been found to be true in methanogenic reactors (70). Similar degrading capacities are found in different gut bacteria (reviewed in reference 71). In roots, many different bacterial genera and species produce hormones, auxins, cytokinins, or gibberellins (reviewed in references 56 and 72). Our research group found that riboflavin is produced and excreted by different strains from several species of *Methylobacterium*, *Rhizobium*, *Sinorhizobium*, and *Bacillus*, both in rice and alfalfa root exudates and in pure cultures in minimal medium (our unpublished data). *In vitro* excretion of riboflavin by a large diversity of bacteria, including *Chromobacterium violaceum* and *Pantoea agglomerans*, was reported earlier (73), and both riboflavin and lumichrome (which is derived from riboflavin) stimulate root respiration (74). Additionally, many different plant-associated bacteria inhibit pathogenic fungi or bacteria (reviewed in reference 56).

Minority species present in the microbiota may help cover some of the host-specific needs. Methanogens, methylotrophs, and nitrogen-fixing bacteria are minor components in guts and rhizospheres (11, 75–78); however, they have important ecological roles. In some roots and guts, nitrogen fixation provides nitrogen to plants (79) and insects (80–82).

GUT AND ROOT BACTERIA ENHANCE THE METABOLIC CAPACITIES OF THEIR HOSTS

It is remarkable that gut bacteria are rich in sugar hydrolases (83) and other catabolic genes, such as those for tannin (84), cholesterol (85), or mucin (gut glycosylated proteins) (86). Similarly, capacities to degrade polyphenols, polysaccharides, protocatechuate, and proteins and to solubilize phosphate and weather rocks (50, 54, 87, 88) are prevalent among different rhizospheric bacteria. Mimosine-degrading bacteria are found in mimosa plants that produce mimosine (89), and cows that have such bacteria in their rumen are capable of degrading it (90). Alginate-degrading bacteria are found in abalone and human guts of algae consumers in Japan (91). The outstanding degrading capacities of root bacteria are the basis of rhizoremediation of polluting substances (92, 93) and are also evidenced in medical drug transformation or degradation in the human gut (94–96). Interestingly, in bioremediation, the abilities of bacteria to degrade soil pollutants may be triggered by flavonoids (97).

Gut and rhizospheric bacteria produce vitamins as riboflavin, as stated above. Vitamin B_{12} is an exclusive product of prokaryotes (98), and it is produced by plant root and gut bacteria (99–102). Essential amino acids and vitamins B and K are produced by gut bacteria (reviewed in reference 58). An alcohol dehydrogenase from the commensal bacterium *Acetobacter pomorum* modulates *Drosophila* developmental and metabolic homeostasis via insulin signaling (103). While root bacteria produce plant hormones that have effects on plant growth (reviewed in reference 56), gut bacteria seem to regulate animal behavior (104, 105).

GUT AND ROOT MICROBIOTAS COMPETE WITH PATHOGENS

Gut and root microbiotas suppress pathogens (reviewed in references 56 and 106). The human control of root bacteria has been envisaged as a manner to promote plant growth and health with benefits to agriculture (93, 107). Bacterial inoculants in agriculture and forestry are considered equivalent to probiotics (beneficial microbes provided as supplements) for animal health. Probiotics stimulate host defense systems and the competitive exclusion of pathogens, as plant growth-promoting rhizobacteria do (108). Seeds may harbor a reservoir of probiotics for their seedlings (54, 109). Prebiotics are added nutrients used to stimulate desirable bacteria in humans (110). We may even speculate that prebiotics were invented by roots, as some substances from their exudates stimulate bacterial growth selectively (89, 111, 112).

For over one hundred years, inoculants have been provided to plants in agricultural fields with variable success. Recently, a large number of commercial products whose effects are not always desirable have appeared to promote plant growth. Similarly, an increased number of probiotics and prebiotics whose effects have not been completely evaluated in different human populations are coming to the market. Gut gene expression in response to probiotics varies from person to person (113). In many cases, clinical benefits have been obtained in patients with specific probiotic strains (114).

Experience with plants has shown that appropriate use and regulation of probiotics (inoculants) is difficult to achieve. Undesirable genetic characteristics, such as denitrifying capacities, have been identified among inoculants (115). Strains used as probiotics should not contain glucosaminidase or glucuronidase genes that seem to have roles in producing toxic substances in the gut (reviewed in reference 116), but these recommendations may not be easily followed.

SIMILAR BACTERIUM-HOST INTERACTIONS IN GUTS AND ROOTS

Differential gene expression of bacteria in hosts. Bacterium-plant interactions have been studied for many years, and a molecular ping pong between rhizobia and plants that may serve as a model to analyze insect or human gut symbioses is known (reviewed in references 1 and 117). In rhizobium-plant molecular dialogue, *Rhizobium* NodD receptors, which bind root exudate molecules, function as transcriptional regulators that induce the expression of several genes, including *nod* genes and secretion systems (reviewed in references 117 and 118). Extrusion pumps are inducible by flavonoids that are present in root exudates but

do not require NodD genes (119). Many ABC transporter systems are induced by the respective substrate or other molecules from roots (111, 120).

In roots, bacteria have a differential gene expression that supposedly allows them to adapt to the root environment. Genes involved in root exudate usage, root attachment, and survival are induced in bacteria colonizing roots (120, 121). *In vitro* expression technology (IVET) (122), proteomic analysis, microarray and RNA Seq transcriptomics, and genetic analysis have revealed rhizobial (120, 121, 123), *Pseudomonas* (124, 125), *Streptomyces* (126), and other bacterial genes expressed on roots or rhizospheres. Similarly, bacteria may differentially express genes when in guts. Gut bacteria are exposed to bile salts that solubilize diet fat, have antimicrobial activities (127), and regulate bacterial gene expression. An efflux transporter of the multidrug resistance type (MDR) was induced in *Bifidobacterium* by bile (128). Different bile substances have been identified to control gene expression in bifidobacteria (129). Other bile-inducible genes have been found in *Lactobacillus plantarum* (130). Lastly, human gut bacteria transform bile salts (131). Gut bacteria can also modify dietary flavonoids (132) that have significant effects on animal physiology. Analogously, in roots, flavonoids produced by plants are signal molecules in bacteria (133) and are also transformed by bacteria *in vitro*, though this has not been shown *in vivo*. Plant phytoalexins are antimicrobials that are expelled from *Rhizobium etli*, *Bradyrhizobium japonicum*, and *Agrobacterium* by MDR efflux pumps that are inducible by root-exudated flavonoids (20, 119, 134).

Interestingly, gut and root microbiotas may follow the circadian cycles of their hosts. This was observed in nitrogen-fixing bacteria that fixed more during the daytime on rice roots (135). Epithelial cell proliferation, gastrointestinal motility, and other gut processes follow biological rhythms. In the gastrointestinal tract, there are large amounts of melatonin, which is a key hormone in the clock biological regulation (136). The Burmese python's microbiota is responsive to host cycles of feeding and fasting (137).

Host gene expression regulated by microbiotas. Outstandingly, gut and root bacteria modify gene expression in animal (138, 139) and plant (140) hosts, respectively. Gut gene expression is also modified by probiotics (113) that modify gut bacterial gene expression as well (141). Gut genes expressed in the presence of the gut bacterium *Bacteroides thetaiotaomicron* are involved in xenobiotic catabolism, in angiogenesis, in gut barrier epithelium maintenance, and in immunity development (139), with very complex host molecular responses (142).

Plants and humans can sense bacterially produced acylhomoserine lactones (AHLs), different volatiles, microbe-associated molecular patterns (MAMPS) (72, 143), and other bacterial molecules unknown at present. Root gene expression is differently modified by acylhomoserine lactones from pathogenic or symbiotic bacteria (144). In turn, plant products may act like quorum-sensing signals in bacteria (145). In recent years, specific regulatory roles of *N*-acylhomoserine lactones have become apparent, because plants responded with either a systemic resistance response or a hormonal regulated growth response to the presence of AHL-producing bacteria colonizing the root surface. Also in the animal/human systems, a specific perception of AHL compounds, produced by Gram-negative, mostly pathogenic bacteria, was found in many tissues, including the gut system, leading to immunomodulatory effects (146). In plants, root genes induced by rhizospheric bacteria are involved in oxidative and defense responses, in plant secondary metabolism, or in signaling (140). Plants may detect bacterial cyclopeptides through auxin sensing pathways (147). In a more specialized symbiosis, a cascade of signaling processes occurs inside root cells in the presence of rhizobia or Nod factors (148).

Control of microbiotas. A *Drosophila* mutant with increased levels of antimicrobial peptides showed deregulated balances of gut populations (149), with smaller numbers of *Commensalibacter intestini* (an acetic acid bacterium present in normal gut) bacteria (150) and increased numbers of *Gluconobacter morbifer* cells that caused gut cell apoptosis and early insect death (149). It is interesting to note that *C. intestini* antagonizes *G. morbifer*, which is a normal gut member, but with detrimental effects when present in large numbers; thus, *C. intestini* contributes to gut homeostasis and host fitness (151). Similarly, among root microbiotas, there are plant-pathogenic bacteria that normally would not affect the plants when kept in low numbers by other plant community strains or plant antimicrobials. Lipopolysaccharide *Rhizobium* mutants that were more sensitive to maize antimicrobial benzoxazinones had reduced rhizospheric colonization (152). Antimicrobial peptides constitute a line of defense in plants as effectors of innate immunity and regulate not only bacteria but also methanogenic archaea in guts (153). Gut immunity determines bacterial composition; reciprocally, bacteria modulate host immunity in guts (154, 155). Carbohydrate binding proteins (lectins) from guts and roots bind bacteria, form aggregates, and may have antibacterial effects (156, 157).

In addition to bacterium-host interactions, bacterium-bacterium interactions may determine community composition and its function (158). Those that occur in the mouth (159) may guide research in gut and root symbioses. In *Rhizobium*, mutants in quorum sensing are affected in rhizosphere colonization (160). Acylhomoserine lactones may be degraded by rhizospheric bacteria causing interference with quorum signals that regulate gene expression in other bacteria (161). This may have a role in protecting plants from pathogens but may also affect mutualistic interactions.

EVOLUTIONARY PATHWAYS

Lateral gene transfer in guts and roots. In roots, root nodules, and guts, lateral transfer of genetic material between different bacteria has been evidenced (2, 162, 163), seemingly promoted by close contacts in high-density populations. The presence of similar catabolic or antibiotic resistance genes in various gut bacterial genera has been explained as acquisitions by lateral gene transfers (91). It has been suggested that starch catabolism genes have been transferred from gut to bacteria (164).

There are many more phages than bacteria in the gut (13), and some may be involved in lateral gene transfer among gut bacteria (165). Lateral transfer of genetic material is mediated by plasmids or genomic island mobilization in rhizobia and other rhizospheric bacteria (54, 166), but phages may have a role as well.

Specialized symbiont evolution from root and gut bacteria. It has been suggested that gut bacteria gave rise to endosymbiotic bacteria in insects (167) based on similarities of gut bacteria and insect endosymbionts (168). Correspondingly, rhizospheric bacteria may have preceded nodule and endophytic bacteria in plants (169). Insect endosymbionts and nodule rhizobia are selected

symbionts that occupy intracellularly host-specialized structures and attain high numbers with a determined functional role. However, transmission modes of plant- and insect-specialized symbionts (reviewed in reference 46) and their genome sizes (rhizobial genome sizes reviewed in references 121 and 170) are different.

CONCLUSIONS

The comparison of plant and gut microbial ecologies may help to guide research toward the understanding of such complex symbioses. Literature on the subject is so extensive that only a few references were used to illustrate the commonalities of gut and root microbiotas. Interested readers are referred to recent literature (171–175). Plants use their "guts" (roots) outwards, and this simplifies their study in comparison to study of animal guts. Gut and root microbiotas significantly impact health, development, and fitness of their respective hosts.

ACKNOWLEDGMENTS

We acknowledge grants PAPIIT IN205412 and CONACyT 2249-5.

We thank Michael Dunn for critically reading the paper.

REFERENCES

1. **Badri DV, Weir TL, van der Lelie D, Vivanco JM.** 2009. Rhizosphere chemical dialogues: plant-microbe interactions. Curr. Opin. Biotechnol. **20**:642–650.
2. **Dillon RJ, Dillon VM.** 2004. The gut bacteria of insects: nonpathogenic interactions. Annu. Rev. Entomol. **49**:71–92.
3. **Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, Takami H, Morita H, Sharma VK, Srivastava TP, Taylor TD, Noguchi H, Mori H, Ogura Y, Ehrlich DS, Itoh K, Takagi T, Sakaki Y, Hayashi T, Hattori M.** 2007. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res. **14**:169–181.
4. **Marchesi JR.** 2010. Prokaryotic and eukaryotic diversity of the human gut. Adv. Appl. Microbiol. **72**:43–62.
5. **Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI.** 2007. The human microbiome project. Nature **449**:804–810.
6. **Donovan SE, Purdy KJ, Kane MD, Eggleton P.** 2004. Comparison of *Euryarchaea* strains in the guts and food-soil of the soil-feeding termite *Cubitermes fungifaber* across different soil types. Appl. Environ. Microbiol. **70**:3884–3892.
7. **Fricke WF, Seedorf H, Henne A, Krüer M, Liesegang H, Hedderich R, Gottschalk G, Thauer RK.** 2006. The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H_2 for methane formation and ATP synthesis. J. Bacteriol. **188**:642–658.
8. **Friedrich MW, Scmitt-Wagner D, Leuders T, Brune A.** 2001. Axial differences in community structure of *Crenarchaeota* and *Euryarchaeota* in the highly compartmentalized gut of the soil-feeding termite *Cubitermes orthognathus*. Appl. Environ. Microbiol. **67**:4880–4890.
9. **Hara K, Shinzato N, Seo M, Oshima T, Yamagishi A.** 2002. Phylogenetic analysis of symbiotic archaea living in the gut of xylophagous cockroaches. Microbes Environ. **17**:185–190.
10. **Horz HP, Conrads G.** 2010. The discussion goes on: what is the role of *Euryarchaeota* in humans? Archaea **2010**:967271.
11. **Jarrell KF, Walters AD, Bochiwal C, Borgia JM, Dickinson T, Chong JP.** 2011. Major players on the microbial stage: why archaea are important. Microbiology **157**:919–936.
12. **Simon HM, Dodsworth JA, Goodman RM.** 2000. Crenarchaeota colonize terrestrial plant roots. Environ. Microbiol. **2**:495–505.
13. **Minot S, Grunberg S, Wu GD, Lewis JD, Bushman FD.** 2012. Hypervariable loci in the human gut virome. Proc. Natl. Acad. Sci. U. S. A. **109**:3962–3966.
14. **Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD.** 2011. The human gut virome: inter-individual variation and dynamic response to diet. Genome Res. **21**:1616–1625.
15. **Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI.** 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature **466**:334–338.
16. **Swanson MM, Fraser G, Daniell TJ, Torrance L, Gregory PJ, Taliansky M.** 2009. Viruses in soils: morphological diversity and abundance in the rhizosphere. Ann. Appl. Biol. **155**:51–60.
17. **Nam YD, Chang HW, Kim KH, Roh SW, Kim MS, Jung MJ, Lee SW, Kim JY, Yoon JH, Bae JW.** 2008. Bacterial, archaeal, and eukaryal diversity in the intestines of Korean people. J. Microbiol. **46**:491–501.
18. **Pandey PK, Siddharth J, Verma P, Bavdekar A, Patole MS, Shouche YS.** 2012. Molecular typing of fecal eukaryotic microbiota of human infants and their respective mothers. J. Biosci. **37**:221–226.
19. **Parfrey LW, Walters WA, Knight R.** 2011. Microbial eukaryotes in the human microbiome: ecology, evolution, and future directions. Front. Microbiol. **2**:153.
20. **Parniske M.** 2008. Arbuscular mycorrhiza: the mother of plant root endosymbioses. Nat. Rev. Microbiol. **6**:763–775.
21. **Bertaux J, Schmid M, Chemidlin Prevost-Boure N, Churin JL, Hartmann A, Garbaye J, Frey-Klett P.** 2003. *In situ* identification of intracellular bacteria related to *Paenibacillus* spp. in the mycelium of the ectomycorrhizal fungus *Laccaria* bicolor S238N. Appl. Environ. Microbiol. **69**:4243–4248.
22. **Bianciotto V, Lumini E, Bonfante P, Vandamme P.** 2003. '*Candidatus* Glomeribacter gigasporarum' gen. nov., sp. nov., an endosymbiont of arbuscular mycorrhizal fungi. Int. J. Syst. Evol. Microbiol. **53**:121–124.
23. **Scheublin TR, Sanders IR, Keel C, van der Meer JR.** 2010. Characterisation of microbial communities colonising the hyphal surfaces of arbuscular mycorrhizal fungi. ISME J. **4**:752–763.
24. **Stingl U, Radek R, Yang H, Brune A.** 2005. "Endomicrobia": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. Appl. Environ. Microbiol. **71**:1473–1479.
25. **Biagi E, Candela M, Fairweather-Tait S, Franceschi C, Brigidi P.** 2012. Ageing of the human metaorganism: the microbial counterpart. Age **34**: 247–267.
26. **Mihajlovski A, Doré J, Levenez F, Alric M, Brugère JF.** 2010. Molecular evaluation of the human gut methanogenic archaeal microbiota reveals an age-associated increase of the diversity. Environ. Microbiol. Rep. **2**:272–280.
27. **Tiihonen K, Ouwehand AC, Rautonen N.** 2010. Human intestinal microbiota and healthy ageing. Ageing Res. Rev. **9**:107–116.
28. **Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI.** 2012. Human gut microbiome viewed across age and geography. Nature **486**: 222–227.
29. **Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, Henrissat B, Knight R, Gordon JI.** 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science **332**:970–974.
30. **Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI.** 2009. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci. Transl. Med. **1**:6ra14. doi: 10.1126/scitranslmed.3000322.
31. **Hooper LV, Midtvedt T, Gordon JI.** 2002. How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annu. Rev. Nutr. **22**:283–307.
32. **Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI.** 2008. Evolution of mammals and their gut microbes. Science **320**:1647–1651.
33. **Hong PY, Wheeler E, Cann IKO, Mackie RI.** 2011. Phylogenetic analysis of the fecal microbial community in herbivorous land and marine iguanas of the Galápagos Islands using 16S rRNA-based pyrosequencing. ISME J. **5**:1461–1470.
34. **Rosenblueth M, Martínez-Romero E.** 2006. Bacterial endophytes and their interactions with hosts. Mol. Plant Microbe Interact. **19**:827–837.
35. **Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, del Rio TG, Edgar RC, Eickhorst T, Ley RE, Hugenholtz P, Tringe SG, Dangl JL.** 2012. Defining the core *Arabidopsis thaliana* root microbiome. Nature **488**: 86–90.
36. **Hartmann A, Schmid M, van Tuinen D, Berg G.** 2009. Plant-driven selection of microbes. Plant Soil **321**:235–257.
37. **Chiarini L, Giovannelli V, Bevivino A, Dalmastri C, Tabacchioni S.** 2000. Different portions of the maize root system host *Burkholderia ce-*

pacia populations with different degrees of genetic polymorphism. Environ. Microbiol. **2**:111–118.
38. **Doornbos RF, van Loon LC, Bakker PAHM.** 2012. Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. A review. Agron. Sustain. Dev. **32**:227–243.
39. **Rovira AD.** 1969. Plant root exudates. Bot. Rev. **35**:35–57.
40. **Chow J, Lee SM, Shen Y, Khosravi A, Mazmanian SK.** 2010. Host-bacterial symbiosis in health and disease. Adv. Immunol. **107**:243–274.
41. **Raaijmakers JM, Paulitz TC, Steinberg C, Alabouvette C, Moenne-Loccoz Y.** 2009. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. Plant Soil **321**:341–361.
42. **Bertin C, Yang X, Weston LA.** 2003. The role of root exudates and allelochemicals in the rhizosphere. Plant Soil **256**:67–83.
43. **Turroni F, Marchesi JR, Foroni E, Gueimonde M, Shanahan F, Margolles A, van Sinderen D, Ventura M.** 2009. Microbiomic analysis of the bifidobacterial population in the human distal gut free. ISME J. **3**:745–751.
44. **Marteau P, Pochart P, Doré J, Béra-Maillet C, Bernalier A, Corthier G.** 2001. Comparative study of bacterial groups within the human cecal and fecal microbiota. Appl. Environ. Microbiol. **67**:4939–4942.
45. **Nava GM, Friedrichsen HJ, Stappenbeck TS.** 2011. Spatial organization of intestinal microbiota in the mouse ascending colon. ISME J. **5**:627–638.
46. **Kikuchi Y, Hosokawa T, Fukatsu T.** 2008. Diversity of bacterial symbiosis in stinkbugs, p 39–63. *In* Dijk TV (ed), Microbial Ecology Research Trends. Nova Science Publishers Inc., New York, NY.
47. **Brune A, Emerson D, Breznak JA.** 1995. The termite gut microflora as an oxygen sink—microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. Appl. Environ. Microbiol. **61**:2681–2687.
48. **Timmers RA, Rothballer M, Strik DP, Engel M, Schulz S, Schloter M, Hartmann A, Hamelers B, Buisman C.** 2012. Microbial community structure elucidates performance of Glyceria maxima plant microbial fuel cell. Appl. Microbiol. Biotechnol. **94**:537–548.
49. **Asahara T, Shimizu K, Nomoto K, Hamabata T, Ozawa A, Takeda Y.** 2004. Probiotic bifidobacteria protect mice from lethal infection with Shiga toxin-producing *Escherichia coli* O157:H7. Infect. Immun. **72**:2240–2247.
50. **Rodriguez H, Gonzalez T, Goire I, Bashan Y.** 2004. Gluconic acid production and phosphate solubilization by the plant growth-promoting bacterium *Azospirillum* spp. Naturwissenschaften **91**:552–555.
51. **Sharma S, Sharma S, Singh RK, Vaishampayan A.** 2008. Colonization behavior of bacterium *Burkholderia cepacia* inside the *Oryza sativa* roots visualized using green fluorescent protein reporter. World J. Microbiol. Biotechnol. **24**:1169–1175.
52. **Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA.** 2012. The application of ecological theory toward an understanding of the human microbiome. Science **336**:1255–1262.
53. **Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA.** 2005. Diversity of the human intestinal microbial flora. Science **308**:1635–1638.
54. **López-López A, Rogel MA, Ormeño-Orrillo E, Martínez-Romero J, Martínez-Romero E.** 2010. *Phaseolus vulgaris* seed-borne endophytic community with novel bacterial species such as *Rhizobium endophyticum* sp. nov. Syst. Appl. Microbiol. **33**:322–327.
55. **Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J, Weissenbach J, Ehrlich SD, Bork P.** 2011. Enterotypes of the human gut microbiome. Nature **473**:174–180.
56. **Friesen ML, Porter SS, Stark SC, von Wettberg EJ, Sachs JL, Martinez-Romero E.** 2011. Microbially mediated plant functional traits. Annu. Rev. Ecol. Evol. Syst. **42**:23–46.
57. **Rosenblueth M, Martínez-Romero E.** 2004. *Rhizobium etli* endophytic populations and their competitiveness for root maize colonization. Arch. Microbiol. **181**:337–344.
58. **Macfarlane S, Macfarlane GT.** 2003. Food and the large intestine, p 24–51. *In* Fuller R, Perdigon G (ed), Gut flora, nutrition, immunity and health, Blackwell Publishing, Oxford, United Kingdom.
59. **Kleerebezem M, Hols P, Bernard E, Rolain T, Zhou M, Siezen RJ, Bron PA.** 2010. The extracellular biology of the lactobacilli. FEMS Microbiol. Rev. **34**:199–230.
60. **Kikuchi Y, Hosokawa T, Fukatsu T.** 2011. An ancient but promiscuous host-symbiont association between *Burkholderia* gut symbionts and their heteropteran hosts. ISME J. **5**:446–460.
61. **Johnston-Monje D, Raizada MN.** 2011. Conservation and diversity of seed associated endophytes in *Zea* across boundaries of evolution, ethnography and ecology. PLoS One **6**:e20396. doi:10.1371/journal.pone.0020396.
62. **Pereira P, Ibánez F, Rosenblueth M, Etcheverry M, Martínez-Romero E.** 2011. Analysis of the bacterial diversity associated with the roots of maize (*Zea mays* L.) through culture-dependent and culture-independent methods. ISRN Ecol. **2011**:938546. doi:10.5402/2011/938546.
63. **Damiani C, Ricci I, Crotti E, Rossi P, Rizzi A, Scuppa P, Esposito F, Bandi C, Daffonchio D, Favia G.** 2008. Paternal transmission of symbiotic bacteria in malaria vectors. Curr. Biol. **18**:R1087–R1088.
64. **Jiménez E, Fernández L, Marín ML, Martín R, Odriozola JM, Nueno-Palop C, Narbad A, Olivares M, Xaus J, Rodríguez JM.** 2005. Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. Curr. Microbiol. **51**:270–274.
65. **Jiménez E, Marín ML, Martín R, Odriozola JM, Olivares M, Xaus J, Fernández L, Rodríguez JM.** 2008. Is meconium from healthy newborns actually sterile? Res. Microbiol. **159**:187–193.
66. **Mshvildadze M, Neu J, Shuster J, Theriaque D, Li N, Mai V.** 2010. Intestinal microbial ecology in premature infants assessed with non-culture-based techniques. J. Pediatr. **156**:20–25.
67. **Hunt KM, Foster JA, Forney LJ, Schütte UM, Beck DL, Abdo Z, Fox LK, Williams JE, McGuire MK, McGuire MA.** 2011. Characterization of the diversity and temporal stability of bacterial communities in human milk. PLoS One **6**:e21313. doi:10.1371/journal.pone.0021313.
68. **Martin R, Jiménez E, Heilig H, Fernandez L, Marín ML, Zoetendal EG, Rodriguez JM.** 2009. Isolation of bifidobacteria from breast milk and assessment of the bifidobacterial population by PCR-denaturing gradient gel electrophoresis and quantitative real-time PCR. Appl. Environ. Microbiol. **75**:965–969.
69. **Martín R, Langa S, Reviriego C, Jiménez E, Marín ML, Xaus J, Fernández L, Rodríguez JM.** 2003. Human milk is a source of lactic acid bacteria for the infant gut. J. Pediatr. **143**:754–758.
70. **Fernández A, Huang S, Seston S, Xing J, Hickey R, Criddle C, Tiedje J.** 1999. How stable is stable? Function versus community composition. Appl. Environ. Microbiol. **65**:3697–3704.
71. **Pérez-Chaia AP, Oliver G.** 2003. Intestinal microflora and metabolic activity, p 77–98. *In* Fuller R, Perdigon G (ed), Gut flora, nutrition, immunity and health. Blackwell Publishing, Oxford, United Kingdom.
72. **Ortíz-Castro R, Contreras-Cornejo HA, Macías-Rodríguez L, López-Bucio J.** 2009. The role of microbial signals in plant growth and development. Plant Signal Behav. **4**:701–712.
73. **Phillips DA, Martínez-Romero E, Yang GP, Joseph JM.** 2000. Release of nitrogen: a key trait in selecting bacterial endophytes for agronomically useful nitrogen fixation, p 205–217. *In* Ladha JK, Reddy PM (ed), The quest for nitrogen fixation in rice. IRRI, Manila, Philippines.
74. **Phillips DA, Joseph CM, Yang GP, Martinez-Romero E, Sanborn JR, Volpin H.** 1999. Identification of lumichrome as a sinorhizobium enhancer of alfalfa root respiration and shoot growth. Proc. Natl. Acad. Sci. U. S. A. **96**:12275–12280.
75. **Gibson GR, Cummings JH, Macfarlane GT.** 1988. Use of a three-stage continuous culture system to study the effect of mucin on dissimilatory sulfate reduction and methanogenesis by mixed populations of human gut bacteria. Appl. Environ. Microbiol. **54**:2750–2755.
76. **Ladha JK, Barraquio WL, Watanabe I.** 1983. Isolation and identification of nitrogen-fixing *Enterobacter cloacae* and *Klebsiella planticola* associated with rice plants. Can. J. Microbiol. **29**:1301–1308.
77. **Madhaiyan M, Poonguzhali S, Kwon SW, Sa TM.** 2009. *Methylophilus rhizosphaerae* sp. nov., a restricted facultative methylotroph isolated from rice rhizosphere soil. Int. J. Syst. Evol. Microbiol. **59**:2904–2908.
78. **St-Pierre B, Wright AD.** 27 April 2012, posting date. Diversity of gut methanogens in herbivorous animals. Animal. http://dx.doi.org/10.1017/S1751731112000912.
79. **Ormeño-Orrillo E, Hungria M, Martinez-Romero E.** Dinitrogen-fixing prokaryotes. *In* Rosenberg E, DeLong EF, Stackebrandt E, Lory S,

Thompson F (ed), The prokaryotes, vol 1. Symbiotic associations, biotechnology, applied microbiology, 4th ed, in press. Springer, New York, NY.
80. **Behar A, Yuval B, Jurkevitch E.** 2005. Enterobacteria-mediated nitrogen fixation in natural populations of the fruit fly *Ceratitis capitata*. Mol. Ecol. **14**:2637–2643.
81. **Desai MS, Brunes A.** 2012. Bacteroidales ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites. ISME J. **6**:1302–1313.
82. **Ohkuma M.** 2008. Symbioses of flagellates and prokaryotes in the gut of lower termites. Trends Microbiol. **16**:345–352.
83. **Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA.** 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat. Rev. Microbiol. **6**:121–131.
84. **Osawa R, Kuroiso K, Goto S, Shimizu A.** 2000. Isolation of tannin-degrading lactobacilli from humans and fermented foods. Appl. Environ. Microbiol. **66**:3093–3097.
85. **Gérard P, Lepercq P, Leclerc M, Gavini F, Raibaud P, Juste C.** 2007. *Bacteroides* sp. strain D8, the first cholesterol-reducing bacterium isolated from human feces. Appl. Environ. Microbiol. **73**:5742–5749.
86. **Derrien M, Collado MC, Ben-Amor K, Salminen S, de Vos WM.** 2008. The mucin degrader *Akkermansia muciniphila* is an abundant resident of the human intestinal tract. Appl. Environ. Microbiol. **74**:1646–1648.
87. **Calvaruso C, Turpault MP, Frey-Klett P.** 2006. Root-associated bacteria contribute to mineral weathering and to mineral nutrition in trees: a budgeting analysis. Appl. Environ. Microbiol. **72**:1258–1266.
88. **Puente ME, Bashan Y, Li CY, Lebsky VK.** 2004. Microbial populations and activities in the rhizoplane of rock-weathering desert plants. I. Root colonization and weathering of igneous rocks. Plant Biol. **6**:629–642.
89. **Soedarjo M, Hemscheidt TK, Borthakur D.** 1994. Mimosine, a toxin present in leguminous trees (*Leucaena* spp.), induces a mimosine-degrading enzyme activity in some *Rhizobium* strains. Appl. Environ. Microbiol. **60**:4268–4272.
90. **Allison MJ, Hammond AC, Jones RJ.** 1990. Detection of ruminal bacteria that degrade toxic dihydroxypyridine compounds produced from mimosine. Appl. Environ. Microbiol. **56**:590–594.
91. **Thomas F, Barbeyron T, Tonon T, Génicot S, Czjzek M, Michel G.** 2012. Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine Flavobacteria to their independent transfers to marine Proteobacteria and human gut Bacteroides. Environ. Microbiol. **14**:2379–2394.
92. **Kuiper I, Lagendijk EL, Bloemberg GV, Lugtenberg BJ.** 2004. Rhizoremediation: a beneficial plant-microbe interaction. Mol. Plant Microbe Interact. **17**:6–15.
93. **Lugtenberg BJ, Kravchenko LV, Simons M.** 1999. Tomato seed and root exudate sugars: composition, utilization by *Pseudomonas* biocontrol strains and role in rhizosphere colonization. Environ. Microbiol. **1**:439–446.
94. **Haiser HJ, Turnbaugh PJ.** 2012. Is it time for a metagenomic basis of therapeutics? Science **336**:1253–1255.
95. **Mikov M.** 1994. The metabolism of drugs by the gut flora. Eur. J. Drug Metab. Pharmacokinet. **19**:201–207.
96. **Sousa T, Paterson R, Moore V, Carlsson A, Abrahamsson B, Basit AW.** 2008. The gastrointestinal microbiota as a site for the biotransformation of drugs. Int. J. Pharm. **363**:1–25.
97. **Pham TT, Tu Y, Sylvestre M.** 2012. Remarkable ability of *Pandoraea pnomenusa* B356 biphenyl dioxygenase to metabolize simple flavonoids. Appl. Environ. Microbiol. **78**:3560–3570.
98. **Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS.** 2003. Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. J. Biol. Chem. **278**:41148–41159.
99. **Albert MJ, Mathan VI, Baker SJ.** 1980. Vitamin B12 synthesis by human small intestinal bacteria. Nature **283**:781–782.
100. **Campbell GR, Taga ME, Mistry K, Lloret J, Anderson PJ, Roth JR, Walker GC.** 2006. *Sinorhizobium meliloti bluB* is necessary for production of 5,6-dimethylbenzimidazole, the lower ligand of B12. Proc. Natl. Acad. Sci. U. S. A. **103**:4634–4639.
101. **Morita H, Toh H, Fukuda S, Horikawa H, Oshima K, Suzuki T, Murakami M, Hisamatsu S, Kato Y, Takizawa T, Fukuoka H, Yshimura T, Itoh K, O'Sullivan D, McKay L, Ohno H, Kikuchi J, Masaoka T, Hattori M.** 2008. Comparative genome analysis of *Lactobacillus reuteri* and *Lactobacillus fermentum* reveal a genomic island for reuterin and cobalamin production. DNA Res. **15**:151–161.
102. **Ramotar K, Conly JM, Chubb H, Louie TJ.** 1984. Production of menaquinones by intestinal anaerobes. J. Infect. Dis. **150**:213–218.
103. **Shin SC, Kim SH, You H, Kim B, Kim AC, Lee KA, Yoon JH, Ryu JH, Lee WJ.** 2011. *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. Science **334**:670–674.
104. **Bercik P, Denou E, Collins J, Jackson W, Lu J, Jury J, Deng Y, Blennerhasset P, Macri J, McCoy K, Verdu EF, Collins SM.** 2011. The intestinal microbiota affect central levels of brain-derived neurotropic factor and behavior in mice. Gastroenterology **141**:599–609.
105. **Díaz Heijtz RD, Wang S, Anuar F, Qiuan Y, Björkholm B, Samuelsson A, Hibberd ML, Frossberg H, Pettersson S.** 2011. Normal gut microbiota modulates brain development and behavior. Proc. Natl. Acad. Sci. U. S. A. **108**:3047–3052.
106. **Kane M, Case LK, Kopaskie K, Kozlova A, MacDearmid C, Chervonsky AV, Golovkina TV.** 2011. Successful transmission of a retrovirus depends on the commensal microbiota. Science **334**:245–249.
107. **Jung SC, Martínez-Medina A, López-Raez JA, Pozo MJ.** 2012. Mycorrhiza-induced resistance and priming of plant defenses. J. Chem. Ecol. **38**:651–664.
108. **Saxelin M, Tynkkynen S, Mattila-Sandholm T, de Vos WM.** 2005. Probiotic and other functional microbes: from markets to mechanisms. Curr. Opin. Biotechnol. **16**:204–211.
109. **Puente ME, Li CY, Bashan Y.** 2009. Rock-degrading endophytic bacteria in cacti. Environ. Exp. Bot. **66**:389–401.
110. **Gibson GR, Rastall RA, Fuller R.** 2003. The health benefits of probiotics and prebiotics, p 52–76. *In* Fuller R, Perdigon G (ed), Gut flora, nutrition, immunity and health. Blackwell Publishing, Oxford, United Kingdom.
111. **Rosenblueth M, Hynes MF, Martínez-Romero E.** 1998. *Rhizobium tropici teu* genes involved in specific uptake of *Phaseolus vulgaris* bean-exudate compounds. Mol. Gen. Genet. **258**:587–598.
112. **Tepfer D, Goldmann A, Pamboukdjian N, Maille M, Lepingle A, Chevalier D, Dénarié J, Rosenberg C.** 1988. A plasmid of *Rhizobium meliloti* 41 encodes catabolism of two compounds from root exudate of *Calystegium sepium*. J. Bacteriol. **170**:1153–1161.
113. **van Baarlen P, Troost F, van der Meer C, Hooiveld G, Boekschoten M, Brummer RJM, Kleerebezem M.** 2011. Human mucosal in vivo transcriptome responses to three lactobacilli indicate how probiotics may modulate human cellular pathways. Proc. Natl. Acad. Sci. U. S. A. **108**(Suppl 1):4562–4569.
114. **Floch MH, Walker WA.** 2008. Advances in clinical use of probiotics. J. Clin. Gastroenterol. **42**:S45.
115. **Zimmer W, Stephan MP, Bothe H.** 1984. Denitrification by *Azospirillum brasilense* Sp 7. Arch. Microbiol. **138**:206–211.
116. **Delgado S, O'Sullivan E, Fitzgerald G, Mayo B.** 2008. *In vitro* evaluation of the probiotic properties of human intestinal *Bifidobacterium* species and selection of new probiotic candidates. J. Appl. Microbiol. **104**:1119–1127.
117. **Peix A, Velázquez E, Silva LR, Mateos PF.** 2010. Key molecules involved in beneficial infection process in rhizobia-legume symbiosis, p 55–80. *In* Khan MH, Zaidi A, Musarrat J (ed), Microbes for legume improvement. Springer, Vienna, Austria.
118. **Downie JA.** 2010. The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. FEMS Microbiol. Rev. **34**:150–170.
119. **González-Pasayo R, Martínez-Romero E.** 2000. Multiresistance genes of *Rhizobium etli* CFN42. Mol. Plant Microbe Interact. **13**:572–577.
120. **Ramachandran VK, East AK, Karunakaran R, Downie JA, Poole PS.** 2011. Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. Genome Biol. **12**:R106.
121. **López-Guerrero MG, Ormeño-Orrillo E, Acosta JL, Mendoza-Vargas A, Rogel MA, Ramírez MA, Rosenblueth M, Martínez-Romero J, Martínez-Romero E.** 2012. Rhizobial extrachromosomal replicon variability, stability and expression in natural niches. Plasmid **68**:149–158.
122. **Ramos-González MI, Campos MJ, Ramos JL.** 2005. Analysis of *Pseudomonas putida* KT2440 gene expression in the maize rhizosphere: *in vitro* expression technology capture and identification of root-activated promoters. J. Bacteriol. **187**:4033–4041.
123. **Karunakaran R, Ramachandran VK, Seaman JC, East AK, Mouhsine B.** 2009. Transcriptomic analysis of rhizobium leguminosarum biovar

viciae in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. J. Bacteriol. **191**:4002–4014.
124. **Espinosa-Urgel M, Ramos JL.** 2001. Expression of a *Pseudomonas putida* aminotransferase involved in lysine catabolism is induced in the rhizosphere. Appl. Environ. Microbiol. **67**:5219–5224.
125. **Kim YC, Miller CD, Anderson AJ.** 2000. Superoxide dismutase activity in *Pseudomonas putida* affects utilization of sugars and growth on root surfaces. Appl. Environ. Microbiol. **66**:1460–1467.
126. **Langlois P, Bourassa S, Poirier GG, Beaulieu C.** 2003. Identification of *Streptomyces coelicolor* proteins that are differentially expressed in the presence of plant material. Appl. Environ. Microbiol. **69**:1884–1889.
127. **Begley M, Gahan CGM, Hill C.** 2005. The interaction between bacteria and bile. FEMS Microbiol. Rev. **29**:625–651.
128. **Gueimonde M, Garrigues C, van Sinderen D, de los Reyes-Gavilán CG, Margolles A.** 2009. Bile-inducible efflux transporter from *Bifidobacterium longum* NCC2705, conferring bile resistance. Appl. Environ. Microbiol. **75**:3153–3160.
129. **Ruiz L, Alvarez-Martín P, Mayo B, de los Reyes-Gavilán CG, Gueimonde M, Margolles A.** 2012. Controlled gene expression in *Bifidobacteria* by use of a bile-responsive element. Appl. Environ. Microbiol. **78**:581–585.
130. **Bron PA, Marco M, Hoffer SM, Van Mullekom E, de Vos MW, Kleerebezem M.** 2004. Genetic characterization of the bile salt response in *Lactobacillus plantarum* and analysis of responsive promoters *in vitro* and *in situ* in the gastrointestinal tract. J. Bacteriol. **186**:7829–7835.
131. **Ridlon JM, Kang DJ, Hylemon PB.** 2006. Bile salt biotransformations by human intestinal bacteria. J. Lipid Res. **47**:241–259.
132. **Blaut M, Schoefer L, Braune A.** 2003. Transformation of flavonoids by intestinal microorganisms. Int. J. Vitam. Nutr. Res. **73**:79–87.
133. **Cooper J.** 2004. Multiple responses of rhizobia to flavonoids during legume root infection. Adv. Bot. Res. **41**:1–62.
134. **Palumbo JD, Kado CI, Phillips DA.** 1998. An isoflavonoid-inducible efflux pump in *Agrobacterium tumefaciens* is involved in competitive colonization of roots. J. Bacteriol. **180**:3107–3113.
135. **Sims GK, Dunigan EP.** 1984. Diurnal and seasonal variations in nitrogenase activity (C_2H_2 reduction) of rice roots. Soil Biol. Biochem. **16**:15–18.
136. **Hoogerwerf WA.** 2006. Biologic clocks and the gut. Curr. Gastroenterol. Rep. **8**:353–359.
137. **Costello EK, Gordon JI, Secor SM, Knight R.** 2010. Postprandial remodeling of the gut microbiota in Burmese pythons. ISME J. **4**:1375–1385.
138. **Comelli EM, Simmering R, Faure M, Donnicola D, Mansourian R, Rochat F, Corthesy-Theulaz I, Cherbut C.** 2008. Multifaceted transcriptional regulation of the murine intestinal mucus layer by endogenous microbiota. Genomics **91**:70–77.
139. **Hooper LV, Wong MH, Thelin A, Hansson L, Falk PG, Gordon JI.** 2001. Molecular analysis of commensal host-microbial relationships in the intestine. Science **291**:881–884.
140. **Rudrappa T, Czymmek KJ, Paré PW, Bais HP.** 2008. Root-secreted malic acid recruits beneficial soil bacteria. Plant Physiol. **148**:1547–1556.
141. **McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL, Henrissat B, Oozeer R, Cools-Portier S, Gobert G, Chervaux C, Knights D, Lozupone CA, Knight R, Duncan AE, Bain JR, Muehlbauer MJ, Newgard CB, Heath AC, Gordon JI.** 2011. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Sci. Transl. Med. **3**:106.
142. **Macpherson AJ, Harris NL.** 2004. Interactions between commensal intestinal bacteria and the immune system. Nat. Rev. Immunol. **4**:478–485.
143. **Farag MA, Ryu CM, Sumner LW, Paré PW.** 2006. GC-MS SPME profiling of rhizobacterial volatiles reveals prospective inducers of growth promotion and induced systemic resistance in plants. Phytochemistry **67**:2262–2268.
144. **Mathesius M, Mulders S, Gao M, Teplitski M, Caetano-Anollés G, Rolfe BG, Bauer WD.** 2003. Extensive and specific responses of a eukaryote to bacterial quorum-sensing signals. Proc. Natl. Acad. Sci. U. S. A. **100**:1444–1449.
145. **Gao M, Teplitski M, Robinson JB, Bauer WD.** 2003. Production of substances by *Medicago truncatula* that affect bacterial quorum sensing. Mol. Plant Microbe Interact. **16**:827–834.
146. **Teplitski M, Mathesius U, Rumbaugh KP.** 2011. Perception and degradation of *N*-acyl homoserine lactone quorum sensing signals by mammalian and plant cells. Chem. Rev. **111**:100–116.
147. **Ortiz-Castro R, Díaz-Pérez C, Martínez-Trujillo M, del Río RE, Campos-García J, López-Bucio J.** 2011. Transkingdom signaling based on bacterial cyclodipeptides with auxin activity in plants. Proc. Natl. Acad. Sci. U. S. A. **108**:7253–7258.
148. **Oldroyd GE, Murray JD, Poole PS, Downie JA.** 2011. The rules of engagement in the legume-rhizobial symbiosis. Annu. Rev. Genet. **45**:119–144.
149. **Ryu JH, Kim SH, Lee HY, Bai JY, Nam YD, Bae JW, Lee DG, Shin SC, Ha EM, Lee WJ.** 2008. Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in *Drosophila*. Science **319**:777–782.
150. **Kim EK, Kim SH, Nam HJ, Choi MK, Lee KA, Choi SH, Seo YY, You H, Kim B, Lee WJ.** 2012b. Draft genome sequence of *Commensalibacter intestini* A911T, a symbiotic bacterium isolated from *Drosophila melanogaster* intestine. J. Bacteriol. **194**:1246.
151. **Roh SW, Nam YD, Chang HW, Kim KH, Kim MS, Ryu JH, Kim SH, Lee WJ, Bae JW.** 2008. Phylogenetic characterization of two novel commensal bacteria involved with innate immune homeostasis in *Drosophila melanogaster*. Appl. Environ. Microbiol. **74**:6171–6177.
152. **Ormeño-Orrillo E, Rosenblueth M, Luyten E, Vanderleyden J, Martínez-Romero E.** 2008. Mutations in lipopolysaccharide biosynthetic genes impair maize rhizosphere and root colonization of *Rhizobium tropici* CIAT899. Environ. Microbiol. **10**:1271–1284.
153. **Bang C, Schilhabel A, Weidenbach K, Kopp A, Goldmann T, Gutsmann T, Schmitz RA.** 2012. Effects of antimicrobial peptides on methanogenic archaea. Antimicrob. Agents Chemother. **56**:4123–4130.
154. **Lee WJ.** 2009. Bacterial-modulated host immunity and stem cell activation for gut homeostasis. Genes Dev. **23**:2260–2265.
155. **Round JL, Mazmanian SK.** 2009. The gut microbiota shapes intestinal immune responses during health and disease. Nat. Rev. Immunol. **9**:313–323.
156. **Peumans WJ, Van Damme EJ.** 1995. Lectins as plant defense proteins. Plant Physiol. **109**:347–352.
157. **Vaishnava S, Yamamoto M, Severson KM, Ruhn KA, Yu X, Koren O, Ley R, Wakeland EK, Hooper LV.** 2011. The antibacterial lectin RegIIIΓ promotes the spatial segregation of microbiota and host in the intestine. Science **334**:255–258.
158. **Gibson GR, Wang X.** 1994. Regulatory effects of bifidobacteria on the growth of other colonic bacteria. J. Appl. Microbiol. **77**:412–420.
159. **Kreth J, Merritt J, Qi F.** 2009. Bacterial and host interactions of oral streptococci. DNA Cell Biol. **28**:397–403.
160. **Edwards A, Frederix M, Wisniewski-Dyé F, Jones J, Zorreguieta A, Downie JA.** 2009. The cin and rai quorum-sensing regulatory systems in *Rhizobium leguminosarum* are coordinated by ExpR and CinS, a small regulatory protein coexpressed with CinI. J. Bacteriol. **191**:3059–3067.
161. **Jafra S, Przysowa J, Czajkowski R, Michta A, Garbeva P, Van der Wolf JM.** 2006. Detection and characterization of bacteria from the potato rhizosphere degrading *N*-acyl-homoserine lactone. Can. J. Microbiol. **52**:1006–1015.
162. **Kroer N, Barkay T, Sorensen S, Weber D.** 1998. Effect of root exudates and bacterial metabolic activity on conjugal gene transfer in the rhizosphere of a marsh plant. FEMS Microbiol. Ecol. **25**:375–384.
163. **Shoemaker NB, Vlamakis H, Hayes K, Salyers AA.** 2001. Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. Appl. Environ. Microbiol. **67**:561–568.
164. **Arias MC, Danchin EGJ, Coutinho PM, Henrissat B, Ball S.** 2012. Eukaryote to gut bacteria transfer of a glycoside hydrolase gene essential for starch breakdown in plants. Mob. Genet. Elements **2**:81–87.
165. **Stern A, Mick E, Tirosh I, Sagy O, Sorek R.** 2012. CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. Genome Res. **22**:1985–1994.
166. **Sullivan JT, Ronson CW.** 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proc. Natl. Acad. Sci. U. S. A. **95**:5145–5149.
167. **Husník F, Chrudimský T, Hypša V.** 2011. Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches. BMC Biol. **9**:87.
168. **Fukatsu T, Hosokawa T.** 2002. Capsule-transmitted gut symbiotic bac-

terium of the Japanese common plataspid stinkbug, *Megacopta punctatissima*. Appl. Environ. Microbiol. **68**:389–396.
169. **López-López A, Rosenblueth R, Martínez J, Martínez-Romero E.** 2010. Rhizobial symbioses in tropical legumes and non-legumes. Soil Biol. **21**: 163–184.
170. **McCutcheon JP, Moran NA.** 2012. Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. **10**:13–26.
171. **de Bruijn FJ.** 2013. Molecular microbial ecology of the rhizosphere, vol I and II. Wiley-Blackwell, Hoboken, New Jersey.
172. **Dessaux Y, Hinsinger P, Lemanceau P.** 2007. Rhizosphere: achievements and challenges. Springer Science Press, Berlin, Germany.
173. **Fuller R, Perdigon G.** 2003. Gut flora, nutrition, immunity and health. Blackwell Publishing, Oxford, United Kingdom.
174. **Pinto R, Varanini Z, Nannipieri P.** 2007. The rhizosphere: biochemistry and organic substances at the soil-plant interface. Taylor and Francis Group/CRC Press, Boca Raton, FL.
175. **Sadowsky MJ, Whitman RL.** 2011. The fecal bacteria. ASM Press, Washington, DC.

Conclusiones

Encontramos que la mayoría de las especies representativas de los diferentes grupos filogenéticos de *Phaseolus* nodulan preferentemente con cepas de *Bradyrhizobium*. También proponemos que ocurrió un reemplazo de simbiontes de *Bradyrhizobium* a *Rhizobium* en un grupo filogenético de *Phaseolus*. La extensión del reemplazo se logró limitar a *P. vulgaris* y a especies relacionadas filogenéticamente pero excluyendo a *P. acutifolius* y *P. parvifolius*.

El reemplazo de simbiontes no está ligado al proceso de domesticación. La nodulación con *Rhizobium* no se correlaciona con el proceso de domesticación debido a que se encontró que las especies domesticas *P. lunatus y P. acutifolius* nodulan preferentemente con *Bradyrhizobium*.

El mecanismo de reconocimiento molecular entre *Phaseolus* y cepas de *Rhizobium* es específico. La cepa CCGE510 obtenida de nódulos de campo de *P. albescens* establece simbiosis inefectivas con *P. vulgaris* a pesar de que son especies filogenéticamente cercanas. Las plantas de *P. vulgaris* inoculadas con la cepa CCGE510 presentaron un color amarillento y sus nódulos eran verdes, reflejando una ineficiente fijación de nitrógeno. La cepa CCGE510 es un simbionte eficiente en *P. albescens* pero se comporta como una bacteria saprófita en *P. vulgaris*.

El reemplazo de simbiontes pudo haber sido facilitado por la transferencia horizontal de los genes *nodZ* y *nolL* de *Bradyrhizobium* a *Rhizobium*. Las bases genéticas del reemplazo de simbiontes en *Phaseolus* quedan por explorar. Del lado de las bacterias, la adquisición de genes que fucosilan y acetilan al factor Nod pueden ser responsables del aumento de afinidad de *Rhizobium* por *Phaseolus*.

En base a los genes *nod* predichos en los genomas de *Rhizobium* sp. CCGE510 y *Bradyrhizobium* sp. CCGE-LA001 podemos inferir que sus factores Nod pueden ser fucosilados. Se ha encontrado que los factores Nod fucosilados son más eficientes para inducir la formación de nódulos y son preferidos por *P. vulgaris* (Laeremans et al., 1999).

El genoma de *Bradyrhizobium* sp. CCGELA001 codifica para una diversidad de genes *nod* involucrados en otras modificaciones químicas tales como acetilación y sulfatación. Los factores Nod producidos por *Bradyrhizobium* sp. CCGELA001 pueden ser más variables y por tanto podrían ser reconocidos por una mayor diversidad de plantas.

Globalmente, los resultados obtenidos ayudan a comprender mejor las bases genéticas de la interacción entre rizobios y *Phaseolus* y los mecanismos de especificidad simbiótica.

También comprobamos que las técnicas filogenómicas y de comparación de métricos de conservación de secuencias genómicas son útiles para resolver las relaciones evolutivas de los diferentes linajes de rizobios de la familia *Rhizobiaceae.*

Finalmente, la comparación de las características de raíces e intestinos nos ha permitido avanzar en el entendimiento del funcionamiento de las comunidades bacterianas en esos hábitats.

Perspectivas

Es de interés obtener un genoma completo para la cepa *Bradyrhizobium* sp. CCGELA001. Hemos enviado ADN genómico a secuenciación utilizando la plataforma de PacBio. Realizaré el ensamble de este genoma cuando se cuente con la secuencia.

Se desea continuar realizando análisis filogenómicos para cepas que están siendo secuenciadas en el laboratorio y que representan a linajes nuevos de *Rhizobium* tales como *R. popolucense* que se aisló de la región de Los Tuxtlas en Veracruz.

La secuenciación de los genomas de otras cepas de *Bradyrhizobium* obtenidas de nódulos de campo de otras especies silvestres de *Phaseolus* podría ser relevante para analizar el contenido y la evolución de genes *nod*. La abundancia de genes *nod* involucrados en la adición de diferentes modificaciones químicas al factor Nod podría llegar a ser una característica común de los bradyrizobios de *Phaseolus*. Esto podría ayudar a comprender el éxito que ha tenido *Bradyrhizobium* para nodular una mayor cantidad de plantas hospederas.

Se planean analizar las historias evolutivas de otros genes implicados en simbiosis tales como el gen *nodM* que codifica para la enzima glucosamina sintasa que produce los precursores para la síntesis de la estructura de los factores Nod. Los genes *nodM* se localizan en islas simbióticas y en plásmidos simbióticos. Es probable que el gen *nodM* haya surgido a partir de duplicaciones génicas y de transferencias laterales de otras enzimas glucosamina sintasa tales como las codificadas por el gen *glsM* que se encuentra localizado en los cromosomas de distintos rizobios.

Actualmente existen grupos trabajando en la secuenciación de genomas de distintas especies de *Phaseolus* en el Laboratorio Nacional de Genómica para la Biodiversidad. Cuando los genomas se encuentren disponibles se podrán analizar las relaciones filogenéticas de distintos genes implicados en el reconocimiento de los rizobios, tales como los genes R y los receptores del factor Nod que pudieran o no brindar información sobre el reemplazo de simbiontes.

El reemplazo de simbiontes de *Sinorhizobium* a *Bradyrhizobium* en nódulos de soya ha sido explicado por variantes de los genes R implicados en el reconocimiento de patógenos (Yang et al., 2010). Mediante el análisis de genomas de *Phaseolus* se podrían reconocer variantes de algún gen R similar al de la soya. Si se llegan a identificar genes R candidatos podrían generarse plantas de *P. vulgaris* que expresen esos genes en raíces transformadas y comprobar que la afinidad por cepas de *Rhizobium* se sustituya por cepas de *Bradyrhizobium*.

Bibliografía

Aguilar OM, Riva O, Peltzer E. 2004. Analysis of *Rhizobium etli* and of its symbiosis with wild *Phaseolus vulgaris* supports coevolution in centers of host diversification. Proc. Natl. Acad. Sci. USA. 101: 13548–13553.

Caballero-Mellado J, Martínez-Romero E. 1999. Soil fertilization limits the genetic diversity of *Rhizobium* in bean nodules. Symbiosis. 26: 111–121.

Cárdenas L, Domínguez J, Quinto C, López-Lara IM, Lugtenberg BJ, Spaink HP, Rademaker GJ, Haverkamp J, Thomas-Oates JE. 1995. Isolation, chemical structures and biological activity of the lipo-chitin oligosaccharide nodulation signals from *Rhizobium etli*. Plant Mol. Biol. 29: 453–464.

Delgado-Salinas A, Turley T, Richman A, Lavin M. 1999. Phylogenetic analysis of the cultivated and wild species of *Phaseolus* (*Fabaceae*). Syst. Bot. 24: 438–460.

Delgado-Salinas A, Bibler R, Lavin M. 2006. Phylogeny of the genus *Phaseolus* (*Leguminosae*): a recent diversification in an ancient landscape. Syst. Bot. 31: 779–791.

Dénarié J, Debellé F, Promé JC. 1996. *Rhizobium* lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. Ann. Rev. Biochem. 65: 503–535.

Folch-Mallol JL, Marroqui S, Sousa C, Manyani H, López-Lara IM, van der Drift KM, Haverkamp J, Quinto C, Gil-Serrano A, Thomas-Oates J, Spaink HP, Megías M. 1996. Characterization of *Rhizobium tropici* CIAT899 nodulation factors: the role of *nodH* and *nodPQ* genes in their sulfation. Mol. Plant-Microbe Interact. 9: 151–163.

Freytag GF, Debouck DG. 2002. Taxonomy, distribution, and ecology of the genus *Phaseolus* (*Leguminosae–Papilionoideae*) in North America, Mexico and Central America. Sida. 23: 1–300.

Geurts R, Bisseling T. 2002. *Rhizobium* Nod factor perception and signalling. Plant Cell. 14 Suppl: S239–249.

Laeremans T, Snoeck C, Mariën J, Verreth C, Martínez-Romero E, Promé JC, Vanderleyden J. 1999. *Phaseolus vulgaris* recognizes *Azorhizobium caulinodans* Nod factors with a variety of chemical substituents. Mol. Plant. Microbe Interact. 12: 820–824.

Lerouge P, Roche P, Faucher C, Maillet F, Truchet G, Promé JC, Dénarié J. 1990. Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. Nature 344: 781–784.

Madsen EB, Madsen LH, Radutoiu S, Olbryt M, Rakwalska M, Szczyglowski K, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J. 2003. A receptor kinasegene of the LysM type is involved in legume perception of rhizobial signals. Nature. 425: 637–640.

Maillet F, Poinsot V, André O, Puech-Pagès V, Haouy A, Gueunier M, Cromer L, Giraudet D, Formey D, Niebel A, Martinez EA, Driguez H, Bécard G, Dénarié J. 2011. Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. Nature. 469: 58–63.

Martínez-Romero E, Pardo MA, Cevallos MA, Palacios R. 1985. Reiteration of nitrogen fixation gene sequences and specificity of *Rhizobium* in nodulation and nitrogen fixation in *Phaseolus vulgaris*. J. Gen. Microbiol. 131: 1779–1786.

Martínez-Romero E. 2009. Coevolution in *Rhizobium*-legume symbiosis?. DNA Cell Biol. 28: 361–370.

Ormeño-Orrillo E, Vinuesa P, Zúñiga-Dávila D, Martínez-Romero E. 2006. Molecular diversity of native bradyrhizobia isolated from lima bean (*Phaseolus lunatus* L.) in Peru. Syst. Appl. Microbiol. 29: 253–62.

Parker MA, 2002. Bradyrhizobia from wild *Phaseolus*, *Desmodium*, and *Macroptilium* species in Northern Mexico. Appl. Environ. Microbiol. 68: 2044–2048.

Parniske M. 2008. Arbuscular mycorrhiza: the mother of plant root endosymbioses. Nat. Rev. Microbiol. 6: 763–775.

Poupot R, Martinez-Romero E, Promé JC. 1993. Nodulation factors from *Rhizobium tropici* are sulfated or nonsulfated chitopentasaccharides containing an N-methyl-N-acylglucosaminyl terminus. Biochemistry. 32: 10430–10435.

Poupot R, Martinez-Romero E, Gautier N, Promé JC. 1995. Wild type *Rhizobium etli*, a bean symbiont, produces acetyl-fucosylated, N-methylated, and carbamoylated nodulation factors. J. Biol. Chem. 270: 6050–6055.

Segovia L, Piñero D, Palacios R, Martínez-Romero E. 1991. Genetic structure of a soil population of nonsymbiotic *Rhizobium leguminosarum*. Appl. Environ. Microbiol. 57: 426–433.

Segovia L, Young JPW, Martínez-Romero E. 1993. Reclassification of American *Rhizobium leguminosarum* biovar phaseoli type I strains as *Rhizobium etli* sp. nov. Int. J. Syst. Bacteriol. 43: 374–377.

Servín-Garcidueñas LE, Zayas-Del Moral A, Ormeño-Orrillo E, Rogel MA, Delgado-Salinas A, Sánchez F, Martínez-Romero E. 2014. Symbiont shift towards *Rhizobium* nodulation in a group of phylogenetically related *Phaseolus* species. Mol. Phylogenet. Evol. 79: 1–11.

Somasegaran P, Hoben HJ, Lewinson L. 1991. Symbiotic interactions of *Phaseolus acutifolius* and *P. acutifolius* × *P. vulgaris* hybrid progeny in symbiosis with *Bradyrhizobium* spp. and *Rhizobium leguminosarum* bv. phaseoli. Can. J. Microbiol. 37: 497–503.

Souza V, Eguiarte L, Avila G, Cappello R, Gallardo C, Montoya J, Piñero D. 1994. Genetic structure of *Rhizobium etli* biovar phaseoli associated with wild and cultivated bean plants (*Phaseolus vulgaris* and *Phaseolus coccineus*) in Morelos, Mexico. Appl. Environ. Microbiol. 60:1260–1268.

Torres Tejerizo G, Del Papa MF, Soria-Diaz ME, Draghi W, Lozano M, Giusti Mde L, Manyani H, Megías M, Gil Serrano A, Pühler A, Niehaus K, Lagares A, Pistorio M. 2011. The nodulation of alfalfa by the acid-tolerant *Rhizobium* sp. strain LPU83 does not require sulfated forms of lipochitooligosaccharide nodulation signals. J. Bacteriol. 193: 30–39.

Yang S, Tang F, Gao M, Krishnan HB, Zhu H. 2010. R gene-controlled host specificity in the legume-rhizobia symbiosis. Proc. Natl. Acad. Sci. USA. 107: 18735–18740.

Capítulo III

Resultados adicionales

Diversidad genómica de simbiontes de insectos y artrópodos nativos de México

Introducción y antecedentes

En el laboratorio de la Dra. Esperanza Martínez, se creó una línea de investigación relacionada con el estudio de los simbiontes obligados y de las bacterias intestinales de insectos y artrópodos en el año 2007. En el laboratorio se han estudiado por muchos años las interacciones que ocurren entre bacterias fijadoras de nitrógeno y sus plantas hospederas como se mencionó en el capítulo II. El descubrimiento de que algunas bacterias intestinales son capaces de fijar nitrógeno para sus insectos hospederos (Pinto-Tomás et al., 2009) llevó al planteamiento de estudiar interacciones novedosas entre bacterias e insectos.

La microbiota intestinal humana y la de otros mamíferos son altamente complejas ya que contienen cientos de miles de especies diferentes de microorganismos, muchos de los cuáles no se han caracterizado ni se han podido cultivar (Lozupone et al., 2012). La mayoría de las microbiotas intestinales de insectos presentan una diversidad limitada por lo que la descripción completa de algunas de estas comunidades ha sido posible (Engel & Moran, 2013; Jing et al., 2014).

Algunos ejemplos de microbiotas intestinales reducidas incluyen las presentes en las polillas gitanas que contienen aproximadamente 14 especies (Broderick et al., 2004), las de las larvas de las mariposas blancas de la col que comprenden entre 6 y 15 especies (Robinson et al., 2010) y las de las moscas *Drosophila* que están conformadas por entre una y 30 especies (Broderick & Lemaitre, 2012).

La composición de la microbiota intestinal de los insectos es dependiente de la dieta (Broderick et al., 2004; Robinson et al., 2010; Colman et al., 2012; Wong et al., 2014; Yun et al., 2014; Pérez-Cobas et al, 2015). En los insectos existen otras asociaciones simbióticas más estrechas con las bacterias. Una gran cantidad de insectos presentan bacteriomas en su cavidad abdominal. Los bacteriomas son estructuras especializadas en donde se mantienen poblaciones bacterianas de entre uno y tres simbiontes obligados.

La mayoría de los simbiontes obligados de insectos no se pueden cultivar debido a reducciones drásticas de sus genomas que ocasionaron la pérdida de su capacidad metabólica para crecer en vida libre. Sin embargo, sus genomas han conservado genes involucrados en la biosíntesis de aminoácidos y por tanto, una de sus funciones esenciales es proveer a sus hospederos con aminoácidos que están limitados en sus dietas (Martínez-Cano et al., 2015). Gran parte del conocimiento actual sobre el potencial metabólico de los simbiontes obligados de insectos se debe a los avances de las técnicas de secuenciación masiva de ADN y de las herramientas bioinformáticas empleadas en estudios metagenómicos, transcriptómicos y proteómicos.

En el laboratorio de la Dra. Esperanza Martínez se planteó como primer objetivo descubrir la diversidad microbiana de insectos nativos de México. Se siguieron enfoques genómicos, metagenómicos y transcriptómicos para comprender las interacciones bacteria-insecto a partir del estudio del potencial genómico de las principales bacterias simbióticas.

Los primeros análisis de diversidad bacteriana que se realizaron en el laboratorio fueron en los insectos escama del género *Llaveia* (Rosas-Pérez et al., 2014) y en los insectos conocidos como cochinillas del carmín asociadas a plantas cactáceas y pertenecientes al género *Dactylopius* (Ramírez-Puebla et al., 2010). Adicionalmente se realizaron censos de diversidad bacteriana de especies de las familias *Ortheziidae*, *Monophlebidae*, *Diaspididae* y *Coccidae* que permitieron comprender algunos eventos evolutivos que han ocurrido entre simbiontes de insectos escama (Rosenblueth et al., 2012).

En *Llaveia* se detectó a una flavobacteria que corresponde a un género candidato que fue nombrado como *'Candidatus* Walczuchella monophlebidarum'. Mediante análisis bioinformáticos se logró obtener el genoma completo de 309,299 pares de bases de longitud de la flavobacteria. El análisis del contenido genético de este genoma sugiere que la flavobacteria probablemente provee a sus insectos hospederos con aminoácidos esenciales que no se encuentran presentes en su dieta a base de savia de plantas. También se detectó la presencia de una γ-proteobacteria que pudiera estar involucrada en el reciclaje de nitrógeno y en la biosíntesis de precursores de aminoácidos (Rosas-Pérez et al., 2014).

En *Dactylopius* se detectó a una β-proteobacteria en los huevecillos, en la hemolinfa y en tejidos especializados conocidos como bacteriocitos que albergan a simbiontes obligados de insectos. La β-proteobacteria descubierta corresponde a un género candidato que se nombró como '*Candidatus* Dactylopiibacterium carminicum'. Otras bacterias detectadas están relacionadas con los géneros *Massilia*, *Herbaspirillum*, *Acinetobacter*, *Mesorhizobium*, y *Sphingomonas* que se detectan comúnmente en tejidos de plantas por lo que es probable que su presencia en estas cochinillas sea resultado de adquisiciones de bacterias que se encuentran en las plantas (Ramírez-Puebla et al., 2010).

Algunos insectos y artrópodos no poseen tejidos especializados para albergar a bacterias simbióticas esenciales. En cambio poseen comunidades bacterianas en sus intestinos que participan en una gran cantidad de mecanismos celulares incluidos el reciclaje y aprovechamiento de nutrientes, defensa del hospedero, y en mecanismos de inmunidad. En el laboratorio, Luis Bolaños está analizando la diversidad bacteriana de los intestinos de dos especies alacranes que habitan en Morelos como parte de su tesis doctoral. Ha encontrado que los alacranes presentan comunidades bacterianas diversas que incluyen la presencia de bacterias relacionadas al género *Spiroplasma* (Bolaños et al., 2015).

Dentro de esta línea de investigación del laboratorio propuse analizar las comunidades bacterianas de las mariposas monarca. Las mariposas monarca realizan una migración anual a lo largo de Norteamérica para arribar a bosques de oyamel del occidente de México con la finalidad de entrar en un estado de dormancia que les permite sobrevivir el invierno. Las mariposas monarca han sufrido las consecuencias de pérdida de hábitat por deforestación y por cambio de uso de suelo para cosechas a gran escala, además de los efectos del uso masivo de pesticidas. Las poblaciones migratorias de mariposas monarca se han visto diezmadas en los últimos años y no se conocen las causas exactas aunque pudieran estar relacionadas con los problemas mencionados anteriormente. La importancia del proyecto reside en estudiar con enfoques moleculares a las comunidades bacterianas de insectos que entran en estado de dormancia y que reducen su dieta al mínimo por largos periodos de tiempo y que además se encuentran amenazados.

En este capítulo se presentan resultados iníciales del estudio de la diversidad microbiana de los intestinos de las mariposas monarca. El capítulo también incluye resultados de proyectos de colaboración sobre otros insectos y artrópodos de México.

Planteamiento del problema

La diversidad de la microbiota intestinal de especies de mariposas ha sido poco explorada y no se han realizado estudios en mariposas que presentan estados de dormancia. La microbiota de las mariposas monarca no se había analizado utilizando técnicas moleculares.

Hipótesis

La diversidad de la microbiota de las mariposas monarca es limitada debido a la reducción de su metabolismo y de su dieta, y por tanto sus simbiontes principales están enriquecidos.

Objetivos

Realizar censos de diversidad microbiana de la microbiota intestinal de mariposas monarca en estado de dormancia mediante técnicas moleculares y de cultivo.

Obtener el genoma del simbionte dominante de los intestinos de las mariposas.

Contribuir en otros proyectos de investigación del laboratorio relacionados con el estudio de bacterias simbiontes de insectos y artrópodos.

Resultados y discusión

Diversidad bacteriana de la microbiota intestinal de mariposas monarca

Los resultados preliminares de los censos de diversidad de la microbiota intestinal de mariposas monarca obtenidos mediante técnicas metagenómicas y de amplificación de genes ribosomales se presentan a continuación.

Colecta de mariposas monarca

Se colectaron mariposas de las colonias de hibernación que se localizan en la Reserva de la Biosfera de la Mariposa Monarca, en el Estado de Michoacán (Figura 1). Las mariposas se muestrearon durante dos años consecutivos para detectar la persistencia de las comunidades bacterianas. Durante enero del 2010 se colectaron mariposas en la Sierra Chivati-Huacal. En enero del 2011 se colectaron mariposas en el área protegida de El Rosario. Las mariposas se transportaron al laboratorio y fueron procesadas para los estudios al siguiente día de su colecta.

Figura 1. Mariposas monarca en los bosques de encimo de la Sierra Chivati-Huacal dentro de la Reserva de la Biosfera de la Mariposa Monarca.

Censos de diversidad bacteriana de la microbiota intestinal de mariposas monarca

En este estudio se amplificaron y se secuenciaron genes ribosomales 16S rRNA bacterianos para evaluar la diversidad de la microbiota intestinal de las mariposas monarca. En los censos de diversidad encontramos que las microbiotas intestinales presentan una abundancia de genes ribosomales 16S rRNA de una α-protebacteria que corresponde al género *Commensalibacter*. En todas las mariposas monarca analizadas se pudieron detectar secuencias correspondientes al género *Commensalibacter* y son las hembras quienes presentan una mayor abundancia de éstas secuencias (Figure 2). Sorprendentemente las secuencias de genes 16S rRNA obtenidas de mariposas monarca que corresponden a *Commensalibacter* fueron 100% idénticas.

Commensalibacter intestini ha sido aislada de los intestinos de la mosca *Drosophila melanogaster* y es la única especie caracterizada del género *Commensalibacter* (Roh et al., 2008). Se ha reportado que las poblaciones de *C. intestini* disminuyen en los intestinos de líneas de laboratorio de *D. melanogaster* genéticamente modificadas para producir concentraciones elevadas de péptidos antimicrobianos. En estos experimentos también se observó que las poblaciones de la bacteria patógena *Gluconobacter morbifer* se incrementaban y causaban la inflamación de los intestinos de *Drosophila* provocando finalmente la muerte a las moscas. Por lo tanto se ha propuesto que *D. melanogaster* regula las poblaciones bacterianas en sus intestinos mediante la producción de péptidos antimicrobianos mientras que *C. intestini* es capaz de mantener el control de poblaciones de bacterias patogénicas (Ryu et al., 2008).

Las secuencias de genes 16S rRNA de *Commensalibacter* de las mariposas monarca y de *C. intestini* están estrechamente relacionadas ya que presentan una identidad de secuencia nucleotídica de 98%. En análisis filogenéticos las secuencias de *Commensalibacter* de *Drosophila* y de mariposas monarcas forman grupos independientes que están más cercanamente relacionados entre sí que con otras secuencias recuperadas de otros insectos como abejas y abejorros.

Otras secuencias de genes ribosomales 16S rRNA obtenidos de las librerías genéticas corresponden a una α-protebacteria del género *Asaia*, a γ-proteobacterias de los géneros *Pseudomonas* y *Serratia* y a β-proteobacterias de los géneros *Xylophilus*, *Polaromonas*, *Herbaspirillum* y *Kinetoplastibacterium*. También se detectaron secuencias que corresponden a bacterias del ácido láctico de los géneros *Lactococcus*, *Carnobacterium* y *Enterococcus* que corresponden al filo Firmicutes (Figura 2).

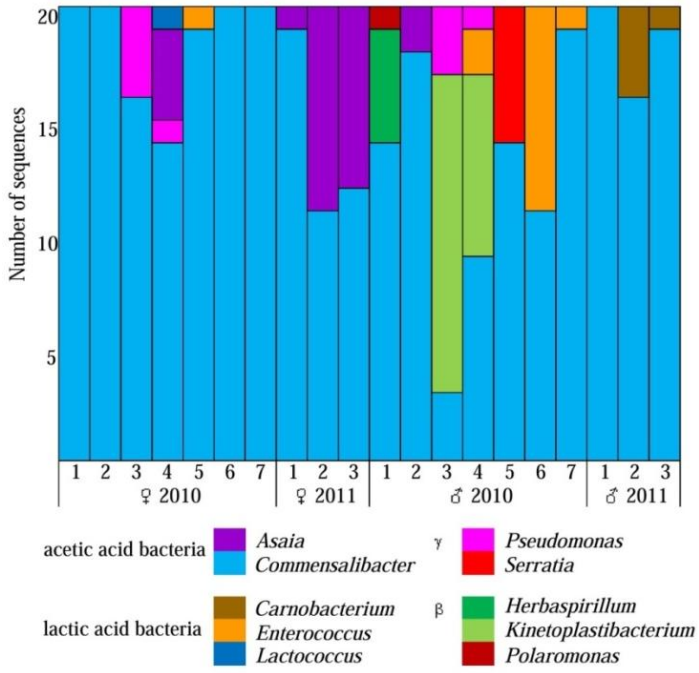


Figura 2. Diversidad bacteriana de la microbiota de mariposas monarca. Los censos de diversidad fueron obtenidos mediante la amplificación y construcción de librerías de genes ribosomales 16S rRNA.

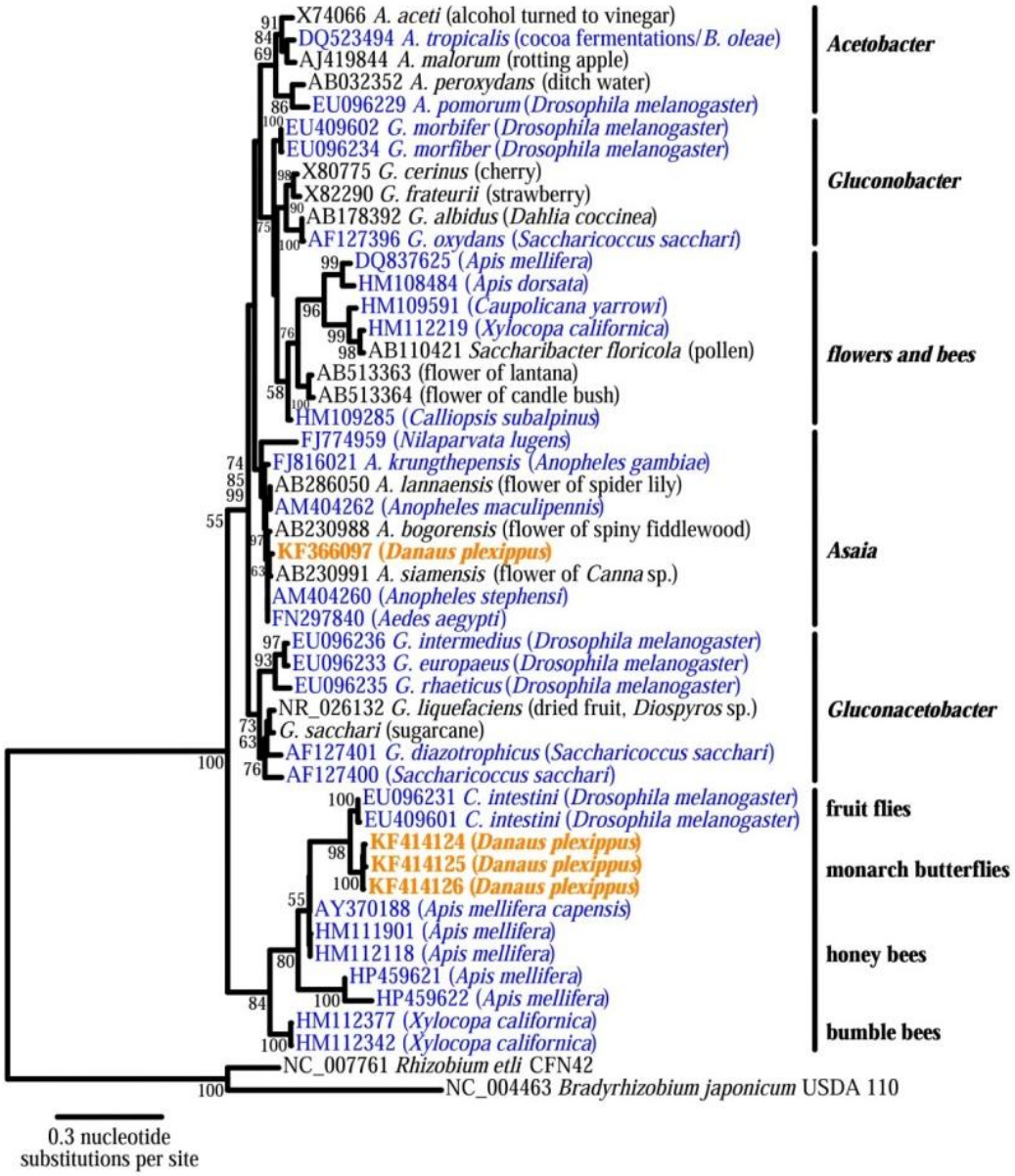


Figura 3. Reconstrucción filogenética de genes ribosomales 16S rRNA de acetobacterias asociadas a insectos. El análisis filogenético corresponde a un análisis de máxima verosimilitud realizado en PhyML (Guindon et al., 2010) utilizando el modelo de sustitución GTR+I+G. Los valores de soporte de las ramas corresponden a 100 réplicas de Bootstrap. En naranja se presentan las secuencias obtenidas de mariposas monarca. En azul se presentan secuencias recuperadas de microbiotas intestinales de otros insectos.

Artículo:

Servín-Garcidueñas LE, Sánchez-Quinto A, Martínez-Romero E. 2014. Draft genome sequence of *Commensalibacter papalotli* MX01, a symbiont identified from the guts of overwintering Monarch butterflies. Genome Announc. 2: pii:e00128-14.

La cepa MX01 perteneciente al género *Commensalibacter* se aisló a partir de muestras de intestinos de mariposas monarca utilizando medio de cultivo LMDA (Lee et al., 1975). El genoma de la cepa MX01 fue secuenciado y ensamblado. El genoma recuperado presenta una longitud de 2.3 megabases y un contenido de G+C de 36.7%. El genoma es de tamaño reducido y presenta un contenido bajo de G+C cuando se compara con otros genomas de acetobacterias (Tabla 1). La cepa MX01 corresponde a una especie nueva que fue nombrada como *Commensalibacter papalotli*. Las descripción sobre la secuenciación del genoma de *C. papalotli* MX01 se presenta en detalle en el artículo. Las posiciones filogenómicas de los genomas de *Commensalibacter* se presentan en la Figura 4.

Tabla 1. Características de genomas representativos de acetobacterias.

Especie con genoma secuenciado	**Longitud del genoma (Mbp)**	**Contenido de G+C**
Gluconacetobacter europaeus LMG 18890	4.23	61.2
Gluconacetobacter diazotrophicus PAl 5	3.91	66.3
Gluconobacter oxydans H24	3.82	56.2
Acetobacter tropicalis NBRC 101654	3.72	55.5
Acetobacter aceti ATCC 23746	3.69	57.0
Gluconacetobacter xylinus NBRC 3288	3.51	60.6
Gluconobacter frateurii NBRC 101659	3.15	55.7
Acetobacter pasteurianus NBRC 101655	3.02	52.7
Gluconobacter oxydans 621H	2.92	60.8
Gluconobacter morbifer G707	2.89	59.0
Acetobacter pasteurianus LMG 1262	2.89	53.1
Acetobacter pomorum DM001	2.88	52.3
Granulibacter bethesdensis CGDNIH1	2.71	59.1
Commensalibacter intestini A911	2.45	36.8
Saccharibacter floricola DSM 15669	2.38	51.2
***Commensalibacter papalotli* MX01**	**2.33**	**36.6**

genomeA
Journals.ASM.org

Draft Genome Sequence of *Commensalibacter papalotli* MX01, a Symbiont Identified from the Guts of Overwintering Monarch Butterflies

Luis E. Servín-Garcidueñas, Andrés Sánchez-Quinto, Esperanza Martínez-Romero

Center for Genomic Sciences, Department of Ecological Genomics, National University of Mexico, Cuernavaca, Morelos, Mexico

We report the draft genome sequence of *Commensalibacter papalotli* strain MX01, isolated from the intestines of an overwintering monarch butterfly. The 2,332,652-bp AT-biased genome of *C. papalotli* MX01 is the smallest genome for a member of the *Acetobacteraceae* family and provides the first evidence of plasmids in *Commensalibacter*.

Received 4 February 2014 **Accepted** 13 February 2014 **Published** 6 March 2014

Citation Servín-Garcidueñas LE, Sánchez-Quinto A, Martínez-Romero E. 2014. Draft genome sequence of *Commensalibacter papalotli* MX01, a symbiont identified from the guts of overwintering monarch butterflies. Genome Announc. 2(2):e00128-14. doi:10.1128/genomeA.00128-14.

Address correspondence to Luis E. Servín-Garcidueñas, luis.e.servin@gmail.com.

Adult monarch butterflies feed on nectar that provides sugars and other nutrients. Monarch butterflies migrate to Mexican forests for overwintering. Overwintering monarchs reduce their metabolism and limit their feeding. Few studies have described gut bacteria in butterflies (1–8), and there are no studies on the molecular identification of symbionts from monarch butterflies.

Commensalibacter is a newly described genus of acetic acid bacteria (9). 16S rRNA gene sequences related to the *Commensalibacter* genus have been recovered from the guts of *Drosophila* species (9, 10), honey bees, and bumble bees (11, 12), as well as from *Heliconius erato* butterflies (1). The type strain *Commensalibacter intestini* A911 was isolated from *Drosophila* intestines (9), and its draft genome sequence has been reported (13). Here, we report the genome of a *Commensalibacter* symbiont isolated from a monarch butterfly.

A female monarch was collected in January 2010 from the Monarch Butterfly Biosphere Reserve in Mexico. Its gut contents were inoculated on Lee's multi-differential agar (LMDA) medium (14) and were incubated at 25°C for 1 week. Colonies were identified by amplifying and sequencing 16S rRNA genes. Genomic DNA from *Commensalibacter papalotli* strain MX01 was isolated using the UltraClean microbial DNA kit (Mo-Bio Laboratories, Inc., Carlsbad, CA). The genome was sequenced using the Illumina HiSeq 2000 platform with a paired-end library. A total of 49,933,356 100-bp reads were generated. Twelve contigs with a sequence length of 2,332,652 bp were assembled *de novo* using Velvet version 1.2.10 (15). The genome coverage reached 2,117.64-fold, and the N_{50} length is 1,547,573 bp. Genome annotation was performed by the NCBI Prokaryotic Genomes Annotation Pipeline version 2.0 (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/). Average nucleotide identity (ANI) values were calculated as previously proposed (16) using the ANI calculator from the Kostas lab (http://enve-omics.ce.gatech.edu/ani/).

The genome of strain MX01 has a G+C content of 36.66% and consists of a chromosome and two complete circular plasmids. The chromosome is contained in 10 contigs, with a length of 2,310,436 bp. Plasmid pA is 15,425 bp and codes for a VapBC toxin-antitoxin system, plasmid replication and partitioning (encoded by *repA*, *parA*, and *parB*), a putative glycosyltransferase, a putative carbohydrate-selective porin (OprB), and proteins for iron acquisition and metabolism, including a transporter, a lipoprotein, a permease, and a polyferredoxin. Plasmid pB is 6,791 bp and codes for a RepB protein, an integrase, and hypothetical proteins. A total of 2,105 genes were predicted, including 2,060 protein-coding genes, 3 rRNAs, and 43 tRNAs. Function predictions were assigned to 1,718 protein-coding genes.

The 16S rRNA gene from strain MX01 has 98% sequence identity with the corresponding gene of *C. intestini* A911^T. Genome comparisons between strains MX01 and A911^T revealed an ANI value of 82.95%. An ANI boundary of 95 to 96% has been useful for taxonomically circumscribing prokaryotic species (17). Strain MX01 then corresponds to a novel *Commensalibacter* species. The name *C. papalotli* MX01 is proposed. The word "papalotl" means butterfly in the Mexican Náhuatl language. A detailed characterization of *C. papalotli* MX01 is currently in progress.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. ATSX00000000. The version described in this paper is version ATSX01000000.

ACKNOWLEDGMENTS

This research was supported by CONACyT 154453. L.E.S.-G. received a Ph.D. scholarship from CONACyT.

We thank the Programa de Doctorado en Ciencias Biomédicas from UNAM, and David Romero, who provided a research permit as Director of the Center for Genomic Sciences, UNAM. Rosendo Antonio Caro-Gómez provided a permit for samplings at the Monarch Butterfly Biosphere Reserve (MBBR) in Mexico. Felipe Martínez, Katia Lemus, and Genaro Mondragón from the MBBR provided crucial support for samplings. The Ministry of Environment and Natural Resources of Mexico also provided a permit for samplings.

REFERENCES

1. **Hammer TJ, McMillan WO, Fierer N.** 2014. Metamorphosis of a butterfly-associated bacterial community. PLoS One **9:**e86995. http://dx.doi.org/10.1371/journal.pone.0086995.
2. **Kim JY, Lee J, Shin NR, Yun JH, Whon TW, Kim MS, Jung MJ, Roh SW, Hyun DW, Bae JW.** 2013. *Orbus sasakiae* sp. nov., a bacterium isolated from the gut of the butterfly *Sasakia charonda*, and emended description of the genus *Orbus*. Int. J. Syst. Evol. Microbiol. **63:** 1766–1770. http://dx.doi.org/10.1099/ijs.0.041871-0.
3. **Kingsley VV.** 1972. Persistence of intestinal bacteria in the developmental stages of the monarch butterfly (*Danaus plexippus*). J. Invertebr. Pathol. **20:**51–58. http://dx.doi.org/10.1016/0022-2011(72)90081-X.
4. **Narita S, Nomura M, Kato Y, Fukatsu T.** 2006. Genetic structure of sibling butterfly species affected by *Wolbachia* infection sweep: evolutionary and biogeographical implications. Mol. Ecol. **15:**1095–1108. http://dx.doi.org/10.1111/j.1365-294X.2006.02857.x.
5. **Robinson CJ, Schloss P, Ramos Y, Raffa K, Handelsman J.** 2010. Robustness of the bacterial community in the cabbage white butterfly larval midgut. Microb. Ecol. **59:**199–211. http://dx.doi.org/10.1007/s00248-009-9595-8.
6. **Russell JA, Funaro CF, Giraldo YM, Goldman-Huertas B, Suh D, Kronauer DJ, Moreau CS, Pierce NE.** 2012. A veritable menagerie of heritable bacteria from ants, butterflies, and beyond: broad molecular surveys and a systematic review. PLoS One **7:**e51027. http://dx.doi.org/10.1371/journal.pone.0051027.
7. **Steinhaus EA.** 1941. A study of the bacteria associated with thirty species of insects. J. Bacteriol. **42:**757–790.
8. **Tagami Y, Miura K.** 2004. Distribution and prevalence of *Wolbachia* in Japanese populations of *Lepidoptera*. Insect Mol. Biol. **13:**359–364. http://dx.doi.org/10.1111/j.0962-1075.2004.00492.x.
9. **Roh SW, Nam YD, Chang HW, Kim KH, Kim MS, Ryu JH, Kim SH, Lee WJ, Bae JW.** 2008. Phylogenetic characterization of two novel commensal bacteria involved with innate immune homeostasis in *Drosophila melanogaster*. Appl. Environ. Microbiol. **74:**6171–6177. http://dx.doi.org/10.1128/AEM.00301-08.
10. **Chandler JA, Lang JM, Bhatnagar S, Eisen JA, Kopp A.** 2011. Bacterial communities of diverse *Drosophila* species: ecological context of a host-microbe model system. PLoS Genet. **7:**e1002272. http://dx.doi.org/10.1371/journal.pgen.1002272.
11. **Jeyaprakash A, Hoy MA, Allsopp MH.** 2003. Bacterial diversity in worker adults of *Apis mellifera capensis* and *Apis mellifera scutellata* (Insecta: Hymenoptera) assessed using 16S rRNA sequences. J. Invertebr. Pathol. **84:** 96–103. http://dx.doi.org/10.1016/j.jip.2003.08.007.
12. **Martinson VG, Danforth BN, Minckley RL, Rueppell O, Tingek S, Moran NA.** 2011. A simple and distinctive microbiota associated with honey bees and bumble bees. Mol. Ecol. **20:**619–628. http://dx.doi.org/10.1111/j.1365-294X.2010.04959.x.
13. **Kim EK, Kim SH, Nam HJ, Choi MK, Lee KA, Choi SH, Seo YY, You H, Kim B, Lee WJ.** 2012. Draft genome sequence of *Commensalibacter intestini* A911T, a symbiotic bacterium isolated from *Drosophila melanogaster* intestine. J. Bacteriol. **194:**1246. http://dx.doi.org/10.1128/JB.06669-11.
14. **Lee SY, Jangaard O, Coors JH.** 1975. Lee's multi-differential agar (LMDA); a culture medium for enumeration and identification of brewing bacteria. J. Am. Soc. Brewing Chem. **33:**18–25.
15. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
16. **Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM.** 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. **57:** 81–91. http://dx.doi.org/10.1099/ijs.0.64483-0.
17. **Richter M, Rosselló-Móra R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. **106:** 19126–19131. http://dx.doi.org/10.1073/pnas.0906412106.

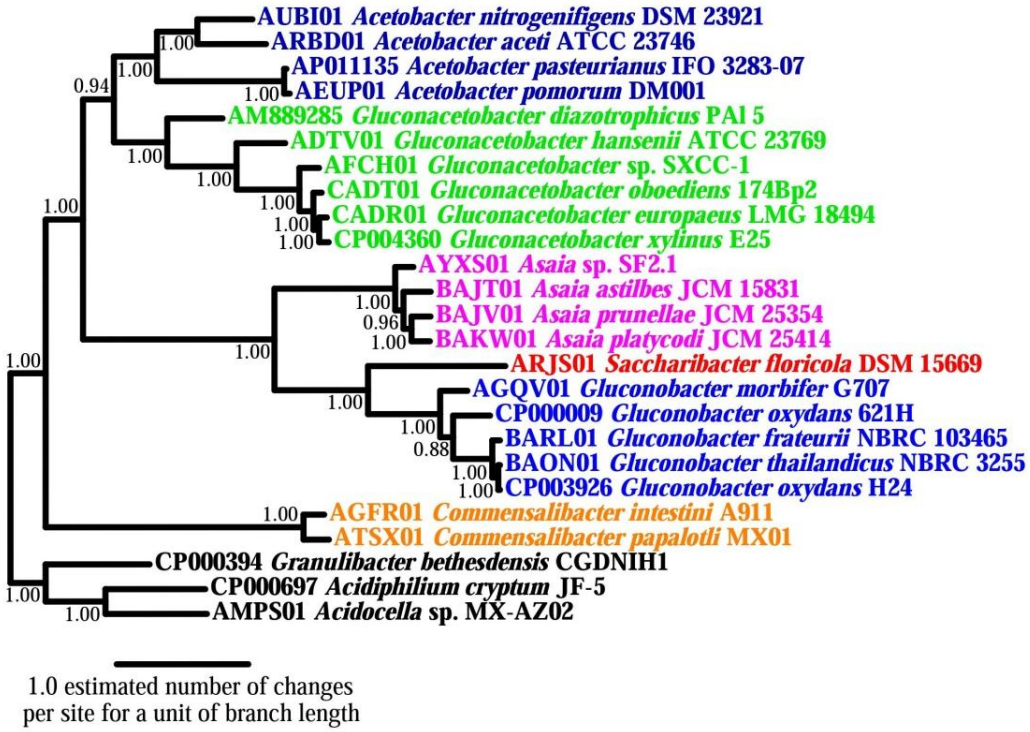


Figura 4. Reconstrucción filogenómica de genomas representativos de acetobacterias. El análisis filogenómico se realizó utilizando la metodología de PhyloPhlAn (Segata et al., 2013) y está basado en 400 proteínas conservadas. Los genomas correspondientes a diferentes géneros se muestran en colores distintos. Los genomas de *Commensalibacter* se ubican en una rama basal al resto de las acetobacterias que se han detectado en insectos.

Censo metagenómico de diversidad bacteriana

Una muestra de ADN total obtenido de los intestinos completos de tres mariposas hembras fue secuenciada para evitar los sesgos inherentes a las técnicas de amplificación de genes ribosomales. La muestra de ADN sirvió como proyecto piloto de la Unidad Universitaria de Secuenciación Masiva de la UNAM.

La muestra de ADN se secuenció sin ningún tipo de amplificación o enriquecimiento y se generaron 2.5 Gb de secuencias metagenómicas. Se obtuvieron un total de 69,820,263 lecturas de 35 pares de bases que no pudieron ser ensambladas debido a la gran cantidad de lecturas cortas y el poder computacional requerido. Por lo tanto, se utilizaron las secuencias de los genes ribosomales 16S rRNA de las librerías de clonas para identificar lecturas mediante búsquedas de BLAST que correspondieran a genes ribosomales 16S rRNA. De éstas búsquedas, se lograron recuperar lecturas que correspondían al gen 16S rRNA completo de '*Commensalibacter papalotli*'. Así se usaron los genomas de la mariposa monarca y de *C. papalotli* para mapear lecturas y estimar su abundancia en el metagenoma (Tabla 2). Los datos metagenómicos preliminares indican que la diversidad de la microbiota intestinal de las mariposas monarca es restringida y sustentan la abundancia de *Commensalibacter*.

Tabla 2. Estadísticas de metagenoma intestinal preliminar de las mariposas monarca.

Número de lecturas metagenómicas	69,820 263
Longitud de las lecturas	35 pares de bases
Total de secuencia del metagenoma	2 443.70 megabases
Número de lecturas que mapean con el genoma de la mariposa	60,529 901
Porcentaje de lecturas que mapean con el genoma de la mariposa	86.69 %
Total de secuencia que mapea contra el genoma de la mariposa	2 118.54 Mb
Cobertura de secuencia que mapea contra el genoma de la mariposa	7.9 X
Número de lecturas que mapean con el genoma de *C. papalotli*	326 609
Porcentaje de lecturas que mapean con el genoma *C. papalotli*	0.46 %
Total de secuencia que mapea contra el genoma de *C. papalotli*	11.43 Mb
Cobertura de secuencia que mapea contra el genoma de *C. papalotli*	5 X

Detección de Commensalibacter en tejidos de intestino de mariposas

Se identificó la presencia de *C. papalotli* en muestras de tejidos de intestinos de mariposas monarca mediante técnicas de FISH (*Fluorescent In Situ Hybridization*) y microscopía electrónica y con focal. La sonda específica utilizada para la técnica de FISH se diseñó a partir de la secuencia del gen ribosomal 16S rRNA de *C. papalotli*. La técnica de FISH empleada fue la reportada por Parsley et al., 2010.

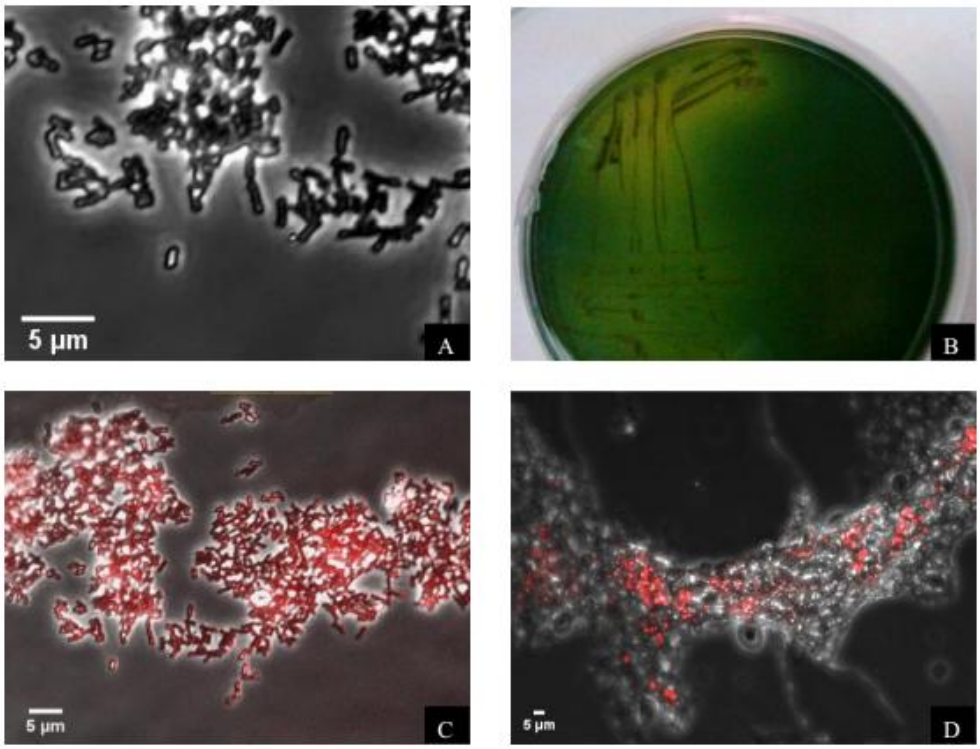


Figura 5. (A) Células de *Commensalibacter papalotli* MX01. (B) Caja de cultivo de *C. papalotli* MX01 en medio LMDA (Lee et al., 1975). La presencia de halos de color amarillo indica la producción de ácidos orgánicos. (C) Células cultivadas de *C. papalotli* MX01 y marcadas con la sonda fluorescente de FISH. (D) Células de *C. papalotli* MX01 en muestras de intestinos pre-tratados con la sonda fluorescente de FISH.

Diferencias genómicas entre genomas de Commensalibacter

Los genomas de *Commensalibacter* recuperados de la microbiota de la mosca *D. melanogaster* y de las mariposas monarca presentan una gran cantidad de bloques genómicos conservados mediante el alineamiento global de sus genomas. Sin embargo, también existe una gran cantidad de rearreglos de los bloques conservados (Figura 6A).

Una de las diferencias más significativas entre ambos genomas es la presencia de dos plásmidos en el genoma de *C. papalotli*. El plásmido pA codifica para genes relacionados con la biosíntesis de polisacáridos y el transporte de hierro además de genes involucrados en replicación de plásmidos y una par de genes toxina/antitoxina *vapBC.* La función del plásmido pB permanece enigmática debido a que los genes que codifica no tienen asignación funcional a excepción de un gen *repB*, una transposasa y una posible excicionasa (Figura 6B). Es interesante notar que el plásmido pA presenta un gen *repA* pero carece de un gen *repB* que sí está presente en el plámido pB. Los genes *repA* y *repB* forman parte del sistema de reparto de plásmidos. Es probable que la replicación de los plásmidos pA y pB sea coordinada.

En el genoma de *C. intestini* existen dos contigs que presentan genes que codifican para proteínas de replicación y proteínas Mob relacionadas con la replicación y movilización de plásmidos (Figura 6C). Sin embargo debido al estatus fragmentado del genoma de *C. intestini* no se puede determinar si son fragmentos de secuencia de plásmidos o regiones integradas en el cromosoma principal.

Algunos otros genes específicos para cada uno de los genomas de *Commensalibacter* se presentan en las Tablas 3 y 4.

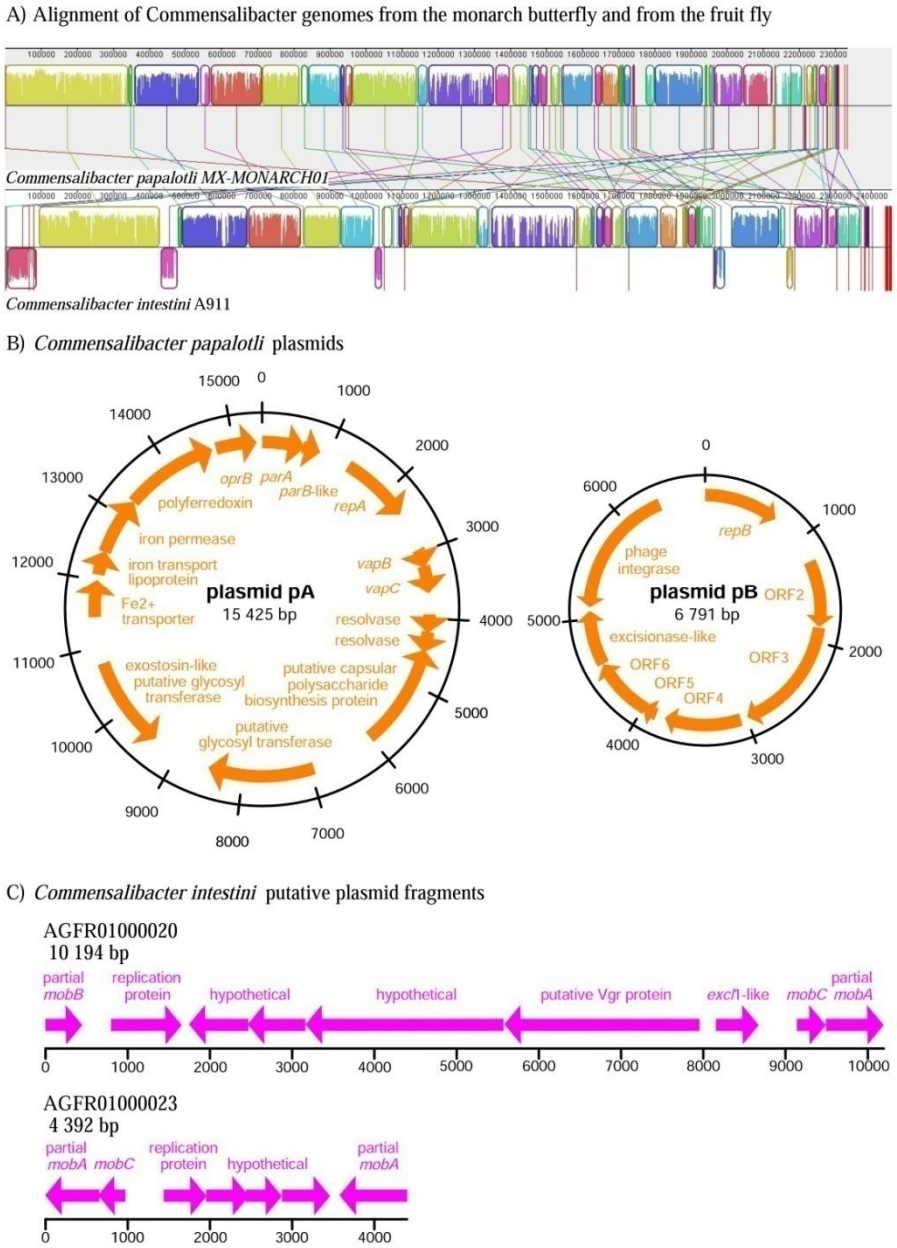


Figura 6. Diferencias genómicas entre los genomas de *Commensalibacter*. (A) Alineamiento global de los genomas de *C. intestini* y *C. papalotli*. (B) Representación de los genomas de los plásmidos del genoma de *C. papalotli*. (C) Representación de los contigs parciales de *C. intestini* que pudieran corresponder a secuencias de plásmidos.

Tabla 3. Genes específicos presentes en el genoma de *C. papalotli*.

1,4-alpha-glucan (glycogen) branching enzyme, GH-13-type (EC 2.4.1.18)
4-carboxymuconolactone decarboxylase domain
methylthioadenosine nucleosidase (EC 3.2.2.16)
5-deoxy-glucuronate isomerase (EC 5.3.1.-)
5-Hydroxyisourate hydrolase (HIUase) (EC 3.5.2.17)
5-keto-2-deoxygluconokinase (EC 2.7.1.92)
6-aminohexanoate-dimer hydrolase(EC:3.5.1.46)
6-phosphofructokinase (EC 2.7.1.11), PF08013 family
ABC transporter involved in cytochrome c biogenesis, CcmB subunit
acid phosphatase/vanadium-dependent haloperoxidase related
Arabinose operon regulatory protein
Arsenic resistance protein ArsH
Arsenical resistance operon repressor
Arsenical-resistance protein ACR3
Beta-galactosidase (EC 3.2.1.23)
Beta-lactamase (EC 3.5.2.6)
Beta-lactamase class C and other penicillin binding proteins
Capsular polysaccharide biosynthesis protein-like protein
Chorismate--pyruvate lyase (EC 4.1.3.40)
Conserved ATP-binding protein YghS
Conserved protein YghT, with nucleoside triphosphate hydrolase domain
C-terminal domain of CinA type S
Cytochrome oxidase biogenesis protein Sco1/SenC/PrrC, putative copper metallochaperone
Epi-inositol hydrolase (EC 3.7.1.-)
Error-prone repair protein UmuD
Error-prone, lesion bypass DNA polymerase V (UmuC)
FIG002473: Protein YcaR in KDO2-Lipid A biosynthesis cluster
FIG003437: hypothetical with DnaJ-like domain
FMN-dependent NADH-azoreductase
Fumarylacetoacetate hydrolase family protein
Glycogen synthase, ADP-glucose transglucosylase (EC 2.4.1.21)
Histidinol-phosphatase [alternative form] (EC 3.1.3.15)
HTH transcriptional regulator MerR family
Hydroxyethylthiazole kinase (EC 2.7.1.50)
Inhibitor of vertebrate lysozyme precursor

Inner membrane protein YihY, formerly thought to be RNase BN
Inososе dehydratase (EC 4.2.1.44)
Lactose permease
L-arabinose isomerase (EC 5.3.1.4)
Large exoproteins involved in heme utilization or adhesion
Large exoproteins involved in heme utilization or adhesion
Large exoproteins involved in heme utilization or adhesion
lemA protein
L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4)
Lysyl-lysine 2,3-aminomutase
Lysyl-tRNA synthetase (class II) (EC 6.1.1.6)
Metal-dependent hydrolase YbeY, involved in rRNA and/or ribosome maturation and assembly
Mg(2+) transport ATPase, P-type (EC 3.6.3.2)
Molybdopterin biosynthesis protein MoeB
Myo-inositol 2-dehydrogenase (EC 1.1.1.18)
NifU-like domain protein
Nitrogen regulation protein NR(I)
Peptide chain release factor 2
Phage tail fiber protein
Phosphatidylinositol-specific phospholipase C (EC 4.6.1.13)
Phosphomethylpyrimidine kinase (EC 2.7.4.7)
Predicted transcriptional regulator of the myo-inositol catabolic operon
programmed frameshift-containing
Putative activity regulator of membrane protease YbbK
putative ROK family transcriptional regulator
Putative stomatin/prohibitin-family membrane protease subunit YbbK
Ribonuclease E (EC 3.1.26.12)
S-adenosylhomocysteine nucleosidase (EC 3.2.2.9)
Shikimate 5-dehydrogenase I gamma (EC 1.1.1.25)
Thiaminase II (EC 3.5.99.2)
Transcriptional repressor of the lac operon
Transketolase, C-terminal section (EC 2.2.1.1)
Transketolase, N-terminal section (EC 2.2.1.1)
UbiA prenyltransferase
Ureidoglycolate/malate/sulfolactate dehydrogenase family (EC 1.1.1.-)

Continuación Tabla 3. Genes específicos presentes en el genoma de *C. papalotli*.

Tabla 4. Genes específicos presentes en el genoma de *C. intestini*.

Arginine/ornithine antiporter ArcD
Arginine decarboxylase (EC 4.1.1.19)
Ornithine decarboxylase (EC 4.1.1.17)
Para-aminobenzoate synthase, amidotransferase component (EC 2.6.1.85)
Aspartate racemase (EC 5.1.1.13)
N-acetylglucosamine kinase of eukaryotic type (EC 2.7.1.59)
N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25)
NAD-dependent malic enzyme (EC 1.1.1.38)
3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35)
3-hydroxybutyryl-CoA epimerase (EC 5.1.2.3)
Acetyl-CoA acetyltransferase (EC 2.3.1.9)
Enoyl-CoA hydratase (EC 4.2.1.17)
Transcriptional regulator, PadR family
UPF0301 protein YqgE
2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase (EC 1.14.13.-)
2-octaprenyl-6-methoxyphenol hydroxylase (EC 1.14.13.-)
3-ketoacyl-CoA thiolase (EC 2.3.1.16)
Flavodoxin 2
NAD(P)H oxidoreductase YRKL (EC 1.6.99.-)
Putative phosphatase YieH
DNA-damage-inducible protein J
Hydroxyaromatic non-oxidative decarboxylase protein C (EC 4.1.1.-)
Hydroxyaromatic non-oxidative decarboxylase protein D (EC 4.1.1.-)
Dihydroneopterin aldolase (EC 4.1.2.25)
3-demethylubiquinol 3-O-methyltransferase (EC 2.1.1.64)
3-polyprenyl-4-hydroxybenzoate carboxy-lyase (EC 4.1.1.-)
YrbA protein
GDP-mannose pyrophosphatase YffH
dNTP triphosphohydrolase, broad substrate specificity, subgroup 2
Phosphate transport system permease protein PstA (TC 3.A.1.7.1)
Translation elongation factor P Lys34:lysine transferase
YafQ toxin protein
Adenylate cyclase (EC 4.6.1.1)
Ferredoxin--NADP(+) reductase (EC 1.18.1.2)
Glutaredoxin

Uncharacterized monothiol glutaredoxin ycf64-like
Alpha-ketoglutarate-dependent taurine dioxygenase (EC 1.14.11.17)
Membrane fusion protein of RND family multidrug efflux pump
Transcriptional regulator, MerR family
DNA-methyltransferase
4-hydroxyphenylpyruvate dioxygenase
Glutamate decarboxylase
Probable glutamate/gamma-aminobutyrate antiporter
Dienelactone hydrolase
Carbonic anhydrase
Lipopolysaccharide biosynthesis protein WzxC
Colicin E5 immunity protein
Phenazine biosynthesis protein PhzF
Bacteriophage tail sheath protein
Putative phage tail tube protein
Phage tail/DNA circulation protein
Bacteriophage tail protein
Prophage baseplate assembly protein V
Bacteriophage protein GP46
Phage terminase large subunit
Phage FluMu protein gp47
VgrG protein
glycerophosphoryl diester phosphodiesterase
Xylulose kinase
Polygalacturonase
Enoyl-CoA hydratase
Chaperone protein TorD
L-rhamnose-proton symporter
Exopolysaccharide biosynthesis protein-like protein
Putative NADPH-quinone reductase
Large exoproteins involved in heme utilization or adhesion
Altronate hydrolase
3-polyprenyl-4-hydroxybenzoate carboxy-lyase UbiX
Calcium/calmodulin-dependent protein kinase kinase 2

Continuación Tabla 4. Genes específicos presentes en el genoma de *C. intestini*.

Artículo:

Servín-Garcidueñas LE, Martínez-Romero E. 2014. Complete mitochondrial genome recovered from the gut metagenome of overwintering monarch butterflies, *Danaus plexippus* (L.) (Lepidoptera: *Nymphalidae*, *Danainae*). Mitochondrial DNA. 25: 427–428.

El metagenoma intestinal de las mariposas monarca contiene una sobrerrepresentación de secuencias que corresponden al genoma mitocondrial de las mariposas monarca. Mediante análisis bioinformáticos se logró reconstruir el genoma completo de la mitocondria que no había sido reportado en la descripción del genoma completo de la mariposa monarca (Zhan et al., 2011). Las características del genoma mitocondrial se describen en detalle en el artículo.

Mitochondrial DNA

http://informahealthcare.com/mdn
ISSN: 1940-1736 (print), 1940-1744 (electronic)

Mitochondrial DNA, 2014; 25(6): 427–428
 DOI: 10.3109/19401736.2013.809441

informa
healthcare

MITOGENOME ANNOUNCEMENT

Complete mitochondrial genome recovered from the gut metagenome of overwintering monarch butterflies, *Danaus plexippus* (L.) (Lepidoptera: Nymphalidae, Danainae)

Luis E. Servín-Garcidueñas and Esperanza Martínez-Romero

Centro de Ciencias Genómicas, UNAM, Cuernavaca, Morelos, Mexico

Abstract

We present a 15,314 bp mitochondrial genome (mitogenome) sequence from monarch butterflies overwintering in Mexico. The complete mitogenome was generated by next generation sequencing techniques and was reconstructed by iterative assembly of reads from a metagenomic study of pooled butterfly gut DNA. The mitogenome codes for 13 putative protein coding genes, 22 tRNA genes, the large and small rRNA genes, and contains the A + T-rich sequence corresponding to the control region. The consensus sequence presented here has a depth of coverage of 142-fold and only three putative single nucleotide polymorphisms could be detected. The recovered *D. plexippus* mitogenome represents the second analyzed for the subfamily Danainae and accordingly, the closest available sequenced mitogenome was found to be the one corresponding to *Euploea mulciber* (Lepidoptera: Nymphalidae, Danainae).

Keywords

Danaus plexippus, lepidoptera, metagenomics, mitochondrial genome, Nymphalidae

History

Received 15 May 2013
Accepted 21 May 2013
Published online 5 July 2013

We present a mitogenome sequence of the monarch butterfly with the GenBank accession number KC836923. The mitogenome was designated with the acronym DPMMX (***D****anaus* ***p****lexippus* **m**itogenome -**Mex**ico) as it is a consensus sequence obtained by metagenomic sequencing of gut DNA from overwintering butterflies.

Butterflies were collected during January 2010 from the Sierra Chivati-Huacal at the Monarch Butterfly Biosphere Reserve in Mexico. Total DNA was purified from the guts of three female butterflies using the UltraClean microbial DNA kit (MoBio Laboratories, Inc., Carlsbad, CA). Metagenomic sequencing was performed on an Illumina GAIIx platform producing 2.51 Gbp. Reads were assembled using Velvet 1.2.07, Cambridge, UK (Zerbino & Birney, 2008). A set of 13 contigs were predicted by BLAST searches to be of mitochondrial origin. Gaps were closed iteratively by mapping and reassembling reads to these contigs using Maq 0.7.1, Cambridge, UK (Li et al., 2008) and Velvet. Additionally, PCRs were performed to amplify and sequence regions within gaps or homopolimeric regions and for validation. The primers used were MDs10-modified 5′-AACAGGATCAAATAATCCATTAGG-3′ and MDs11 5′-AAATTACCTTAGGGATAACAGCG-3′, MDs14 5′-TCGTGGATTATCAATTATTAAACAGATTCC-3′ and MDs2L-modified 5′-GCTGTAATACCTACTGCTC-3′ (Cameron & Whiting, 2008) and 12SB 5′-AAACTAGGATTAGATACCC-3′ (Skerratt et al., 2002) and 16SB-modified 5′-CACCGGTTTGAACTCAGATCA-3′ (Bybee et al., 2004). Open reading frames were predicted using GeneMark.hmm2.0, Atlanta, GA (Besemer & Borodovsky, 1999). Features were manually verified using Artemis, Cambridge, UK (Carver et al., 2005).

The length of DPMMX was determined to be 15,314 bp and its nucleotide composition is biased exhibiting 81.4% A + T content. The average coverage of the circular genome was 142-fold. Only three candidate single nucleotide polymorphisms (SNPs) were detected by Maq (T323, C1712 and G1732). The mitogenome codes for 13 protein coding genes, 22 tRNA genes, the large and small rRNA genes, and contains a control region with an A + T content of 94.5% (Figure 1) exhibiting some conserved lepidopteran motifs including the ''ATAGA'' and 19 bp poly (T) stretch. The protein encoding genes have typical mitochondrial ATN start codons, except for cox1, which contains the unusual CGA codon and which also shows an incomplete stop codon.

Blast searches were performed using DPMMX as a query against the MonarchBase (Zhan & Reppert, 2013), a database repository for the data of the *D. plexippus* genome (Zhan et al., 2011). A contig of 24,802 bp could be detected with accession number AGBW01003356.1 corresponding to the non-described mitogenome. This unedited contig shows redundancy at its ends and should be circular and its sequence is conserved with DPMMX albeit with dissimilarities. Major differences are an additional 1954 bp region containing repetitive TA bases and a non-conserved sequence not present in DPMMX and the absence of a 221 bp region conserved among other sequenced lepidopteran mitogenomes including DPMMX. Differences could be attributed to natural variation, sequencing errors or misassembles at homopolimeric regions. In conclusion, the mitogenome of *D. plexippus* presents a conserved structure and shows few polymorphisms in accordance with the low genetic differentiation

Correspondence: Luis E. Servín-Garcidueñas, Centro de Ciencias Genómicas, Av. Universidad S/N, CP 62210, Cuernavaca, Morelos, Mexico. Tel/Fax: (777) 3291897. E-mail: luis.e.servin@gmail.com

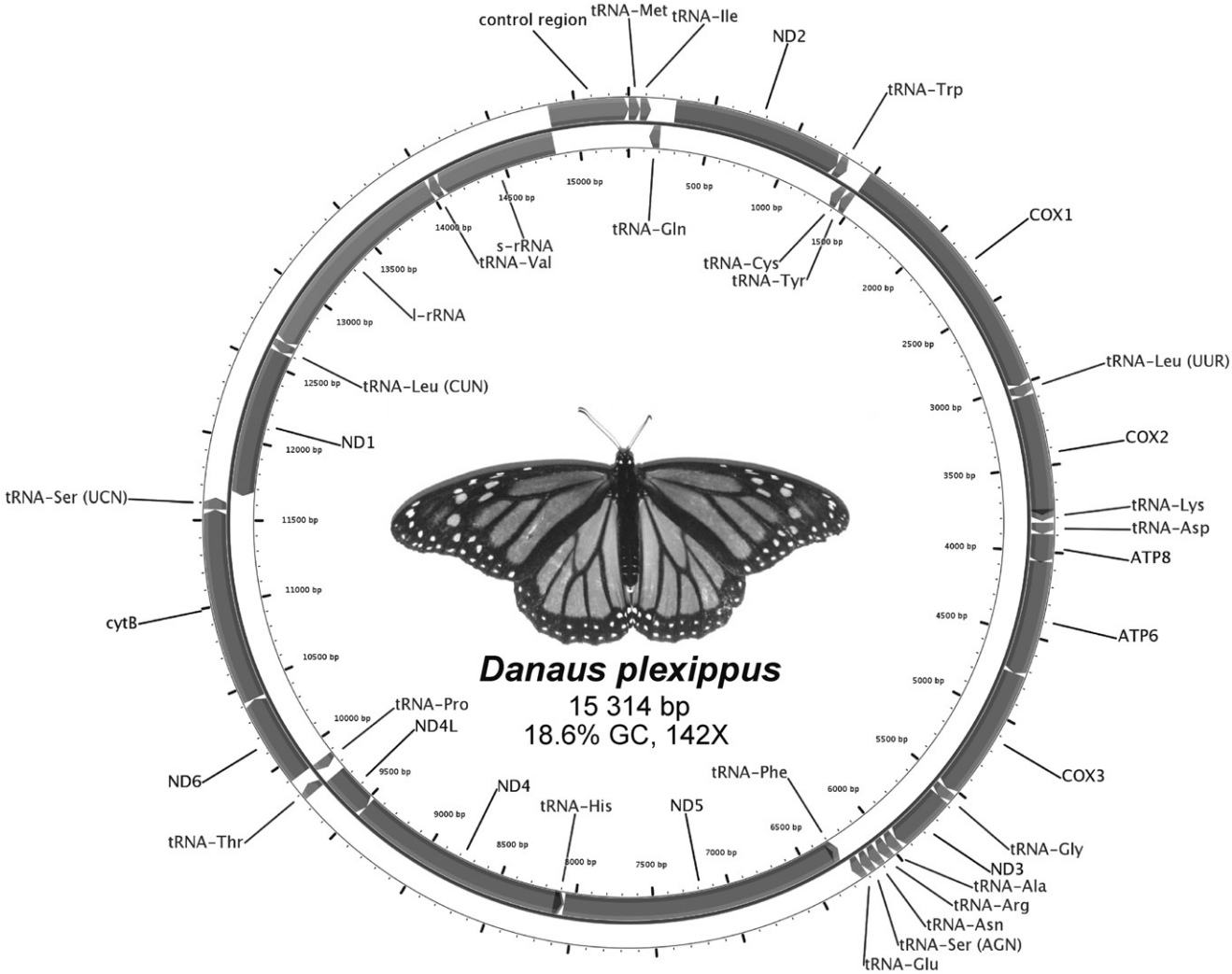


Figure 1. Circular map of the DPMMX version of the mitochondrial genome of *Danaus plexippus*.

reported for monarch butterflies (Brower & Boyce, 1991; Brower & Jeansonne, 2004; Lyons et al., 2012).

Acknowledgements

We are thankful to the Programa de Doctorado en Ciencias Biomédicas from UNAM, Dr Ricardo Grande, MSc. Verónica Jiménez and Dr Enrique Morett from the University Unit for Massive Sequencing, UNAM for sample sequencing and Dr David Romero who provided a research permit as Director of the Center for Genomic Sciences, UNAM. Director Rosendo Antonio Caro-Gomez provided a permit for samplings at the Monarch Butterfly Biosphere Reserve (MBBR) in Mexico. Felipe Martínez Meza, Katia Lemus Ramírez and Genaro Mondragón Contreras from the MBBR provided crucial support for samplings. The Ministry of Environment and Natural Resources of Mexico also provided a permit for samplings.

Declaration of interest

This research was supported by PAPIIT IN205412 from DGAPA, UNAM. L.E.S.G. received a PhD scholarship from CONACYT (Mexico). The authors report no conflict of interest. The authors alone are responsible for the content and writing of the paper.

References

Besemer J, Borodovsky M. (1999). Heuristic approach to deriving models for gene finding. Nucleic Acids Res 27:3911–20.

Brower AVZ, Boyce TM. (1991). Mitochondrial DNA variation in monarch butterflies. Evolution 45:1281–6.

Brower AVZ, Jeansonne MM. (2004). Geographical populations and ''subspecies'' of New World monarch butterflies (Nymphalidae) share a recent origin and are not phylogenetically distinct. Ann Entomol Soc Am 97:519–23.

Bybee SM, Taylor SD, Nelson CR, Whiting MF. (2004). A phylogeny of robber flies (Diptera: Asilidae) at the subfamilial level: Molecular evidence. Mol Phylogenet Evol 30:789–97.

Cameron SL, Whiting MF. (2008). The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. Gene 408:112–23.

Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. (2005). ACT: The Artemis comparison Tool. Bioinformatics 21:3422–3.

Li H, Ruan J, Durbin R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18: 1851–8.

Lyons JI, Pierce AA, Barribeau SM, Sternberg ED, Mongue AJ, De Roode JC. (2012). Lack of genetic differentiation between monarch butterflies with divergent migration destinations. Mol Ecol 21: 3433–44.

Skerratt LF, Campbell NJH, Murrell A, Walton S, Kemp D, Barker SC. (2002). The mitochondrial 12S gene is a suitable marker of populations of *Sarcoptes scabiei* from wombats, dogs and humans in Australia. Parasitol Res 88:376–9.

Zerbino DR, Birney E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–9.

Zhan S, Merlin C, Boore JL, Reppert SM. (2011). The monarch butterfly genome yields insights into long-distance migration. Cell 147:1171–85.

Zhan S, Reppert SM. (2013). MonarchBase: the monarch butterfly genome database. Nucleic Acids Res 41:D758–63.

Artículo:

Ramírez-Puebla ST, Servín-Garcidueñas LE, Ormeño-Orrillo E, Vera-Ponce de León A, Rosenblueth M, Delaye L, Martínez J, Martínez-Romero E. 2015. Species in *Wolbachia*? Proposal for the designation of '*Candidatus* Wolbachia bourtzisii', '*Candidatus* Wolbachia onchocercicola', '*Candidatus* Wolbachia blaxteri', '*Candidatus* Wolbachia brugii', '*Candidatus* Wolbachia taylori', '*Candidatus* Wolbachia collembolicola' and '*Candidatus* Wolbachia multihospitum' for the different species within *Wolbachia* supergroups. Syst. Appl. Microbiol. doi: 10.1016/j.syapm.2015.05.005.

En este trabajo se secuenciaron muestras de ADN metagenómico obtenido de tejidos de *Dactylopius coccus*. Se realizaron análisis bioinformáticos que lograron recuperar genomas parciales de dos bacterias del género *Wolbachia*. Análisis filogenómicos ubicaron a las wolbachias de *D. coccus* dentro de los súper grupos A y B. Mediante los análisis filogenómicos se propuso la reclasificación taxonómica del género *Wolbachia* y se propusieron especies nuevas. Los resultados completos se presentan en el artículo.

Systematic and Applied Microbiology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Species in *Wolbachia*? Proposal for the designation of '*Candidatus* Wolbachia bourtzisii', '*Candidatus* Wolbachia onchocercicola', '*Candidatus* Wolbachia blaxteri', '*Candidatus* Wolbachia brugii', '*Candidatus* Wolbachia taylori', '*Candidatus* Wolbachia collembolicola' and '*Candidatus* Wolbachia multihospitum' for the different species within *Wolbachia* supergroups

Shamayim T. Ramírez-Puebla[a], Luis E. Servín-Garcidueñas[a], Ernesto Ormeño-Orrillo[a], Arturo Vera-Ponce de León[a], Mónica Rosenblueth[a], Luis Delaye[b], Julio Martínez[a], Esperanza Martínez-Romero[a,*]

[a] *Centro de Ciencias Genómicas, UNAM, Cuernavaca, Morelos, Mexico*
[b] *Departamento de Ingeniería Genética, CINVESTAV-Irapuato, Irapuato, Guanajuato, Mexico*

ARTICLE INFO

Article history:
Received 22 October 2014
Received in revised form 21 May 2015
Accepted 27 May 2015

Keywords:
Wolbachia supergroups
Wolbachia Candidatus
Wolbachia taxonomy
ANI
DDH

ABSTRACT

Wolbachia are highly extended bacterial endosymbionts that infect arthropods and filarial nematodes and produce contrasting phenotypes on their hosts. *Wolbachia* taxonomy has been understudied. Currently, *Wolbachia* strains are classified into phylogenetic supergroups. Here we applied phylogenomic analyses to study *Wolbachia* evolutionary relationships and examined metrics derived from their genome sequences such as average nucleotide identity (ANI), *in silico* DNA–DNA hybridization (DDH), G + C content, and synteny to shed light on the taxonomy of these bacteria. Draft genome sequences of strains *w*DacA and *w*DacB obtained from the carmine cochineal insect *Dactylopius coccus* were included. Although all analyses indicated that each *Wolbachia* supergroup represents a distinct evolutionary lineage, we found that some of the analyzed supergroups showed enough internal heterogeneity to be considered as assemblages of more than one species. Thus, supergroups would represent supraspecific groupings. Consequently, *Wolbachia pipientis nomen* species would apply only to strains of supergroup B and we propose the designation of '*Candidatus* Wolbachia bourtzisii', '*Candidatus* Wolbachia onchocercicola', '*Candidatus* Wolbachia blaxterii', '*Candidatus* Wolbachia brugii', '*Candidatus* Wolbachia taylorii', '*Candidatus* Wolbachia collembolicola' and '*Candidatus* Wolbachia multihospitis' for other supergroups.

Introduction

Wolbachia is a genus of endosymbiotic bacteria that are wide spread in nature. *Wolbachia* endosymbionts do not have a free-living phase and are under confinement to particular hosts. It is estimated that *Wolbachia* may be found in 40% of arthropod species [106], while a previous report calculated 60% [44]. *Wolbachia* endosymbionts have been found associated with nematodes from the Onchocercidae family [22,54]. Interactions with their hosts range from parasitism to mutualism. In arthropods are mostly considered as parasites since they may manipulate host reproduction by mechanisms like parthenogenesis, feminization, male killing, and cytoplasmic incompatibility [12,80,100]. However, *Wolbachia* symbiosis has been implicated in host fitness [15,94], or as being necessary for oogenesis [25]; in nematodes they are regarded as mutualistic and essential for survival [58]. *Wolbachia* symbiosis is outstanding as it may cause host speciation events [11].

Some insects and their endosymbionts have a parallel evolutionary history, and cospeciation events have been described for both host and bacteria, especially primary endosymbionts [1,10,21,79,87]. For endosymbionts that have cospeciated with

* Corresponding author. Tel.: +52 777 3131697.
E-mail address: emartine@ccg.unam.mx (E. Martínez-Romero).

http://dx.doi.org/10.1016/j.syapm.2015.05.005
0723-2020/

their hosts, endosymbionts in different hosts would be distinct species. It seems that cospeciation is rare in *Wolbachia* and insects as their phylogenies are usually not congruent [1,49,86,88]. Thus, adaptations to different hosts would not necessarily mean bacterial speciation. *Wolbachia* infections in insects may be recent in some cases [41,48], implying a short host–symbiont interaction that would not lead to speciation. Recent *Wolbachia* acquisitions may come from horizontal transfers from close or even distant insects [41]. In filarial nematode–*Wolbachia* associations, congruence between *Wolbachia* phylogenies and those of their host has been documented [9,17,32]. In this case, cospeciation between bacteria and their worm hosts seems to have occurred and a single origin of this symbiosis for supergroups C and D has even been proposed [32]. *Wolbachia* have become essential for nematode development and play an important role in host embryogenesis [58]. Nematodes treated with antibiotics cannot reach adulthood [13,93].

Wolbachia pipientis Hertig 1936 [43], was first observed in cells of the *Culex pipiens* mosquito [42]. Heterogeneity within *Wolbachia* has been revealed by sequence analysis of 16S rRNA genes and protein-coding genes, resulting in its distribution into sixteen phylogenetic supergroups, ten of which are found in arthropods (A, B, E, H, I, K, M, N, P, Q), five in nematodes (C, D, J, L, O) and one comprising both arthropod and nematode endosymbionts (F) [4]. The strains of *Wolbachia* detected in Australian spiders [81], were designed as Supergroup G but it was later revealed that it has a *wsp* gene that is a recombinant between those of A and B supergroups rather than being a distinct new supergroup [8]. A phylogenetic tree based on a multilocus analysis has been recently published giving insight about the relationships between *Wolbachia* supergroups [37]. A consensus of whether supergroups represent lineages of *W. pipientis* or distinct species has not been reached. Sequence divergence between supergroups seemed to indicate that each represented a species [78], however, other studies have indicated that they do not represent isolated genetic entities [7,99], as would be expected from *bona fide* species [61]. *Wolbachia* have been described as highly recombinogenic bacteria [6,7,99]. Multiple infections with different *Wolbachia* are frequent in the same insect individual [98,104], affording the opportunity for recombination between different strains, including not closely related ones [6,104]. Nevertheless, a recent study found that recombination is far higher within supergroups than between them [30]. Recombination events between supergroups are limited to small DNA fragments.

Endosymbiont confinement in a host leads to an inevitable dependence on the host. This is evident upon inspection of endosymbiont genomes, which generally lack many functions required for independent living. Classic taxonomy relying on phenotypic characterization of pure cultures as well as establishing genomic relatedness by DNA–DNA hybridization (DDH) experiments could not be applied to non-cultivable endosymbionts like *Wolbachia*. In the genomic era, however, metrics based on genome sequences like ANI (average nucleotide identity) and *in silico* DDH can be used as replacements for wet lab DDH [3,38], thus allowing the use of similar taxonomic criteria for both cultivable and non-cultivable prokaryotes. Furthermore, it is increasingly acknowledged that phenotypes should not be given as much importance for species delineation as they currently are [20,70,96].

Here, we evaluated the diversity of *Wolbachia* by performing phylogenomic analyses and by analyzing genome-derived metrics like ANI, *in silico* DDH, G+C content and synteny in order to shed light on the taxonomy of these endosymbionts. Additionally, we increased the genomic database of *Wolbachia* by reporting sequences from two strains recently obtained from the carmine cochineal insect *Dactylopius coccus*.

Materials and methods

Genome sequences

Sequences of all reported *Wolbachia* genomes were retrieved from GenBank database, except those of strains *w*Di and *w*Ls, which were available at http://nematodes.org/genomes/index_filaria.html [22]. Genomes of *Wolbachia* strains *w*DacA (Bioproject PRJNA274701) and *w*DacB (Bioproject PRJNA274698) were sequenced by a metagenomic approach from dissected cochineal insects of *Dactylopius coccus*. Detailed functional analyses of those genomic sequences will be reported elsewhere (Ramírez-Puebla et al., in preparation). For G+C content determination, contigs of each genome were concatenated, the number of G plus C nucleotides counted and the sum divided by the genome length. Genome of strain *w*Men was obtained from the *Strepsiptera* Genome Project [68], and genomes of strains *w*Fol, *w*Oc and *w*Cte were only deposited like Sequence Read Archive (SRA) so they were not included in G+C determination because they were not completely sequenced [36].

Phylogenomic analyses

Predicted proteomes were obtained from annotated genomes deposited at GenBank if available. The RAST server was also used for annotating and comparing whole genome sequences [5]. The AMPHORA2 pipeline [103], was used to identify a set of 31 conserved bacterial proteins from complete or draft genomes. Sequencing reads were obtained from the Sequence Read Archive (SRA) database to obtain phylogenetic markers for strains *w*Men, *w*Fol, *w*Oc, and *w*Cte. Reads were mapped against individual marker genes obtained from fully sequenced *Wolbachia* genomes using the runMapping option from Newbler (Roche). The obtained mapped reads were processed to obtain the markers for these strains by performing tblastn searches against reference protein sequences corresponding to the markers from other sequenced strains. Protein sequences were concatenated using the EMBOSS union web tool (http://emboss.bioinformatics.nl/cgi-bin/emboss/union). The concatenated sequences were then aligned using MUSCLE v.3.8.31 [29], and the resulting alignment was processed with Gblocks [18], to obtain conserved protein blocks and eliminate poorly aligned positions and divergent regions. The edited alignment contained 9151 amino acid positions. A maximum-likelihood analysis was then performed using the JTT substitution model under PhyML 3.0 [40]. Branch support values are based on 100 bootstrap replicates. The genomes from *Ehrlichia canis* (GenBank CP000107) and *Anaplasma marginale* (GenBank CP001079) were used as outgroups.

In silico *DDH and ANI calculations*

DDH estimates were computed using the Genome-to-Genome Distance Calculator version 2.0 [65], as recommended by Auch et al. [2,3], and Meier-Kolthoff et al. [65]. BLAST+ was used for alignment and formula 2 for genome distance estimation. ANI values were calculated as previously proposed [38], using the ANI calculator from the Kostas lab (http://enve-omics.ce.gatech.edu/ani/) with default parameters.

Synteny

Syntenic blocks between ten finished *Wolbachia* genomes were identified by BLASTN. Only blocks at least 3000 bp in length and with 80% identity or higher were used to construct the graphs using the Artemis comparison tool [16].

Results and discussion

Genome-based relationships

Predicted evolutionary relationships between all 34 complete or almost complete *Wolbachia* genomes currently available and our two *Wolbachia* strains from *D. coccus*, *w*DacA and *w*DacB (Table 1) based on a set of conserved proteins is shown in Fig. 1. The distinctiveness of *Wolbachia* supergroups was evidenced by each forming a different and well-supported clade. The A and B supergroups clustered together in a branch separated from the C, D and F supergroups as previously observed by Nikoh et al. [69], for a set of six genomes using 52 ribosomal proteins. The phylogenetic reconstructions were also in agreement with a previous analysis obtained with a set of 90 orthologous genes from only eleven sequenced *Wolbachia* strains [22]. All the *Wolbachia* strains associated with *Culex* mosquitoes (*W. pipientis* representatives) tightly clustered in a single clade within supergroup B. The *Wolbachia* strain *w*Bol1-b from *Hypolimnas bolina* was also phylogenetically close to the *W. pipientis* strains. *Wolbachia* strains from *D. coccus* clustered into distinct supergroups. *w*DacA was resolved as the most phylogenetically distant strain within supergroup A, whereas *w*DacB was a member of supergroup B having *w*VitB from *Nasonia vitripennis* as its closest sequenced relative. The strain *w*Fol associated with the springtail *Folsomia candida* was found placed as the most distant of all the analyzed *Wolbachia* supergroups as reported previously [36].

G + C content

The G + C content of *Wolbachia* genomes ranged from 32.1% to 38.4% (Table 1) evidencing an enrichment of AT nucleotides as is common in endosymbionts [67]. Mean G + C contents of supergroups A, B, C and D were 35.5%, 34.0%, 32.4% and 33.4%, respectively. The sequenced representative of supergroup F had a G + C content of 36.3%. Although the analyzed genomes may not completely comprise the natural variation present in *Wolbachia* supergroups, it is worth noting that each supergroup seems to have a characteristic G + C content (Table 1).

In silico *DDH and ANI estimates*

DDH is the "gold standard" for species delineation in prokaryotes but it is not applicable for non-cultivable bacteria like *Wolbachia*. ANI represents a suitable surrogate for wet lab DDH as correlation analyses indicate that strains showing ANI higher or equal than 95–96% shared DDH values higher or equal than 70% and are thus considered to be of the same species [38]. Genome sequences also allow the estimation of *in silico* DDH values, and

Table 1
Characteristics of the sequenced *Wolbachia* genomes used in this work.

Wolbachia strain	Host species	GenBank accession number	Genome status	Number of contigs	Genome size (bp)	G + C%	Super group	Reference
*w*Mel	*Drosophila melanogaster*	AE017196	Complete	1	1,267,782	35.2	A	[102]
*w*MelPop	*Drosophila melanogaster*	AQQE00000000	WGS	80	1,239,155	35.1	A	[101]
*w*Ri	*Drosophila simulans* Riverside	CP001391	Complete	1	1,445,873	35.2	A	[52]
*w*Ha	*Drosophila simulans*	CP003884	Complete	1	1,295,804	35.1	A	[30]
*w*Sim	*Drosophila simulans*	AAGC00000000	WGS	629	1,063,100	35.4	A	[84]
*w*Au	*Drosophila simulans*	LK055284	Complete	1	1,268,461	35.2	A	[91]
*w*Rec	*Drosophila recens*	JQAM00000000	WGS	43	1,126,656	35.2	A	[66]
*w*Suzi	*Drosophila suzukii*	CAOU02000000	WGS	110	1,415,350	35.2	A	[89]
*w*Dwi	*Drosophila willistoni*	AAQP00000000	WGS	260	1,145,915	38.4	A	Remington et al. (unpublished)
*w*Ana	*Drosophila ananassae*	AAGB00000000	WGS	464	1,440,750	35.7	A	[84]
*w*Gmm	*Glossina morsitans morsitans*	AWUH00000000	WGS	241	1,019,510	35.2	A	[14]
*w*Uni	*Muscidifurax uniraptor*	ACFP00000000	WGS	256	867,873	35.2	A	[52]
*w*DacA	*Dactylopius coccus*	PRJNA274701	WGS	456	933,576	35.0	A	This study
*w*No	*Drosophila simulans*	CP003883	Complete	1	1,301,823	34.0	B	[30]
*w*Pip_Pel	*Culex quinquefasciatus*	AM999887	Complete	2	1,482,455	34.2	B	[51]
*w*Pip_JHB	*Culex quinquefasciatus*	ABZA00000000	WGS	21	1,543,661	34.2	B	[85]
*w*Pip_Mol	*Culex molestus*	HG428761	Complete	1	1,340,443	33.9	B	[74]
*w*Pip	*Culex pipiens molestus*	CACK00000000	WGS	888	1,479,531	34.3	B	Sinkins et al. (unpublished)
*w*Di	*Diaphorina citri*	AMZJ00000000	WGS	124	1,240,904	34.0	B	[83]
*w*Bol1-b	*Hypolimnas bolina*	CAOH00000000	WGS	144	1,377,933	33.9	B	[28]
*w*AlbB	*Aedes albopictus*	CAGB00000000	WGS	156	1,162,431	33.7	B	[63]
*w*DacB	*Dactylopius coccus*	PRJNA274698	WGS	321	1,282,277	34.2	B	This study
*w*VitB	*Nasonia vitripennis*	AERW00000000	WGS	523	1,107,643	33.9	B	[50]
*w*Cte	*Ctenocephalides felis*	SRR1222150	Raw data	–	–	–	B	[36]
*w*Oo	*Onchocerca ochengi*	HE660029	Complete	1	957,990	32.1	C	[24]
*w*Ov	*Onchocerca volvulus* strain Cameroon	HG810405	Complete	1	960,618	32.1	C	Cotton et al. (unpublished)
*w*Di	*Dirofilaria immitis*	PRJEB4154[a]	WGS	2	921,012	32.7	C	[22]
*w*Bm strain TRS	*Brugia malayi*	AE017321	Complete	1	1,080,084	34.2	D	[35]
*w*Bn	*Wuchereria bancrofti*	ADHD00000000	WGS	763	1,052,327	34.0	D	[26]
*w*Ls	*Litomosoides sigmodontis*	PRJEB4155[a]	WGS	10	1,048,936	32.1	D	[22]
*w*Fol	*Folsomia candida*	SRR1222159	Raw data	–	–	–	E	[36]
*w*Cle	*Cimex lectularius*	AP013028	Complete	1	1,250,060	36.3	F	[69]
*w*Oc	*Osmia caerulescens*	SRR1221705	Raw data	–	–	–	F	[36]
*w*Men	*Mengenilla moldrzyki*	SRX095325	WGS	–	–	–	F	[68]

[a] Accessions numbers correspond to the European Nucleotide Archive database as submitted by the original authors.

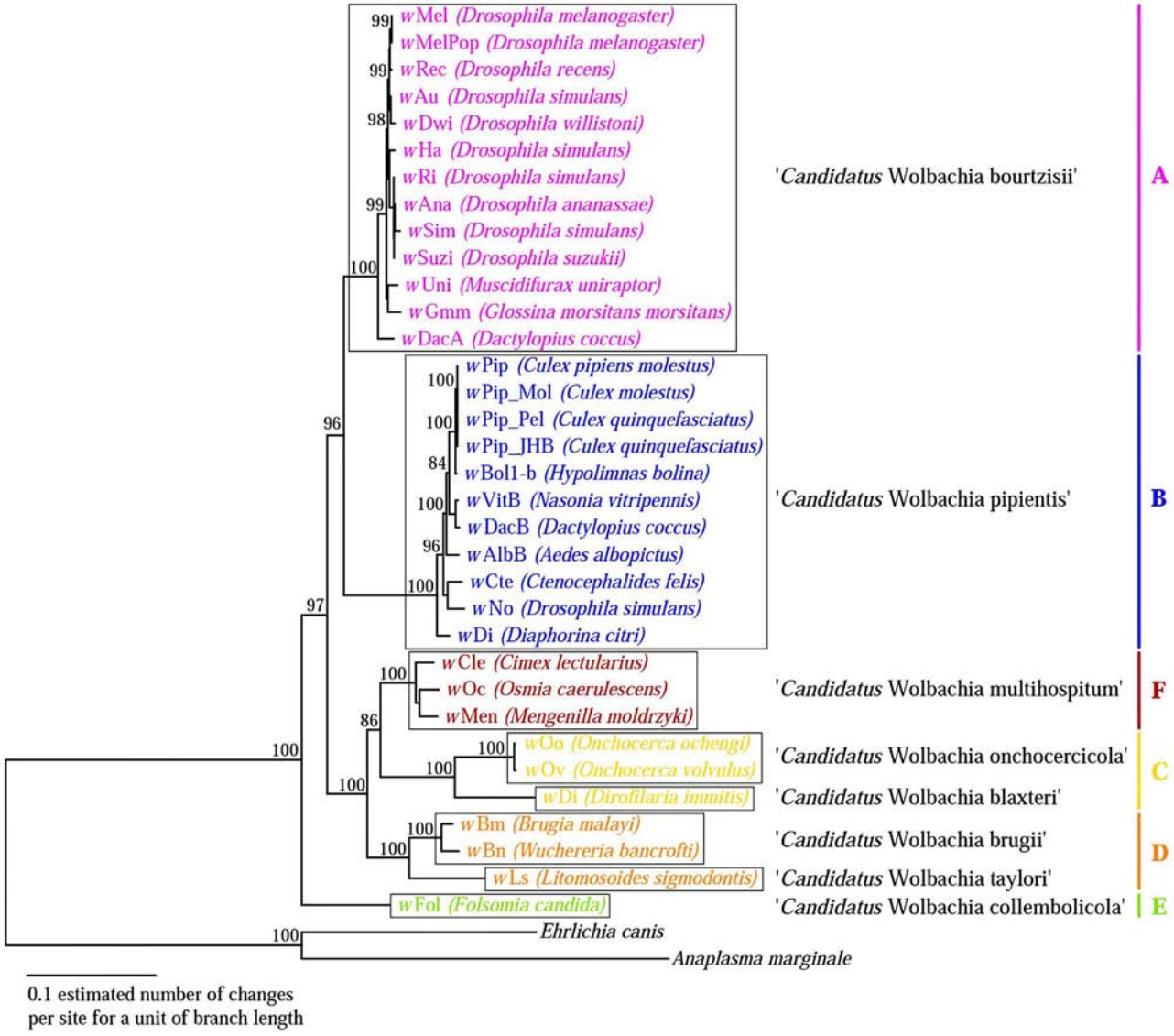


Fig. 1. Phylogenomic tree showing evolutionary relationships between *Wolbachia* strains inferred with PhyML based on a concatenated alignment of 31 marker proteins detected with AMPHORA2 and analyzed with the JTT substitution model. Hosts for each strain are indicated in forceps. Strains included in the new designations for each species name are boxed. Proposed species names are shown next to the boxes. Supergroups are shown to the right. Numbers at the branch points represent bootstrap support values based on 100 replicates. The scale bar represents the estimated number of amino acid changes per site for a unit of branch length.

these correlate very well with wet lab DDH [3]. Based in these criteria other bacteria as *Ensifer* and *Rhizobium* have been reclassified [71].

In silico DDH and ANI values were calculated for all pairs of analyzed *Wolbachia* genomes (Tables 2 and 3, respectively). Strains from different supergroups showed maximum ANI and *in silico* DDH values of 86.8% and 34.6%, respectively indicating that *Wolbachia* is comprised of several different species as previously suggested based on fewer ANI comparisons [75]. Within each supergroup, most members showed ANI values over 96% (Table 3) and *in silico* DDH values over 70% (Table 2), relatedness levels that are consistent with single species. However, in some supergroups there were strains with enough differences to put them below or near the borderline for species delineation. In supergroup A *Dactylopius* strain *w*DacA, and in supergroup B *Drosophila simulans* strain *w*No and *Diaphorina citri* strain *w*Di, showed *in silico* DDH values <62%, below the species circumscription level with all members of their own supergroups and ANI values just above of the species cut-off level (Table 2). In the phylogenomic analyses, strains *w*DacA, *w*No, *w*Di occupied peripheral positions within their supergroups (Fig. 1). *w*AlbB from *Aedes albopictus* and *w*DacB from *D. coccus* in supergroup B, also showed *in silico* DDH values below or very close to 70% with their supergroup neighbors, although these differences were not reflected by low ANI values (Table 2).

Within supergroup B, *in silico* DDH and ANI values were high among representative sequenced strains of *Wolbachia pipientis* (*w*Pip strains). Values were also high between *w*Bol1-b and *W. pipientis*. In contrast, comparisons between *w*Pip strains and other members of supergroup B produced values that were just over or slightly under the cutoff limits recommended for species delineation. The genome from strain *w*VitB had ANI value of around 97.4% when compared with the *w*Pip strain (Table 3) but most of their DDH estimates were below 70% (Table 2).

Clear examples of the existence of different species within supergroups were observed for nematode strains *w*Di of

Table 2
In silico DNA–DNA hybridization (DDH).

		in silico DNA-DNA[a] hybridization																														
Strain	Supergroup	*w* Mel	*w* MelPop	*w* Rec	*w* Au	*w* Dwi	*w* Uni	*w* Gmm	*w* Ana	*w* Ri	*w* Suzi	*w* Sim	*w* Ha	*w* DacA	*w* Pip	*w* Pip_Mol	*w* Pip_Pel	*w* Pip_JHB	*w* Bol1•b	*w* VitB	*w* DacB	*w* Di	*w* AlbB	*w* No	*w* Oo	*w* Ov	*w* Di	*w* Bm	*w* Bn	*w* Ls	*w* Cle
w Mel	A	100.0	77.7	89.3	93.9	70.5	75.8	71.2	72.8	71.9	72.1	75.1	73.3	57.0																		
w MelPop	A		100.0	95.3	93.6	92.6	76.2	71.7	72.8	72.1	72.3	74.9	73.2	57.6																		
w Rec	A			100.0	90.6	94.0	75.4	72.0	73.0	70.9	71.2	74.3	75.2	57.6																		
w Au	A				100.0	64.7	75.6	70.8	72.4	70.8	71.2	74.3	73.4	57.0																		
w Dwi	A					100.0	75.4	70.5	73.4	72.0	71.9	74.4	72.6	57.0																		
w Uni	A						100.0	74.5	73.5	73.7	73.8	73.3	73.6	58.6																		
w Gmm	A							100.0	67.1	65.1	66.1	67.9	69.5	56.4																		
w Ana	A								100.0	98.0	97.9	95.6	72.0	57.1																		
w Ri	A									100.0	98.6	85.2	69.3	55.9																		
w Suzi	A										100.0	90.7	69.5	55.9																		
w Sim	A											100.0	72.1	56.9																		
w Ha	A												100.0	56.2																		
w DacA	A													100.0																		
w Pip	B														100.0	100.0	96.7	96.4	83.1	71.5	69.5	58.8	63.7	57.3								
w Pip_Mol	B															100.0	84.4	67.6	81.5	70.8	68.5	58.2	63.0	56.6								
w Pip_Pel	B																100.0	86.3	78.0	68.6	67.5	56.7	61.9	55.2								
w Pip_JHB	B																	100.0	77.2	68.8	67.5	56.3	61.3	55.0								
w Bol1•b	B																		100.0	69.8	68.0	58.1	61.9	57.6								
w VitB	B																			100.0	83.4	61.5	66.0	61.1								
w DacB	B																				100.0	59.6	62.6	58.6								
w Di	B																					100.0	59.0	54.5								
w AlbB	B																						100.0	58.9								
w No	B																							100.0								
w Oo	C																								100.0	96.1	29.2					
w Ov	C																									100.0	29.1					
w Di	C																										100.0					
w Bm	D																											100.0	69.8	36.5		
w Bn	D																												100.0	38.7		
w Ls	D																													100.0		
w Cle	F																														100.0	

[a] Values <70% are shaded.

Table 3
Average nucleotide identity (ANI) percentage values between Wolbachia strains.

		ANI[a] percentages for strain																														
Strain	Supergroup	*w* Mel	*w* MelPop	*w* Rec	*w* Au	*w* Dwi	*w* Uni	*w* Gmm	*w* Ana	*w* Ri	*w* Suzi	*w* Sim	*w* Ha	*w* DacA	*w* Pip	*w* Pip_Mol	*w* Pip_Pel	*w* Pip_JHB	*w* Bol1•b	*w* VitB	*w* DacB	*w* Di	*w* AlbB	*w* No	*w* Oo	*w* Ov	*w* Di	*w* Bm	*w* Bn	*w* Ls	*w* Cle
w Mel	A	100.0	100.0	99.6	99.6	99.4	97.9	97.3	97.4	97.8	97.7	97.9	97.7	95.0																		
w MelPop	A		100.0	99.6	99.6	99.4	97.9	97.4	97.5	97.8	97.7	97.8	97.6	94.9																		
w Rec	A			100.0	99.6	99.5	97.9	97.4	97.5	97.6	97.6	97.6	97.8	94.9																		
w Au	A				100.0	99.9	98.0	97.4	97.4	97.6	97.6	97.7	97.6	94.9																		
w Dwi	A					100.0	97.8	97.2	97.2	97.5	97.4	97.7	97.5	94.5																		
w Uni	A						100.0	97.6	97.7	97.6	97.6	97.6	97.5	95.0																		
w Gmm	A							100.0	96.9	96.9	96.9	96.9	97.0	94.6																		
w Ana	A								100.0	99.9	99.9	99.8	97.7	94.7																		
w Ri	A									100.0	100.0	99.9	97.7	94.7																		
w Suzi	A										100.0	99.9	97.7	94.7																		
w Sim	A											100.0	97.5	94.8																		
w Ha	A												100.0	94.8																		
w DacA	A													100.0																		
w Pip	B														100.0	100.0	99.9	99.9	99.1	97.4	97.3	95.8	96.3	95.3								
w Pip_Mol	B															100.0	99.8	99.8	99.0	97.2	97.2	95.7	96.1	95.2								
w Pip_Pel	B																100.0	99.9	98.9	97.2	97.1	95.6	96.1	95.1								
w Pip_JHB	B																	100.0	98.9	97.1	97.2	95.6	96.1	95.2								
w Bol1•b	B																		100.0	97.2	97.1	95.7	96.2	95.3								
w VitB	B																			100.0	98.8	95.7	96.2	95.6								
w DacB	B																				100.0	95.5	96.1	95.4								
w Di	B																					100.0	95.4	94.8								
w AlbB	B																						100.0	95.3								
w No	B																							100.0								
w Oo	C																								100.0	99.4	86.1					
w Ov	C																									100.0	86.1					
w Di	C																										100.0					
w Bm	D																											100.0	96.3	88.8		
w Bn	D																												100.0	88.8		
w Ls	D																													100.0		
w Cle	F																														100.0	

[a] Values <96% are shaded.

Dirofilaria immitis and *w*Ls in supergroups C and D, respectively (Tables 2 and 3). In these cases, both strains occupied peripheral positions within their supergroups (Fig. 1). Also, *in silico* DDH and ANI values supported the differentiation of these strains from other members of their respective supergroups supporting that they belong to different species.

Cospeciation between nematodes and *Wolbachia* is reflected in the fact that more distantly related nematodes [32], also have more distantly related *Wolbachia*. In supergroup D, *Wolbachia* from related nematode genera *Brugia* and *Wuchereria* are more closely related to each other than to *Wolbachia* from the more distant genus *Litomosoides* (Fig. 1, Tables 2 and 3). It is worth noting that *Wolbachia* infecting *Brugia* and *Wuchereria* are at the borderline for species definition by *in silico* DDH. Similarly, in supergroup C, *Wolbachia* from two species of *Onchocerca* are related but showed very low *in silico* DDH and ANI values to the endosymbiont of *D. immitis* (Fig. 1, Tables 2 and 3). Taken together, these data seem to indicate that each nematode genus harbors a different *Wolbachia* species. Nevertheless, the presence of supergroup F *Wolbachia* in both nematodes and insects, and its loss in some filarial species in

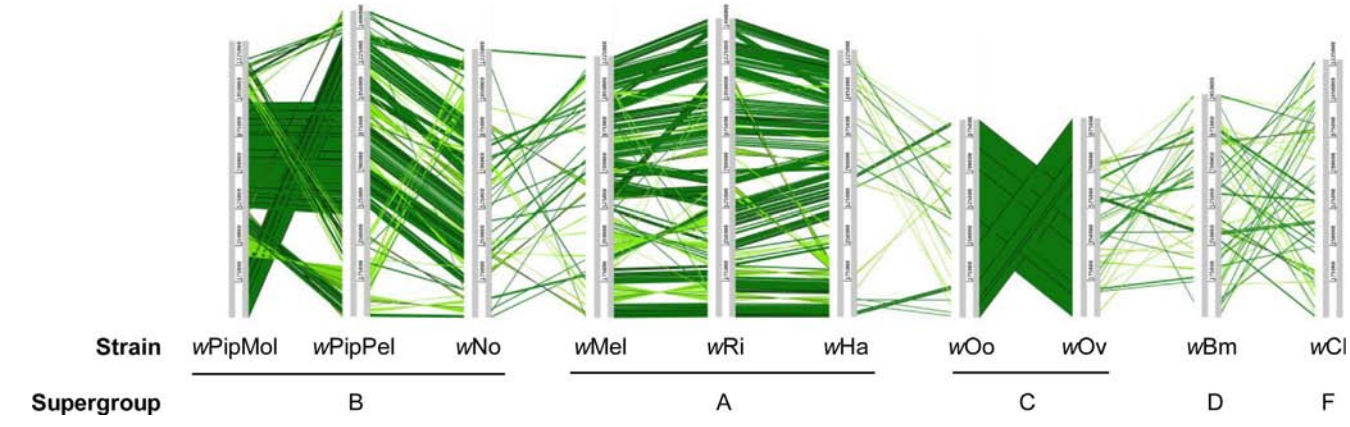


Fig. 2. Synteny between complete genomes of *Wolbachia* strains. Sequence blocks oriented in the same or inverted directions are shown in dark or light green, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

the onchocercids, contravene the idea of cospeciation in general [33].

Genome synteny

Structural genome comparisons in other bacteria like rhizobia have shown that chromosome synteny is very well conserved within a species and less maintained between different species [39,77]. Synteny was used as a further criterion to distinguish *Wolbachia* species. It has been observed that levels of genome synteny are higher within than between *Wolbachia* supergroups [30], as it is evidenced in Fig. 2 for ten finished genomes. *Wolbachia* genomes have high levels of repetitive DNA and mobile genetic elements that lead to DNA rearrangements that diminish synteny even between related strains [30,35,52,102]. Genome rearrangements in other organisms represent recombination barriers and could lead to genetic isolation [76]. Strain *w*No showing significant divergence by *in silico* DDH and ANI values is less syntenic with its supergroup siblings. It would be worth investigating if a speciation process could start within a supergroup by *Wolbachia* strains developing novel genomic rearrangements, as discussed previously [30].

Conclusions

We showed here that *Wolbachia* supergroups represent distinct evolutionary lineages based on phylogenomics, G + C content, ANI, *in silico* DDH and synteny. Our results support the previous proposal that *Wolbachia* from different supergroups should be considered as genetically distinct clades not only from implications related to host confinement and their biology [72], but on the basis of molecular evidence [30,53,78]. Furthermore, we found heterogeneity within supergroups. The more divergent strains within each supergroup were recovered as outliers in the phylogenomic analyses. Not all strains, however, seem to have accumulated enough nucleotide sequence differences to show ANI values lower than 95–96%, used to delineate different species, with other distant strains in their supergroups (Table 3). Nevertheless, within a supergroup significant genome content differences were evidenced by low *in silico* DDH (less than 70%) and genome synteny among supergroup members was not always high. Thus, our novel analyses indicate that different species may occur inside a supergroup. Consequently, supergroups would have a supraspecific status.

Given the evidence reviewed and presented here, the name *W. pipientis* (Hertig 1936) [43], should be applied only to supergroup B strains. As *Wolbachia* are still uncultivable, the proper designation for supergroup B strains should be '*Candidatus* Wolbachia pipientis'. In order to distinguish the different species within *Wolbachia*, we propose the designation '*Candidatus* Wolbachia bourtzisii' for strains in supergroup A, '*Candidatus* Wolbachia onchocercicola' for endosymbionts of genus *Onchocerca* in supergroup C, '*Candidatus* Wolbachia blaxterii' for endosymbionts of *D. immitis* in supergroup C, '*Candidatus* Wolbachia brugii' for endosymbionts of nematodes from *Brugia* and *Wuchereria* species in supergroup D, '*Candidatus* Wolbachia taylorii' for endosymbionts of nematodes from *Litomosoides* species in supergroup D, '*Candidatus* Wolbachia collembola' for endosymbionts of Collembola arthropods in supergroup E and '*Candidatus* Wolbachia multihospitis' for *Wolbachia* strains hosted by nematodes and arthropods in supergroup F.

Description of '*Candidatus* Wolbachia bourtzisii'

'*Candidatus* Wolbachia bourtzisii' (bourt.zi'si.i. N.L. gen. n. *bourtzisii*, of Bourtzis, in honor of Kostas Bourtzis, as a recognition for his studies on *Wolbachia* and other bacteria associated with arthropods).

The description of the species '*Candidatus* Wolbachia bourtzisii' is based on the studies reported by Louis and Nigro [62], Sacchi et al. [82], Texeira et al. [94], and Zhukova and Kiseleva [105]. Cell size is 0.5 μm in *D. simulans*, and 0.5–1.0 μm in *D. melanogaster*. Cells are roundish and less frequently rod shaped and are surrounded by three enveloping membranes. The first is the plasmatic membrane and the second represents the outer part of the cell wall. The third one, closely related to the cytoplasm of the host cell, forms a vacuole for each single microorganism. Ribosomes and nucleic acid fibrils are observed in the cytoplasm. In *D. melanogaster* individual bacterial cells are distributed throughout the host cell cytoplasm, occasionally occurring as small groups. Bacteria occur in the ovarioles in high numbers and in germline cells like cytocytes, oogonia, oocytes and nurse cells.

The percentage of apoptotic cells in germaria are increased in *D. melanogaster* infected with *w*MelPop. Tetracycline treatments accelerated the time to death in *D. melanogaster* infected with *Drosophila* C virus (DCV) as the bacteria confer resistance to DCV by interfering with virus proliferation. The DNA G + C content is between 35.0 and 38.4 mol% as calculated from genomic sequences. Most strains exhibit a DNA G + C content of 35.2 mol%.

Description of '*Candidatus* Wolbachia onchocercicola'

'*Candidatus* Wolbachia onchocercicola' [on.cho.cer.ci'co.la. N.L. fem. n. Onchocerca a filarial nematode genus; L. suffix – *cola* (from L. masc. or fem. n. *incola*), a dweller, an inhabitant; N.L. fem. n. *onchocercicola*, a dweller of *Onchocerca*].

The description of the species '*Candidatus* Wolbachia onchocercicola' is based on the studies reported by Determan et al. [27], Egyed et al. [31], Horeauf et al. [46], Kozek and Marroquin [55], and Langworthy et al. [59]. Cell size is 0.3 up to 0.8 μm in diameter and 1.5 up to 1.8 μm in length. Cells are generally round or spherical shaped. Bacteria are located in the cytoplasm surrounded by a membrane-bound vacuole. In *Onchocerca lupi* each vacuole contains only one bacterium surrounded by a double membrane. In contrast, in *Onchocerca volvulus* they often form clusters and in *Onchocerca ochengi* some of them contain up to seven bacteria. Wolbachia live in the subcutaneous and connective tissues of their hosts, usually enclosed in fibrous cysts or nodules. In adults and larvae bacterial cells occur in the lateral cords, and in germinal tissues in females. Depletion of the endosymbiont by oxytetracycline in *O. ochengi* results in the death of adults and microfilaria. Also, there is a decline in the quantity of embryos and an increase in the proportion of embryos showing abnormal morphology. In *O. volvulus* doxycycline treatment blocks embryogenesis. The DNA G + C content is 32.1 mol% as calculated from genomic sequences.

Description of '*Candidatus* Wolbachia blaxteri'

'*Candidatus* Wolbachia blaxteri' (blax'ter.i. N.L. gen. n. *blaxteri*, of Blaxter, in honor of Mark Blaxter, in recognition of his molecular studies on nematodes and their associated *Wolbachia* symbionts).

The description of the species '*Candidatus* Wolbachia baxteri' is based on the studies reported by Kozek [55,56], McLaren et al., [64], and Sironi et al. [90]. Cell size is 0.3–1.0 μm in diameter and 4.5 μm in length. Cells are spherical or ovoid shaped. Bacteria are contained in an individual membrane-bounded host vacuole. Some bacterial cells have condensations of dense material within their cytoplasm. In *D. immitis* bacteria occur in the reproductive tract mainly in the ovary and the proximal region of the uterus, and are also found in oocytes and in all embryonic stages of microfilariae developing in the uterus. In lateral cords of adults, they occur as clusters that can

fill most of the hypodermal tissue. Often they appear to surround the hypodermal nuclei. In embryos, five to ten bacteria per host cell are found. Also, bacteria are abundant in oogonia, eggs and early dividing embryos. Treatment with tetracycline blocks embryo development. The DNA G + C content is 32.7 mol% as calculated from genomic sequences.

Description of '*Candidatus* Wolbachia brugii'

'*Candidatus* Wolbachia brugii' (bru'gi.i. N.L. gen. n. *brugii*, of Brug, named after S. L. Brug, a Dutch parasitologist who first described the filarial nematode *Brugia malayi*, a model for the study of *Wolbachia*-nematode relationships).

The description of the species '*Candidatus* Wolbachia brugii' is based on the studies reported by Fischer et al. [34], and Landmann et al. [57,58], and Taylor et al. [92]. Cell size is 0.5 μm up to 1 μm. Cells are spherical or have an elongated shape and are surrounded with a double membrane. Bacteria are contained within membrane-bound vacuoles. In *Brugia malayi* clusters of bacteria are detected in microfilaria. In larvae L2, bacterial cells are detected in the hypodermis and in L3 and L4 larvae in the cells of lateral chord, in high numbers. In adult female worms, bacteria are commonly found in the lateral hypodermal cords, in hypodermis, and close to or inside the ovaries. Bacteria are also seen in the cells surrounding the basal lamina of the oviduct. In adult male worms, microfilariae, and third-stage larvae bacteria are detected in the lateral cord, but in lower numbers compared with females and dispersed in focal groups or as individual bacteria. They are also detected in testis and the border of vas deferens. In *Wuchereria bancrofti* bacteria show a similar distribution as in *B. malayi*, in small clusters or as a single bacterium.

Tetracycline treatments dramatically reduce the endosymbiont population in female adults of *B. malayi*. Pyknotic nuclei are observed throughout the ovaries and uteri in the female germline. Microfilaria resulting from a completed embryogenesis after antibiotic treatments, showed defects as abnormal muscle quadrants. Apoptotic nuclei are detected in the ovaries of treated females and become more numerous as the uteri is filled with embryos. The DNA G + C content is between 34.0 and 34.2 mol% as calculated from genomic sequences.

Description of '*Candidatus* Wolbachia taylori'

'*Candidatus* Wolbachia taylori' (tay'lo.ri. N.L. gen. n. *taylori*, of Taylor, in honor of Mark J. Taylor, in recognition of his studies on the role of *Wolbachia*–nematode symbiosis in human diseases and his search for treatments).

The description of the species '*Candidatus* Wolbachia taylori' is based on the studies reported by Chagas-Moutinho et al. [19], and Horeauf et al. [45]. Cell size is approximately1 μm and round shaped. Cells present a reduced cell wall and not a typical septum during cell division. Cells are surrounded by a host-derived vacuolar membrane. In *Litomosoides chagasfilhoi*, bacterial cells occur in regions of the hypoderm, in the oocytes, early-stage embryos and complete developed intrauterine microfilariae close to the cell host nucleus. In other filarial tissues, bacteria are found in intracellular vacuoles associated to the nuclear envelope. They are also observed in proximity to the endoplasmic reticulum. TEM suggested a single bacterium per vacuole.

Depletion by tetracycline results in infertility by blocking female worm development and early embryogenesis in *Litomosoides sigmodontis*. The DNA G + C content is 32.1 mol% as calculated from genomic sequences.

Description of '*Candidatus* Wolbachia collembolicola'

'*Candidatus* Wolbachia collembolicola' [col.lem.bo.li'co.la. N.L. n. pl. *Collembola* a lineage of hexapods; L. suffix – *cola* (from L. masc. or fem. n. *incola*), a dweller, an inhabitant; N.L. fem. n. *collembolicola*, a dweller of *Collembola*].

The description of the species '*Candidatus* Wolbachia collembolicola' is based on the studies reported by Czarnetzki and Tebbe [23], Pike and Kingcombe [73], Timmermans and Ellers [95], and Vandekerckhove et al. [97]. Cells detected in hexapod species of the order Collembola. Cell size is 0.2 μm up to 1.4 μm. Cells are pleomorphic from curved to almost hairpin-shaped. Cell wall lacks detectable peptidoglycan layer. Periplasmic space is of around 5–15 nm. Cells are surrounded by a host-derived vacuolar membrane. DNA filaments are visible in a rather diffuse network dispersed throughout the cell and interspersed with ribosomes. Cells occur in aggregations and are found mostly in close association with the rough endoplasmic reticulum in the ovaries. Fat bodies and interstitial cells as detected by TEM techniques or restricted to the ovary and brain as detected by FISH techniques.

Infection is obligatory for host offspring survival. The endosymbiont is sensitive to high-dose of rifampicin and heat treatments. High-dose tetracycline treatment is inefficient for removing cell infections. Bacteria obligate role early in the parthenogenetic developmental process includes egg hatching.

Description of '*Candidatus* Wolbachia multihospitum'

'*Candidatus* Wolbachia multihospitis' (mul.ti.hos'pi.tum. L. adj. *multus* many, numerous; L. n. *hospes -itis*, he who entertains a stranger, a host; N.L. gen. pl. n. multihospitum of numerous hosts, referring to the occurrence of the bacterium on various species of arthropods and nematodes).

The description of '*Candidatus* Wolbachia multihospitis' is based on the studies reported by Ferri et al. [33], Hosokawa et al. [47], and Lefoulon et al. [60]. In *Cimex lectularius* cell size is 0.5 up to 1.2 μm. Cells are rod-shaped. In males, bacterial cells are located in the testis-associated bacteriome, whereas in females they are located in bacteriomes and ovaries. Cells are also detected in the nutritive cord and developing oocytes. In the nematode *Madathamugadia hiepei*, they are detected in young and late embryos. In adult females they are observed in the ovaries and the intestinal wall. In contrast with other nematodes they are absent in the hypodermal lateral chord. In *Cercopithifilaria japonica* and *Mansonella perforata* bacteria are located in the epithelial somatic gonad and in the intestinal wall.

Elimination of the endosymbiont by rifampicin treatments in *C. lectularius* resulted in deformed developing eggs, reduction in the adult emergence rate and prolonged nymphal period. The DNA G + C content is 36.3 mol% as calculated from genomic sequences.

Acknowledgements

This work was supported by CONACyT basic science grant 154453. Results presented in this paper are part of the PhD thesis by S. T. Ramírez-Puebla, A. Vera-Ponce de León and L. E. Servín-Garcidueñas. They are students from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and each was recipient of a scholarship from Consejo Nacional de Ciencia y Tecnología (CONACYT), México. We thank Mariana Mateos for discussion and Michael Dunn for critical reading of this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.syapm.2015.05.005

References

[1] Ahmed, M.Z., De Barro, P.J., Ren, S.X., Greeff, J.M., Qiu, B.L. (2013) Evidence for horizontal transmission of secondary endosymbionts in the *Bemisia tabaci* cryptic species complex. PLoS ONE 8, e53084.

[2] Auch, A.F., Klenk, H.P., Göker, M. (2010) Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand. Genomic Sci. 2, 142–148.

[3] Auch, A.F., von Jan, M., Klenk, H.P., Göker, M. (2010) Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand. Genomic Sci. 2, 117–134.

[4] Augustinos, A.A., Santos-García, D., Dionyssopoulou, E., Moreira, M., Papapanagiotou, A., Scarvelakis, M., Doudoumis, V., Ramos, S., Aguiar, A.F., Borges, P.A.V., Khadem, M., Latorre, A., Tsiamis, G., Bourtzis, K. (2011) Detection and characterization of *Wolbachia* infections in natural populations of aphids: is the hidden diversity fully unraveled? PLoS ONE 6, e28695.

[5] Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O. (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9, 75.

[6] Baldo, L., Lo, N., Werren, J.H. (2005) Mosaic nature of the *Wolbachia* surface protein. J. Bacteriol. 187, 5406–5418.

[7] Baldo, L., Bordenstein, S., Wernegreen, J.J., Werren, J.H. (2006) Widespread recombination throughout *Wolbachia* genomes. Mol. Biol. Evol. 23, 437–449.

[8] Baldo, L., Werren, J.L. (2007) Revisiting *Wolbachia* supergroup typing based on WSP: spurious lineages and discordance with MLST. Curr. Microbiol. 55, 81–87.

[9] Bandi, C., Anderson, T.J., Genchi, C., Blaxter, M.L. (1998) Phylogeny of *Wolbachia* in filarial nematodes. Proc. Biol. Sci. 265, 2407–2413.

[10] Baumann, P., Moran, N.A., Baumann, L. (2000) Bacteriocyte associated endosymbionts of insects. In: Dworkin, M. (Ed.), The prokaryotes, Springer, New York, pp. 155–189.

[11] Bordenstein, S.R. (2003) Symbiosis and the origin of species. In: Bourtzis, K., Miller, T.A. (Eds.), Insect Symbiosis, CRC Press, Boca Raton, pp. 283–304.

[12] Bordenstein, S.R., Paraskevopoulos, C., Dunning-Hotopp, J.C., Sapountzis, P., Lo, N., Bandi, C., Tettelin, H., Werren, J.H., Bourtzis, K. (2009) Parasitism and mutualism in *Wolbachia*: what the phylogenomic trees can and cannot say. Mol. Bio. Evol. 26, 231–241.

[13] Bosshardt, S.C., McCall, J.W., Coleman, S.U., Jones, K.L., Petit, T.A., Klei, T.R. (1993) Prophylactic activity of tetracycline against *Brugia pahangi* infection in jirds (*Meriones unguiculatus*). J. Parasitol. 79, 775–777.

[14] Brelsfoard., C., Tsiamis, G., Falchetto, M., Gomulski, L.M., Telleria, E., Alam, U., Doudoumis, V., Scolari, F., Benoit, J.B., Swain, M., Takac, P., Malacrida, A.R., Bourtzis, K., Aksoy, S. (2014) Presence of extensive *Wolbachia* symbiont insertions discovered in the genome of its host *Glossina morsitans morsitans*. PLoS Negl. Trop. Dis. 8, e2728.

[15] Brownlie, J.C., Cass, B.N., Riegler, M., Witsenburg, J.J., Iturbe-Ormaetxe, I., McGraw, E.A., O'Neill, S.L. (2009) Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipientis*, during periods of nutritional stress. PLoS Pathog. 5, e1000368.

[16] Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., Parkhill, J. (2005) ACT: the Artemis Comparison Tool. Bioinformatics 21, 3422–3423.

[17] Casiraghi, M., Bain, O., Guerrero, R., Martin, C., Pocacqua, V., Gardner, S.L., Franceschi, A., Bandi, C. (2004) Mapping the presence of *Wolbachia pipientis* on the phylogeny of filarial nematodes: evidence for symbiont loss during evolution. Int. J. Parasitol. 34, 191–203.

[18] Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

[19] Chagas-Moutinho, V.A., Silva, R., de Souza, W., Motta, M.C. (2015) Identification and ultrastructural characterization of the *Wolbachia* symbiont in *Litomosoides chagasfilhoi*. Parasit. Vector 8, 74.

[20] Chan, J.Z.-M., Halachev, M., Loman, N., Constantinidou, C., Pallen, M. (2012) Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*. BMC Microbiol. 12, 302.

[21] Chen, X., Li, S., Aksoy, S. (1999) Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. J. Mol. Evol. 48, 49–58.

[22] Comandatore, F., Sassera, D., Montagna, M., Kumar, S., Koutsovoulos, G., Thomas, G., Repton, C., Babayan, S.A., Gray, N., Cordaux, R., Darby, A., Makepeace, B., Blaxter, M. (2013) Phylogenomics and analysis of shared genes suggest a single transition to mutualism in *Wolbachia* of nematodes. Genome Biol. Evol. 5, 1668–1674.

[23] Czarnetzki, A.B., Tebbe, C.C. (2004) Detection and phylogenetic analysis of *Wolbachia* in *Collembola*. Environ. Microbiol. 6, 35–44.

[24] Darby, A.C., Armstrong, S.D., Bah, G.S., Kaur, G., Hughes, M.A., Kay, S.M. (2012) Analysis of gene expression from the *Wolbachia* genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. Genome Res. 22, 2467–2477.

[25] Dedeine, F., Vavre, F., Fleury, F., Loppin, B., Hochberg, M.E., Boulétreau, M. (2001) Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. Proc. Natl. Acad. Sci. 98, 6247–6252.

[26] Desjardins, C.A., Cerqueira, G.C., Goldberg, J.M., Dunning-Hotopp, J.C., Haas, B.J., Zucker, J., Ribeiro, J.M.C., Saif, S., Levin, J.Z., Fan, L., Zeng, Q., Russ, C., Wortman, J.R., Fink, D.L., Birren, B.W., Nutman, T.B. (2013) Genomics of *Loa loa*, a *Wolbachia*-free filarial parasite of humans. Nat. Genet. 45, 495–500.

[27] Determann, A., Mehlhorn, H., Ghaffar, F. (1997) Electron microscope observations on *Onchocerca ochengi* and *O. fasciata* (Nematoda: Filarioidea). Parasitol. Res. 83, 591–603.

[28] Duplouy, A., Iturbe-Ormaetxe, I., Beatson, S.A., Szubert, J.M., Brownlie, J.C., McMeniman, C.J., McGraw, E.A., Hurst, G.D.D., Charlat, S., O'Neill, S.L., Woolfit, M. (2013) Draft genome sequence of the male-killing *Wolbachia* strain *w*Bol1 reveals recent horizontal gene transfers from diverse sources. BMC Genomic 14, 20.

[29] Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

[30] Ellegaard, K.M., Klasson, L., Näslund, K., Bourtzis, K., Andersson, S.G. (2013) Comparative genomics of *Wolbachia* and the bacterial species concept. PLoS Genet. 9, e1003381.

[31] Egyed, Z., Sréter, T., Széll, Z., Nyirő, G., Dobos-Kovács, M., Márialigeti, K., Varga, I. (2002) Electron microscopic and molecular identification of *Wolbachia* endosymbionts from *Onchocerca lupi*: implications for therapy. Vet. Parasitol. 106, 75–82.

[32] Fenn, K., Blaxter, M. (2004) Are filarial nematode *Wolbachia* obligate mutualist symbionts? Trends Ecol. Evol. 19, 163–166.

[33] Ferri, E., Bain, O., Barbuto, M., Martin, C., Lo, N., Uni, S., Landmann, F., Baccei, S.G., Guerrero, R., de Souza Lima, S., Bandi, C., Wanji, S., Diagne, M., Casiraghi, M. (2011) New insights into the evolution of *Wolbachia* infections in filarial nematodes inferred from a large range of screened species. PLoS ONE 6, e20843.

[34] Fischer, K., Beatty, W.L., Jiang, D., Weil, G.J., Fischer, P.U. (2011) Tissue and stage-specific distribution of *Wolbachia* in *Brugia malayi*. PLoS Negl. Trop. Dis. 5, e1174.

[35] Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., Bhattacharyya, A., Kapatral, V., Kumar, S., Posfai, J., Vincze, T., Ingram, J., Moran, L., Lapidus, A., Omelchenko, M., Kyrpides, N., Ghedin, E., Wang, S., Goltsman, E., Joukoy, V., Ostrovskaya, O., Tsukerman, K., Mazur, M., Comb, D., Koonin, E., Slatko, B. (2005) The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. PLoS Biol. 3, e121.

[36] Gerth, M., Gansauge, M-T., Weigert, A., Bleidron, C. (2014) Phylogenomic analyses uncover origin and spread of the *Wolbachia* pandemic. Nat. Commun. 5, 5117.

[37] Glowska, E., Dragun-Damian, A., Dabert, M., Gerth, M. (2015) New *Wolbachia* supergroups detected in quill mites (Acari: Syringophilidae). Infect. Genet. Evol. 30, 140–146.

[38] Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M. (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81–91.

[39] Guerrero, G., Peralta, H., Aguilar, A., Díaz, R., Villalobos, M.A., Medrano-Soto, A., Mora, J. (2005) Evolutionary, structural and functional relationships revealed by comparative analysis of syntenic genes in Rhizobiales. BMC Evol. Biol. 55, 55.

[40] Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

[41] Heath, B.D., Butcher, R.D.J., Whitfield, W.G.F., Hubbard, S.F. (1999) Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. Curr. Biol. 9, 313–316.

[42] Hertig, M., Wolbach, S.B. (1924) Studies in *Rickettsia*-like micro-organisms in insects. J. Med. Res. 44, 329–374.

[43] Hertig, M. (1936) The rickettsia, *Wolbachia pipientis* (gen. et sp. n.) and associated inclusions of the mosquito, Culex pipiens. Parasitology 28, 453–486.

[44] Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A., Werren, J.H. (2008) How many species are infected with *Wolbachia*? A statistical analysis of current data. FEMS Microbiol. Lett. 281, 215–220.

[45] Hoerauf, A., Volkmann, L., Nissen-Paehle, K., Schmetz, C., Autenrieth, I., Büttner, D.W., Fleischer, B. (2000) Targeting of *Wolbachia* endobacteria in *Litomosoides sigmodontis*: comparison of tetracyclines with chloramphenicol, macrolides and ciprofloxacin. Trop. Med. Int. Health 5, 275–279.

[46] Hoerauf, A., Mand, S., Adjei, O., Fleischer, B., Büttner, D.W. (2001) Depletion of *Wolbachia* endobacteria in *Onchocerca volvulus* by doxycycline and microfilariadermia after ivermectin treatment. The Lancet 357, 1415–1416.

[47] Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X-Y., Fukatsu, T. (2010) *Wolbachia* as a bacteriocyte-associated nutritional mutualist. PNAS 107, 769–774.

[48] Huigens, M.E., de Almeida, R.P., Boons, P.A.H., Luck, R.F., Stouthamer, R. (2004) Natural interspecific and intraspecific horizontal transfer of parthenogenesis-inducing *Wolbachia* in *Trichogramma* wasps. Proc. Biol. Sci. 271, 509–515.

[49] Jäckel, R., Mora, D., Dobler, S. (2013) Evidence for selective sweeps by *Wolbachia* infections: phylogeny of *Altica* leaf beetles and their reproductive parasites. Mol. Ecol. 22, 4241–4255.
[50] Kent, B.N., Salichos, L., Gibbons, J.G., Rokas, A., Newton, I.L., Clark, M.E., Bordenstein, S.R. (2011) Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. Genome Biol. Evol. 3, 209–218.
[51] Klasson, L., Walker, T., Sebaihia, M., Sanders, M.J., Quail, M.A., Lord, A., Sanders, S., Earl, J., O'Neill, S.L., Thomson, N., Sinkins, S.P., Parkhill, J. (2008) Genome evolution of *Wolbachia* strain *w*Pip from the *Culex pipiens* group. Mol. Biol. Evol. 25, 1877–1887.
[52] Klasson, L., Westberg, J., Sapountzis, P., Näslund, K., Lutnaes, Y., Darby, A.C., Veneti, Z., Chen, L., Braig, H.R., Garrett, R., Bourtzis, K., Andersson, S.G.E. (2009) The mosaic genome structure of the *Wolbachia w*Ri strain infecting *Drosophila simulans*. Proc. Natl. Acad. Sci. U. S. A. 106, 5725–5730.
[53] Konstantinidis, K.T., Roselló-Móra, R. (2015) Classifying the uncultivated microbial majority: a place for metagenomics data in the *Candidatus* proposal. Syst. Appl. Microbiol., http://dx.doi.org/10.1016/j.syapm.2015.01.001
[54] Koutsovoulos, G., Makepeace, B., Tanya, V.N., Blaxter, M. (2014) Palaeosymbiosis revealed by genomic fossils of *Wolbachia* in a Strongyloidean nematode. PLoS Genet. 10, e1004397.
[55] Kozek, W.J., Marroquin, H.F. (1977) Intracytoplasmic bacteria in *Onchocerca volvulus*. Am J. Trop. Med. Hyg. 26, 663–678.
[56] Kozek, W.J. (2005) What is new in the *Wolbachia*/*Dirofilaria* interaction? Vet. Parasitol. 133, 127–132.
[57] Landmann, F., Voronin, D., Sullivan, W., Taylor, M.J. (2011) Anti-filarial activity of antibiotic therapy is due to extensive apoptosis after *Wolbachia* depletion from filarial nematodes. PLoS Pathog. 7, e1002351.
[58] Landmann, F., Foster, J.M., Michalski, M.L., Slatko, B.E., Sullivan, W. (2014) Co-evolution between an endosymbiont and its nematode host: *Wolbachia* asymmetric posterior localization and AP polarity establishment. PLoS Negl. Trop. Dis. 8, e3096.
[59] Langworthy, N.G., Renz, A., Mackenstedt, U., Henkle-du, K., Bronsvoort, M.B.C., Tanya, V.N., Donnelly, M.J., Trees, A.J. (2000) Macrofilaricidal activity of tetracycline against the filarial nematode *Onchocerca ochengi*: elimination of *Wolbachia* precedes worm death and suggests a dependent relationship. Proc. R. Soc. Lond. 267, 1063–1069.
[60] Lefoulon, E., Gavotte, L., Junker, K., Barbuto, M., Uni, S., Landmann, F., Laaksonen, S., Saari, S., Nikander, S., Souza, S.D., Casiraghi, M., Bain, O., Martin, C. (2012) A new type F *Wolbachia* from Splendidofilariinae (Onchocercidae) supports the recent emergence of this supergroup. Int. J. Parasitol. 42, 1025–1036.
[61] Lo, N., Paraskevopoulos, C., Bourtzis, K., O'Neill, S.L., Werren, J.H., Bordenstein, S.R., Bandi, C. (2007) Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. Int. J. Syst. Evol. Microbiol. 57, 654–657.
[62] Louis, C., Nigro, L. (1989) Ultrastructural evidence of *Wolbachia* rickettsiales in *Drosophila simulans* and their relationships with unidirectional cross-incompatibility. J. Invertebr. Pathol. 24, 39–44.
[63] Mavingui, P., Valiente Moro, C., Tran-Van, V., Wisniewski-Dyé, F., Raquin, V., Minard, G., Tran, F.H., Voronin, D., Rouy, Z., Bustos, P., Lozano, L., Barbe, V., González, V. (2012) Whole-genome sequence of *Wolbachia* strain *w*AlbB, an endosymbiont of tiger mosquito vector *Aedes albopictus*. J. Bacteriol. 194, 1840.
[64] McLaren, D.J., Worms, M.J., Laurence, B.R., Simpson, M.G. (1975) Micro-organisms in filarial larvae (Nematoda). Trans. R. Soc. Trop. Med. Hyg. 69, 509–514.
[65] Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.P., Göker, M. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatic 14, 60.
[66] Metcalf, J.A., Jo, M., Bordenstein, S.R., Jaenike, J., Bordenstein, S.R. (2014) Recent genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism. PeerJ 2, e529.
[67] Moran, N.A. (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A. 93, 2873–2878.
[68] Niehuis, O., Hartig, G., Grath, S., Pohl, H., Lehmann, J., Tafer, H., Donath, A., Krauss, V., Eisenhardt, C., Hertel, J., Petersen, M., Mayer, C., Meusemann, K., Peters, R.S., Stadler, P.F., Beutel, R.G., Bornberg-Bauer, E., McKenna, D.D., Misof, B. (2012) Genomic and morphological evidence converge to resolve the enigma of *Strepsiptera*. Curr. Biol. 22, 1309–1313.
[69] Nikoh, N., Hosokawa, T., Moriyama, M., Oshima, K., Hattori, M., Fukatsu, T. (2014) Evolutionary origin of insect-*Wolbachia* nutritional mutualism. Proc. Natl. Acad. Sci. U. S. A. 111, 10257–10262.
[70] Ormeño-Orrillo, E., Martínez-Romero, E. (2013) Phenotypic tests in *Rhizobium* species description: an opinion and (a sympatric speciation) hypothesis. Syst. Appl. Microbiol. 36, 145–147.
[71] Ormeño-Orrillo, E., Servín-Garcidueñas, L.E., Rogel, M.A., González, V., Peralta, H., Mora, J., Martínez-Romero, J., Martínez-Romero, E. (2014) Taxonomy of rhizobia and agrobacteria from the Rhizobiaceae family in light of genomics. Syst. Appl. Microbiol. 38, 287–291.
[72] Pfarr, K., Foster, J., Slatko, B., Hoerauf, A., Eisen, J.A. (2007) On the taxonomic status of the intracellular bacterium *Wolbachia pipientis*: should this species name include the intracellular bacteria of filarial nematodes? Int. J. Syst. Evol. Microbiol. 57, 1677–1678.
[73] Pike, N., Kingcombe, R. (2009) Antibiotic treatment leads to the elimination of *Wolbachia* endosymbionts and sterility in the diplodiploid collembolan *Folsomia candida*. BMC Biol. 7, 54.
[74] Pinto, S.B., Stainton, K., Harris, S., Kambris, Z., Sutton, E.R., Bonsall, M.B., Parkhill, J., Sinkins, S.P. (2013) Transcriptional regulation of *Culex pipiens* mosquitoes by *Wolbachia* influences cytoplasmic incompatibility. PLoS Pathog. 9, e1003647.
[75] Richter, M., Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. 106, 19126–19131.
[76] Rieseberg, L.H. (2001) Chromosomal rearrangements and speciation. Trends Ecol. Evol. 16, 351–358.
[77] Rogel, M.A., Bustos, P., Santamaría, R.I., González, V., Romero, D., Cevallos, M.A., Lozano, L., Castro-Mondragón, J., Martínez-Romero, J., Ormeño-Orrillo, E., Martínez-Romero, E. (2014) Genomic basis of symbiovar mimosae in *Rhizobium etli*. BMC Genomic 15, 575.
[78] Ros, V.I., Fleming, V.M., Feil, E.J., Breeuwer, J.A. (2009) How diverse is the genus *Wolbachia*? Multiple-gene sequencing reveals a putatively new *Wolbachia* supergroup recovered from spider mites (Acari: Tetranychidae). App. Environ. Microbiol. 75, 1036–1043.
[79] Rosenblueth, M., Sayavedra, L., Sámano-Sánchez, H., Roth, A., Martínez-Romero, E. (2012) Evolutionary relationships of flavobacterial and enterobacterial endosymbionts with their scale insect hosts (Hemiptera: Coccoidea). J. Evol. Biol. 25, 2357–2368.
[80] Rousset, F., Bouchon, D., Pintureau, B., Juchault, P., Solignac, M. (1992) *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. Proc. R. Soc. Lond. B 250, 91–98.
[81] Rowley, S.M., Raven, R.J., McGraw, E.A. (2004) *Wolbachia pipientis* in Australian spiders. Curr. Microbiol. 49, 208–214.
[82] Sacchi, L., Genchi, M., Clementi, E., Negri, I., Alma, A., Ohler, S., Sassera, D., Bourtzis, K., Bandi, C. (2010) Bacteriocyte-like cells harbour *Wolbachia* in the ovary of *Drosophila melanogaster* (Insecta, Diptera) and *Zynginidia pullula* (Insecta, Hemiptera). Tissue Cell 42, 328–333.
[83] Saha, S., Hunter, W.B., Reese, J., Morgan, J.K., Marutani-Hert, M., Huang, H., Lindeberg, M. (2012) Survey of endosymbionts in the *Diaphorina citri* metagenome and assembly of a *Wolbachia w*Di draft genome. PLoS ONE 7, e50067.
[84] Salzberg, S.L., Dunning Hotopp, J.C., Delcher, A.L., Pop, M., Smith, D.R., Eisen, M.B., Nelson, W.C. (2005) Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. Genome Biol. 6, R23.
[85] Salzberg, S.L., Puiu, D., Sommer, D.D., Nene, V., Lee, N.H. (2009) Genome sequence of the *Wolbachia* endosymbiont of *Culex quinquefasciatus* JHB. J. Bacteriol. 191, 1725.
[86] Schilthuizen, M., Stouthamer, R. (1997) Horizontal transmission of parthenogenesis-inducing microbes in *Trichogramma* wasps. Proc. R. Soc. Lond. B 264, 361–366.
[87] Schröder, D., Deppisch, H., Obermayer, M., Krohne, G., Stackebrandt, E., Hölldobler, B., Goebel, W., Gross, R. (1996) Intracellular endosymbiotic bacteria of *Camponotus* species (carpenter ants): systematics, evolution and ultrastructural characterization. Mol. Microbiol. 21, 479–489.
[88] Shoemaker, D.D., Machado, C.A., Molbo, D., Werren, J.H., Windsor, D.M., Herre, E.A. (2002) The distribution of *Wolbachia* in fig wasps: correlations with host phylogeny, ecology and population structure. Proc. R. Soc. Lond. B 269, 2257–2267.
[89] Siozios, S., Cestaro, A., Kaur, R., Pertot, I., Rota-Stabelli, O., Anfora, G. (2013) Draft genome sequence of the *Wolbachia* endosymbiont of *Drosophila suzukii*. Genome Announc. 1, e00032–13.
[90] Sironi, M., Bandi, C., Sacchi, L., Di Sacco, B., Damiani, G., Genchi, C. (1995) Molecular evidence for a close relative of the arthropod endosymbiont *Wolbachia* in a filarial worm. Mol. Biochem. Parasitol. 74, 223–227.
[91] Sutton, E.R., Harris, S.R., Parkhill, J., Sinkins, S.P. (2014) Comparative genome analysis of *Wolbachia* strain *w*Au. BMC Genomic 15, 928.
[92] Taylor, M.J., Bilo, K., Cross, H.F., Archer, J.P., Underwood, A.P. (1999) 16S rDNA phylogeny and ultrastructural characterization of *Wolbachia* intracellular bacteria of the filarial nematodes *Brugia malayi*, *B. pahangi*, and *Wuchereria bancrofti*. Exp. Parasitol. 91, 356–361.
[93] Taylor, M.J., Bandi, C., Hoerauf, A.M., Lazdins, J. (2000) *Wolbachia* bacteria of filarial nematodes: a target for control? Parasitol. Today 16, 179–180.
[94] Teixeira, L., Ferreira, A., Ashburner, M. (2008) The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. PLoS Biol. 6, e1000002.
[95] Timmermans, M.J.T.N., Ellers, J. (2009) *Wolbachia* endosymbiont is essential for egg hatching in a parthenogenetic arthropod. Evol. Ecol. 23, 931–942.
[96] Vandamme, P., Peeters, C. (2014) Time to revisit polyphasic taxonomy. Antonie van Leeuwenhoek 106, 57–65.
[97] Vandekerckhove, T.T., Watteyne, S., Willems, A., Swings, J.G., Mertens, J., Gillis, M. (1999) Phylogenetic analysis of the 16S rDNA of the cytoplasmic bacterium *Wolbachia* from the novel host *Folsomia candida* (Hexapoda, Collembola) and its implications for wolbachial taxonomy. FEMS Microbiol. Lett. 180, 279–286.
[98] Werren, J.H., Windsor, D., Guo, L.R. (1995) Distribution of *Wolbachia* among neotropical arthropods. Proc. R. Soc. Lond. B 262, 197–204.
[99] Werren, J.H., Bartos, J.D. (2001) Recombination in *Wolbachia*. Curr. Biol. 11, 431–435.
[100] Werren, J.H., Baldo, L., Clark, M.E. (2008) *Wolbachia*: master manipulators of invertebrate biology. Nat. Rev. Microbiol. 6, 741–751.
[101] Woolfit, M., Iturbe-Ormaetxe, I., Brownlie, J.C., Walker, T., Riegler, M., Seleznev, A., Popovici, J., Rancès, E., Wee, B.A., Pavlides, J., Sullivan, M.J., Beatson, S.A., Lane, A., Sidhu, M., McMeniman, C.J., McGraw, E.A., O'Neill, S.L. (2013)

Genomic evolution of the pathogenic *Wolbachia* strain, *w*MelPop. Genome Biol. Evol. 5, 2189–2204.

[102] Wu, M., Sun, L.V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J.C., McGraw, E.A., Martin, W., Esser, C., Ahmadinejad, N., Wiegand, C., Madupu, R., Beanan, M.J., Brinkac, L.M., Daugherty, S.C., Durkin, A.S., Kolonay, J.F., Nelson, W.C., Mohamoud, Y., Lee, P., Berry, K., Young, M.B., Utterback, T., Weidman, J., Nierman, W.C., Paulsen, I.T., Nelson, K.E., Tettelin, H., O'Neill, S.L., Eisen, J.A. (2004) Phylogenomics of the reprodutive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. 2, 327–341.

[103] Wu, M., Scott, A.J. (2012) Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics 28, 1033–10334.

[104] Yang, X.H., Zhu, D.H., Liu, Z., Zhao, L., Su, C.Y. (2013) High levels of multiple infections, recombination and horizontal transmission of *Wolbachia* in the *Andricus mukaigawae* (Hymenoptera; Cynipidae) communities. PLoS ONE 8, e78970.

[105] Zhukova, M., Kiseleva, E. (2012) The virulent *Wolbachia* strain *w*MelPop increases the frequency of apoptosis in the female germline cells of *Drosophila melanogaster*. BMC Microbiol. 12, S15.

[106] Zug, R., Hammerstein, P. (2012) Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. PLoS ONE 7, e38544.

Artículo:

Bolaños LM, Servín-Garcidueñas LE, Martínez-Romero E. 2015. Arthropod-*Spiroplasma* relationship in the genomic era. FEMS Microbiol. Ecol. 91:1-8.

En el artículo se presenta una revisión sobre el género *Spiroplasma* en base a las características de los genomas disponibles para distintas especies de artrópodos. En este trabajo también se presenta un análisis filogenómico del género *Spiroplasma*. Los datos obtenidos se presentan en detalle en el artículo de revisión.

FEMS Microbiology Ecology, 91, 2015

doi: 10.1093/femsec/fiu008
Advance Access Publication Date: 5 December 2014
Minireview

MINIREVIEW

Arthropod–*Spiroplasma* relationship in the genomic era

Luis M. Bolaños*, Luis E. Servín-Garcidueñas and Esperanza Martínez-Romero

Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México

*Corresponding author: Centro de Ciencias Genómicas, UNAM. Av. Universidad SN, Chamilpa, CP 62210, Cuernavaca, Mor., Mexico. Tel/Fax: +52-777-317-5581. E-mail: lbolanos@lcg.unam.mx

One Sentence Summary: MiniReview focused in Spiroplasma-Arthropod symbiosis in the context of the bacterial sequenced genomes and the elucidation of functional and evolutionary traits shaping these relationships.
Editor: Gerard Muyzer

ABSTRACT

The genus *Spiroplasma* comprises wall-less, low-GC bacteria that establish pathogenic, mutualistic and commensal symbiotic associations with arthropods and plants. This review focuses on the symbiotic relationships between *Spiroplasma* bacteria and arthropod hosts in the context of the available genomic sequences. *Spiroplasma* genomes are reduced and some contain highly repetitive plectrovirus-related sequences. *Spiroplasma*'s diversity in viral invasion susceptibility, virulence factors, substrate utilization, genome dynamics and symbiotic associations with arthropods make this bacterial genus a biological model that provides insights about the evolutionary traits that shape bacterial symbiotic relationships with eukaryotes.

Key words: *Spiroplasma*; symbiosis; comparative genomics

INTRODUCTION

The *Spiroplasma* genus consists of cell-wall-less, helical, low-GC bacteria belonging to the class *Mollicutes*. Spiroplasmas are described as facultative anaerobes that exhibit a wide range of growth temperatures between 5 and 41°C (Konai *et al.*, 1996). These bacteria establish symbiotic associations mainly with arthropods. Associations with dipteran and coleopteran insect orders are frequent and have been largely reported (Wedincamp *et al.*, 1996). Other insect orders where spiroplasmas have been isolated are Hemiptera, Homoptera, Hymenoptera, Lepidoptera and Odonata (Hackett and Clark 1989; Hackett *et al.*, 1990; Watanabe *et al.*, 2014). Also, spiroplasmas have been isolated from non-insect arthropods and plants (Davis, Lee and Worley 1981; Saillard *et al.*, 1987; Wang *et al.*, 2004; Goodacre *et al.*, 2006).

The majority of *Spiroplasma* species described as insect symbionts have no effect on the hosts and are considered as commensal bacteria (Gasparich 2010). In some hosts like shrimp (Nunan *et al.*, 2004), honeybees (Clark 1977), and mosquitoes (Phillips and Humphery-Smith 1995), spiroplasmas have been characterized as pathogens. Pathogenicity is related to the capacity to cross the midgut lumen barrier, hemolymph invasion and therefore colonization of other host tissues, that lead in some cases to host death (Nunan *et al.*, 2005).

A significant effect on hosts like *Drosophila* (Williamson *et al.*, 1999) and other insects (Tabata *et al.*, 2011) is the male-killing phenotype, where maternally inherited *Spiroplasma* kill host male offspring in early stages of development.

Spiroplasma as mutualistic symbionts can be found in *Drosophila* and aphids providing host protection against

Received: 7 July 2014; **Accepted:** 6 October 2014

Table 1. General features of the sequenced *Spiroplasma* genomes.

	S. chryso-picola DF-1	*S. syrphy-dicola* EA-1	*S. citri* Gll3-3X	*S. melli-ferum* IPMB4A	*S. melli-ferum* KC3	*S. dimi-nutum* CUAS-1^T	*S. taiwa-nense* CT-1^T	*S. apis* B31^T	*S. culici-cola* AES-1^T	*S. sabau-diense* Ar-1343^T
Arthropod host	*Chrysops sp.*	*E. arbustorum*	*C. haema-toceps*	*Apis-Mellifera*	*A. Mellifera*	*Culex-annulus*	*C. tritaeni-orhynchus*	*A. Mellifera*	*Aedes-sollicitans*	*stricticus/ A. vexans*
Symbiotic relationship	Commensal	Commensal	Commensal*	Pathogenic	Pathogenic	Commensal	Pathogenic	Pathogenic	Pathogenic	Commensal
Genome Size (bp)	1123 322	1107 344	1525 756	1098 846	1260 174	945 296	1075 140	1160 554	1175 131	1075 953
Chromo-somal contigs	1	1	39	24	4	1	1	1	1	1
G + C content (%)	28.8	29.2	25.9	27.5	27	25.5	23.9	28.3	26.4	30.2
rRNA operon	1	1	1	1	1	1	1	1	1	2
tRNA	29	29	29	29	29	29	29	29	29	30
Number of plasmids	0	0	7	4	0	0	1	0	0	0
Protein-coding genes	1009	1006	1170	920	1046	858	991	997	1071	924

*Commensal: insect is used only as a vector to infect plants.

parasitoid wasps (Xie, Vilchez and Mateos 2010), nematodes (Jaenike *et al.*, 2010; Cockburn *et al.*, 2013) and fungal pathogens (Lukasik *et al.*, 2013). *Spiroplasma kunkelii* increases the survival rate of the leafhooper *Dalbulus maidis* during cold and dry periods when the leafhopper's host plant is not accessible (Ebbert and Nault 1994).

Spiroplasma symbiotic associations with arthropods can be considered as biological model systems to study molecular mechanisms and evolutionary traits that shape contrasting bacterial-host interactions.

In this review, we will focus on arthropod-associated *Spiroplasma* species with complete genome sequences, the biological implications of harboring this bacteria and the information that genomic sequences provide towards understanding of symbiotic relationships between arthropods and spiroplasmas.

GENERAL FEATURES OF *Spiroplasma* GENOMES

It was not until 2010 that the first draft genome sequence of *S. citri* became available (Carle *et al.*, 2010). Currently, there are 10 genomes deposited in either draft or complete assembly in the NCBI genome database. Four sequenced strains are associated with mosquitoes. *Spiroplasma culicicola* AES-1 and *S. taiwanense* CT-1^T are known pathogens that produce tissue damage and increased mortality in their respective mosquito host (Humphery-Smith *et al.*, 1991). Infections by the sequenced strains of *S. sabaudiense* Ar-1343 and *S. diminutum* CUAS-1^T show no significant effects on mosquitoes and are considered commensal bacteria (Abalain-Colloc *et al.*, 1987; Williamson *et al.*, 1996). Three other *Spiroplasma* genomes correspond to honeybee pathogens, including two *S. melliferum* strains and one strain of *S. apis* (Bové *et al.*, 1983). Further, *S. chrysopicola* DF-1 and *S. syrphydicola* EA-1 strains are considered commensals of the deerfly *Chrysops* sp. (Whitcomb *et al.*, 1997) and the syrphyd fly *Eristalis arbustorum* (Whitcomb *et al.*, 1996), respectively. Finally, the sequenced strain of *S. citri* Gll3-3X is the causal agent of citrus stubborn disease in plants and is transmitted by leafhoppers that feed from phloem nutrients. In this case, the insect host act as a vector of the plant pathogen (Bové *et al.*, 2003). It is important to highlight that no genome sequences for mutalistic *Spiroplasma* genomes have been published or released to public repositories.

Spiroplasma genome features are summarized in Table 1. Their genome size ranges from 780 to 2220 kb. This range of genome sizes is wider than those of other mollicutes such as *Mycoplasma* (Carle *et al.*, 1995). G + C content of the sequenced genomes ranges from 23.9% of *S. taiwanense* CT-1^T to 30.2% of *S. sabaudiense* Ar-1343. All the sequenced genomes have one rRNA operon and 29 tRNA genes, except *S. sabaudiense*, which has two complete and identical rRNA operons and 30 tRNA genes with an extra copy of tRNA-Ser gene. Only few spiroplasmas have plasmids. *Spiroplasma citri* Gll3-3X has seven plasmids, the largest being pSci6 (35.3 kb) and the shortest pSciA (7.8 kb) (Saillard *et al.*, 2008). *Spiroplasma melliferum* IPMB4A has four plasmids of 4.7, 5.6, 9.86 and 14.45 kb (Alexeev *et al.*, 2011). *Spiroplasma taiwanense* CT-1^T has only one plasmid of approximately 11 kb (Gasparich and Hackett 1994).

Spiroplasmas have small genomes with large variations in gene content. For example, in the Citri–Chrysopicola–Mirum, the two strains of *S. melliferum* share 864 genes representing 78.4% of the total genes. If this comparison is made with more phylogenetically distant species like *S. citri*, the amount of common genes decreases to 51.7% of the total genes (Lo *et al.*, 2013a). In the Apis clade, *S. diminitum* and *S. taiwanense* share 59% of their genes. Comparisons between genomes of different clades resulted in lower values of common genes. *Spiroplasma melliferum* shares 38.84% of its genes with *S. diminitum* and only 34.5% with *S. taiwanense* (Lo *et al.*, 2013b). Recently, it has been proposed that 'a prokaryotic genus can be defined as a group of species with all pairwise percentage of conserved proteins values higher than 50%' (Qin *et al.*, 2014). In this context, the above-mentioned set of common genes between the *S. melliferum* strains and *S. citri* is on the borderline limit, even when they belong to the Citri–Chrysopicola–Mirum clade. Furthermore, the shared genes between spiroplasmas of different clades are lower than the proposed threshold which belong to the same genus. Reduced genome size is a common feature in spiroplasmas, but gene conservation seems to depend on host selective pressures.

Spiroplasma GENOME SEQUENCES AND THEIR PHYLOGENOMIC RELATIONSHIPS

The Entomoplasmatales order is composed of four clades: Mycoides–Entomoplasmataceae, Apis, Citri–Chrysopicola–Mirum and Ixodetis. *Mycoplasma*, *Mesoplasma* and *Entomoplasma* genera are restricted to Mycoides–Entomoplasmataceae clade. *Spiroplasma* species are distributed in the other three clades.

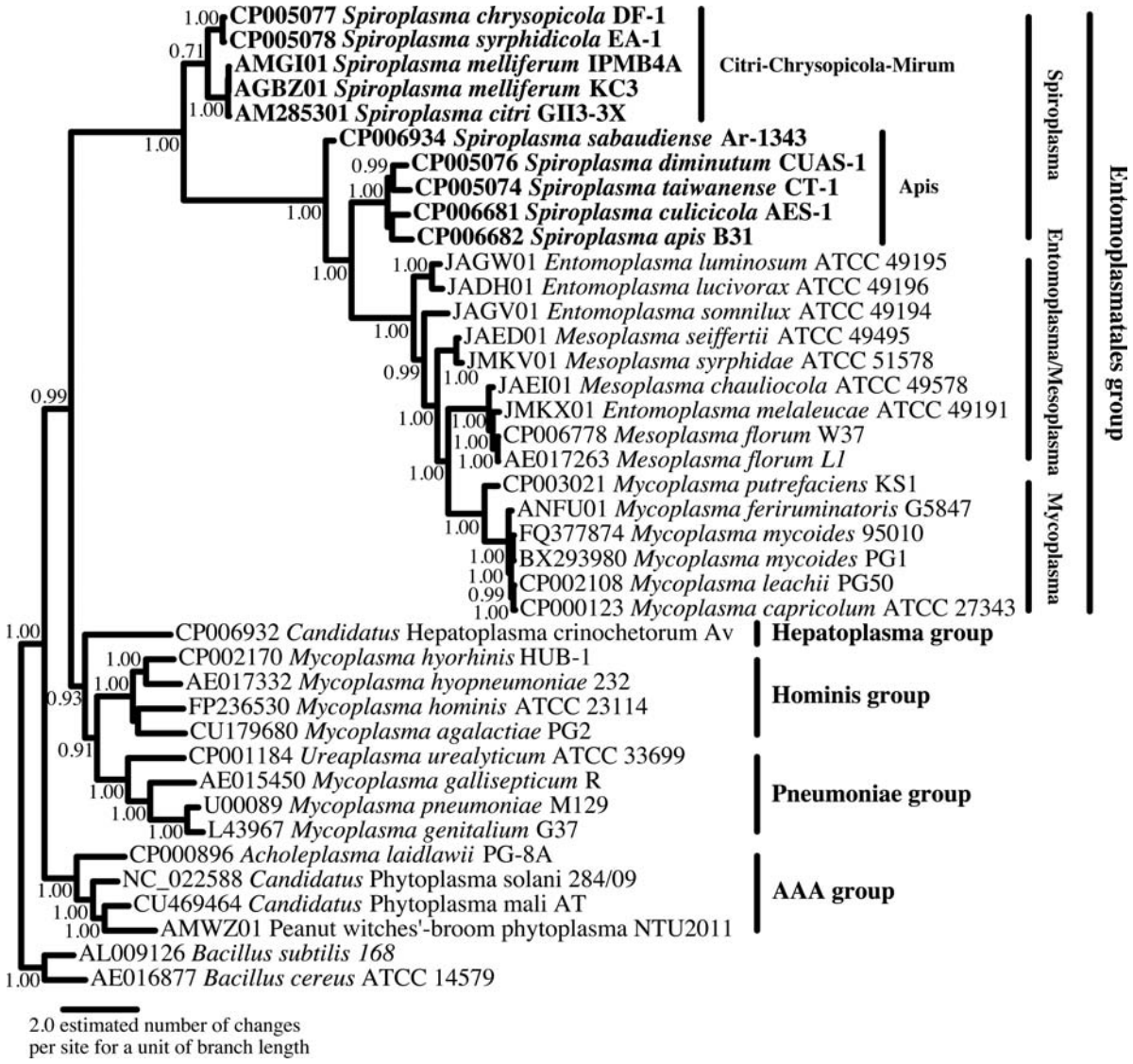


Figure 1. Topology of a phylogenomic analysis showing the predicted evolutionary relationships of sequenced *Spiroplasma* strains within the *Mollicutes*. The tree was reconstructed with PhyloPhlAn using a multisequence alignment of 388 conserved proteins. PhyloPhlAn performs individual alignments from each protein set recovered from the *Mollicutes* input genomes. PhyloPhlAn then concatenates the most discriminative positions in each protein alignment into a single long sequence to reconstruct a phylogenetic tree using FastTree. *Spiroplasma* strains are shown in bold in the tree. Accession numbers are indicated for all sequenced genomes. *Mollicutes* groups are indicated in the tree. Two *Bacillus* strains were used as outgroup. Numbers at the branch points represent SH-like local support values (based on 1000 resamples). The scale bar represents the estimated number of amino acid changes per site for a unit of branch length.

Based on the 16S rRNA gene phylogeny, the genus *Spiroplasma* is not monophyletic (Gasparich 2002).

Spiroplasma strains with available genome sequences correspond to representative members from the Citri–Chrysopicola–Mirum and Apis clades. No *Spiroplasma* genome from the Ixodetis clade is yet available. A phylogenomic approach using multiple amino acid markers (Segata *et al.*, 2013) clearly distinguishes the Citri–Chrysopicola–Mirum clade from the Apis clade, which seems to share a common ancestor with the Entomoplasma/Mesoplasma and Mycoplasma groups (Fig. 1).

Average Nucleotide Identity (ANI) and DNA–DNA Hybridization (DDH) pairwise comparisons between *Spiroplasma* genomes are congruent with the phylogenomic tree. ANI and DDH values highlight the genomic divergence between the Citri–Chrysopicola–Mirum clade from the Apis clade. Species with highest ANI and DDH values are *S. syrphidicola* and *S. chrysopicola* in the Apis clade and *S. citri* with both strains of *S. melliferum* in the Citri–Chrysopicola–Mirum clade. *Spiroplasma sabaudiense* has relatively low values compared with all other sequenced spiroplasmas, but is more closely related to the Apis clade (Tables S1 and S2, Supporting Information). 16S rRNA gene pairwise alignments also revealed striking differences between both clades as they share approximately 90% sequence identity over the entire marker gene (Table S3, Supporting Information). It has been suggested that a cutoff of 94.9% ± 0.4 should define genus boundaries based on 16S rRNA gene sequence identities (Yarza *et al.*, 2008, 2010). The wide range of *Spiroplasma* identities has a minimum of 90.08% and a maximum of 99.67%, which also supports the great divergence between the different species in the genus. Even more revealing is the difference between the highest identities of some Apis clade species against those of the Citri–Chrysopicola–Mirum clade, similar to the observed

ANI and DDH values. For example, both strains of *S. melliferum* are 97.15% identical, and different Apis clade species such as *S. chrysopicola* and *S. syrphidicola*, which are 99.67% identical.

PLECTROVIRAL SEQUENCES

One of the most striking features of some *Spiroplasma* genomes is the presence of a great amount of plectroviral sequences. Plectroviruses are bacteriophages that infect exclusively cell-wall-less bacteria. Genomes of plectroviruses are present in multiple regions of bacterial genomes (Rakonjac 2012). The presence of plectroviral sequences was detected since the first report of the *S. kunkelii* 85 kb genomic sequence (Zhao *et al.*, 2003). Plectroviral sequences have been found in other *Spiroplasma* genomes. In *S. citri*, *S. melliferum* IPMB4A and *S. melliferum* KC3, repetitive plectroviral sequences and proteins of viral origin were found distributed all over their chromosome. The absence of plectroviral sequences in the genomes of the Chrysopicola and Apis clades seems to indicate that susceptibility of viral invasions may be restricted only to the Citri clade. *Spiroplasma chrysopicola* and *S. syrphidicola* genomes do not have any trace of plectroviral sequences (protein-coding or non-coding) unlike the Citri clade genomes. This may be in relation to the presence of antiviral systems in these strains, such as clustered regularly interspaced short palindromic repeats and type 1 and 2 restriction/modification (R/M), which were found in *S. syrphidicola* and *S. chrysopilica*, respectively (Ku *et al.*, 2013). However, the type 1 R/M system is truncated in *S. citri*.

Ku *et al.* (2013) proposed a model for *Spiroplasma* evolution in relation to viral susceptibility. Allegedly, the common ancestor of Chrysipicola and Citri clades had an active antiviral system or systems and therefore was resistant to the virus, similar to *S. chrysipicola* and *S. syrphiciola*. The Chrysipicola clade diverged and the ancestor of the Citri clade lost its antiviral system(s) and began to accumulate viral fragments. The consequences of these viral infections were the increase in genome sizes and higher homologous recombination rates due to the copies of viral fragments, and even non-viral DNA acquisition. Concomitantly, there should be a counterbalance of genomic acquisitions with loss of old fragments as new viral sequences were inserted. The rearrangement effect of the old fragments seems to be untraceable from the present distribution of the viral fragments.

Viruses can help to horizontally transfer virulence related genes among bacteria (Moore and Lindsay 2001). The genomes susceptible to viral infections are those from *Spiroplasma* pathogenic to bees and plants from the Citri clade. Pathogenicity genes were probably acquired by lateral transfer mediated by virus; thus, a correlation of viral infections with pathogenic lifestyle of spiroplasmas has been proposed (Ku *et al.*, 2013). Genome plasticity emerges from constant sequence acquisitions and losses under strong selective pressures. This plasticity may have led *Spiroplasma* to develop mechanisms that allow it to be undetected by the host immune system (Anbutsu and Fukatsu 2010; Herren and Lemaitre 2011).

The *recA* gene is truncated in strains of *S. citri* and *S. melliferum* MC3 (Marais, Bove and Renaudin 1996). RecA mediates homologous recombination, essential for maintaining genomic integrity and generating genetic diversity (Chen, Yang and Pavletich 2008). *Spiroplasma melliferum* IPMB4A genome lacks the machineries for mismatch repair and homologous recombination, *recA* included (Lo *et al.*, 2013a). The loss of a functional RecA in these species seems to be a relatively recent event. It is suggested that genome instability occurred before *recA* loss (Ku *et al.*, 2013).

PATHOGENICITY FACTORS

Spiroplasma pathogenicity in arthropods has been correlated with the ability of the bacteria to cross the epithelial gut lumen barrier. After trespassing the gut tissue, bacteria infect other host tissues via the hemolymph (Kwon, Wayadande and Fletcher 1999). To accomplish this task, pathogens need a set of molecular tools with different specialized functions. Genomic sequencing has provided insights into possible genes associated with the gradual steps of host tissue invasion.

Chitin degradation genes including chitinase A (*chiA*) and a putative chitin deacetylase were proposed to be used for the first step of invasion, which is the permeabilization of bacterial load through the epithelial barrier. The discovery of the protein product of chitin deacetylase gene in the *S. melliferum* KC3 proteome provides evidence that this gene is being expressed (Alexeev *et al.*, 2011). Chitin is a structural biopolymer of various insect cuticles including the gut lumen (Merzendorfer and Zimoch 2003). Chitin degradation causes permeation of structural components of the peritrophic matrix of the gut epithelium. However, *chiA* and putative chitin deacetylase homologues are found in commensal *S. chrysopicola* and *S. syrphidicola*, and not in pathogenic *S. taiwanense*. Chitin is also present in the cell walls of fungi; it remains to be elucidated if *Spiroplasma* chitinases could have an effect on fungal gut microbiota, and consequently on hosts.

Another proposed mechanism of transmissibility and invasion of insect cells is receptor-mediated endocytosis (Özbek *et al.*, 2003; Ammar *et al.*, 2004). Spiralin has been proposed as a protein for intestinal epithelium receptor recognition, along with P89, P58, sc76 and P32 (Ye, Melcher and Fletcher 1997; Yu, Wayadande and Fletcher 2000; Boutareaud *et al.*, 2004; Killiny *et al.*, 2006). Spiralin genes are present among the sequenced Citri–Chrysopicola–Mirum *Spiroplasma* genomes and the gene products account for up to 30% of the total protein mass of spiroplasmas (Wroblewski *et al.*, 1977). Protein sequence pairwise comparisons between spiralins revealed low identities between species. The highest protein identity is 99% between both strains of *S. melliferum*. *Spiroplasma citri* strain identities range from 92.7 to 99% (Khanchezar *et al.*, 2014) . The lowest identity among the five sequenced species is 38% between *S. chrysopicola* and both strains of *S. melliferum*. The highest interspecies identity is 71% between *S. citri* and *S. melliferum* KC3. These values support that spiralin is a highly divergent protein (Foissac *et al.*, 1996; Meng *et al.*, 2010). *In vitro*, spiralin binds glycoproteins from its insect vector (Killiny, Castroviejo and Saillard 2005); although it is not important for pathogenicity in plants, it is essential for *S. citri* infection (Duret *et al.*, 2003). Regularly, spiralin is distributed along the cell. During adhesion of *S. citri* with *Circulifer haematoceps* cells, spiralin relocates to the space of contact and acts as an adhesin, which allow further internalization of bacterial cells into the insect cell (Duret *et al.*, 2014).

Once the intestinal lumen has been crossed, spiroplasmas reach hemolymph, where they start to proliferate. It has been suggested that proliferation is limited by the availability of nutrients, specifically hemolymph lipids (Herren *et al.*, 2014).

Spiroplasma citri (Gaurivaud *et al.*, 2000) and *S. melliferum* (Chang and Chen 1983) can ferment trehalose, the main sugar and carbon source in insect hemolymph. TreB is a transporter involved in the uptake of trehalose and TreA is the enzyme that converts trehalose-6P to glucose-6P. Both genes are present in *S. melliferum*, *S. citri* and *S. diminutum* (non-pathogenic). *Spiroplasma taiwanense* lacks *treA* and *treB*, so it may have alternative metabolic capacities to survive in its host hemolymph.

Genomic comparisons of pathogenic and commensal spiroplasmas of mosquitoes indicated that *L*-α-glycerophosphate oxidase (GlpO) might be a virulence factor. GlpO converts *sn*-glycerol 3-phosphate + O_2 to glycerone phosphate + H_2O_2 (Chang *et al.*, 2014). Previously, it was found that GlpO plays a central role in virulence of *Mycoplasma mycoides* due to the production and translocation of H_2O_2 into the host cell (Bischof, Vilei and Frey 2009). In *Spiroplasma*, the two pathogenic species *S. culicicola* and *S. taiwanense* share a copy of *glpO*, along with the transporter genes *ugpA*, *ugpC* and *ugpE*, which allow the *sn*-glycerol 3-phosphate uptake and a glycerol kinase for glycerol phosphorylation (*glpK*) (Chang *et al.*, 2014). In contrast, the mosquito commensals *S. diminutum* and *S. sabaudiense* lack these genes. Interestingly, *glpO* is conserved in the commensal species *S. chrysopicola* and *S. syrphidicola* from deer flies and syrphid flies. These two species also have an *ugpA* ortholog annotated as a hypothetical protein, but no tissue damage in their hosts has been reported.

COMPARATIVE GENOMICS AND METABOLISM

Spiroplasmas have some common biochemical characteristics such as glucose fermentation, arginine hydrolysis and inability to hydrolyze urea, and the majority require an external source of sterols (Regassa and Gasparich 2006). Like other Mollicutes, spiroplasmas have very limited biosynthetic capabilities (Petzel and Hartman 1990) and are considered fastidious organisms due to the complex nutritional requirements needed to grow in culture. They lack almost all of the genes required for amino acid synthesis. In contrast, they conserve a set of genes that encode transporters like the arginine/ornithine antiporter present in *S. sabaudiense*, *S. citri* and *S. syrphidicola* among other amino acid permeases. But the putative main system to acquire amino acids from the media is the oligopeptide transport system, which is conserved among all the *Spiroplasma* genomes (*oppA, oppB, oppC, oppD, oppF*).

Other important conserved permeases among all the genomes are those for the transport of glucose (*ptsG*) and fructose (*fruA*). Besides the transporters, spiroplasmas have the complete set of genes involved in glycolysis. Furthermore, *S. diminutum* has genes for sucrose uptake (*scrA*), and its conversion to glucose-6P (*scrB*) or fructose-6P (*scrK*). It is important to note that the presence of genes does not necessarily mean that they are expressed. Future transcriptomic analyses are required to confirm the functionality of these genes.

Among the few common biosynthetic capabilities of spiroplasmas are the non-mevalonate pathway for isopentenyl pyrophospate synthesis and the pathway for nucleotide biosynthesis. The non-mevalonate pathway for isopentenyl pyrophospate (I-PP) is composed of seven genes (*dxr, dxs, ispD, ispE, ispF, ispG* and *ispH*). This pathway takes as input pyruvate to produce I-PP, a precursor for the biosynthesis of terpenes. For numerous microbial pathogens, the non-mevalonate pathway is the only source of terpenoids (Rohdich *et al.*, 2002). However, spiroplasmas have no annotated genes involved in the next steps of terpenoid biosynthesis, but only the intermediate *uppS* gene. This gene transforms farnesyl pyrophosphate into undecaprenyl pyrophosphate. Spiroplasmas apparently lack the enzyme that converts I-PP into farnesyl-PP: this missing gene should link both pathways and challenges whether the spiroplasmas can produce undecaprenyl pyrophosphate terpene.

SEX-RATIO DISTORTION MECHANISM

Spiroplasma is widely recognized because of the male-killing phenotype induction in *Drosophila* flies. These bacteria are vertically transmitted maternally and kill male eggs before gastrulation (Counce and Poulson 1962). Several strains of male-killing *Spiroplasma* have been isolated from different species of *Drosophila* (Williamson and Poulson 1979; Pool, Wong and Aquadro 2006), butterflies (Jiggins *et al.*, 2000) and ladybird beetles (Tinsley and Majerus 2006), in addition to other strains that do not express male-killing phenotype in their hosts (Kageyama *et al.*, 2006). The molecular mechanisms underlying this phenotype have begun to be elucidated. Recently, two mechanisms have been described: apoptosis-dependent epidermal cell death and apoptosis-independent neural malformation. *Drosophila* embryos infected with male-killing spiroplasmas develop a remarkable neural malformation. Additionally, *Drosophila* embryos with *Spiroplasma* show an up-regulated, male-specific apoptotic pathway mainly targeted to embryonic epithelial cells (Martin, Chong and Ferree 2013). The two mechanisms seem to be independent because even if the host apoptotic pathway is disrupted, the male-specific neural malformation occurs (Harumoto, Anbutsu and Fukatsu 2014).

An important observation in the study of the male-apoptosis-dependent epidermal cell death mechanism is that *Spiroplasma* abundance is not the factor responsible for the phenotype. The signal that triggers these effects on male embryos should be *Spiroplasma*-derived factor(s) that act(s) selectively. Unfortunately, currently there are no genomic sequences of *Spiroplasma* isolated from *Drosophila* hosts. Transcriptomic studies with male-killing and non-male-killing strains may unveil *Spiroplasma* factors produced in the presence of *Drosophila* embryos.

CONCLUSIONS

The presence of repetitive phage sequences hampered the complete assembly of *Spiroplasma* genomes. Despite the difficulties, *Spiroplasma* genome sequencing projects have elucidated important information on metabolism, pathogenicity and genome dynamics. Other biological aspects such as the molecular male-killing mechanisms or the possible genes involved in mutualistic symbiosis functions have not been revealed by comparative genomics. Hypothetical proteins account for approximately 40% of the total protein coding genes and the majority of the species-specific genes across genome comparisons are annotated as hypothetical. Species-specific genes could be the most important elements to understand the intimate and unique associations that *Spiroplasma* establishes with their hosts. Ongoing *Spiroplasma* genome sequencing projects will enrich phylogenomic and comparative genome analyses. Parallel studies of *in vivo* transcriptomic or proteomic analysis should be done to understand the gene expression dynamics of *Spiroplasma* genes and proteins in the presence of the host. Furthermore, the creation of mutant banks from these strains could help elucidate functions for novel, hypothetical or unknown annotated *Spiroplasma* genes.

SUPPLEMENTARY DATA

Supplementary data is available at FEMSEC online.

ACKNOWLEDGEMENTS

We are thankful to the Programa de Doctorado en Ciencias Biomédicas from UNAM. We also thank M. A. Cevallos, C. Sohlenkamp, M. Dunn and G. E. Gasparich for critically reading the paper, and V. Rodríguez for the graphical abstract picture.

FUNDING

Financial support was from CONACyT (154453). LMB and LES had CONACyT scholarship as graduate students.

Conflict of interest statement. None declared.

REFERENCES

Abalain-Colloc ML, Chastel C, Tully JG, *et al*. *Spiroplasma sabaudiense* sp. nov. from mosquitoes collected in France. *Int J Syst Bacteriol* 1987;**37**:260–5.

Alexeev D, Kostrjukova E, Aliper A, *et al*. Application of *Spiroplasma melliferum* proteogenomic profiling for the discovery of virulence factors and pathogenicity mechanisms in host-associated spiroplasmas. *J Proteome Res* 2011;**11**:224–36.

Ammar ED, Fulton D, Bai X, *et al*. An attachment tip and pili-like structures in insect-and plant-pathogenic spiroplasmas of the class Mollicutes. *Arch Microbiol* 2004;**181**:97–105.

Anbutsu H, Fukatsu T. Evasion, suppression and tolerance of Drosophila innate immunity by a male-killing *Spiroplasma* endosymbiont. *Insect Mol Biol* 2010;**19**:481–8.

Bischof DF, Vilei EM, Frey J. Functional and antigenic properties of GlpO from *Mycoplasma mycoides* subsp. mycoides SC: characterization of a flavin adenine dinucleotide-binding site deletion mutant. *Vet Res* 2009;**40**:1–12.

Boutareaud A, Danet JL, Garnier M, *et al*. Disruption of a gene predicted to encode a solute binding protein of an ABC transporter reduces transmission of *Spiroplasma citri* by the leafhopper *Circulifer haematoceps*. *Appl and Environ Microb* 2004;**70**:3960–7.

Bové JM, Mouches C, Carle-Junca P, *et al*. Spiroplasmas of Group I: the *Spiroplasma citri* cluster. *Yale J Biol Med* 1983;**56**:573–82.

Bové JM, Renaudin J, Saillard C, *et al*. *Spiroplasma citri*, a plant pathogenic mollicute: relationships with its two hosts, the plant and the leafhopper vector. *Annu Rev Phytopathol* 2003;**41**:483–500.

Carle P, Laigret F, Tully JG, *et al*. Heterogeneity of genome sizes within the genus *Spiroplasma*. *Int J Syst Bacteriol* 1995;**45**(1):178–181.

Carle P, Saillard C, Carrère N, *et al*. Partial chromosome sequence of *Spiroplasma citri* reveals extensive viral invasion and important gene decay. *Appl Environ Microb* 2010;**76**:3420–6.

Chang CJ, Chen TA. Nutritional requirements of two flower spiroplasmas and honeybee spiroplasma. *J Bacteriol* 1983;**153**:452–7.

Chang TH, Lo WS, Ku C, *et al*. Molecular evolution of the substrate utilization strategies and putative virulence factors in mosquito-associated *Spiroplasma* species. *Genome Biol Evol* 2014;**6**:500–9.

Chen Z, Yang H, Pavletich NP. Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature* 2008;**453**:489–94.

Clark TB. *Spiroplasma* sp., a new pathogen in honeybees. *J Invertebr Pathol* 1977;**29**:112–3.

Cockburn SN, Haselkorn TS, Hamilton PT, *et al*. Dynamics of the continent-wide spread of a *Drosophila* defensive symbiont. *Ecol Lett* 2013;**16**:609–16.

Counce SJ, Poulson DF. Developmental effects of the sex-ratio agent in embryos of *Drosophila willistoni*. *J Exp Zool* 1962;**151**:17–31.

Davis RE, Lee IM, Worley JF. *Spiroplasma floricola*, a new species isolated from surfaces of flowers of the tulip tree, *Liriodendron tulipifera* L. *Int J Syst Bacteriol* 1981;**31**:456–64.

Duret S, Batailler B, Dubrana MP, *et al*. Invasion of insect cells by *Spiroplasma citri* involves spiralin relocalization and lectin/glycoconjugate-type interactions. *Cell Microbiol* 2014;**16**: 1119–32.

Duret S, Berho N, Danet JL, *et al*. Spiralin is not essential for helicity, motility, or pathogenicity but is required for efficient transmission of *Spiroplasma citri* by its leafhopper vector *Circulifer haematoceps*. *Appl Environ Microb* 2003;**69**:6225–34.

Ebbert MA, Nault LR. Improved overwintering ability in *Dalbulus maidis* (Homoptera: Cicadellidae) vectors infected with *Spiroplasma kunkelii* (Mycoplasmatales: Spiroplasmataceae). *Environ Entomol* 1994;**23**:634–44.

Foissac X, Saillard C, Gandar J, *et al*. Spiralin polymorphism in strains of *Spiroplasma citri* is not due to differences in post-translational palmitoylation. *J. Bacteriol*. 1996;**178**:2934–40.

Gasparich GE. Spiroplasmas: evolution, adaptation and diversity. *Front Biosci* 2002;**7**:d619–40.

Gasparich GE. Spiroplasmas and phytoplasmas: microbes associated with plant hosts. *Biologicals* 2010;**38**:193–203.

Gasparich GE, Hackett KJ. Characterization of a cryptic extrachromosomal element isolated from the mollicute *Spiroplasma taiwanense*. *Plasmid* 1994;**32**:342–3

Gaurivaud P, Laigret F, Garnier M, *et al*. Fructose utilization and pathogenicity of *Spiroplasma citri* characterization of the fructose operon. *Gene* 2000;**252**:61–9.

Goodacre SL, Martin OY, Thomas CG, *et al*. *Wolbachia* and other endosymbiont infections in spiders. *Mol Ecol* 2006;**15**:517–27.

Hackett KJ, Clark TB. Ecology of spiroplasmas. In: Whitcomb RF, Tully JGRF (eds). *The Mycoplasmas*, Vol. **5**. New York, NY: Academic Press, 1989,113–200.

Hackett KJ, Whitcomb RF, Henegar RB, *et al*. Mollicute diversity in arthropod hosts. *Zbl Bakt* 1990;**20**:441–54.

Harumoto T, Anbutsu H, Fukatsu T. Male-killing *Spiroplasma* induces sex-specific cell death via host apoptotic pathway. *PLoS Pathog* 2014;**10**:e1003956.

Herren JK, Lemaitre B. *Spiroplasma* and host immunity: activation of humoral immune responses increases endosymbiont load and susceptibility to certain Gram-negative bacterial pathogens in *Drosophila melanogaster*. *Cell Microbiol* 2011;**13**:1385–96.

Herren JK, Paredes JC, Schüpfer F, *et al*. Insect endosymbiont proliferation is limited by lipid availability. *Elife* 2014;**3**:e02964.

Humphery-Smith I, Grulet O, Le Goff F, *et al*. Spiroplasma (Mollicutes: Spiroplasmataceae) pathogenic for *Aedes aegypti* and *Anopheles stephensi* (Diptera: Culicidae). *J Med Entomol* 1991;**28**:219–22.

Jaenike J, Unckless R, Cockburn SN, *et al*. Adaptation via symbiosis: recent spread of a Drosophila defensive symbiont. *Science* 2010;**329**:212–5.

Jiggins FM, Hurst GDD, Jiggins CD, *et al*. The butterfly *Danaus chrysippus* is infected by a male-killing *Spiroplasma* bacterium. *Parasitology* 2000;**120**:439–46.

Kageyama D, Anbutsu H, Watada M, *et al*. Prevalence of a non-male-killing spiroplasma in natural populations of *Drosophila hydei*. *Appl Environ Microbiol* 2006;**72**:6667–73.

Khanchezar A, Béven L, Izadpanah K, *et al*. Spiralin Diversity Within Iranian Strains of *Spiroplasma citri*. *Current microbiology* 2014;**68**(1):96–104.

Killiny N, Batailler B, Foissac X, *et al.* Identification of a *Spiroplasma citri* hydrophilic protein associated with insect transmissibility. *Microbiology* 2006;**152**:1221–30.

Killiny N, Castroviejo M, Saillard C. *Spiroplasma citri* spiralin acts in vitro as a lectin binding to glycoproteins from its insect vector *Circulifer haematoceps*. *Phytopathology* 2005;**95**: 541–8.

Konai M, Clark EA, Camp M, *et al.* Temperature ranges, growth optima, and growth rates of *Spiroplasma* (Spiroplasmataceae, class Mollicutes) species. *Curr Microbiol* 1996;**32**:314–9.

Ku C, Lo WS, Chen LL, *et al.* Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in *Spiroplasma*. *Genome Biol Evol* 2013;**5**:1151–64.

Kwon MO, Wayadande AC, Fletcher J. *Spiroplasma citri* movement into the intestines and salivary glands of its leafhopper vector, *Circulifer tenellus*. *Phytopathology* 1999;**89**: 1144–51.

Lo WS, Chen LL, Chung WC, *et al.* Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. *BMC Genomics* 2013a;**14**:22.

Lo WS, Ku C, Chen LL, *et al.* Comparison of metabolic capacities and inference of gene content evolution in mosquito-associated *Spiroplasma diminutum* and *Spiroplasma taiwanense*. *Genome Biol Evol* 2013b;**5**:1512–23.

Lukasik P, Guo H, Asch M, *et al.* Protection against a fungal pathogen conferred by the aphid facultative endosymbionts *Rickettsia* and *Spiroplasma* is expressed in multiple host genotypes and species and is not influenced by coinfection with another symbiont. *J Evolution Biol* 2013;**26**:2654–61.

Marais A, Bove JM, Renaudin J. Characterization of the recA gene regions of *Spiroplasma citri* and *Spiroplasma melliferum*. *J Bacteriol* 1996;**178**:7003–9.

Martin J, Chong T, Ferree PM. Male-killing Spiroplasma preferentially disrupts neural development in the *Drosophila melanogaster* embryo. *PloS One* 2013;**8**:e79368

Meng Q, Ou J, Ji H, *et al.* Identification and characterization of spiralin-like protein *SLP25* from *Spiroplasma eriocheiris*. *Veterinary microbiology* 2010;**144**(3):473–77.

Merzendorfer H, Zimoch L. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *J Exp Biol* 2003;**206**:4393–412

Moore PCL, Lindsay JA. Genetic variation among hospital isolates of methicillin-sensitive *Staphylococcus aureus*: evidence for horizontal transfer of virulence genes. *J Clin Microbiol* 2001;**39**:2760–7.

Nunan LM, Lightner DV, Oduori MA, *et al. Spiroplasma penaei* sp. nov., associated with mortalities in *Penaeus vannamei*, Pacific white shrimp. *Int J Syst Evol Micr* 2005;**55**:2317–22.

Nunan LM, Pantoja CR, Salazar M, *et al.* Characterization and molecular methods for detection of a novel spiroplasma pathogenic to *Penaeus vannamei*. *Dis Aquat Organ* 2004;**62**:255–64.

Özbek E, Miller SA, Meulia T, *et al.* Infection and replication sites of *Spiroplasma kunkelii* (Class: Mollicutes) in midgut and Malpighian tubules of the leafhopper *Dalbulus maidis*. *J Invertebr Pathol* 2003;**82**:167–75.

Petzel JP, Hartman PA. Aromatic amino acid biosynthesis and carbohydrate catabolism in strictly anaerobic mollicutes (*Anaeroplasma* spp.). *Syst Appl Microbiol* 1990;**13**:240–7.

Phillips RN, Humphery-Smith I. The histopathology of experimentally induced infections of *Spiroplasma taiwanense* (class: Mollicutes) in *Anopheles stephensi* mosquitoes. *J Invertebr Pathol* 1995;**66**:185–95.

Pool JE, Wong A, Aquadro CF. Finding of male-killing *Spiroplasma* infecting *Drosophila melanogaster* in Africa implies transatlantic migration of this endosymbiont. *Heredity* 2006;**97**:27–32.

Qin QL, Xie BB, Zhang XY, *et al.* A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* 2014;**196**:2210–5.

Rakonjac J. Filamentous Bacteriophages: Biology and Applications. eLS. *Curr Issues Mol Biol* 2012;**13**:51–76.

Regassa LB, Gasparich GE. Spiroplasmas: evolutionary relationships and biodiversity. *Front Biosci* 2006;**11**:2983–3002.

Rohdich F, Hecht S, Gärtner K, *et al.* Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. *P Natl Acad Sci USA* 2002;**99**: 1158–63.

Saillard C, Carle P, Duret-Nurbel S, *et al.* The abundant extrachromosomal DNA content of the *Spiroplasma citri* GII3–3X genome. *BMC Genomics* 2008;**9**:195.

Saillard C, Vignault JC, Bové JM, *et al. Spiroplasma phoeniceum* sp. nov., a new plant-pathogenic species from Syria. *Int J Syst Bacteriol* 1987;**37**:106–15.

Segata N, Börnigen D, Morgan XC, *et al.* PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* 2013;**4**:2304.

Tabata J, Hattori Y, Sakamoto H, *et al.* Male killing and incomplete inheritance of a novel *Spiroplasma* in the moth *Ostrinia zaguliaevi*. *Microb Ecol* 2011;**61**:254–63.

Tinsley MC, Majerus MEN. A new male-killing parasitism: *Spiroplasma* bacteria infect the ladybird beetle *Anisosticta novemdecimpunctata* (Coleoptera: Coccinellidae). *Parasitology* 2006;**132**:757–65.

Wang W, Chen J, Du K, *et al.* Morphology of spiroplasmas in the Chinese mitten crab *Eriocheir sinensis* associated with tremor disease. *Res Microbiol* 2004;**155**:630–5.

Watanabe M, Yukuhiro F, Maeda T, *et al.* Novel strain of *Spiroplasma* found in flower bugs of the genus *Orius* (Hemiptera: Anthocoridae): transovarial transmission, coexistence with *Wolbachia* and varied population density. *Microb Ecol* 2014;**67**:219–28.

Wedincamp J, Jr, French FE, Whitcomb RF, *et al.* Spiroplasmas and Entomoplasmas (Procaryotae: Mollicutes) associated with tabanids (Diptera: Tabanidae) and fireflies (Coleoptera: Lampyridae). *J Invertebr Pathol* 1996;**68**:183–6.

Whitcomb RF, French FE, Tully JG, *et al. Spiroplasma chrysopicola* sp. nov., *Spiroplasma gladiatoris* sp. nov., *Spiroplasma helicoides* sp. nov., and *Spiroplasma tabanidicola* sp. nov., from Tabanid (Diptera: Tabanidae) flies. *Int J Syst Bacteriol* 1997;**47**: 713–9.

Whitcomb RF, Gasparich GE, French FE, *et al. Spiroplasma syrphidicola* sp. nov., from a Syrphid Fly (Diptera: Syrphidae). *Int J Syst Bacteriol* 1996;**46**:797–801.

Williamson DL, Poulson DF. Sex ratio organisms (spiroplasmas) of *Drosophila*. *The Mycoplasmas* 1979;**3**:175–208.

Williamson DL, Sakaguchi B, Hackett KJ, *et al. Spiroplasma poulsonii* sp. nov., a new species associated with male-lethality in *Drosophila willistoni*, a neotropical species of fruit fly. *Int J Syst Bacteriol* 1999;**49**:611–8.

Williamson DL, Tully JG, Rosen L, *et al. Spiroplasma diminutum* sp. nov., from *Culex annulus* mosquitoes collected in Taiwan. *Int J Syst Bacteriol* 1996;**46**:229–33.

Wróblewski H, Johansson KE, Hjérten S, *et al.* Purification and characterization of spiralin, the main protein of the *Spiroplasma citri* membrane. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1977;**465**(2):275–89.

Xie J, Vilchez I, Mateos M. Spiroplasma bacteria enhance survival of *Drosophila hydei* attacked by the parasitic wasp *Leptopilina heterotoma*. *PLoS One* 2010;**5**:e12149.

Yarza P, Ludwig W, Euzéby J, *et al*. Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. *Syst and Appl Microbiol* 2010;**33**:291–9.

Yarza P, Richter M, Peplies J, *et al*. The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* 2008;**31**:241–50.

Ye F, Melcher U, Fletcher J. Molecular characterization of a gene encoding a membrane protein of *Spiroplasma citri*. *Gene* 1997;**189**:95–100.

Yu J, Wayadande AC, Fletcher J. *Spiroplasma citri* surface protein P89 implicated in adhesion to cells of the vector *Circulifer tenellus*. *Phytopathology* 2000;**90**(7):716–722.

Zhao Y, Hammond RW, Jomantiene R, *et al*. Gene content and organization of an 85-kb DNA segment from the genome of the phytopathogenic mollicute *Spiroplasma kunkelii*. *Mol Gen Genomics* 2003;**269**:592–602.

Conclusiones

En el estudio sobre la microbiota de las mariposas monarca encontramos comunidades bacterianas que presentan una diversidad limitada. Una única especie bacteriana del género *Commensalibacter* es dominante en los intestinos de las mariposas monarcas.

El genoma de bajo contenido de G+C de *Commensalibacter papalotli* MX01 se logró obtener gracias a que se pudo cultivar a la bacteria en el laboratorio. El genoma de la cepa MX01 es el más pequeño entre la familia de las acetobacterias. Las propiedades del genoma de *C. papalotli* podrían ser indicativas de que están ocurriendo procesos de especialización para adaptarse a las condiciones de los intestinos de las mariposas.

Es probable que *C. papalotli* tenga un papel importante en la defensa contra patógenos mediante la producción de ácidos orgánicos cuando se encuentran en estado de dormancia, además de participar en la asimilación de nutrientes para las mariposas mediante procesos fermentativos. Estas conclusiones quedan pendientes por analizarse.

Las mariposas monarca tienen el potencial de convertirse en una especie atractiva para estudiar interacciones moleculares entre bacterias e insectos debido a la baja diversidad que presenta su microbiota, al cultivo de sus simbiontes principales y debido a la disponibilidad del genoma parcial de las mariposas (Zhan et al., 2011).

Del análisis de genomas de *Wolbachia* podemos concluir que los análisis filogenómicos y de comparación de secuencias genómicas resolvieron la taxonomía de los supergrupos principales. Los resultados obtenidos indican que los supergrupos de *Wolbachia* representan linajes evolutivos diferentes. Además se encontró evidencia de que algunos supergrupos contienen suficiente heterogeneidad interna para ser considerados como agrupaciones de más de una especie. Una de las contribuciones más relevantes de este estudio fue la designación de especies candidatas para distinguir a los linajes de *Wolbachia*.

De análisis comparativo de *Spiroplasma* podemos concluir que los genomas de este género bacteriano se caracterizan por ser reducidos y por presentar secuencias repetidas relacionadas a plectrovirus. Los proyectos de secuenciación de genomas de *Spiroplasma* han revelado información sobre sus potenciales metabólicos, factores de virulencia, utilización de sustratos, la dinámica de sus genomas y las interacciones simbióticas con sus hospederos. Los estudios filogenómicos y de genómica comparativa ayudarán a comprender mejor las interacciones simbióticas y la taxonomía del género *Spiroplasma*.

Perspectivas

Los resultados sobre la diversidad de la microbiota intestinal de las mariposas monarca se pretenden publicar basados en los resultados de las librería de clonas de genes ribosomales y en el metagenoma.

Se desea realizar una caracterización detallada sobre las propiedades fenotípicas y metabólicas de *C. papalotli* para publicar su descripción formal. También se desean intentar ensayos de inoculación de *C. papalotli* en líneas de *Drosophila* para analizar fenotipos ya que no se pueden mantener mariposas monarca en el laboratorio.

Es necesario hacer un estudio de genómica comparativa entre los genomas de *Commensalibacter* y de otras acetobacterias secuenciadas de vida libre. Los análisis comparativos podrán ayudar a identificar a los genes que se han conservado y a los que se han perdido durante la evolución de *Commensalibacter.*

También es necesario realizar un análisis más detallado para identificar las diferencias entre los potenciales metabólicos de ambos genomas de *Commensalibacter* que pudieran ayudar a identificar adaptaciones de acuerdo a las condiciones específicas de la microbiota intestinal de las moscas *Drosophila* y de la mariposa monarca.

Se planea la secuenciación de un segundo metagenoma intestinal de las mariposas monarca utilizando técnicas de secuenciación más avanzadas que permitan generar ensambles de

genomas bacterianos con una mejor resolución. Además se planea realizar un estudio transcriptómico para analizar con más detalle las posibles funciones que tiene la microbiota intestinal de las mariposas monarca.

Finalmente se desea analizar la diversidad del microbiota de mariposas monarca de otras regiones geográficas y en diferentes etapas de su migración anual para poder hacer comparaciones a un nivel más detallado.

Los análisis sobre el potencial metabólico de los genomas de *Wolbachia* y de otras bacterias simbióticas recuperados del metagenoma de *Dactylopius* serán reportados por Shamayim Tabita Ramirez-Puebla en su tesis doctoral.

La microbiota de especies de alacranes nativos de Morelos ha sido reportada por Luis Bolaños (Bolaños et al., 2015) y será parte de su tesis doctoral.

Bibliografía

Bolaños LM, Rosenblueth M, Castillo-Ramírez S, Figuier-Huttin G, Martínez-Romero E. 2015. Species-specific diversity of novel bacterial lineages and differential abundance of predicted pathways for toxic compound degradation in scorpion gut microbiota. Environ. Microbiol. doi: 10.1111/1462-2920.12939.

Broderick NA, Raffa KF, Goodman RM, Handelsman J. 2004. Census of the bacterial community of the gypsy moth larval midgut by using culturing and culture-independent methods. Appl. Environ. Microbiol. 70: 293–300.

Broderick NA, Lemaitre B. 2012. Gut-associated microbes of *Drosophila melanogaster*. Gut Microbes. 3: 307–321.

Colman DR, Toolson EC, Takacs-Vesbach CD. 2012. Do diet and taxonomy influence insect gut bacterial communities?. Mol. Ecol. 21: 5124–5137.

Engel P, Moran NA. 2013. The gut microbiota of insects - diversity in structure and function. FEMS Microbiol. Rev. 37: 699–735.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59: 307–321.

Jing X, Wong AC, Chaston JM, Colvin J, McKenzie CL, Douglas AE. 2014. The bacterial communities in plant phloem-sap-feeding insects. Mol. Ecol. 23: 1433–1444.

Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. 2012. Diversity, stability and resilience of the human gut microbiota. Nature. 489: 220–230.

Martínez-Cano DJ, Reyes-Prieto M, Martínez-Romero E, Partida-Martínez LP, Latorre A, Moya A, Delaye L. 2015. Evolution of small prokaryotic genomes. Front. Microbiol. 5: 742.

Parsley LC, Newman MM, Liles MR. 2010. Fluorescence in situ hybridization of bacterial cell suspensions. Cold Spring Harb. Protoc. 2010(9): pdb.prot5493.

Pérez-Cobas AE, Maiques E, Angelova A, Carrasco P, Moya A, Latorre A. 2015. Diet shapes the gut microbiota of the omnivorous cockroach *Blattella germanica*. FEMS Microbiol. Ecol. 91: fiv022.

Pinto-Tomás AA, Anderson MA, Suen G, Stevenson DM, Chu FS, Cleland WW, Weimer PJ, Currie CR. 2009. Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. Science. 326: 1120–1123.

Ramírez-Puebla ST, Rosenblueth M, Chávez-Moreno CK, de Lyra MC, Tecante A, Martínez-Romero E. 2010. Molecular phylogeny of the genus *Dactylopius* (Hemiptera: *Dactylopiidae*) and identification of the symbiotic bacteria. Environ. Entomol. 39: 1178–1183.

Robinson CJ, Schloss P, Ramos Y, Raffa K, Handelsman J. 2010. Robustness of the bacterial community in the cabbage white butterfly larval midgut. Microb. Ecol. 59: 199–211.

Roh SW, Nam YD, Chang HW, Kim KH, Kim MS, Ryu JH, Kim SH, Lee WJ, Bae JW. 2008. Phylogenetic characterization of two novel commensal bacteria involved with innate immune homeostasis in *Drosophila melanogaster*. Appl. Environ. Microbiol. 74: 6171–6177.

Rosas-Pérez T, Rosenblueth M, Rincón-Rosales R, Mora J, Martínez-Romero E. 2014. Genome sequence of "*Candidatus* Walczuchella monophlebidarum" the

flavobacterial endosymbiont of *Llaveia axin axin* (Hemiptera: *Coccoidea*: *Monophlebidae*). Genome Biol. Evol. 6: 714–726.

Rosenblueth M, Sayavedra L, Sámano-Sánchez H, Roth A, Martínez-Romero E. 2012. Evolutionary relationships of flavobacterial and enterobacterial endosymbionts with their scale insect hosts (Hemiptera: *Coccoidea*). J. Evol. Biol. 25: 2357–2368.

Ryu JH, Kim SH, Lee HY, Bai JY, Nam YD, Bae JW, Lee DG, Shin SC, Ha EM, Lee WJ. 2008. Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in *Drosophila*. Science. 319: 777–782.

Yun JH, Roh SW, Whon TW, Jung MJ, Kim MS, Park DS, Yoon C, Nam YD, Kim YJ, Choi JH, Kim JY, Shin NR, Kim SH, Lee WJ, Bae JW. 2014. Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. Appl. Environ. Microbiol. 80: 5254–5264.

Zhan S, Merlin C, Boore JL, Reppert SM. 2011. The monarch butterfly genome yields insights into long-distance migration. Cell. 147: 1171–1185.

Apéndice I - Resumen Capítulo I

Diversidad genómica de microorganismos extremófilos de Los Azufres

El presente estudio analizó la diversidad microbiana que reside en el campo geotérmico de Los Azufres usando técnicas metagenómicas. La diversidad genómica de los microorganismos de Los Azufres no se había estudiado hasta ahora. Se emplearon técnicas de secuenciación de ADN ambiental y análisis bioinformáticos para revelar la diversidad microbiana y para generar ensambles de genomas de microorganismos abundantes. En el metagenoma de una solfatara ácida se identificó a una comunidad integrada por arqueas termoacidófilas. La población más abundante corresponde a un género candidato del orden *Sulfolobales*. Asimismo se descubrieron poblaciones de arqueas anaeróbicas del orden *Thermoproteales*. Con el metagenoma se reconstruyeron los genomas de estas poblaciones de arqueas. También se recuperaron los genomas de los virus SMR1 y SMF1. Se descubrió que los sedimentos circundantes a la solfatara ácida contienen además a una cepa nueva de *Sulfolobus acidocaldarius*. Por otra parte, en un metagenoma de sedimentos termales se identificó a una comunidad integrada en su mayoría por una microalga roja de la familia *Cyanidiaceae* y por arqueas conocidas como ABC-plasmas del orden *Thermoplasmatales*. Las secuencias metagenómicas permitieron el ensamble de los genomas de los organelos de la microalga y el ensamble del genoma de una arquea enigmática del filo *Parvarchaeota*. También se identificaron secuencias metagenómicas que corresponden a linajes nuevos de actinobacterias y de otras bacterias de los filos *Chloroflexi* y *Thermotogae*. Finalmente se identificó que las comunidades microbianas de una laguna ácida están compuestas en su mayoría por una población de una microalga verde de la clase *Trebouxiophyceae*. El genoma nuclear de la microalga verde se obtuvo parcialmente del metagenoma de la laguna ácida y los genomas de sus organelos se reconstruyeron por completo. En el metagenoma de la laguna ácida también se identificaron secuencias de bacterias acidófilas. Además se obtuvo de células en cultivo, el primer genoma de bacterias del género *Acidocella*. En conclusión, las manifestaciones termales de Los Azufres contienen comunidades microbianas novedosas con diversidad limitada. Los microorganismos identificados poseen atributos funcionales potenciales que son consistentes con las condiciones geoquímicas.

Apéndice II - Resumen Capítulo II

Diversidad genómica de rizobios de *Phaseolus* nativos de México

En el presente trabajo se analizó la diversidad genética de rizobios de especies representativas del género *Phaseolus*. *Phaseolus vulgaris* ha sido por muchos años la especie modelo para estudiar interacciones moleculares entre *Rhizobium* y *Phaseolus.* La diversidad genética y las interacciones moleculares de simbiontes de otras especies de *Phaseolus* silvestres han recibido menor atención. Mediante el cultivo y la identificación molecular de bacterias de los nódulos de especies de *Phaseolus* pertenecientes a distintos grupos filogenéticos se identificó que la mayoría de las especies nodulan preferentemente con cepas de *Bradyrhizobium*. Además se propuso un reemplazo de simbiontes de *Bradyrhizobium* a *Rhizobium* en un grupo filogenético de *Phaseolus*. El reemplazo de simbiontes pudo haber sido facilitado por la transferencia horizontal de los genes *nodZ* y *nolL* de *Bradyrhizobium* a *Rhizobium*. Se encontró que la diversidad de estos genes en *Rhizobium* es limitada y se sugirió su transferencia de *Bradyrhizobium* al plásmido simbiótico del simbiovar phaseoli de *Rhizobium etli* y de otros simbiontes de *P. vulgaris*. También se obtuvieron cultivos de rizobios aislados de plantas de *Phaseolus* silvestres para realizar la secuenciación de sus genomas con la finalidad de comprender mejor las bases genéticas de su potencial simbiótico. A partir de nódulos de *Phaseolus albescens* se recuperó a la cepa CCGE510 que representa a un nuevo linaje de *Rhizobium*. Las relaciones evolutivas de la cepa CCGE510 se clarificaron a partir de análisis filogenómicos. También se obtuvo la secuencia genómica de la cepa *Bradyrhizobium* sp. CCGE-LA001 que se aisló de nódulos de *Phaseolus microcarpus.* El genoma de la cepa CCGE-LA01 es el primero que se secuencia para un *Bradyrhizobium* que establece simbiosis con *Phaseolus.* Este genoma reveló un repertorio amplio de genes simbióticos con secuencias novedosas. Finalmente se realizó una revisión taxonómica de la familia *Rhizobiaceae* basada en métricas y parámetros derivados de comparaciones de secuencias genómicas. Se pudieron identificar dos superclados dentro de la familia *Rhizobiaceae* que corresponden a los géneros *Rhizobium*/*Agrobacterium* y *Shinella*/*Ensifer*. En conclusión, los resultados obtenidos ayudaron a comprender mejor los mecanismos de especificidad simbiótica y las bases genéticas de la interacción entre rizobios y *Phaseolus*. Por otro lado, se comprobó que las técnicas de filogenómica y de comparación de secuencias genómicas son útiles para resolver las relaciones evolutivas de los diferentes linajes de rizobios.

Apéndice III - Resumen Capítulo III

Diversidad genómica de simbiontes de insectos y artrópodos nativos de México

En este trabajo se planteó descubrir la diversidad microbiana de insectos nativos de México mediante enfoques genómicos y metagenómicos. Uno de las metas fue analizar la microbiota de las mariposas monarca que habitan en los bosques de oyamel del occidente de México durante el invierno. La microbiota de las mariposas monarca no se había analizado utilizando técnicas moleculares. Mediante censos de diversidad microbiana se encontró que la microbiota de las mariposas monarca es de baja complejidad y que contiene una población abundante de bacterias del género *Commensalibacter*. Mediante técnicas de cultivo fue posible aislar a la cepa *Commensalibacter* sp. MX01 que representa a una especie nueva que fue nombrada como *Commensalibacter papalotli*. Se encontró que el genoma de la cepa MX01 presenta bajo contenido de G+C y que representa el genoma más pequeño entre la familia de las acetobacterias. Las propiedades del genoma de la cepa MX01 podrían ser indicativas de que están ocurriendo procesos de especialización para adaptarse a los micro nichos de los intestinos de las mariposas. Concluimos que las mariposas monarca tienen el potencial de convertirse en una especie atractiva para estudiar interacciones moleculares entre bacterias e insectos debido a la baja diversidad que presenta su microbiota, al cultivo de sus simbiontes principales y debido a la disponibilidad del genoma de las mariposas. Por otro lado, mediante enfoques metagenómicos se pudieron reconstruir genomas de bacterias no cultivables del género *Wolbachia* de las cochinillas del carmín que pertenecen a los supergrupos A y B. En una revisión taxonómica del género *Wolbachia* se encontró que los supergrupos representan a diferentes linajes evolutivos. También se realizó una revisión de las características genómicas y de las relaciones evolutivas del género *Spiroplasma* que incluye bacterias que son habitantes comunes de los tejidos de insectos y artrópodos. En conclusión, Las técnicas de filogenómica y de comparación de secuencias genómicas son útiles para resolver las relaciones evolutivas de bacterias simbióticas de insectos y artrópodos. Finalmente, los estudios metagenómicos permiten analizar la composición y el potencial genético de las microbiotas de insectos y artrópodos sobrepasando las limitantes de cultivo de sus simbiontes.

Apéndice IV - Técnicas de PCR

Se presentan los programas de PCR utilizados para la amplificación de genes ribosomales y de otros marcadores moleculares. Todas las reacciones de PCR se hicieron de manera independiente para cada pareja de oligonucleótidos utilizados.

Los oligonucleótidos utilizados para la amplificación de genes ribosomales para cada dominio y de otros marcadores moleculares se presentan en la Tabla 2 del Capítulo 1.

Concentración de soluciones para una reacción de PCR de 20 µl.

oligonucleótido 1	0.2 µl
oligonucleótido 2	0.2 µl
solución de Mg	0.6 µl
solución de buffer	2.0 µl
solución de dNTPs	0.16 µl
Taq polimerasa	0.1 µl
agua esterilizada	16.74 µl

Programa de PCR para la amplificación de genes ribosomales 16S rRNA de bacterias

1 ciclo	94 °C	3 minutos
35 ciclos	94 °C	45 segundos
	57 °C	1 minuto
	72 °C	2 minutos
1 ciclo	72 °C	7 minutos

Programa de PCR para la amplificación de genes ribosomales 16S rRNA de arqueas

1 ciclo	94 °C	3 minutos
35 ciclos	94 °C	1 minuto
	55 °C	2 minuto
	72 °C	3 minutos
1 ciclo	72 °C	7 minutos

Programa de PCR para la amplificación de genes ribosomales 16S rRNA de eucariotas

1 ciclo	94 °C	3 minutos
35 ciclos	94 °C	1 minuto
	52 °C	2 minuto
	72 °C	3 minutos
1 ciclo	72 °C	10 minutos

Los programas de PCR de otros marcadores moleculares (arsenito oxidasa, ITS bacterianos y genes de nitrogenasa) fueron utilizando las condiciones para la amplificación de genes ribosomales 16S rRNA de bacterias. Otros oligonucleótidos y condiciones de PCR especiales se describen en los artículos publicados.