



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE DOCTORADO EN CIENCIAS BIOMÉDICAS
CENTRO DE CIENCIAS GENÓMICAS

**Diversidad genética de las bacterias asociadas al
tracto digestivo de alacranes de las especies
Centruroides limpidus limpidus y *Vaejovis smithi***

QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS

PRESENTA:
LUIS MANUEL BOLAÑOS AVELLANEDA

DIRECTORA DE TESIS:
DRA. ESPERANZA MARTINEZ ROMERO
Centro de Ciencias Genómicas

COMITÉ TUTORAL:
DR. JUAN MIRANDA RÍOS
Instituto de Investigaciones Biomédicas
DR. ALEJANDRO ALAGÓN CANO
Instituto de Biotecnología

CUERNAVACA, MOR. NOVIEMBRE 2015



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

La presente tesis se realizó en el laboratorio de Microbiología Ambiental y Simbiótica del Programa de Ecología Genómica del Centro de Ciencias Genómicas de la UNAM.

Con el apoyo de la beca de manutención nivel doctorado del Consejo Nacional de Ciencia y Tecnología.

Agradecimientos

En primer lugar quiero agradecer a mi familia, quienes a través de todo mi proyecto de vida nunca han dejado de creer en mi y apoyarme incondicionalmente.

A mis padres: Martha Avellaneda y José Luis Bolaños por hacer de una ilusión una realidad a través de su amor y su esfuerzo. Sin dejar de creer en la realización de ésta en ningún momento.

A mi hermana Carmen Elisa por caminar siempre de la mano conmigo ante un sin fin de aventuras en las que siempre hemos compartido los diferentes sabores de la vida y en todo momento ha enriquecido el camino.

A mi tía Carmen Beatriz por ser el elocuente ejemplo, siempre cómplice de mi formación.

A mi familia adoptiva: Marcela, Gustavo, Oti y Gustavo Mendoza, que ante cualquier situación siempre han sido un soporte que me ha motivado a seguir adelante.

A mi familia colombiana que me enseñó que la distancia no se mide en kilómetros y siempre han estado cerca de mí.

A esos personajes y sus respectivas familias, que a través de los años se han convertido en mis hermanos: Oscar, Jorge, Francisco, “Muerto”, Sofía, Adolfo, Moi, Fer de Puebla, Alejandro M., Toño y Aldo.

A Violeta por el amor tan grande y a su familia por el gran apoyo en todo momento y en todas las formas.

Como todo proyecto académico, el desarrollo inherente de éste no podría ser del todo exitoso sin el apoyo de un conjunto de personas. Por lo que agradezco en primer lugar a mi tutora Esperanza Martínez por enseñarme a gozar la ciencia, compartir sus ideas y siempre motivarme a desarrollar las mías.

Al comité tutorial: Alejandro Alagón y Juan Miranda. Quienes semestre a semestre acompañaban y aportaban al crecimiento del proyecto.

A todos los compañeros, maestros y personal administrativo del PDCB y del CCG que hicieron de estos años un período de vida inolvidable y sobretodo divertido.

A todo el laboratorio de Ecología Genómica, que a través de sus aportaciones, ayuda y enseñanzas hicieron que el proyecto llegara a buen puerto. Un agradecimiento especial a Mónica Rosenblueth que ha sido co-tutora de este proyecto.

A los recolectores de alacranes, que voluntariamente aportaban material y siguieron de cerca el desarrollo de este trabajo Ale Zayas, Orlando y Leif Eriksson. Y en general a todo aquel que con entusiasmo recolectó y aportó alacranes.

A la Universidad Nacional Autónoma de México

Al Consejo Nacional de Ciencia y Tecnología

A la sociedad mexicana por construir y mantener estas instituciones

A todos los que de alguna u otra manera me han acompañado y colaborado a través de diferentes etapas de mi vida.

Índice

RESUMEN	5
ABSTRACT	6
INTRODUCCIÓN.....	7
HIPÓTESIS.....	12
OBJETIVO GENERAL.....	12
OBJETIVOS PARTICULARES	12
CAPÍTULO I.....	13
Análisis de la variación de la diversidad de bacterias mediante un enfoque no cultivable del tracto digestivo a través de diferentes tratamientos alimenticios: Colecta peridoméstica, dieta homogenizada con larvas de <i>Tenebrio molitor</i>) y privación de alimento	13
1.1 Colecta e identificación de alacranes	14
1.1.1 Identificación morfológica y genética de los alacranes colectados.....	14
1.1.2 Árboles filogenéticos basados en la región intergénica 12S-16S mitocondrial.....	15
1.2 Resultados: La especie de alacrán es un factor de mayor influencia que la dieta en la composición de la microbiota intestinal de los alacranes.....	16
1.2.1 Descripción de las librerías	16
1.2.2 Análisis de las librerías.....	18
1.3 Análisis de las comunidades intestinales en alacranes durante los primeros dos estadios.....	18
1.4 Conclusiones	20
CAPÍTULO II.....	21
<i>Mycoplasma</i> spp. y SG1 como posibles simbioses de alacranes.....	21
2.1 Análisis filogenéticos de los nuevos linajes relacionados al <i>phylum</i> Tenericutes asociados a alacranes	21
2.2 Revisión de los genomas secuenciados del género <i>Spiroplasma</i>	22
2.3 Conclusiones	23
CAPÍTULO III.....	24
Predicción metagenómica basada en los resultados de las librerías de 16S rRNA.	24
3.1 Introducción y fundamento de las predicciones funcionales	24
3.2.1 Resultados: Existencia de un grupo común y abundante de funciones metabólicas	25
3.2.2 Resultados: Abundancia significativa de vías metabólicas de degradación de compuestos tóxicos en los alacranes <i>C. limpidus</i> recién capturados.	26
3.3 Conclusiones	26
CAPÍTULO IV	27
Secuenciación y análisis metagenómico del intestino de alacranes <i>C. limpidus</i> y <i>V. smithi</i> en estado de privación de alimento.	27
4.1 Análisis del metagenoma del intestino de los alacranes	27

4.2 Comparación de los metagenomas con las predicciones basadas en las librerías de 16S rRNA	27
4.3. Microscopía de homogenizados del intestino de alacrán muestra una cantidad significativamente menor de bacterias en <i>C. limpidus</i> que en <i>V. smithi</i>	28
4.4 Conclusiones	28
DISCUSIÓN	29
PERSPECTIVAS	31
REFERENCIAS	32
ANEXOS	37

Anexo 1 .- Bolaños, L. M., Rosenblueth, M., Castillo-Ramírez, S., Figuier-Huttin, G. and 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. *Environmental Microbiology*. doi: 10.1111/1462-2920.12939 *In press*.

Anexo 2.- Bolaños, L. M., Servín-Garcidueñas, L. E., & Martínez-Romero, E. (2015). Arthropod–Spiroplasma relationship in the genomic era. *FEMS microbiology ecology*, 91(2), 1-8.

RESUMEN

Los alacranes son considerados como “fósiles vivientes” que han conservado características anatómicas ancestrales, se han adaptado a numerosos hábitats y cuentan con características fisiológicas que los hacen únicos dentro de los animales. Sin embargo, la diversidad de su microbiota intestinal y las posibles funciones de ésta no se han estudiado. El objetivo de este estudio ha sido caracterizar la microbiota bacteriana de dos especies de alacrán, *Vaejovis smithi* y *Centruroides limpidus*, con técnicas moleculares independientes de cultivo. Para analizar la diversidad de las comunidades bacterianas a través de diferentes tratamientos alimenticios, se realizaron librerías de genes 16S rRNA. Se encontró que el factor de la especie influye en mayor grado que el de la dieta en la composición de la microbiota intestinal, inclusive se podría definir en que estas comunidades en conjunto son específicas de especie de alacrán. La privación de alimento disminuyó la diversidad en la comunidad bacteriana. Además, los análisis filogenéticos del gen 16S rRNA revelaron linajes bacterianos nuevos, los cuales están restringidos a la especie de alacrán.

Para establecer las posibles funciones de estas comunidades bacterianas, se realizó una predicción bioinformática de funciones basada en los resultados de las librerías del gen 16S rRNA. Los perfiles funcionales revelaron un grupo común y abundante de vías metabólicas en las dos especies. Este grupo estaba conformado principalmente por vías

relacionadas al metabolismo de aminoácidos, carbohidratos y cofactores. Adicionalmente, las comparaciones entre los perfiles funcionales mostraron que las vías de degradación de compuestos tóxicos eran significativamente más abundantes en la microbiota de los alacranes recién capturados de la especie *C. limpidus*.

A manera de comprobación de las predicciones, se realizaron secuenciaciones metagenómicas de tipo “shotgun” de alacranes de las dos especies en estado de privación de alimento. En *V. smithi* aproximadamente el 50% de los genes predichos eran bacterianos y en *C. limpidus* fue menor del 1%, sugiriendo que la abundancia de bacterias o DNA bacteriano en relación con el hospedero es menor en *C. limpidus* que en *V. smithi*. Con el set de genes del microbioma de *V. smithi* se compararon las abundancias de vías metabólicas con la predicción funcional, se encontró que de las 41 vías comunes, solo 2 estaban subrepresentadas en el set “shotgun” y no hubo ninguna diferencia significativa en la comparación de abundancias en todas las vías predichas contra las de la secuenciación. Este estudio provee la primera aproximación al conocimiento de la microbiota bacteriana de los alacranes y de igual manera consolida un punto de referencia para estudios futuros en la microbiota intestinal de otras especies de arácnidos.

ABSTRACT

Scorpions are considered “living fossils” that have conserved ancestral anatomical features, have adapted to numerous habitats and possess unique physiological features among animals. However, their gut microbiota diversity and their possible function have not been studied. The aim of this study was to characterize the bacterial microbiota of two scorpion species, *Vaejovis smithi* and *Centruroides limpidus*, with non-cultivable techniques. The gut bacterial diversity was analyzed through different diet treatments creating 16S rRNA clone libraries. We found that scorpion species variable influences more bacterial composition than diet. Scorpion gut microbiota is species-specific and food deprivation reduces bacterial diversity. 16S rRNA gene phylogenetic analysis revealed novel bacterial lineages showing a low level of sequence identity to any known bacteria. Furthermore, these novel bacterial lineages were each one restricted to a scorpion species. To determine the possible functions of this bacterial communities, we generate a bioinformatics prediction of the functional profiles based on the 16S rRNA clone libraries results. We found that the predicted metagenomic profiles revealed a core set of pathways that were highly abundant in both species and mostly related to amino acid, carbohydrate,

and cofactor metabolism. Additionally, comparisons between predicted metagenomic profiles showed that toxic compound degradation pathways were more abundant in *Centruroides limpidus* recently captured scorpions.

To corroborate the functional predictions, we did metagenomic shotgun sequencing from food-deprived scorpions of both species. From the total gene-calling prediction, 50% of the genes were from bacterial origin in *V. smithi* and only <1% in *C. limpidus*, suggesting that bacterial load or bacterial/host DNA ratio were less in *C. limpidus* than in *V. smithi*. Bacterial genes from *V. smithi* microbiota were compared with the predicted functional profile. We found that only two pathways from the 41 of the total “core” set were underrepresented in the shotgun sequencing. And there was not a significant difference in the pathway abundance comparison between prediction and shotgun sequencing.

This study gives a first insight into the scorpion gut microbiota and provides a reference for future studies on the gut microbiota from other arachnid species.

INTRODUCCIÓN

Alacranes

Los alacranes son artrópodos pertenecientes a la clase Arachnida. Hasta la fecha se han descrito más de 2000 especies, distribuidas en todos los continentes - excepto en la Antártica - (Savory, 1977; Prendini, 2011; Rein, 2012) y han radiado a diferentes hábitats; desde áreas desérticas, bosques, sabanas, selvas e inclusive nevados con más de 5 500 m. de altitud (Polis, 1990).

Los alacranes son los arácnidos más antiguos que se han fechado, se originaron en el Silúrico hace aproximadamente 430 millones de años. (Dunlop, 2010). Además, resulta sorprendente que desde su origen, a través de su historia evolutiva sobre la tierra no han modificado gran parte de sus características morfológicas (Hoffman, 1993), por lo que se consideran fósiles vivientes.

En la actualidad los alacranes han sido objeto de estudio debido a la toxicidad de su veneno y el riesgo que representa hacia el humano, sin embargo los alacranes poseen otras características fisiológicas interesantes que resaltan sobre las de otros artrópodos. Una característica relevante para la colonización diferentes hábitats es la adaptación a condiciones extremas de temperatura, existen especies que pueden sobrevivir en temperaturas cercanas al congelamiento durante semanas y recuperarse en pocas horas. Las

especies *Hadrurus arizonensis* (Norte América), *Centruroides sculpturatus* (Norte América), *Leiurus quinquestriatus* (África) y *Buthotus minax* (África) que habitan en los desiertos, logran soportar temperaturas más altas que la mayoría de los otros artrópodos que ahí habitan (Hadley, 1974).

Los alacranes también tienen un uso muy eficiente del agua. (Hadley, 1970). Excretan productos de desecho nitrogenados insolubles, como guanina, ácido úrico y xantina. Son tan eficientes, que logran sobrevivir al obtener el agua de forma indirecta de su comida.

Dentro de las características sobresalientes de los alacranes se ha reportado la existencia del fenómeno de partenogénesis como forma de reproducción, la cual es asexual dado que los gametos de una hembra se desarrollan a embriones, sin la necesidad de que el macho la fertilice. Estos casos se han reportado en las especies *Tityus serrulatus* (Toscano-Gadea, 2004) y *Liocheles australasiae* (Yamazaki, *et al.*, 2004). Se ha demostrado que esta forma de reproducción asexual es inducida por la bacteria *Wolbachia* en avispas parasitarias de huevos, como *Telenomus Hawaii* (Arakaki, *et al.*, 2000) y *Leptopilina clavipes* (Pannebakker, 2005), las dos pertenecientes al orden Himenóptera. Este tipo de reproducción inducido por *Wolbachia* se ha reportado en el ácaro *Bryobia praetiosa*, perteneciente a los arácnidos (Weeks, *et al.*, 2001). Sobre la presencia de *Wolbachia* en alacranes, existen dos reportes; el primero en especies de alacranes del género *Opisththalmus* del sur de África (Baldo, *et al.*, 2007), en el cual las cepas encontradas pertenecen al clado F de *Wolbachia*. Las cepas de este clado se encuentran en artrópodos y nemátodos, incluyendo termitas (*Microcerotermes sp.* and *Kaloterme flavicollis*), nemátodos filariales (*Mansonella spp.*), un gorgojo (*Rhinocyllus conicus*) y dos grillos (*Orocharis saltator* y *Hapithus agitator*) (Panaram, *et al.*, 2007). El segundo caso reportado es en *Tityus serrulatus* (Buthidae), en el que encuentran una secuencia similar a las que están presentes en *Drosophila innubila* y Hymenopteros como *Nasonia longicornis* y *Nasonia giraulti* (Suesdek-Rocha *et al.*, 2006).

Como un antecedente directo a este estudio está la búsqueda de endosimbiontes en alacranes de la familia Vaejovidae (Bryson, 2014). En este caso de estudio, no lograron obtener amplicones de 16S rRNA con “*primers*” específicos para endosimbiontes “canónicos” como: *Wolbachia*, *Cardinium*, *Spiroplasma* y *Rickettsia*.

Para el presente estudio se escogieron dos especies de alacranes con una amplia distribución y abundancia en el estado de Morelos. Estas no presentan una estacionalidad

marcada (Cordova, 2005), por lo que se tuvo acceso a ellas durante todo el año. Estas dos especies son:

→ *Centruroides limpidus* (Karsch, 1879). Pertenece a la familia *Buthidae* y su distribución en México se concentra en los estados de Michoacán, Guerrero, México, Morelos, Puebla y Oaxaca (Fet, *et al.*, 2000). Estos alacranes miden entre 5.8 y 7.1 cm. de longitud; son de coloración amarilla (por lo cual se conocen junto con otras especies de la familia *Buthidae*, con el nombre popular de alacranes “güeros”) con un patrón de 4 líneas longitudinales. Las hembras suelen ser más grandes que los machos. El número de dientes pectinales en las hembras varía de 19 a 23 y en los machos de 22 a 26. (Hoffman, 1932). Se distingue de otros *Centruroides* spp. de la depresión del Balsas por el patrón de pigmentación del caparazón cefalotorácico. Esta especie tiene un gran impacto en la salud pública, ya que son responsables de la mayoría de los casos de alacranismo en la ciudad y en la entidad. Las toxinas de esta especie tienen una dosis letal media (LD50) de 1.72 mg/Kg (Padilla *et al.*, 2003) y se encuentra dentro de las especies más peligrosas para el humano en México.

→ *Vaejovis smithi* (Pocock, 1902). Esta especie pertenece a la familia *Vejovidae* y su distribución se concentra únicamente en el estado de Morelos, México. (Fet, *et al.*, 2000). Son alacranes que miden entre 4.5 y 7.0 cm de longitud. Se caracterizan por una coloración parda, con diferentes patrones de manchas longitudinales en el dorso del mesosoma. Las hembras son significativamente más grandes que los machos, tienen entre 16 y 19 dientes pectinales; mientras que los machos, tienen de 22 a 25. Principalmente se encuentran en las zonas altas de la ciudad de Cuernavaca (su rango de distribución en el estado de Morelos, es de los 1,930 a 2,480 msnm.). Se encuentra bajo piedras y troncos, no presenta estacionalidad y se ha colectado junto con *C. limpidus* y *C. margaritatus*.

Sistema digestivo de los alacranes

Morfológicamente el sistema digestivo de los alacranes se caracteriza por ser un tracto sencillo que no presenta diferenciaciones a lo largo del cuerpo (figura 1), desde el aparato bucal hasta el ano. Sin embargo, se puede dividir en 3 partes: estomodeo, mesenterón y proctodeo.

- a) Estomodeo: Consiste de la faringe bucal y el esófago. La digestión inicia en la cavidad preoral, donde las secreciones de las glándulas maxilares empiezan el proceso digestivo y antes de que los jugos alimenticios de la presa entren a la faringe, son

filtrados por un cepillo maxilar. La faringe es un pequeño compartimiento, que se ha modificado para convertirse en un órgano succionador (Polis, 1990).

- b) Mesenterón: Consiste en el estómago, intestino y hepatopáncreas. El estómago, es un pequeño segmento dilatado entre el esófago y el diafragma, ubicado en el cefalotórax. El intestino es el segmento más largo del sistema digestivo y se extiende desde el diafragma, hasta el cuarto segmento del metasoma. El intestino se puede dividir en dos partes: anterior y posterior. El anterior se encuentra en el mesosoma y el posterior en el metasoma (cola). El intestino anterior se conecta con el hepatopáncreas. El hepatopáncreas es la principal glándula digestiva y ocupa la mayoría de la cavidad del mesosoma, consta de células digestivas (que producen enzimas) y de absorción, cuando las células de absorción se llenan de desecho, este es regresado al intestino en forma de gránulos. El hepatopáncreas también tiene función de almacenamiento y esto permite que los alacranes puedan sobrevivir por largos periodos sin comer. El intestino posterior es un tubo de pared delgada de epitelio cilíndrico y uniforme (Polis, 1990).
- c) Proctodeo: consiste del intestino trasero y el ano. El intestino trasero se ubica en el quinto segmento del metasoma. La apertura anal se encuentra entre el quinto segmento del meta soma y el telson (Polis, 1990).

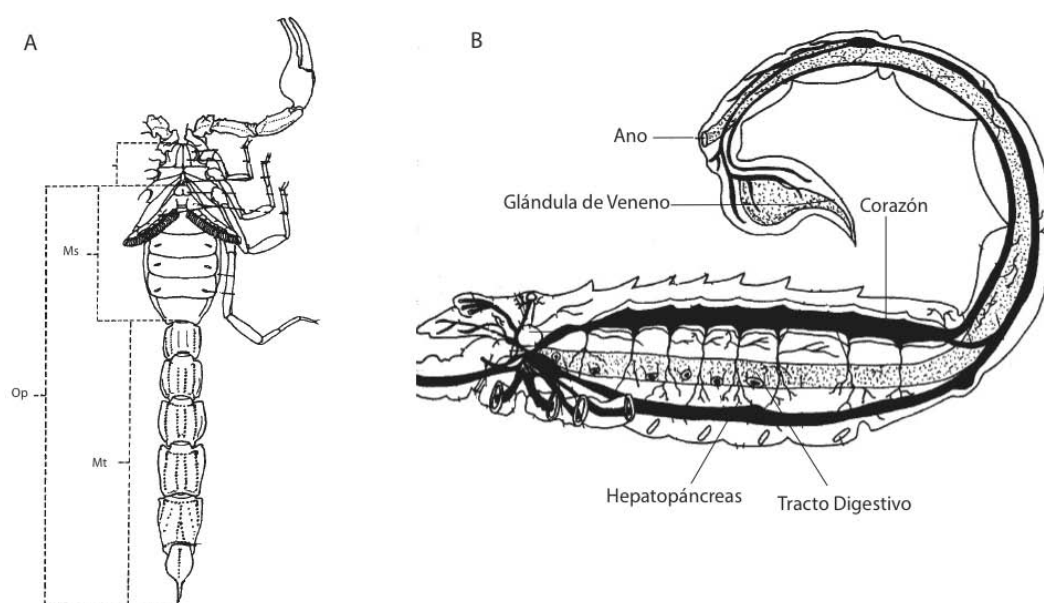


Figura 1.- Anatomía de los alacranes y órganos relevantes. A) Vista ventral. Op, opistosoma; Ms, mesosoma; Mt, metasoma. B) Vista transversal. Se señalan algunos órganos relevantes. (Polis, G. A. 1990)

Dieta y hábitos alimenticios de los alacranes

Los alacranes son depredadores generalistas. Dentro de la amplia gama de animales que pueden cazar se encuentran insectos y otros arácnidos (inclusive, otros alacranes). Pero también pueden incorporar a su dieta isópodos, gasterópodos e inclusive algunos vertebrados pequeños como lagartijas (Polis, G. A. 1990).

Los alacranes son animales de hábitos nocturnos, sin embargo la mayoría de las especies adoptan por esperar en sus madrigueras a que sean las presas quienes caigan en el lugar donde están con el fin de ahorrar energía. Este comportamiento adhiere una variable más; lo inconstante de los periodos de alimentación, la cual bajo condiciones ambientales de pocas presas podría causar grandes etapas de inanición. Estos largos intervalos de tiempo sin alimento son compensados con la capacidad de bajar sus tasas metabólicas. De hecho, junto con algunas arañas, los alacranes comparten el record de las tasas metabólicas más bajas dentro de los artrópodos (Lighton *et al.*, 2001).

Microbiota intestinal

El término microbiota se utiliza para definir una comunidad de microorganismos que habitan en diferentes tejidos de animales y que han desarrollado una relación estrecha con el hospedero.

En lo que respecta a la microbiota intestinal de artrópodos, la mejor estudiada es la comunidad de bacterias que se encuentran en los intestinos de las termitas. Esto se debe a que esta comunidad bacteriana, tiene la capacidad de usar la celulosa como fuente de alimento. La celulosa es un polisacárido compuesto por cientos de unidades de D-glucosa y es el componente estructural de las plantas. El humano no es capaz de digerir la celulosa, pero las termitas y los animales rumiantes sí, debido a las bacterias simbióticas que habitan el intestino de estos grupos de animales (Dillon, *et al.*, 2004). A partir de estos descubrimientos, se ha generado un gran interés biotecnológico por las comunidades de bacterias intestinales de los insectos. Aún así, los reportes de análisis independientes de cultivo, son escasos en los insectos y casi nulos en otros artrópodos.

Esta falta de estudios en los intestinos de los artrópodos hace difícil establecer la importancia de las diferentes variables como la dieta, especie, localización geográfica o fisiología del intestino, sobre la comunidad microbiana. Sin embargo - de los pocos estudios en artrópodos que se han realizado - se ha establecido que un factor determinante en la naturaleza de la comunidad intestinal, es el tipo de dieta del hospedero (Broderick, *et al.*, 2004).

Dentro de las aportaciones de la microbiota intestinal a la parte nutricional de los artrópodos, se encuentran: mejorar la habilidad de sobrevivir con dietas subóptimas, mejorar la eficiencia de la digestión, adquirir enzimas digestivas y proveer vitaminas y aminoácidos (Dillon, *et al.*, 2004). Sin embargo, se ha propuesto que la microbiota intestinal no solo tenga una función nutricional sino también tenga funciones importantes en la fisiología del hospedero, tales como: la maduración del sistema inmune (Masmanian *et al.*, 2005) y la influencia en el comportamiento del hospedero (Rochellys *et al.*, 2010).

Los artrópodos son el grupo más diverso de animales sobre la tierra (Ødegard, 2000) y lo mismo se refleja en los distintos tipos de sus dietas. Dentro de éstas, podemos encontrar artrópodos hematófagos, fitófagos, omnívoros y carnívoros. Por lo tanto, se espera que dentro de la diversidad de los artrópodos, encontremos una amplia diversidad de relaciones simbióticas de comunidades bacterianas asociadas a los tractos digestivos.

HIPÓTESIS

Dado que los alacranes tienen una dieta generalista, existirá una gran variación individual en la composición de la microbiota, sin embargo habrá un grupo común de bacterias a través de las especies y tratamientos que pudieran establecer una relación estrecha con el epitelio intestinal.

OBJETIVO GENERAL

Determinar la composición bacteriana de la microbiota intestinal de los alacranes *Centruroides limpidus* y *Vaejovis smithi* y establecer su posible función.

OBJETIVOS PARTICULARES

- Analizar la variación de la microbiota intestinal de las dos especies de alacranes bajo diferentes dietas: dieta homogenizada en el laboratorio a base de larvas de tenebrio (rica en lípidos y carbohidratos), privación de alimento y la dieta original no controlada

de los individuos recién capturados, en busca de un grupo común que pudiera tener una relación estrecha con el epitelio intestinal de los alacranes.

- Analizar la posición filogenética de nuevos linajes bacterianos.
- Secuenciar y analizar los genomas que constituyan el microbioma intestinal de las especies de alacrán *C. limpidus* y *V. smithi*.

CAPÍTULO I

Análisis de la variación de la diversidad de bacterias mediante un enfoque no cultivable del tracto digestivo a través de diferentes tratamientos alimenticios: Colecta peridoméstica, dieta homogenizada con larvas de *Tenebrio molitor* y privación de alimento

En este primer capítulo se aborda uno de los ejes principales de esta investigación; el análisis de la variación de la microbiota intestinal bajo diferentes dietas y especies de alacranes. Las principales preguntas a responder son: a) ¿Qué factores son más importantes para la composición de la microbiota en los alacranes? b) ¿Existe una bacteria o un grupo común de bacterias presentes en varias condiciones que podamos definir como la microbiota nativa?

Mediante un método no cultivable, que involucra: la extracción de DNA total de los intestinos de los alacranes, amplificación del gen 16S rRNA de las bacterias con “*primers*” generales, clonación de los amplicones y secuenciación de los mismos; se analizó la composición de la microbiota en las dos especies de alacranes *C. limpidus* y *V. smithi*. Adicionalmente se analizó la variación de las comunidades bacterianas a través de diferentes condiciones alimenticias, las cuales son:

1. Colecta Peridoméstica: Este grupo lo componen los alacranes que han sido colectados en diferentes lugares de la ciudad de Cuernavaca, mayormente en casas y jardines. La extracción de DNA total se hace máximo dos días después de capturados, por lo cual no tenemos ninguna información a priori de su alimentación, pero reflejará que bacterias están presentes naturalmente en los alacranes. En este grupo esperamos una gran diversidad individual y una mayor cantidad de filotipos de bacterias.
2. Dieta homogenizada con larvas de tenebrio: Este grupo lo componen los alacranes que han sido colectados de la misma forma que los de la colecta peridoméstica, pero a

diferencia de los anteriores, estos se mantienen en el laboratorio con una dieta controlada que consiste de un tenebrio cada siete días durante un mes. La extracción de DNA total se realizó 6 días después de que comieron el último tenebrio.

3. **Privación de alimento:** Este grupo lo componen las librerías de los genes 16S rRNA de alacranes que han sido colectados de la misma forma que los de la colecta peridoméstica, pero a diferencia de estos, los especímenes estuvieron en el laboratorio sin ningún tipo de comida durante 31 días (solo suministrando agua a través de algodones estériles). Pasados los 31 días de privación de alimento se realizaron las extracciones de DNA total de los intestinos.

Definir a la microbiota nativa y diferenciarla de la transitoria no ha sido fácil en ningún modelo, sin embargo en insectos se han propuesto algunos criterios como son: presencia en todos los individuos de la población, que tengan una persistencia temporal y que la tasa de crecimiento en un nicho particular del intestino sea igual o mayor a la tasa de eliminación del mismo (Dillon, *et al.*, 2004). Extrapolando estos criterios a los arácnidos y aunado a la propuesta de otro criterio en este trabajo, que es: la presencia de las bacterias a través de diferentes variantes de la dieta (sobre todo en animales generalistas donde las fluctuaciones en la composición de la microbiota pueden ser más drásticas) determinamos los factores que influyen en la composición de bacterias en el intestino de los alacranes y se propone un grupo de bacterias que podrían considerarse como la microbiota nativa.

1.1 Colecta e identificación de alacranes

La colecta se realizó en diferentes puntos de la ciudad de Cuernavaca y Oaxtepec, mayormente en casas, jardines y terrenos con escombro. Los alacranes de la especie *V.smithi* se localizaron principalmente en 3 puntos geográficos, los cuales se encuentran en zonas altas de la ciudad de Cuernavaca con coordenadas: a) 18.950291,-99.268942 b) 18.985032,-99.240425 c) 1.970382,-99.23624. Los especímenes de *C. limpidus* se colectaron mayormente en 2 puntos, uno en la ciudad de Cuernavaca con coordenadas 18.870067,-99.208131 el cual es limítrofe con el municipio de Emiliano Zapata y en Oaxtepec 18.885011,-98.981752.

1.1.1 Identificación morfológica y genética de los alacranes colectados

La identificación taxonómica se realizó a través de claves dicotómicas sobre especies de alacranes presente en el estado de Morelos (Athanasiadis 2005). Se escogieron

los alacranes pertenecientes a las especies *C. limpidus* y *V. smithi*. Sin embargo en las colectas se identificaron también *C. margaritatus*, *V. mexicanus* y *C. balsasensis*, los cuales fueron descartados para esta fase del proyecto.

En la tabla 1 (anexo 1) se muestra la cantidad de alacranes de cada especie que fueron sujetos directamente a la disección y extracción de DNA total (colecta peridoméstica) y los que fueron sujetos a tratamiento en el laboratorio. Los alacranes mostrados no representan el total de la colecta, representan aquellos de los cuales se logró obtener una cantidad adecuada de DNA para amplificar el gen 16S ribosomal de bacterias para la creación de librerías y en algunos individuos la amplificación de un marcador mitocondrial que posteriormente sirvió para corroborar la identificación taxonómica.

El marcador mitocondrial es un espacio intergénico entre los genes 12S y 16S rRNA, cuyos “*primers*” se anclan en el 12S y 16S respectivamente y tiene una longitud esperada de aproximadamente 1500 pares de bases. En un estudio previo (Bolaños, 2010) se usaron amplicones del 18S rRNA cromosomal de los alacranes, los cuales, dado su alto nivel de conservación entre especies e inclusive géneros de alacranes no tienen una señal filogenética que pudiera servir para identificación, por lo que se descartaron.

1.1.2 Árboles filogenéticos basados en la región intergénica 12S-16S mitocondrial

Con las secuencias obtenidas de la región intergénica mitocondrial y el mismo segmento obtenido de la secuenciación de mitocondrias de otras especies de alacranes; se realizaron árboles filogenéticos de máxima verosimilitud con mil réplicas de *bootstrap* y enraizados con las secuencias obtenidas de dos mitocondrias secuenciadas de la araña marina *Ammothea carolinensis* y la araña *Habronattus oregonensis*

Es de destacar que las secuencias obtenidas de los diferentes individuos de cada especie son 99% idénticas entre si mismas. Aunque no existe reportado algún porcentaje de identidad usado como corte para definir especies, en el árbol filogenético se observan características claras para soportar la identificación de la especie como: la formación de clados definidos monofileticamente y el alto porcentaje de similitud (en los marcadores no se distinguió alguno que fuera divergente de los demás). Los individuos de estudio están identificados adecuadamente, soportando la evidencia morfológica con evidencia molecular (figura 2).

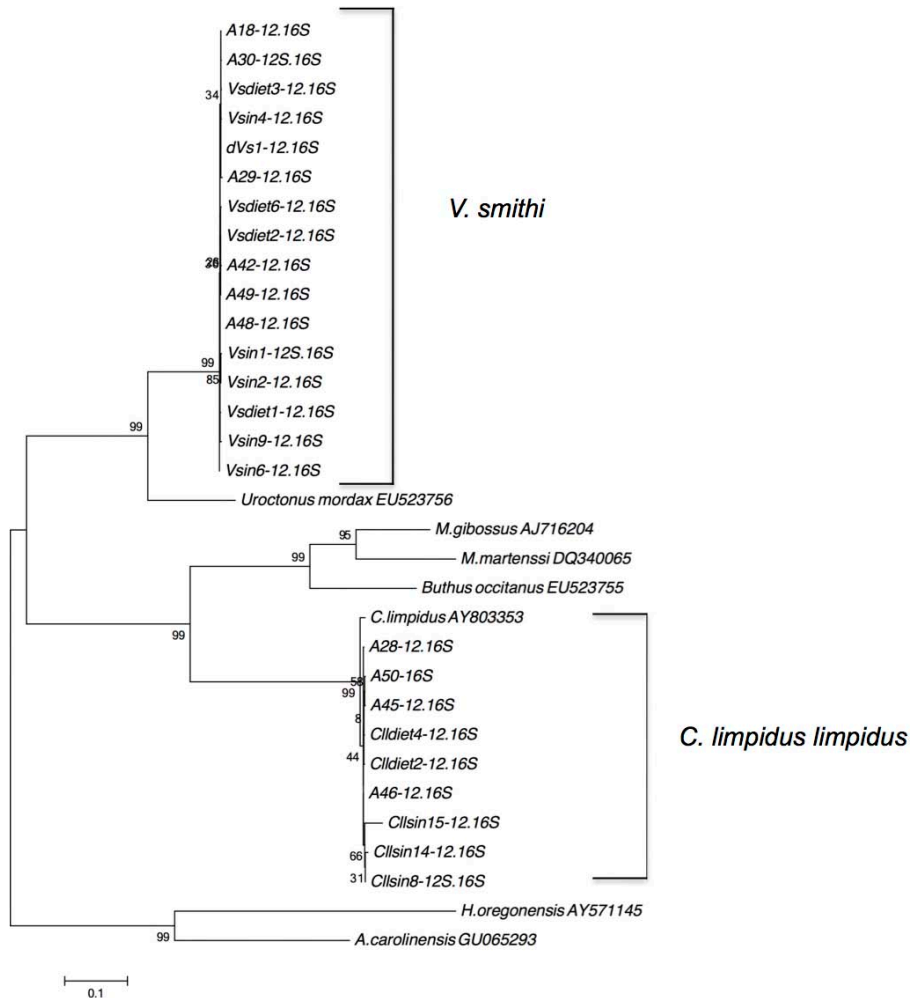


Figura 2.- Árbol filogenético de secuencias de la región intergénica 12S-16S mitocondrial de los alacranes. PhyML con 1000 réplicas de bootstrap y modelo de sustitución GTR G+I.

1.2 Resultados: La especie de alacrán es un factor de mayor influencia que la dieta en la composición de la microbiota intestinal de los alacranes.

1.2.1 Descripción de las librerías

Se analizaron en total 34 librerías de los productos de PCR del gen 16S rRNA de intestinos individuales de *C. limpidus* y *V. smithi* en las 3 diferentes condiciones: colecta peridoméstica, homogenización de la dieta y privación de alimento. De las librerías se obtuvieron un total de 1365 clones categorizadas mediante un análisis de restricción de DNA ribosomal amplificado (ARDRA, por sus siglas en inglés), de estas se secuenciaron 180 cuyos amplicones eran secuencias casi completas del gen 16S rRNA (~1400 pb.). Después de filtrar las quimeras, secuencias de baja calidad y repetidas idénticas, se obtuvo

un conjunto de datos de 114 secuencias. Se definieron 55 unidades operacionales taxonómicas (OTUs, por sus siglas en inglés) con una distancia de 0.03.

Los OTUs se clasificaron al nivel taxonómico de clase de la siguiente manera: 23.63%, 16.36% y 18.18% del total correspondían a Alpha, Beta y Gammaproteobacteria, respectivamente. Bacilli y Clostridia (Firmicutes) representaron 12.72% y 7.27%, respectivamente; seguidos por Actinobacteria (12.72%), Mollicutes (3.63%) y Spirochaetes (1.81%). No fue posible clasificar 2 OTUs: OTU 4 con una identidad del 90% con *Spiroplasma veloscicrescens* y OTU 1 con una identidad de 79% con *Spiroplasma lampyridicola*; esta la designamos como Scorpion Group 1 (SG1).

Comparaciones pareadas entre los índices de diversidad Chao1 de cada grupo mostraron que la microbiota de los alacranes *V. smithi* en dieta homogenizada y en privación de alimento es significativamente menor que la de los recién capturados ($p < 0.05$). En los individuos *C. limpidus*, aquellos en privación de alimento tienen un índice de diversidad significativamente menor que los otros dos grupos ($p < 0.05$).

De los 55 OTUs, solo 8 se encontraban presentes en las dos especies de alacrán. En *C. limpidus* se encontraron 6 OTUs con presencia en más de un tratamiento y 7 en *V. smithi*. Esto indica que hay una alta variación de bacterias entre los diferentes grupos. Dentro de los OTUs comunes en más de un grupo, se encuentran secuencias relacionadas con los géneros *Streptococcus*, *Methylobacterium*, *Agrobacterium*, *Stenotrophomonas*, *Pseudomonas* y *Ochrobactrum*. Notablemente, diferentes bacterias relacionadas a estos géneros se han reportado como colonizadoras de los tractos intestinales de diferentes artrópodos, mayoritariamente insectos (Engel & Moran, 2013; Yun *et al.*, 2014).

Cuatro OTUs destacaron entre los demás por su presencia y baja identidad con secuencias de la base de datos. La presencia del OTU que denominamos “Scorpion Group 1” o SG1 se encontró totalmente sesgado a estar presente en *V. smithi* sin importar el tratamiento (94.4% de los alacranes) y no se encontró en ningún *C. limpidus*. Este filotipo tiene un 79% de identidad con la secuencia mas cercana en la base de datos nr del Centro Nacional de Información Biotecnológica (NCBI por sus siglas en inglés), la cual es *Spiroplasma lampyridicola*.

Otros dos OTUs distribuidos ampliamente son OTU2 y OTU3, a los que nos referiremos como MVs y MCl (*Mycoplasma*-like) de *V. smithi* y *C. limpidus* respectivamente. MVs estuvo presente en 14% de los individuos recién capturados, 16% en los de dieta homogenizada y 80% en aquellos que estuvieron en privación de alimento, de los alacranes *V. smithi*. MCl, estuvo presente en el 50% de los individuos recién capturados, 20% en los de dieta homogenizada y 80% en los privados de alimento de los

alacranes *C. limpidus* y en un espécimen de *V. smithi* de los capturados recientemente. Las secuencias de MVs y MCI tienen una identidad de 93.5% entre ellos y de 86% con *Mycoplasma hyorhinis* GDL-1, que es la secuencia más cercana en la base de datos.

1.2.2 Análisis de las librerías

Para responder la pregunta ¿Qué factor (especie o tratamiento) explica más la variación en la composición de la microbiota de los alacranes? Se realizaron dos análisis: Agrupamiento por distancias euclidianas y por distancias filogenéticas. El agrupamiento jerárquico por distancias euclidianas reveló que la especie de alacrán es el factor principal que divide los grupos, seguido por los tratamientos. Esta forma de agrupamiento solo considera la presencia/ausencia de los OTUs, por lo que se realizó un análisis en el cual se consideró también las posiciones filogenéticas de las secuencias que conformaban los grupos. Se realizó el análisis tomando en cuenta la abundancia de las clonas y sin considerarla. El análisis de coordenadas principales basado en distancias filogenéticas reveló que las muestras se agrupaban preferencialmente por la especie de alacrán. Si las muestras eran etiquetadas por el tipo de tratamiento no se creaba ningún patrón distintivo de agrupamiento. (Anexo 1 figuras 1,2)

Estos resultados sugieren que la especie de alacrán explica mayormente la variación e influye más que el tratamiento en determinar la composición bacteriana del intestino.

1.3 Análisis de las comunidades intestinales en alacranes durante los primeros dos estadios.

Dado que existen muchos ejemplos de transmisión vertical de bacterias en artrópodos, se realizó una exploración de las bacterias presentes en los dos primeros estadios de los alacranes.

Todos los alacranes son vivíparos a diferencia de los demás arácnidos. Los alacranes emergen de la madre a través del opérculo genital e inmediatamente después las crías suben al dorso de la madre y se establecen ahí. Estos alacranes permanecen en el dorso hasta su primera muda (fin del primer estadio) y posteriormente bajan de la madre y se dispersan. Durante el primer estadio en el dorso de la madre, los alacranes juveniles no se alimentan, en cambio, absorben la reserva de nutrientes con la que nacen y adquieren agua a través de la madre y el ambiente húmedo que se forma (la cutícula es permeable). Dadas estas características fisiológicas y la vulnerabilidad de los alacranes juveniles en el primer estadio, es importante conocer si están heredando alguna bacteria o comunidad de

bacterias que establezcan algún tipo de simbiosis basado en los beneficios funcionales que el alacrán pudiera adquirir de estas, como: protección contra patógenos mediante la producción de antibióticos, producción y metabolismo de nutrientes mediante la fijación de nitrógeno, degradación de polímeros o producción de vitaminas o aminoácidos, entre otros.

En una primera aproximación para establecer las bacterias presentes en los primeros estadios de los alacranes y la comparación de estas con las presentes en los alacranes adultos, se realizaron dos librerías del gen 16S rRNA de dos individuos *V. smithi*, uno en el primer estadio y otro en el segundo estadio (1 día después de bajar del dorso de la madre). En las dos librerías aproximadamente el 40% de los amplicones del 16S rRNA correspondían a un filotipo con una identidad del 99% a una secuencia de *Streptomyces* sp. Otra secuencia presente en las dos muestras en alta abundancia (aproximadamente 20% de las librerías), tiene una identidad del 99% con *Brevibacterium* sp. J3. Estos dos filotipos en total componen aproximadamente el 60% en cada una de las muestras, y los dos pertenecen al *phylum* Actinomycetales (Fig. 3). Resulta interesante que los filotipos dominantes en los primeros dos estadios juveniles pertenezcan a los Actinomycetales, ya que es ampliamente reconocido que este *phylum* de bacterias se caracteriza por su capacidad de síntesis de compuestos antibióticos. En este caso, quizás se estén seleccionando filotipos que confieran algún tipo de defensa a patógenos dada la inmadurez del sistema inmune y en general de los alacranes. Hasta el momento estos son datos preliminares y un mayor muestreo en individuos y estadios serán necesarios para determinar la composición y posible función de la microbiota en etapas juveniles de los alacranes.

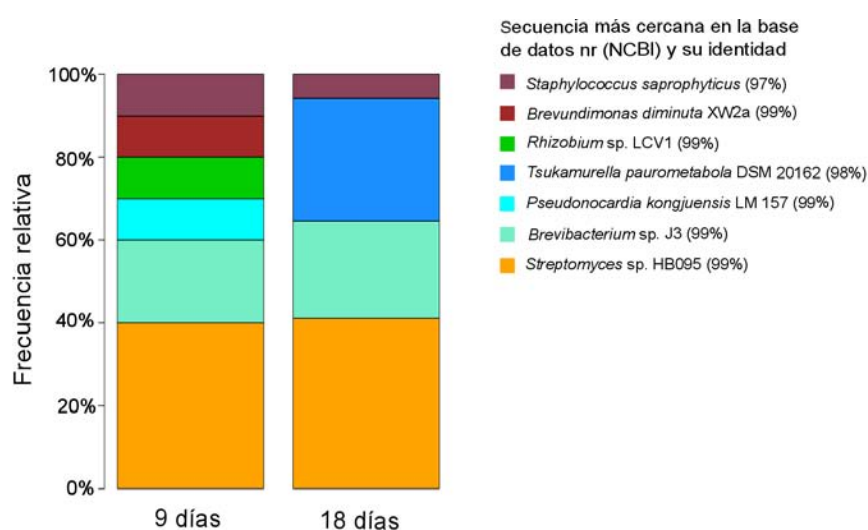


Figura 3.- Frecuencia relativa de clonas y su identidad con las secuencias más cercanas. Las librerías se crearon usando el DNA total del intestino de dos alacranes *V. smithi* en los dos primeros estadios. Las dos muestras están dominadas por Actinomycetales

1.4 Conclusiones

La especie de alacrán tiene un mayor efecto en la composición de la microbiota que los diferentes tratamientos que se analizaron en este estudio.

El estado de privación de alimento disminuye significativamente la diversidad bacteriana en relación con los otros dos tratamientos.

Se encontraron linajes bacterianos específicos de cada especie con identidades bajas a las secuencias reportadas en la base de datos (en el próximo capítulo se profundizará más en el aspecto filogenético de estos).

CAPÍTULO II

***Mycoplasma* spp. y SG1 como posibles simbioses de alacranes.**

En el capítulo anterior, uno de los resultados más relevantes fue encontrar lo que al parecer es un nuevo *phylum* de bacterias muy divergente a lo reportado en la literatura y en las bases de datos al cual hemos denominado SG1. Este nuevo *phylum* tiene un sesgo a encontrarse en la especie *V.smithi*, dado que el 95% de los especímenes estudiados sin importar el grupo del tratamiento al que pertenecían, lo tenían, al contrario de lo que sucedió en *C. limpidus* donde solo se encontró en un individuo del total de analizados.

Otro resultado importante es haber encontrado dos secuencias que forman un nuevo clado dentro del género *Mycoplasma*, las cuales por identidad de secuencia no solo podrían ser un nuevo clado, si no un nuevo género de bacterias dentro de los Mollicutes. Este nuevo grupo de bacterias, no solo resulta interesante por su posición filogenética o lo novedoso dentro de los micoplasmas, si no que al parecer cada integrante del clado es específico a cada especie de alacrán. A partir de esta nueva idea estas secuencias adquieren una gran importancia y abren nuevas preguntas como: a) De ser simbioses de los alacranes ¿Que función podrían estar cumpliendo estas bacterias? b) ¿En que parte del intestino están localizadas? c) ¿Podrían ser bacterias intracelulares? y d) El hecho de que de las especies analizadas, cada una tenga su propio *Mycoplasma* sp., ¿Se podría extrapolar a otras especies de alacranes y que exista algún efecto de coespeciación ancestral?

A continuación se describen los resultados de los análisis filogenéticos de estos OTUs divergentes.

2.1 Análisis filogenéticos de los nuevos linajes relacionados al *phylum* Tenericutes asociados a alacranes

Se realizaron reconstrucciones filogenéticas basadas en el gen 16S rRNA determinar la posición de los nuevos linajes dentro del *phylum* Tenericutes y en el dominio de las bacterias (anexo 1, figura 3a y 3b).

En lo que respecta a la posición filogenética de estos nuevos linajes dentro de los Tenericutes, SG1 es posicionado como un clado hermano del grupo *Pneumoniae* de los micoplasmas. Sin embargo, es importante considerar que este filotipo se encuentra en una rama larga con bajo soporte de *bootstrap*, lo cual indica una alta divergencia o evolución acelerada con respecto a los demás grupos del *phylum* por lo que se necesita una aproximación con secuencias más lejanas para determinar si este fenómeno es real o podría

ser una fenómeno de atracción por rama larga.

Por otro lado, los micoplasmas especie específicos MCI y MVs son posicionados como un clado hermano al grupo Hominis de los micoplasmas. El soporte de la divergencia de este clado es alta (>60) y es clara la posición como linajes diferentes en un mismo clado dentro de los Tenericutes.

Con la intención de explorar la posición filogenética de SG1 en relación a otros *phyla* y determinar si se agrupa con los Tenericutes o como un *phylum* nuevo, se realizó la reconstrucción filogenética de estos linajes nuevos en el contexto del dominio de las bacterias. Esta reconstrucción dio como resultado el posicionamiento de SG1 como una rama basal del *phylum* Tenericutes y de igual forma que en la anterior reconstrucción este filotipo se caracteriza por formar una rama larga. Dada la divergencia y la nula existencia de secuencias con mayor parentesco a este filotipo resulta impreciso determinar si SG1 está formando un nuevo *phylum* o pertenece a los Tenericutes.

Dentro de la historia evolutiva de los Tenericutes se profundizará más en el próximo subíndice en el fenómeno de los grandes intervalos de identidad de los genes 16S rRNA que rompen con los promedios generales calculados para bacterias que establecen umbrales de identidad para determinar especies, géneros o *phyla* basados en estos marcadores.

2.2 Revisión de los genomas secuenciados del género *Spiroplasma*

En los últimos años el avance en las tecnologías de secuenciación masiva de ácidos nucleicos ha permitido tener un mayor conocimiento de bacterias endosimbiontes no cultivables o que son extremadamente difíciles de cultivar dada su dependencia casi total del hospedero. El análisis de los genomas de estas, ha permitido proponer funciones, conocer de manera más profunda las bases de las relaciones simbióticas que se establecen y proponer posibles orígenes evolutivos de estas. Dentro de los géneros que han despertado mayor interés en los insectos dada la gran gama de relaciones simbióticas que establecen, los spiroplasma han sido poco estudiados desde el punto de vista genómico. No fue hasta el 2010 que se realizó el primer “draft” de un genoma de spiroplasma, *Spiroplasma citri*. A partir de esa fecha tan solo 10 genomas más se han secuenciado.

Dado que la secuencia más cercana en las bases de datos con respecto a SG1 fue un *Spiroplasma*, se decidió realizar una revisión sobre las relaciones simbióticas artrópodo-*Spiroplasma* en el contexto del conocimiento generado a partir de los primeros genomas secuenciados de este género de bacterias (Anexo 2).

Un punto importante que se encontró en esta revisión y que afecta de manera directa la interpretación de nuestras reconstrucciones filogenéticas, es el hecho de que los spiroplasma no satisfacen los umbrales establecidos para determinar los cortes entre niveles taxonómicos (especie, género, *phylum*, etc.). El intervalo de identidades entre especies de este género de bacterias es desde el 91% hasta el 99.9%, cuando el umbral propuesto para considerar a los filotipos de la misma especie es del 97%. Lo mismo pasa a nivel genómico con los cálculos de ANIs y DDHs. Esto sugiere que los diferentes linajes de *Spiroplasma* han divergido mucho entre ellos a nivel de secuencias genómicas, pero conservando las características bioquímicas comunes al género. Por lo tanto, dados los valores de identidad de los nuevos filotipos encontrados (en especial SG1), es aún más complicado entender su posición evolutiva y taxonómica con respecto a las secuencias relativamente cercanas.

2.3 Conclusiones

Nuevos linajes especie-específicos se encontraron formando parte de la comunidad de bacterias intestinales de los alacranes. Estos linajes están relacionados con el *phylum* Tenericutes. Claramente dos filotipos MCl y MVs están formando un clado hermano el grupo Hominis de los micoplasmas.

SG1, cuya presencia está sesgada completamente a los alacranes de la especie *V. smithi*, forma una rama larga cuya posición en el árbol del dominio bacteriano parece que es un grupo basal de los Tenericutes. Sin embargo al solo tener el gen 16S rRNA no podemos establecer una filogenia con un mayor soporte de este filotipo altamente divergente a las secuencias descritas en las bases de datos.

CAPÍTULO III

Predicción metagenómica basada en los resultados de las librerías de 16S rRNA.

3.1 Introducción y fundamento de las predicciones funcionales

Hacer perfiles de comunidades bacterianas con base en el gen 16S rRNA es una técnica que se ha usado exhaustivamente y ha sido clave para la descripción de nuevos linajes bacterianos que no se han logrado cultivar en el laboratorio. Sin embargo la información que se obtiene de estos estudios se limitaba a conocer la composición taxonómica de la comunidad.

A partir del auge de la secuenciación masiva de ácidos nucleicos, una nueva forma de estudiar las comunidades microbianas fue el uso de la metagenómica, la cual tiene como objetivo hacer un muestreo amplio de todo el ADN de una muestra compuesta de varios organismos para poder tener acceso a las capacidades funcionales de los diferentes filotipos y no solo de la composición de la comunidad basada en un gen marcador. Sin embargo al estudiar las comunidades mediante esta aproximación no cultivable se tienen dos grandes retos: la complejidad de las muestras y la poca representación de datos genómicos comparables de filotipos nuevos.

En lo que respecta a las muestras complejas, podemos establecer que mientras mayor diversidad de filotipos tengamos va a ser más compleja la muestra. Mientras más compleja sea una muestra tendremos una mayor cantidad de genomas diferentes y se necesitará una cobertura de secuenciación alta, lo cual incrementa los costos y el poder de cómputo necesario para obtener buenos ensamblajes y por lo tanto una buena aproximación a las funciones. No solo la cantidad de filotipos es una fuente de complejidad, también el cómo están distribuidas las proporciones de abundancia de estos filotipos en una comunidad, si se tiene una diferencia muy alta entre las proporciones de organismos muy abundantes y no abundantes, necesitaríamos tener una cobertura exponencialmente alta para poder aproximarnos a determinar las funciones de estos grupos minoritarios.

Desde el punto de vista de la bioinformática, el reto consiste en hacer ensamblajes *de novo*, esto quiere decir sin ninguna referencia. En una analogía sencilla es como si quisiéramos armar varios rompecabezas en los que se tienen las fichas mezcladas sin tener como marco de referencia las imágenes de estos armados. Por lo tanto mientras más genomas haya secuenciados de organismos filogenéticamente muy cercanos a los que

componen la comunidad se podrán separar mejor las piezas que los componen y será más fácil ensamblarlos.

Dados estos retos, en los últimos años se han hecho varios intentos de desarrollar un sistema de predicción de funciones con base en el marcador 16S rRNA y en la abundancia de genomas filogenéticamente cercanos a estos genes marcadores. Langille *et al.* en el año 2013 desarrollaron un método en el que se aprovecha la gran cantidad de genomas secuenciados en los últimos años. Este método se materializó en una tubería de algoritmos bioinformáticos (Anexo 1 materiales y Métodos) cuyo componente principal es el programa PICRUST (Langille *et al.*, 2013) que hace una reconstrucción del ancestro común de los genomas cercanos a el gen 16S rRNA y por métodos de inferencia se calcula la probabilidad de que cierto gen se encuentre en el organismo del que provino ese 16S rRNA.

Este método puede darnos una aproximación para superar al primer reto en cuanto a la complejidad de la muestra. Sin embargo, el segundo reto de comparación bioinformática con genomas ya secuenciados no se aborda desde esta perspectiva. Si tomamos en cuenta la posición filogenética de algún 16S rRNA en el que no haya genomas secuenciados cercanos, este organismo no se considera en la predicción. Por otra parte, es interesante destacar que mientras más genomas secuenciados cercanos tengamos a los organismos de donde provienen los genes marcadores, se tendrá un mayor soporte en la probabilidad de que cierta función este presente o no.

A manera de tener una primera aproximación a las funciones de las comunidades bacterianas y como varían estas en los diferentes tratamientos, realizamos una predicción con la “tubería” desarrollada por Langille *et al.* A continuación se muestra el resumen de resultados obtenidos. Las Alpha y Gamma proteobacterias componen el 95.5% de los ORFs bacterianos de *V. smithi*. Y dentro de las funciones más abundantes el metabolismo de aminoácidos y carbohidratos componen el 24.5 % de las asignaciones.

3.2.1 Resultados: Existencia de un grupo común y abundante de funciones metabólicas

Se definió un grupo de funciones comunes y que estaban abundantemente representadas en los 6 grupos. (Fig. 4a anexo 1). 41 vías metabólicas cumplieron la abundancia y presencia en los 6 grupos, dentro de estas se encuentran vías relacionadas con la síntesis y degradación de vitaminas, carbohidratos y aminoácidos. Además, otras vías presentes están relacionadas con la fijación de carbono, biosíntesis de estreptomicina y degradación de cloroalquenos.

Estas funciones, serían potencialmente la contribución de las bacterias a la comunidad y al hospedero.

3.2.2 Resultados: Abundancia significativa de vías metabólicas de degradación de compuestos tóxicos en los alacranes *C.limpidus* recién capturados.

Por otro lado, se determinaron las funciones cuya abundancia es significativamente diferente entre los 6 grupos (Fig. 4b anexo 1). Sobre las predicciones se realizó una prueba no paramétrica de suma de rangos de Wilcoxon seguido de un análisis de discriminación lineal. Dentro de las diferencias por cada grupo que se trató, es notable la cantidad de vías metabólicas relacionadas con la degradación de compuestos tóxicos en *C. limpidus* provenientes de la colecta peridoméstica (recientemente capturados). Entre este conjunto de vías se encuentran aquellas en las que se degradan limoneno, pineno, bisfenol, naftaleno, DDT e hidrocarburos aromáticos policíclicos. De manera interesante en este grupo de *C. limpidus* se encontró diferencialmente abundante la biosíntesis de hormonas esteroides.

Este sesgo de abundancia de funciones hacia el grupo de los alacranes *C. limpidus* de la colecta peridoméstica, realza una posible diferencia entre la biología de las dos especies de alacrán, en la que *V. smithi* podría no ser tan susceptible a estos compuestos tóxicos por tener ciertas capacidades metabólicas propias que *C. limpidus* no tiene y por lo tanto se seleccionan bacterias en el tracto digestivo que tengan estas capacidades de degradación.

3.3 Conclusiones

Mediante predicciones de las capacidades funcionales de las comunidades previamente descritas por el gen 16S rRNA, se analizaron las vías metabólicas comúnmente abundantes en los 6 grupos de alacranes y aquellas que tenían una abundancia diferencialmente significativa entre los grupos. Se encontró que existe un grupo de funciones comunes relacionadas a la síntesis y degradación de vitaminas, carbohidratos y aminoácidos.

Por otro lado se encontró que en *C. limpidus* de la colecta peridoméstica existe una abundancia significativamente mayor en vías metabólicas relacionadas a la degradación de compuestos tóxicos.

Estos resultado a partir de las predicciones es una aproximación para entender el potencial metabólico de las comunidades bacterianas y como se ve afectada por la variación taxonómica, dada por el gen 16S rRNA, entre los diferentes grupos.

CAPÍTULO IV

Secuenciación y análisis metagenómico del intestino de alacranes *C. limpidus* y *V. smithi* en estado de privación de alimento.

4.1 Análisis del metagenoma del intestino de los alacranes

Se realizó el metagenoma de individuos deprivados de alimento de las dos especies con dos plataformas de secuenciación 454 GS FLX Titanium e Illumina HiSeq 1000 72x72 ambos con lecturas pareadas.

La secuenciación se realizó con 3 objetivos principales: acceder al potencial funcional de la comunidad de los alacranes en estado de privación de alimento, comparar los resultados de las predicciones con la secuenciación y obtener la mayor cantidad posible de secuencias que pudieran pertenecer a los nuevos linajes descritos por el gen 16S rRNA. En cuanto al último punto, cabe resaltar que es en el estado de privación de alimento donde hay una abundancia relativa mayor de los nuevos linajes, al parecer estos se enriquecen en las muestras al disminuir la cantidad de otros filotipos.

Los resultados de la secuenciación se encuentran en Materiales y Métodos del Anexo 1. Se generaron ensamblajes independientes para cada plataforma de secuenciación y solo se tomó el set de *contigs* bacterianos de las dos muestras para el análisis.

Para *V. smithi* se obtuvieron 9,048 marcos de lectura abierta (ORFs) de origen bacteriano, en contraste para *C. limpidus* solo 28. Este primer resultado, sugiere una cantidad de células bacterianas significativamente menor en *C. limpidus* que en *V. smithi* en el tratamiento de privación de alimento. (Fig. S3 Anexo1)

Dada esta gran diferencia de genes, no se pudo hacer una comparación entre ellos y solo se usó el set de *V. smithi* para los siguientes análisis.

El set de *contigs* bacterianos de *V. smithi* está dominado por Alpha y Gamma proteobacterias. Y las funciones dominantes son el metabolismo de aminoácidos y carbohidratos, donde el 24.5% de los ORFs predichos caen en estas categorías.

4.2 Comparación de los metagenomas con las predicciones basadas en las librerías de 16S rRNA

Se comparó la abundancia de las funciones del metagenoma de *V. smithi* con las predicciones (Fig. 4a anexo1). De las 41 categorías metabólicas conservadas que definimos como común en los 6 grupos, el patrón de estas funciones en el metagenoma de *V. smithi* se asemejaba mucho a las predicciones, no hubo diferencias significativas en las

abundancias y solo 2 de las categorías no alcanzaron el umbral impuesto para pertenecer a este grupo: degradación de lisina y transportadores ABC.

Se agregó la información del metagenoma al set de datos para comparar si había alguna diferencia significativa en cuanto a la abundancia de alguna función en este con respecto a las 6 predicciones, sin embargo no existió ninguna categoría significativamente más abundante.

4.3. Microscopía de homogenizados del intestino de alacrán muestra una cantidad significativamente menor de bacterias en *C. limpidus* que en *V. smithi*.

Finalmente, con el fin de comprobar si la diferencia en la cantidad de *contigs* bacterianos en los ensamblajes metagenómicos era un fenómeno real de menor abundancia bacteriana y no algún artefacto de la secuenciación o ensamble; se realizó un conteo de células bacterianas de homogenizados de intestino de las dos especies de alacrán por medio de microscopía confocal (Fig. 5 anexo 1).

Los resultados muestran una cantidad significativamente menor en las 3 muestras de *C. limpidus* comparadas con las de *V. smithi*. Inclusive si se extrapola la cantidad de bacterias por tejido, el conteo de bacterias es casi un orden de magnitud menor en *C. limpidus*.

4.4 Conclusiones

Hay menor cantidad de bacterias en el intestino de *C. limpidus* en relación con *V. smithi*. Por lo que para obtener una buena muestra representativa de las funciones en *C. limpidus* se tendría que hacer una secuenciación de mayor cobertura o buscar otros métodos que no sea la secuenciación de la muestra compleja (con DNA del hospedero).

Las predicciones funcionales correlacionaron ampliamente con lo que se obtuvo de la secuenciación metagenómica de *V. smithi*. Es importante destacar que por la poca cobertura bacteriana en *C. limpidus* no se pudo hacer esta comparación con las predicciones.

DISCUSIÓN

Los alacranes han sido objeto de estudio por varias razones como: las diversas características fisiológicas que sobresalen de la mayoría de artrópodos, los pocos cambios morfológicos que han tenido desde su origen como colonizadores terrestres o el potencial biotecnológico de los diferentes péptidos que componen su veneno. Sin embargo, hasta el momento no había estudios sobre posibles simbioses bacterianas en los alacranes. Dados los beneficios que se pueden adquirir de establecer relaciones simbióticas con otros organismos, es importante conocer si los alacranes han debido algo de su éxito evolutivo a organismos simbioses.

La mayoría de los estudios de la microbiota intestinal de artrópodos se han dirigido a la clase de insectos. Sólo unos pocos estudios han comenzado a dilucidar la composición y las propiedades funcionales de la microbiota intestinal de los arácnidos (Narasimhan *et al.*, 2014).

En este estudio analizamos como varía la microbiota intestinal de los alacranes en 3 diferentes tratamientos alimenticios. Se encontró que la diversidad individual es baja, de 1 a 9 filotipos. Esta diversidad bacteriana relativamente baja se ha observado anteriormente en otras comunidades microbianas del intestino de artrópodos (Koch & Schmid-Hempel, 2011), con la excepción de algunas termitas (Warnecke *et al.*, 2007). Del total de OTUs encontrado en las bibliotecas de clonas, casi el 60% pertenecía a un par de clases dentro de las proteobacterias. Esta dominancia de las proteobacterias en la microbiota intestinal es extendida en los artrópodos (Engel & Moran, 2013). Además, se encontraron secuencias relacionadas a los Tenericutes en la mayoría de los alacranes estudiados. Estas secuencias del gen 16S rRNA tienen una muy baja identidad con cualquier otra reportada en las bases de datos y una podría representar un nuevo *phylum* de bacterias. El hallazgo de estos linajes bacterianos novedosos en sólo dos especies de alacranes garantiza más estudios para develar más de la probable diversidad bacteriana que se encuentre en la microbiota intestinal de otras especies de alacranes. Además, se necesita una caracterización más completa de esta nueva diversidad bacteriana mediante la secuenciación de sus genomas y pruebas bioquímicas de posibles aislados en medios de cultivo.

La privación de alimentos causó una reducción significativa en la diversidad bacteriana en el intestino de ambas especies de escorpión. Esto no era completamente

inesperado, ya que se ha demostrado previamente que, en otros artrópodos, la condición de privación de alimento reduce la diversidad bacteriana intestinal (Dillon *et al.*, 2010). Por otro lado, este tratamiento produjo el enriquecimiento de algunos taxones. Por ejemplo, hemos observado un aumento proporcional de las secuencias tipo *Mycoplasma* en ambas especies. Creemos que este efecto puede ser debido a una reducción en el número de bacterias transitorias y la persistencia de los micoplasmas bajo condiciones de estrés.

Las predicciones funcionales bacterianas muestran un considerable número de vías principalmente relacionados al metabolismo de aminoácidos, carbohidratos y vitaminas. Sin embargo, encontramos una abundancia diferencial en los diferentes grupos de tratamiento. Las vías metabólicas para la degradación de compuestos tóxicos fueron más abundantes en *C. limpidus* recientemente capturados, lo que sugiere que esto podría ser una función importante de la microbiota en el entorno natural de esta especie. Se sabe que algunos compuestos como los terpenoides son tóxicos para algunos insectos (Tripathi *et al.*, 2003). Las plantas producen terpenoides tales como limoneno y pineno. Estos están involucrados en repeler plagas y patógenos de insectos; además, son atractores de insectos para el control herbívoro, la polinización y dispersión de semillas (Rodríguez *et al.*, 2011; Pontin *et al.*, 2015). Otras sustancias como el benzoato y naftaleno, y algunos hidrocarburos aromáticos policíclicos también están presentes en las plantas y tienen propiedades similares (Dudareva *et al.*, 2004; Wöll *et al.*, 2013; Piskorski *et al.*, 2011). Así que esperamos que muchos insectos que se alimentan de plantas pueden albergar estos compuestos tóxicos; y como los alacranes se alimentan de estos insectos, puede ser que se estén seleccionando las bacterias que son capaces de degradar estos compuestos. En líneas con nuestros hallazgos, se ha demostrado que la microbiota intestinal ayuda a degradar estos compuestos (Adams *et al.*, 2011). El abuso y el uso excesivo de insecticidas distribuidos en el medio ambiente pueden conducir también a la selección de bacterias capaces de degradar estos compuestos tóxicos en *C. limpidus* escorpiones. El hecho de ser significativamente más abundantes estas vías en *C. limpidus* que en *V. smithi*, sugiere que los alacranes *V. smithi* podrían contar con genes propios que tuvieran funciones de detoxificación.

La disminución de la cantidad de bacterias en alacranes *C. limpidus* privados de alimento podría ser una estrategia para conservar energía y nutrientes, además de bajar las tasas metabólicas en periodos largos de privación de alimento. Hemos observado que los alacranes *C. limpidus* son más propensos a morir en este estado de privación de alimento. Además los alacranes *C. limpidus* son menos activos en esta condición de privación de

alimento, por lo tanto parecen ser menos tolerantes a estos periodos de inanición y esto afecta negativamente a su microbiota.

En general, nuestro estudio proporciona una caracterización inicial de la microbiota intestinal de escorpión que establece un punto de referencia para otros estudios sobre la microbiota intestinal de los escorpiones y otras especies de arácnidos.

PERSPECTIVAS

A manera de perspectiva se ha propuesto secuenciar los genomas de los nuevos linajes mediante técnicas de secuenciación a partir de células individuales. En dado caso de tener los genomas de estas bacterias, será una herramienta fundamental para entender su ecología y evolución.

Dado que los nuevos linajes bacterianos son especie específicos, se ha propuesto un estudio de exploración en otras especies y analizar como han radiado las bacterias en los hospederos. Inclusive se ha propuesto que este muestreo se podría generalizar a otros arácnidos. Teniendo los marcadores moleculares de bacterias y alacranes se podría hacer un estudio filogenético para revisar si ha habido coespeciación de estos nuevos linajes de las bacterias con los alacranes e inclusive trazar perdidas, ganancias y el posible origen de esta relación simbiótica.

REFERENCIAS

- Adams, A. S., Adams, S. M., Currie, C. R., Gillette, N. E., & Raffa, K. F.** (2010). Geographic variation in bacterial communities associated with the red turpentine beetle (Coleoptera: Curculionidae). *Environmental entomology*, *39*(2), 406-414.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J.** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Research*, **25**: 3389–3402.
- Arakaki, N., Noda, H., Yamagishi, K.** (2000) *Wolbachia*-induced parthenogenesis in the egg parasitoid *Telenomus nawai*. *Entomologia Experimentalis et Applicata*, **96**: 177–184
- Baldo, L., Prendini, L., Corthals, A., Werren, J. H.** (2007) *Wolbachia* Are Present in Southern African Scorpions and Cluster with Supergroup F. *Current Microbiology*, **55**: 367–373
- Broderick, N. A., Raffa K. F., Goodman, R. M., Handelsman, J.** (2004) Census of the Bacterial Community of the Gypsy Moth Larval Midgut by Using Culturing and Culture-Independent Methods. *Applied and Environmental Microbiology*, **70**: 293–300
- Córdova- Athanasiadis. M.** (2005) Escorpiofauna (Arachnida: Scorpiones) del estado de Morelos, México. Tesis de Licenciatura en Biología. Facultad de Ciencias Biológicas, Universidad Autónoma del estado de Morelos.
- Dillon, R. J., Dillon, V. M.** (2004) The Gut Bacteria of Insects: Nonpathogenic Interactions. *Annual Review of Entomology*, **49**:71–92
- Dudareva, N., Pichersky, E., Gershenzon, J.** (2004) Biochemistry of plant volatiles. *Plant Physiology*, *135*(4), 1893-1902.
- Dunlop, J. A.** (2010) Geological history and phylogeny of Chelicerata. *Arthropod Structure & Development*, *39*:124-142
- Engel, P., & Moran, N. A.** (2013). The gut microbiota of insects—diversity in structure and function. *FEMS microbiology reviews*, *37*(5), 699-735.
- Felsenstein, J.** (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.
- Fet, V., W.D. Sissom, G. Lowe y M.E. Braunwalder (Eds).** (2000) Catalog of the Scorpions of the World (1758- 1998). The New York Entomological Society. Nueva York. E.U.A.
- Gasparich, G.E. Whitcom, R.F. Dodge, D. French, F.E. Glass, J. Williamson, D.L.** (2004) The genus *Spiroplasma* and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the *Mycoplasma mycoides* clade. *International Journal of Systematic and Evolutionary Microbiology*, *54*:893-918
- Guindon, S., Gascuel, O.** (2003) A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**: 696–704

- Hadley, N. F.** (1974) Adaptational Biology of Desert Scorpions. *Journal of Arachnology*, **2**:11-23
- Hadley, N. F.** (1970) Water Relations of the Desert Scorpion, *Hadrurus Arizonensis*. *Journal of Experimental Biology*, **S3**:547-558
- Hoffmann, A.** (1993) El Maravilloso Mundo de los Arácnidos. Fondo de Cultura Económica. México D.F, México.
- Hoffmann, C. C.** (1932) Monografías para la entomología médica de México. Monografía num. 2. Los escorpiones de México. Segunda parte: Buthidae. *Anales del Instituto de Biología, Universidad Nacional Autónoma de México*, 243-361.
- Huber, T., Faulkner, G., Hugenholtz, P.** (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics*, **20**: 2317–2319
- Kaltenpoth, M., Gottler, W., Herzner, G., and Strohm, E.** (2005) Symbiotic Bacteria Protect Wasp Larvae from Fungal Infestation. *Current Biology*, **15**: 475–479
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori M.** (2004). The KEGG resource for deciphering the genome. *Nucleic Acid Research* **32**
- Kikuchi, Y., Hosokawa, T., Nikoh, N., Meng, X., Kamagata, Y., and Fukatsu, T.** (2009) Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biology*, **7**:2
- Koch, H., & Schmid-Hempel, P.** (2011). Bacterial communities in central European bumblebees: low diversity and high specificity. *Microbial ecology*, **62**(1), 121-133.
- Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., ... & Huttenhower, C.** (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology*, **31**(9), 814-821.
- Larkin M., Blackshields G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins D.G.** (2007) Clustal W and Clustal X version 2.0 *Bioinformatics*, **23**(21):2947-2948
- Liaw A., and R Development Core Team** (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN3-900051-07-0, URL <http://www.R-project.org/>.
- Lighton, J. R. B., Brownell, P. H., Joos, B., Turner, R. J.** (2001) Low Metabolic Rate in Scorpions: Implications for Population Biomass and Cannibalism. *Journal of Experimental Biology*. **204**:607–613
- Lourenço, W.R., Sissom, W.D.** (2000) Escorpiones. 115-135 pp. *En: Córdova-Athanasiadis. M.* (2005) Escorpiofauna (Arachnida: Scorpiones) del estado de Morelos, México. Tesis de Licenciatura en Biología. Facultad de Ciencias Biológicas, Universidad Autónoma del estado de Morelos.

- Lozupone, C., and Knight, R.** (2005) UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology*. 71(12): 8228–8235
- Mazmanian, S. K., Liu, C. H., Tzianabos, A. O., Kasper, D. L.,** (2005) An Immunomodulatory Molecule of Symbiotic Bacteria Directs Maturation of the Host Immune System. *Cell*, **122**:107–118
- Meyer, F., Paarmann, D. D'Souza, M. Olson, R. Glass, E. M. Kubal, T. Paczian, M. Rodriguez, A. Stevens, R. Wilke, A. Wilkening, J. Edwards. R. A.,** (2008) The Metagenomics RAST server – A public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, 9:386
- Narasimhan, S., Rajeevan, N., Liu, L., Zhao, Y. O., Heisig, J., Pan, J., ... & Fikrig, E.** (2014). Gut Microbiota of the Tick Vector *Ixodes scapularis* Modulate Colonization of the Lyme Disease Spirochete. *Cell host & microbe*, *15*(1), 58-71.
- Ødegaard, F.** (2000) How many species of arthropods? Erwin's estimate revised. *Biological Journal of the Linnean Society*, 71: 583–597
- Ohkuma, M., Kudo, T.** (1996) Phylogenetic Diversity of the Intestinal Bacterial Community in the Termite *Reticulitermes speratus*. *Applied and Environmental Microbiology*, **62**:461-468
- Padilla, A., Govezenskyb, T., Possani, L. D., Larralde, C.** (2003) Experimental envenoming of mice with venom from the scorpion *Centruroides limpidus limpidus*: differences in mortality and symptoms with and without antibody therapy relating to differences in age, sex and strain of mouse. *Toxicon*, **41**: 959–965
- Panaram, K., Marshall, J. L.** (2007) F supergroup *Wolbachia* in bush crickets: what do patterns of sequence variation reveal about this supergroup and horizontal transfer between nematodes and arthropods?. *Genética*, 130:53-60.
- Pannebakker, B. A., Schidlo, N. S., Boskamp, G. J. F., Dekker, L., Van Dooren, T. J. M., Buekeboom, L. W., Zwaan, B. J., P. M. Brakefield, M., Van Alphen, J. J. M.** (2005) Sexual functionality of *Leptopilina clavipes* (Hymenoptera: Figitidae) after reversing *Wolbachia*-induced parthenogenesis. *Journal of Evolutionary Biology*, **18**: 1019–1028
- Piskorski, R., Ineichen, S., Dorn, S.** (2011). Ability of the oriental fruit moth *Grapholita molesta* (Lepidoptera: Tortricidae) to detoxify juglone, the main secondary metabolite of the non-host plant walnut. *Journal of chemical ecology*, *37*(10), 1110-1116.
- Pocock, R.I.** (1902) Arachnida, Scorpiones, Pedipalpi and Solifugae. *Biología Centrali-Americana*. Taylor y Francis, Londres. 71 p.
- Polis, G. A.** (1990) The biology of Scorpions. Stanford University Press. E.U.A.
- Pontin, M., Bottini, R., Burba, J. L., & Piccoli, P.** (2015). *Allium sativum* produces terpenes with fungistatic properties in response to infection with *Sclerotium cepivorum*. *Phytochemistry*. S0031-9422(15)00050-3 doi: 10.1016/j.phytochem.2015.02.003.

- Rodríguez, A., San Andrés, V., Cervera, M., Redondo, A., Alquézar, B., Shimada, T., Gadea, J., Rodrigo, M., Zacarías, L., Palou, L., López, M.M., Castañera, P., Peña, L.** (2011) The monoterpene limonene in orange peels attracts pests and microorganisms. *Plant Signal Behavior*, 6(11),1820-1823
- Heijtz, R. D., Wang, S., Anuar, F., Qian, Y., Björkholm, B., Samuelsson, A., ... & Pettersson, S.** (2010) Normal gut microbiota modulates brain development and behavior. *Proceedings of the National Academy of the United States of America*, 108:3047-3052
- Savory, T.** (1977) *Arachnida*, 2nd ed. London and New York: Academic Press. *En:* Polis, G. A. (1990) *The biology of Scorpions*. Standford University Press. E.U.A.
- Schloss, P. D., Delalibera, I. Jr., Handelsman, J., Raffa, K. F.** (2006) Bacteria Associated with the Guts of Two Wood-Boring Beetles: *Anoplophora glabripennis* and *Saperda vestita* (Cerambycidae). *Environmental Entomology*, 35(3): 625-629
- Schloss, P.D., Handelsman, J.** (2004) Status of the Microbial Census. *Microbiology and Molecular Biology Reviews*, 68: 686–691
- Schloss, P. D., Larget, B. R., Handelsman, J.** (2004) Integration of microbial ecology and statistics: a test to compare gene libraries. *Applied and Environmental Microbiology*, 70:5485-5492.
- Schloss, P. D., Westcott, S. L.** (2009) Introducing mothur: A Computational Toolbox for Describing and Comparing Microbial Communities. Obtenido en: http://schloss.micro.umass.edu/mothur/Main_Page. Departamento de Microbiología de la Universidad de Massachusets.
- Servin, A. L.** (2004) Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. *FEMS Microbiology Review*, 28: 405-440.
- Thimm, T. Hoffman, A, Borkott, H. Munch, H.J. Tebbe, C.C.** (1998) The gut of the soil microarthropod *Folsomia candida* (Collembola) is a frequently changeable but selective habitat and a vector for microorganisms. *Applied and Environmental Microbiology*, 64:2660–2669
- Tripathi, A. K., Prajapati, V., Khanuja, S. P. S., & Kumar, S.** (2003). Effect of d-limonene on three stored-product beetles. *Journal of economic entomology*, 96(3), 990-995.
- Toscano-Gadea, C. A.,** (2004) Confirmation of Parthenogenesis in *Tutyus trivittatus* Kraepelin 1898 (SCORPIONES, BUTHIDAE). *The Journal of Arachnology*, 32:866–869
- Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T. H., Stege, J. T., ... & Leadbetter, J. R.** (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*, 450(7169), 560-565.
- Weeks, A. R., Breeuwer, J. A.** (2001) Wolbachia-induced parthenogenesis in a genus of phytophagous mites. *Proceedings Biological Sciences*, 268(1482): 2245–2251
- Weisburg, W. G., Barns, S. M., Pelletier, D. A., Lane, D. J.** (1991) 16S Ribosomal DNA Amplification for Phylogenetic Study. *Journal of Bacteriology*, 173(2): 697-703

- Whitcomb, R. F.** (1983). Culture and media for spiroplasmas. *Methods Mycoplasma* 1, 147–158.
- Wöll, S., Kim, S. H., Greten, H. J. & Efferth, T.** (2013). Animal plant warfare and secondary metabolite evolution. *Natural products and bioprospecting*, 3(1), 1-7.
- Yamazaki, K., Makioka, T.** (2004) Parthenogenesis Through Five Generations in the Scorpion *Liocheles australasiae*. Fabricius 1775 (SCORPIONES, ISCHNURIDAE). *The Journal of Arachnology*, 32:852–8
- Yuasa, T., Takahashi O., Mayama, S.** (2004) PCR “primers” for the amplification of the nuclear small subunit ribosomal DNA sequences from polycystine radiolarians. *Japan. Journal Protozoology*, 37(2): 133-137
- Yun, J.H., Roh, S.W., Whon, T.W., Jung, M.J., Kim, M.S., Park, D.S., and Bae, J.W.** (2014) Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Applied and Environmental Microbiology*, 80: 5254–5264.

Species-specific diversity of novel bacterial lineages and differential abundance of predicted pathways for toxic compound degradation in scorpion gut microbiota

Luis M. Bolaños,^{1*} Mónica Rosenblueth,¹
Santiago Castillo-Ramírez,² Gilles Figuiet-Huttin¹
and Esperanza Martínez-Romero¹

¹Programa de Ecología Genómica y

²Programa de Genómica Evolutiva, Centro de Ciencias
Genómicas, Universidad Nacional Autónoma de México,
Cuernavaca, Morelos 62210, México.

Summary

Scorpions are considered 'living fossils' that have conserved ancestral anatomical features and have adapted to numerous habitats. However, their gut microbiota diversity has not been studied. Here, we characterized the gut microbiota of two scorpion species, *Vaejovis smithi* and *Centruroides limpidus*. Our results indicate that scorpion gut microbiota is species-specific and that food deprivation reduces bacterial diversity. 16S rRNA gene phylogenetic analysis revealed novel bacterial lineages showing a low level of sequence identity to any known bacteria. Furthermore, these novel bacterial lineages were each restricted to a different scorpion species. Additionally, our results of the predicted metagenomic profiles revealed a core set of pathways that were highly abundant in both species, and mostly related to amino acid, carbohydrate, vitamin and cofactor metabolism. Notably, the food-deprived *V. smithi* shotgun metagenome matched almost completely the metabolic features of the prediction. Finally, comparisons among predicted metagenomic profiles showed that toxic compound degradation pathways were more abundant in recently captured *C. limpidus* scorpions. This study gives a first insight into the scorpion gut microbiota and provides a reference for future studies on the gut microbiota from other arachnid species.

Introduction

Scorpions are among the oldest animals on Earth (Dunlop, 2010) and are considered to be living fossils. To date, over 2000 species of scorpions have been described, and these colonize a worldwide range of habitats (Prendini, 2011; Rein, 2012). Normally, scorpions feed on insects, small vertebrates or other scorpions, and are able to tolerate different stresses, such as food (Polis, 1990) and water deprivation (Hadley, 1970), and extreme temperatures (Hadley, 1974). Scorpions are efficient predators and have the ability to consume a large amount of food in one meal (Polis, 1990). In nature, two important aspects of scorpions feeding behaviour are the variability in prey types and the inconstancy of the time intervals between each feeding period.

In Mexico, there are more than 200 described species (González-Santillán, 2001; Francke, 2014). *Vaejovis smithi* (Pocock, 1902) and *Centruroides limpidus* (Karsch, 1879) scorpions are found in the state of Morelos, located in South-Central Mexico, and have different geographical distributions, with *V. smithi* occurring in more temperate and higher areas. As far as their venoms are concerned, that of *C. limpidus* is highly toxic with an LD₅₀ of 1.72 mg kg⁻¹ in mice (Padilla *et al.*, 2003), whereas *V. smithi* belongs to a non-neurotoxic genus (Rowe *et al.*, 2011).

Symbiotic bacteria are commonly found in arthropods and affect reproduction (Werren *et al.*, 2008; Sharon *et al.*, 2010), behaviour (Ridley *et al.*, 2012), longevity and immunity of their hosts (Guo *et al.*, 2014). Therefore, symbionts are recognized as key elements driving the evolution and adaptation of arthropods (Klepzig *et al.*, 2009). Clearly, arthropod gut microbiota are involved in many important and diverse physiological processes, such as host peritrophic membrane integrity maintenance (Narasimhan *et al.*, 2014), polysaccharide degradation (Brune and Ohkuma, 2011; Lee *et al.*, 2014), nitrogen fixation and recycling (Ohkuma, 2008; Morales-Jiménez *et al.*, 2013), toxin degradation (Kikuchi *et al.*, 2012), and protection against pathogens (Gonzalez-Ceron *et al.*, 2003; Dillon *et al.*, 2005; Cirimotich *et al.*, 2011). Remarkably, gut bacterial community composition in arthropods is

Received 26 February, 2015; accepted 31 May, 2015. *For correspondence. E-mail lbolanos@lcg.unam.mx; Tel. +527 773 175 581; Fax +527 773 175 581.

Table 1. Statistical analysis of 16S rRNA gene clone libraries.

	<i>Vaejovis smithi</i>			<i>Centruroides limpidus</i>		
	Recently captured	Laboratory-fed	Food-deprived	Recently captured	Laboratory-fed	Food-deprived
No. of individual libraries	7	6	5	6	5	5
Accession numbers	KM978218- KM978243	KM978244- KM978265	KM978266- KM978280	KM978281- KM978305	KM978306- KM978324	KM978325- KM978331
Statistics						
Mean of clones per library	47.1 ± 25.2	38.5 ± 7.7	39.4 ± 15.4	38.8 ± 17.7	35.4 ± 9.86	39.4 ± 6.7
No. of unique representative sequences	26	22	15	25	19	7
No. of OTUs ^a	18	15	11	20	16	4
Diversity index and measure						
Chao1	53	37	29	37.5	35.5	7
Faith's PD	1.40158	1.13899	0.89077	1.23759	1.34444	0.40619

a. OTUs were defined at 97% sequence identity.

influenced by the host diet (Huang *et al.*, 2013), developmental stage (Vasanthakumar *et al.*, 2008; Wong *et al.*, 2011), geographical location (Corby-Harris *et al.*, 2007; Adams *et al.*, 2010) and species (Schloss *et al.*, 2006; Tang *et al.*, 2012; Wong *et al.*, 2013).

To assess the gut microbiota of scorpions, we used a culture-independent approach to analyse the taxonomic diversity and metabolic capabilities from recently captured *V. smithi* and *C. limpidus* specimens. Furthermore, we compared these results to food-deprived and laboratory-fed scorpions to better understand the dynamics of the scorpion's gut microbiota and to analyse the effects of starvation and a homogeneous type of diet on it.

Results

Gut microbiota are species-specific and food deprivation reduces bacterial diversity

To study the gut bacterial diversity from the scorpion specimens, we created 16S rRNA libraries. First, a total of 34 libraries were constructed with 16S rRNA polymerase chain reaction (PCR) products from individual guts of both scorpion species under three different treatments: recently captured, laboratory-fed and food-deprived. Out of a total of 1365 clones categorized by amplified ribosomal DNA restriction analysis (ARDRA), 180 almost full-length 16S rRNA genes (~1400 bp.) were sequenced. We removed 16 chimeric, 10 low-quality sequences and identical sequences from each library. Identical sequences were considered for the clone count in further analysis. The total set of 16S rRNA was composed of 114 sequences, and the source and characteristics of these sequences are shown in Table 1. Using this dataset, we defined 55 operational taxonomic units (OTUs) at a 0.03 distance. We used the Ribosomal Database Project to classify these OTUs and found that 23.63%, 16.36% and 18.18% of the total OTUs corresponded to *Alpha*-, *Beta*-

and *Gamma*proteobacteria, respectively. Bacilli and Clostridia (Firmicutes phylum) represented 12.72% and 7.27%, respectively, followed by Actinobacteria (12.72%), Mollicutes (3.63%) and Spirochaetes (1.81%). Two OTUs could not be classified: OTU4 with 90% identity to *Spiroplasma veloscicrescens* and OTU1 with 79% identity to *Spiroplasma lampyridicola*; we designated the latter as Scorpion Group 1 (SG1).

Recently captured specimens from *V. smithi* and *C. limpidus* had the most diverse microbiota with a total of 18 and 20 OTUs, respectively (Table 1). Pairwise comparisons between the Chao1 diversity indexes showed that microbiota from *V. smithi* laboratory-fed and food-deprived scorpions were significantly less diverse than that from the recently captured group (chi-square test, $P < 0.05$). In *C. limpidus*, bacteria from food-deprived scorpions were significantly less diverse than bacteria from laboratory-fed and the recently captured scorpions (chi-square test, $P < 0.05$), while the comparison of bacteria from laboratory-fed and recently captured scorpions was not statistically significant (chi-square test, $P = 0.842$). To further understand the bacterial diversity, we calculated the Faith's phylogenetic diversity (i.e. the sum of all the branches on a phylogenetic tree of a group of organisms) for the six groups (Table 1). Food-deprived scorpions of both species had lower measures relative to the other groups, with food-deprived *C. limpidus* being the lowest of the six groups (0.4062). The remaining four groups (laboratory-fed and recently captured of both species) had relatively similar measures with no statistical difference between them (chi-square test, P -value = 0.9985). Rarefaction curves showed great variance in the number of OTUs present among treatments independently of how deeply the library was sampled. However, in food-deprived individuals, the maximum numbers of phylotypes were four and two in *V. smithi* and *C. limpidus*, respectively (Fig. S1). Taken together, these results imply that, for both species, the recently captured

and laboratory-fed specimens had a more diverse microbiota. A possible explanation is that transient gut bacteria might have been eliminated under food deprivation.

We analysed the number of OTUs that are shared between the two species, regardless of the treatment. Of the 55 total OTUs, only eight were shared between both scorpion species. Six and seven OTUs were found in more than one treatment in *C. limpidus* and *V. smithi*, respectively. Evidently, this indicates that there is a great bacterial variation among species and among treatments. However, SG1 (related to the phylum Tenericutes) was found in 94.4% of the *V. smithi* scorpions regardless of the treatment. Clearly, the presence of this sequence is not affected by the treatment, and its absence in *C. limpidus* could suggest a strong association with *V. smithi*. Two other widely distributed OTUs were also related to the Mollicutes class. OTU2, here referred to as *Mycoplasma*-like from *V. smithi* (MVs), was present in 14% of the recently captured individuals, 16% of laboratory-fed specimens and 80% of food-deprived *V. smithi* scorpions. OTU3, here referred to as *Mycoplasma*-like from *C. limpidus* (MCI), was present in 50% of the recently captured specimens, 20% of laboratory-fed individuals, 80% of food-deprived *C. limpidus* scorpions and in one specimen of *V. smithi* that was recently captured. The MV and MCI sequences have an identity of 93.5% between them. The identity to the closest match in the National Center for Biotechnology Information non-redundant database was 89% and 88% to *Mycoplasma hyorhina* GDL-1 for MVs and MCI, respectively. Seemingly, each species of scorpion has its own *Mycoplasma*-like lineage that persists across the treatments and even when scorpions are food-deprived (Fig. 1). Other frequently observed OTUs, present in more than one treatment group, are those related to *Streptococcus*, *Methylobacterium*, *Agrobacterium*, *Stenotrophomonas*, *Pseudomonas* and *Ochrobactrum*. Notably, all of them have been reported to colonize the intestinal tracts of many arthropods (Vasanthakumar *et al.*, 2008; Geib *et al.*, 2009; Shelomi *et al.*, 2013).

We evaluated which of the two factors (scorpion species or diet treatment) has the major influence on the bacterial composition of the gut microbiota. Hierarchical clustering of the OTUs revealed that scorpion species is the main factor that divides the groups, followed by diet treatment (Fig. 1). Weighted and unweighted principal coordinate analysis (PCoA), based on UNIFRAC distances also show that samples are clustered preferentially by scorpion species (Fig. 2A). Principal coordinate analysis in which samples were tagged by treatment did not create a distinctive clustering pattern of recently captured, laboratory-fed or food-deprived scorpions (Fig. 2B). These results indicate that scorpion species explains

most of the variation, and this is more influential than diet treatment in determining gut bacterial composition, although food deprivation reduces bacterial diversity.

Phylogenetic analysis reveals novel bacterial lineages

To gain further information on SG1 and the other previously unreported OTUs, we used seqmatch from the Ribosomal Database Project and BLAST searches against Greengenes database to classify the novel bacterial sequences that would correspond to SG1, *Mycoplasma*-like sequences and OTU4. The 16S rRNA of these novel lineages showed low levels of sequence identity to any other known bacteria in public databases and a distant relationship with the Mycoplasmatales and the Entomoplasmatales orders in Tenericutes phylum. Therefore, we conducted a phylogenetic analysis using type strain sequences of the 16S rRNA gene from the main taxa of the Tenericutes along with the novel sequences (Fig. 3A). SG1 is placed as a sister clade of the Pneumoniae group on a rather long branch that precludes a finer phylogenetic affiliation. The scorpion species-specific *Mycoplasma*-like MCI and MVs were placed as a new sister clade to the Hominis group, while OTU4 seems to have a common origin and diverged from them. MCI and MVs were well differentiated from one another and even more so from OTU4. Finally, in order to determine if SG1 represents a new bacterial phylum or just a basal member of the Tenericutes, we constructed a 16S rRNA gene phylogenetic tree of bacteria with available sequences from 54 bacterial phyla (including MCI, MVs, and SG1) and using archaeal sequences as an outgroup (Fig. 3B). Regardless of the high sequence divergence of the SG1, 16S rRNA phylogeny placed SG1 as a basal long branch in the Tenericutes phylum and not as an isolated branch representing a new phylum. The placement as a new basal long branch in the Tenericutes phylum tree and the sequence divergence supports the consideration of SG1 as a new bacterial lineage. Regarding the phylogenetic placement of the *Mycoplasma*-like sequences, they form a subclade within Tenericutes, representing a new lineage in this phylum. In summary, our phylogenetic analysis clearly demonstrates the presence of previously not described bacterial diversity within our dataset.

Conservation of functional capabilities and overrepresentation of toxic compound degradation pathways in recently captured C. limpidus scorpions

We carried out a functional analysis to test whether the microbial composition of each treatment group provided different metabolic capabilities to the scorpions. Notably, over the last years, there have been attempts to predict

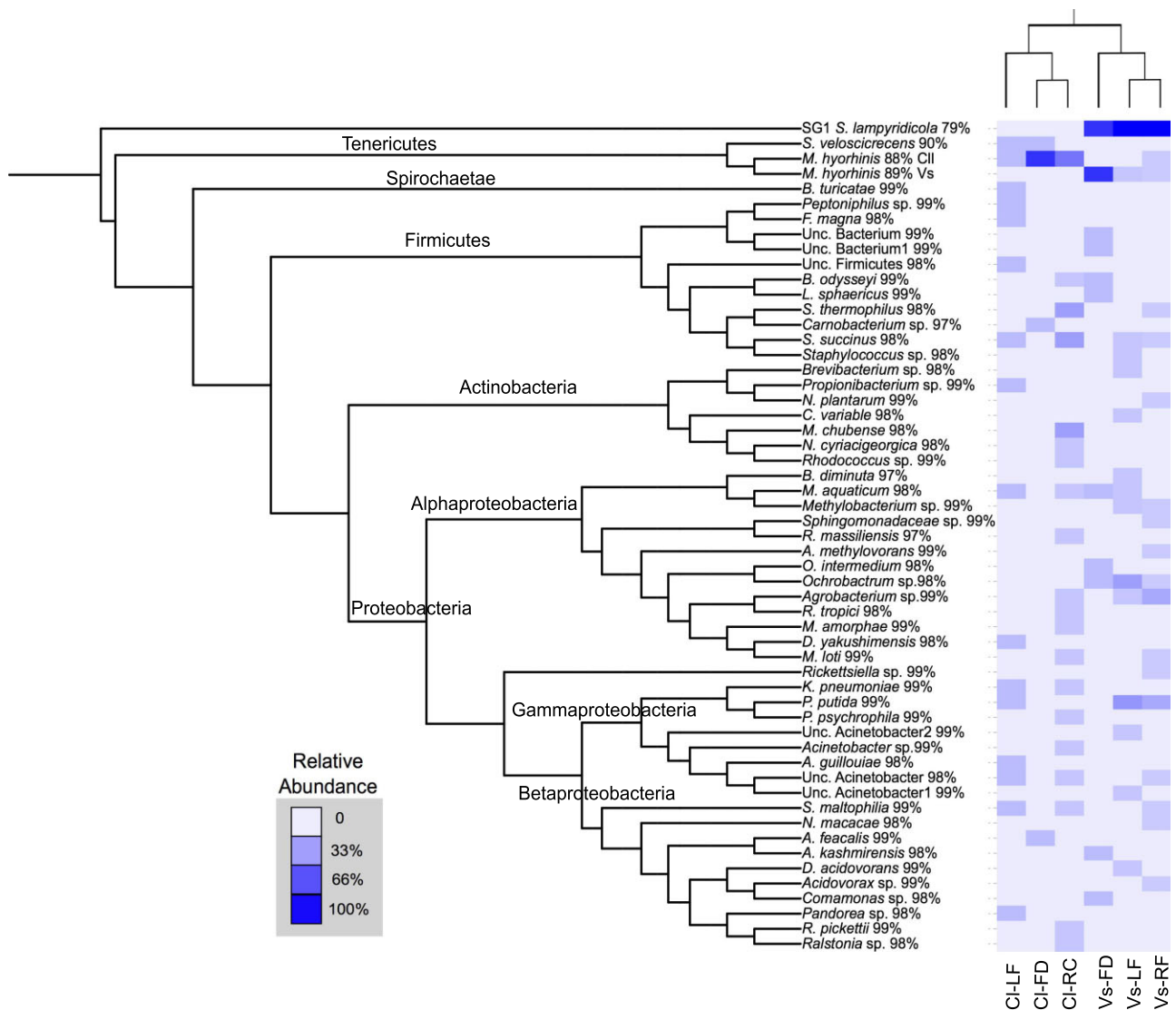


Fig. 1. Relative abundance heat map for the operational taxonomic units (OTUs) across treatment groups from both scorpion species. The range of colours indicates the OTUs' relative abundance for each group: from light blue (presence in 0% of specimens) to dark blue (presence in 100% of specimens). The 55 OTUs phylogenetic tree (RAxML 100 bootstrap replicates) is shown on the left hand side; the tip of the branches indicates the name of the closest sequence in the database and their identity. Euclidian distance clustering dendrogram is shown on top of the graph. CI, *C. limpidus*; Vs, *V. smithi*; RC, recently captured; LF, laboratory-fed; FD, food-deprived.

the genomic diversity based on the ribosomal diversity (Zaneveld *et al.*, 2010). For instance, Langille and colleagues recently developed PICRUST, a method to predict functional profiles from 16S rRNA sequences (Langille *et al.*, 2013). We used this approach and extrapolated the presence of gene content and metabolic pathways from the genomes sequences of related taxa based on the 16S clone library analysis (see *Experimental procedures*). The PICRUST programme was used to calculate KEGG orthologue gene abundance by a hidden state prediction algorithm based on the phylogenetic position of the 16S rRNA gene and the available sequenced genomes in the clade. The whole dataset takes into account the abun-

dance of clones of the treatment groups. It is important to highlight that sequences that did not fall in any OTU referenced in the Greengenes database with at least 97% identity were excluded from the prediction; therefore SG1, OTU4, MCI and MVs were not included. Comparison of the predicted gut microbiome of recently captured, food-deprived and laboratory-fed scorpions showed no significant differences at the gene metabolic categorization (Fig. S2A and B); this was true for both scorpion species. The more abundant functional gene categories across the different treatments in both scorpion species were membrane transport, amino acid metabolism and carbohydrate metabolism.

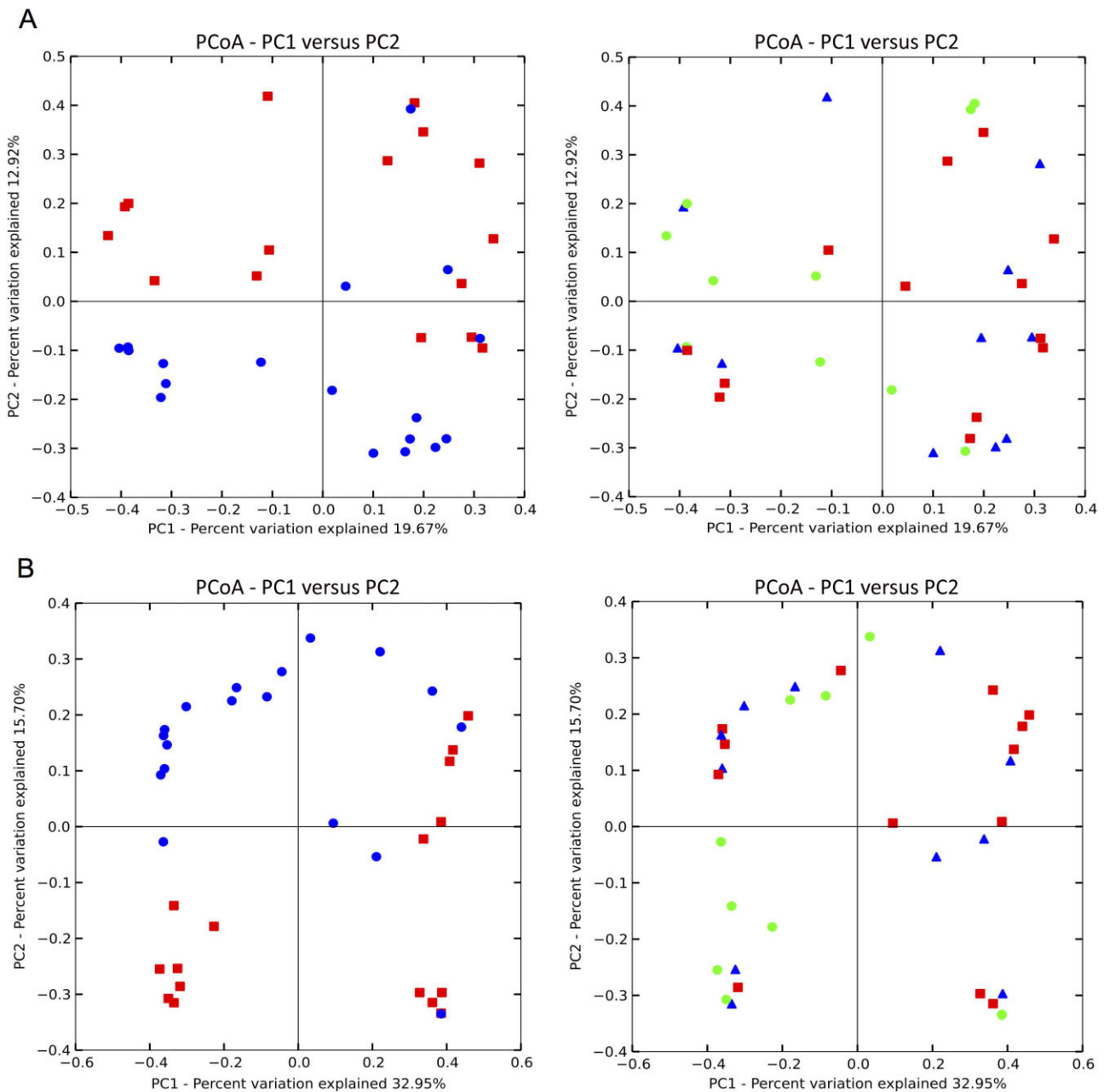


Fig. 2. Principal coordinate analysis of the 16S rRNA gene clone libraries. The ordination was constructed using weighted and unweighted UNIFRAC distances. Only the first two principal components are shown. Samples tagged for scorpion species (left panels): clone libraries from *C. limpidus* are shown in red squares and those from *V. smithi* in blue circles. Samples tagged for treatment group (right panels): recently captured group in red squares, laboratory-fed group in blue triangles and food-deprived group in green circles: (A) unweighted PCoA; (B) weighted PCoA.

The abundance of the orthologous gene families was used to reconstruct whole pathways using the MinPath approach (Ye and Doak, 2009) implemented in HUMANN (Abubucker *et al.*, 2012). We define the core cluster of pathways from scorpion gut microbiota prediction profiles as follows: we considered a pathway to be a core pathway when its relative abundance was over 0.01 and was present in the six groups. Forty-one metabolic pathways

were considered as core and mainly composed of pathways related to amino acid, carbohydrate, vitamin and cofactor metabolism (which together accounted approximately for 68% of core pathways). Among the remaining pathways, we found functions such as chloroalkane and chloroalkene degradation for detoxification, streptomycin biosynthesis, terpenoid backbone biosynthesis, and carbon fixation in prokaryotes (Fig. 4A).

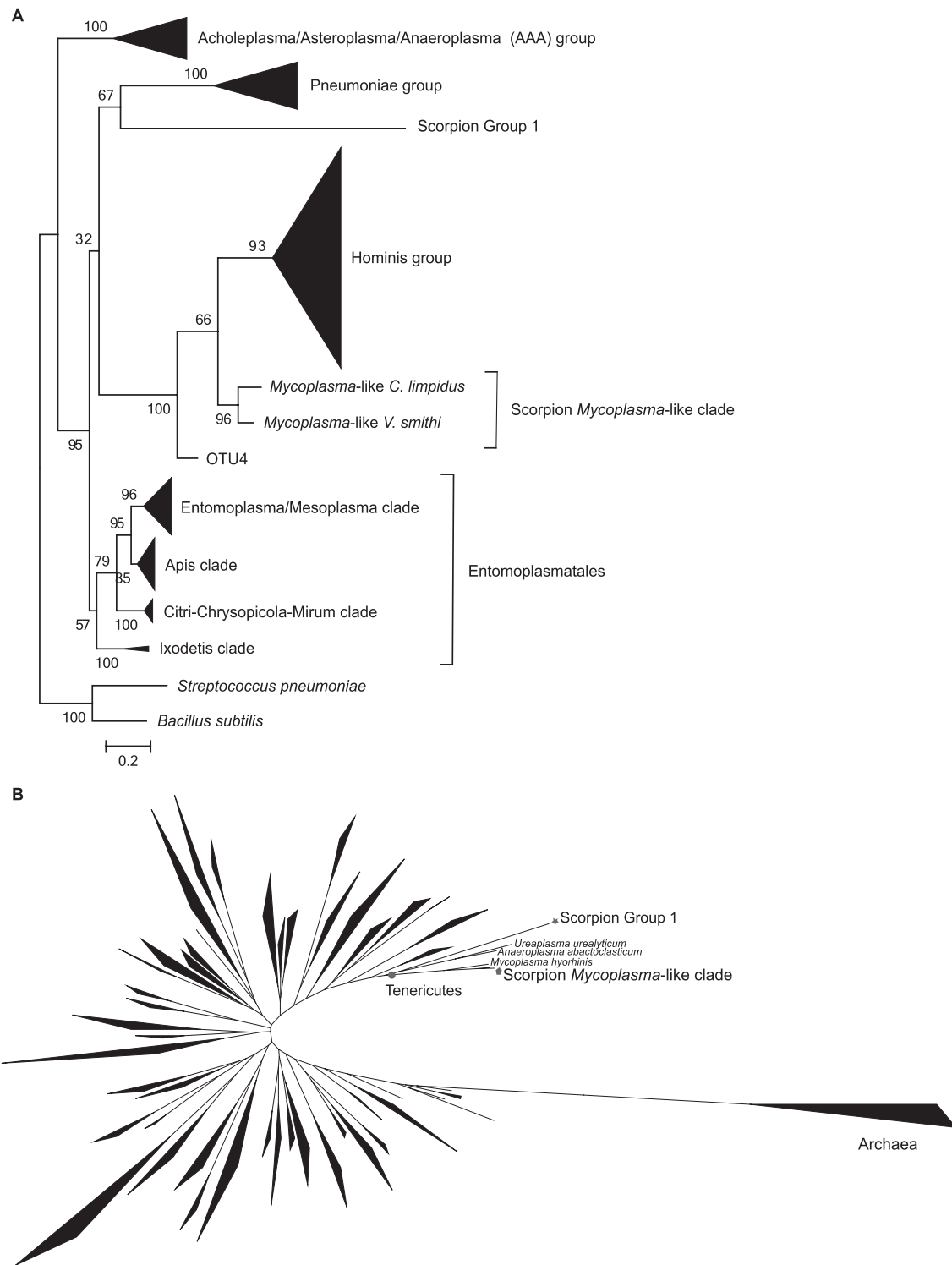


Fig. 3. Maximum likelihood phylogeny of the 16S rRNA gene sequences. Phylogenetic trees define the position of the novel lineages in the Tenericutes phylum and the bacteria domain.

A. Phylogenetic tree of the Tenericutes phylum constructed with 171 different species and the new lineages found in the scorpions. SG1 forms a long branch and shares a common ancestor with the pneumoniae group. MCI and MVs form a new sister clade to the Hominis group. *Streptococcus pneumoniae* and *Bacillus subtilis* were used as outgroups.

B. Phylogenetic tree of bacteria with three to five representative sequences for each phylum rooted with five archaeal sequences. The filled circle indicates the origin of the Tenericutes phylum. The star denotes the SG1 group, which forms a basal long branch of the Tenericutes phylum. MCI, MVs and OTU4 form the scorpion *Mycoplasma*-like clade (a pentagon) within the Tenericutes phylum.

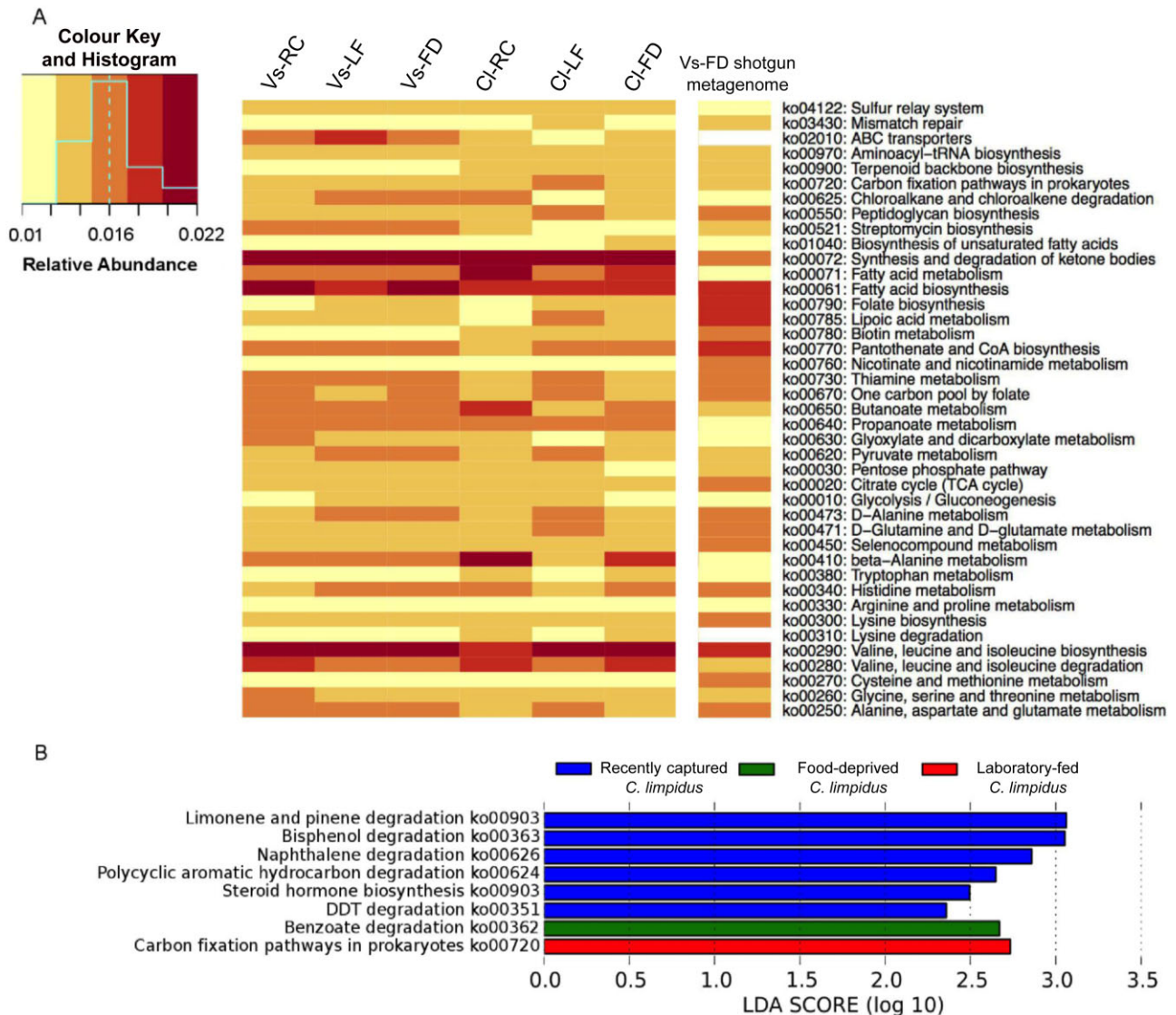


Fig. 4. Relative abundance heat map for the core pathways across treatments groups of both species. The range of colours indicates the average pathway relative abundance for each group: from light yellow (0.01) to dark red (0.022).

A. Set of pathways determined as the core cluster of metabolic pathways present in all the groups with a relative abundance greater than 0.01. The heat map is divided between predictions (left) and the shotgun metagenome from food-deprived scorpions (right). White colour (blank space) indicates that the pathway has a relative abundance below 0.01.

B. Linear discriminant analysis (LDA) of the pathways' relative abundance between diet-species groups. Degradation of toxic compounds such as bisphenol, limonene, pinene, naphthalene, dichlorodiphenyltrichloroethane and polycyclic aromatic hydrocarbons were more abundant in the recently captured *C. limpidus* scorpions. Only pathway categories meeting an LDA significant threshold of 2.0 are shown.

To determine which metabolic pathways are significantly more abundant under each treatment and species, we applied a non-parametric Wilcoxon sum-rank test, followed by a linear discriminant analysis (LDA) to assess the effect size of each of the differentially abundant pathway with the LEfSe programme (Segata *et al.*, 2011). Comparisons between the metabolic pathway profiles of the six groups revealed that some degradation pathways for toxic compounds (bisphenol, limonene, pinene, naphthalene, dichlorodiphenyltrichloroethane and

polycyclic aromatic hydrocarbons) and steroid hormone biosynthesis were more abundant in the recently captured group of *C. limpidus* scorpions compared with the other groups (Fig. 4B). The carbon fixation pathways not only were among the most abundant in all groups (present in the core pathways), but also were overrepresented in the laboratory-fed *C. limpidus*.

We did not find significant differences in gut bacterial metabolic functions between groups as far as the functional classification at gene level is concerned. However,

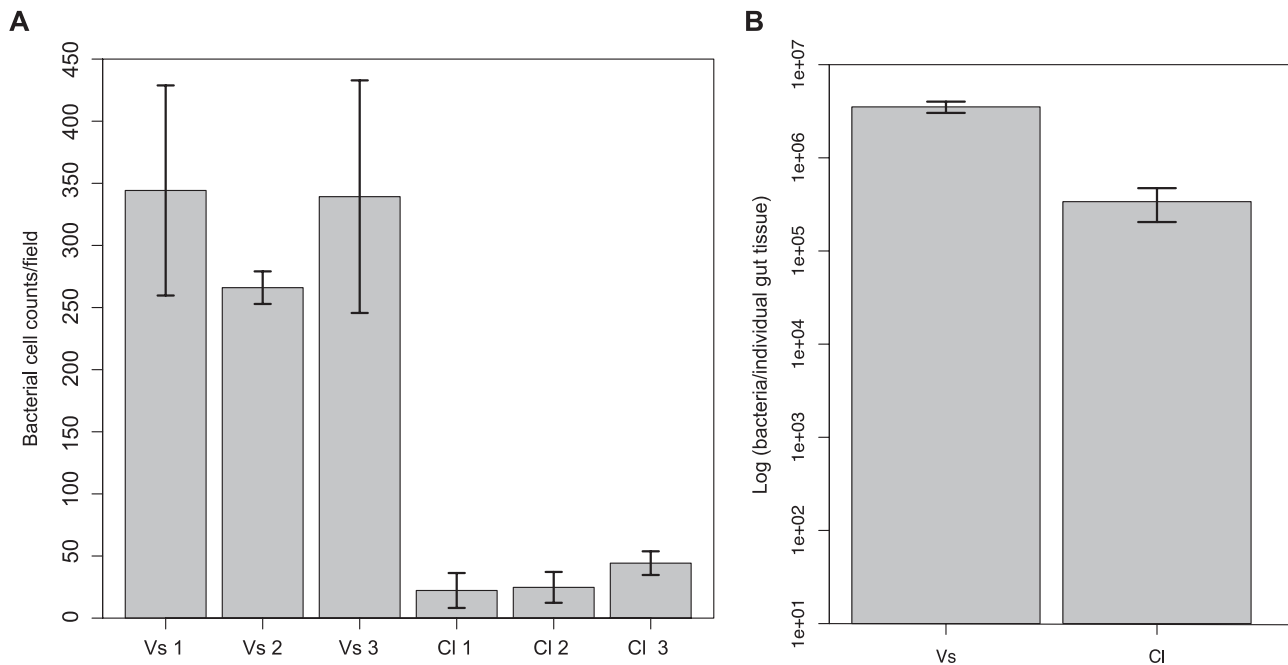


Fig. 5. Total bacteria quantification in food-deprived scorpions counted with confocal microscopy. (A) Total bacteria per field in each scorpion gut tissue. Three samples of each species were quantified. (B) Comparison between *V. smithi* and *C. limpida* food-deprived groups. The average of the total counts (average of the fields) from the three samples is plotted. *C. limpida* specimens have significantly reduced bacterial cell counts. Error bars represent standard deviations. Cl, *C. limpida*; Vs, *V. smithi*.

at a pathway level, the presence of overrepresented pathways was biased to the *C. limpida* species. The difference in the bacterial composition between scorpion species resulted in significant differential abundance of some pathways related to toxic compound degradation, hormone biosynthesis and carbon fixation. The defined core cluster of pathways was enriched in amino acid, carbohydrate, vitamin and cofactor metabolism, suggesting that these functions are conserved features in the gut microbiota regardless of species or treatment. Among the latter capabilities, the category of synthesis and degradation of ketone bodies is overrepresented in the core cluster. These results imply that although species and treatment influence the bacterial taxonomic composition, some gut microbiota functional capabilities are conserved. Nevertheless, species and treatment affect some traits, like enrichment of toxic compound degradation pathways in the recently captured *C. limpida* scorpions.

Shotgun metagenomic sequencing of food-deprived scorpions and the dominance of Proteobacteria in the *V. smithi* microbiome

In order to infer and compare the gut microbiomes from both scorpion species in a food-deprived state and to validate the functional prediction, whole shotgun metagenomic sequencing was performed. Independent

assemblies of data generated by 454 GS FLX Titanium and Illumina technologies together yielded 12 138 contigs for *V. smithi* (N50 = 1553) and 12 042 contigs for *C. limpida* (N50 = 513). Both sets of contigs were restricted to have a minimal length of 250 bp. Gene predictions on the contigs estimated 18 009 open reading frames (ORFs) for *V. smithi* and 11 803 ORFs for *C. limpida*. We obtained a similar number of contigs for both species, although N50 and ORF predictions indicated that *C. limpida* dataset was underrepresented. Then, in order to focus exclusively on bacterial gene contents, we conducted BLASTP searches (of the predicted ORFs in MEGAN taxonomic classifier) and discarded all the ORFs that had a match with eukaryotic sequences. We kept 9048 predicted bacterial ORFs for *V. smithi* and only 28 for *C. limpida* (Fig. S3A); therefore, there were much fewer bacterial reads in *C. limpida* than in *V. smithi* metagenomes from the food-deprived samples. This might suggest that there were fewer bacteria in *C. limpida* guts. To corroborate this, we quantified the bacterial cells from those samples via confocal microscopy. Notably, there were considerably less bacterial cells inside the guts from *C. limpida* than in *V. smithi* food-deprived scorpions (chi-square test, P -value < 2.2×10^{-16} , the difference is almost one order of magnitude Fig. 5B). Of note, we found homogeneous bacterial load within the different individuals of each species (Fig. 5A). Thus, the low number of bacterial reads

detected in *C. limpidus* very likely is due to the considerably lower number of bacteria in *C. limpidus* compared with *V. smithi*.

Alpha- and *Gammaproteobacteria* accounted for 95.5% of the bacterial ORFs in the *V. smithi* dataset (Fig. S3B). In each of the classes, only one genus was the major contributor; for example, *Stenotrophomonas* accounted for 90.2% of the ORFs in the *Alphaproteobacteria*, whereas *Brevundimonas* explained 75.5% of the ORFs in the *Gammaproteobacteria*. A total of 4292 predicted bacterial ORFs from the *V. smithi* metagenome were associated with KEGG pathways, where carbohydrate metabolism and amino acid metabolism accounted for the 24.5% of the assignments (Fig. S3C).

To validate the functional predictions, we compared the *V. smithi* deprivation shotgun metagenome ORFs with the predicted functional profile of that group. We reconstructed whole metabolic pathways using the MinPath approach in HUMANN based on the BLASTP results of the bacterial ORFs. First, we compared the abundance of the 'core' pathways determined from the predictions with the abundance of the same 41 pathways in the shotgun metagenome. From the shotgun metagenome set, only two categories – lysine degradation and ABC transporters – have lower abundance than 0.001 (Fig. 4A). This suggests that the core cluster of pathways determined from the predictions is a good proxy to determine the conserved metabolic features. As a second approach to validate the metagenome predictions, we conducted an LDA of pathway abundance between *V. smithi* food-deprived predicted metagenome and the shotgun sequencing. The analysis showed no significant differences in any of the pathway categories (the threshold on the logarithmic LDA score was 2.0).

Our analysis of the shotgun metagenomic sequencing of food-deprived scorpions reveals a lower bacterial abundance in *C. limpidus*, compared with *V. smithi*, and a high frequency of Proteobacteria in the *V. smithi* microbiome. Furthermore, the food-deprived *V. smithi* shotgun metagenome matched almost completely the metabolic features of the prediction, suggesting that predictions based on 16S rRNA sequences are a practical and reliable approach to determine the metagenomic profiles of the scorpion gut microbiota.

Discussion

Although scorpion species have evolved to survive in a wide variety of environmental conditions and have a clear medical importance, their gut microbiota had not been studied. Clearly, the majority of the studies of arthropod gut microbiota have targeted the insect class. Only few studies have started to elucidate the composition and functional properties of arachnid gut microbiota

(Narasimhan *et al.*, 2014). Thus, the main aim of this study was to characterize the gut microbiota of scorpions.

Notably, our 16S rRNA clone library analysis revealed great variation among individual samples; the number of OTUs within the libraries ranged from 1 to 8. This relatively low bacterial diversity has been previously observed in many other gut microbial communities of arthropods (Koch and Schmid-Hempel, 2011), with the exception of some termites (Warnecke *et al.*, 2007). Of the total OTUs found in the 16S rRNA clone libraries survey, almost 60% belonged to just a few classes within the Proteobacteria (*Alpha*, *Beta*, *Gamma*), whereas the remaining OTUs belonged to the Firmicutes, Actinobacteria and Spirochaetes phyla – all these are known to be common colonizers of insect guts (Engel and Moran, 2013; Yun *et al.*, 2014). The bacterial diversity, however, does not seem to be the same for the two scorpion species. We found that each scorpion species gut microbiota has a differential bacterial species composition. For instance, a novel bacterial lineage, namely SG1, was only found in the microbiota of *V. smithi* scorpions, present in approximately 94% of the specimens. Remarkably, this lineage is only 79% identical to the closest sequence (*Spiroplasma lampyridicola*) in the NCBI nr database. Furthermore, it could be the case that this lineage might not even be a member of the phylum Tenericutes, as SG1 sits on a long branch in the phylogenetic tree and has a basal position. Thresholds used to describe new bacterial lineages proposed that a 16S rRNA sequence with an identity equal or lower than 75% with another known type should be considered a different bacterial phylum and 95% a different genus (Yarza *et al.*, 2014). However, it is known that some Tenericutes, such as *Spiroplasma* and *Mycoplasma*, disatisfy these cut-offs (Bolaños *et al.*, 2014). Although SG1 might constitute a new bacterial phylum, more gene markers – or even complete genome sequences – along with biochemical characterization are needed to firmly establish the taxonomic level of these novel bacteria. Additionally, we found another putative novel clade of bacteria belonging to the Mollicutes class. The proposed clade contains two different lineages, each one found in only one of the two different scorpion species. MVs and MCI are approximately 89% and 88% identical to *Mycoplasma hyorhinis* GDL-1, respectively, and are a sister clade to the Hominis group. Mollicutes bacteria are found in insects, and we are starting to elucidate the symbiotic interactions with non-insect arthropods. For instance, *Candidatus Hepatoplasma crinochetorum* colonizes the hepatopancreatic gland of *Porcellio scaber* isopod (Wang *et al.*, 2004). Furthermore, '*Candidatus Hepatoplasma*' group seems to be common in many terrestrial isopods (Fraune and Zimmer, 2008). The finding of these novel bacterial lineages in just two scorpion species warrants further studies to unveil more of the bacterial

diversity likely to be found in the gut microbiota of other scorpion species. Additionally, a more comprehensive characterization of this novel bacterial diversity by next-generation sequencing and biochemical tests for cultivable isolates is needed to better understand how this unreported bacterial diversity is organized.

Food deprivation caused a significant reduction in bacterial diversity in the gut of both scorpion species. This was not completely unexpected, as it has been shown previously that, in locusts, starvation reduced gut bacterial diversity (Dillon *et al.*, 2010). On the other hand, this treatment produced the enrichment of some taxa. For instance, we noted a proportional increase of *Mycoplasma*-like sequences in both species. We think this effect may be due to a reduction in the number of transient bacteria and the persistence of some taxa (in the case of the *Mycoplasma*-like bacteria) under stressful conditions. The fact that these novel *Mycoplasma*-like organisms were still found under the starvation condition could indicate a stable relationship with scorpions.

Functional bacterial predictions show a considerable number of core pathways; mainly related to amino acid, carbohydrate, vitamin and cofactor metabolism. However, we found a differential abundance of categories in the different treatment groups. Pathways for toxic compound degradation were more abundant in recently captured *C. limpidus* scorpions, suggesting that this could be an important function of the microbiota in this species' natural environment. It is known that some compounds, such as terpenoids, are toxic to some insects (Tripathi *et al.*, 2003). Plants produce terpenoids, such as limonene and pinene. These are involved in repelling insect pests and pathogens; in addition, they are attractors to insects for herbivore control, pollination and seed dispersal (Rodríguez *et al.*, 2011; Pontin *et al.*, 2015). Other substances, such as benzoate and naphthalene, and some polycyclic aromatic hydrocarbon are also present in plants and have similar properties (Dudareva *et al.*, 2004; Piskorski *et al.*, 2011; Wöll *et al.*, 2013). So, we expect that many insects that feed on plants may harbour them, and as scorpions feed on these insects they need bacteria that are able to degrade these compounds. In line with our findings, it has been shown that gut microbiota help degrade these compounds (Adams *et al.*, 2011). The abuse and overuse of insecticides spread in the environment may lead also to the selection of bacteria capable of degrading these toxic compounds in *C. limpidus* scorpions.

Shotgun metagenomic sequencing of samples from food-deprived specimens showed a lower bacterial abundance and diversity in *C. limpidus* compared with *V. smithi*. Bacterial contigs from *C. limpidus* were scarcely found and were not included in the analysis to determine the metabolic pathway abundance comparison. Microscopy cell counts showed a very low abundance of bacteria

in food-deprived *C. limpidus*, which would explain the scarcity of reads in its metagenome. We observed that *C. limpidus* scorpions had a larger mortality rate after capture and in the food-deprived treatment compared with *V. smithi* (data not shown). Additionally, *C. limpidus* scorpions are less active after food deprivation; thus, they seem less tolerant to the deprivation treatment, which seemingly affects negatively their gut microbiota. The decrease of bacterial loads in *C. limpidus* could be a strategy to conserve energy or nutrients and decrease metabolic rates for long food deprivation periods.

Shotgun metagenome of *V. smithi* matched almost completely the core pathways abundance and showed no significant difference of pathways abundance compared with the metagenome prediction.

On the whole, our study provides an initial characterization of the scorpion gut microbiota that sets a point of reference for further studies on the gut microbiota of scorpions and other arachnid species.

Experimental procedures

Scorpions, feeding treatments and dissection

C. limpidus and *V. smithi* specimens (16 and 18, respectively) were collected in the urban area of Cuernavaca, Morelos. The specimens of each species were split into three groups: recently captured scorpions, food-deprived and laboratory-fed. The groups are as follows: six recently captured individuals, five food-deprived and five laboratory-fed individuals for *C. limpidus*; and seven recently captured scorpions, five food-deprived and six laboratory-fed individuals for *V. smithi*. Recently captured scorpions were dissected 1 or 2 days after being sampled; food-deprived scorpions were kept individually isolated in the laboratory without any food for 30 days, and the laboratory-fed group was kept in the laboratory with one *Tenebrio molitor* larvae every 7 days for 30 days. Scorpions were anaesthetized by placing them in closed containers with chloroform, and surface was disinfected by rinsing it with three cycles of 70% ethanol–sterile water rinses. Gut dissections were performed under sterile conditions, during which the hepatopancreatic gland was removed.

DNA extraction

Each gut tissue (lacking the hepatopancreatic gland) was placed in 100 µl of sterile PBS. The tissue was macerated with sterile polypropylene micro pestle inside a 1.5 ml tube. The tubes were centrifuged at low speed to pellet the macerated of gut tissue. DNA from the pellets was extracted using the UltraClean Microbial DNA Isolation Kit (MoBio Laboratories, Solana Beach, CA, USA) according to the manufacturer's directions. DNA was observed in a 0.8% agarose gel (90 V, 35 min).

PCR amplification and 16S rRNA gene clone libraries

DNA extracted from each tissue was used as template for PCR amplification of community 16S rRNA using the uni-

versal primers fD1 and rD1 (Weisburg *et al.*, 1991). Final concentrations for 20 µl PCR reactions were as follows: 1 µl DNA (25 ng µl⁻¹), 0.2 nM of each primer, 0.2 mM dNTPs, 5 U of *Taq* DNA Polymerase (Invitrogen, Carlsbad, CA, USA), 1X *Taq* polymerase buffer and 1.5 mM MgCl. The reaction conditions were 94°C for 3 min, 30 cycles at 94°C for 50 s, 56°C for 1 min, 72°C for 2 min and a final extension at 72°C for 10 min. Near full-length (approximately 1400 bases) 16S rRNA gene sequences were amplified in triplicate. PCR products were cloned into pCR 4-TOPO (Invitrogen) cloning vector and transformed in *Escherichia coli* DH5α competent cells. Transformed cells were plated on Luria–Bertani agar with carbenicillin 100 µg ml⁻¹ and kanamycin 30 µg ml⁻¹, and incubated overnight at 37°C. Inserts from transformed colonies were amplified with M13R and M13F primers. Correct size amplicons were subjected to an ARDRA using restriction enzymes *Rsa*I (10 U) and *Hind*III (10 U) (New England Biolabs, Ipswich, Massachusetts, USA). Finally, representative isolates from each ARDRA pattern were sequenced.

16S rRNA sequence analysis

Sequences were checked for chimeras using Bellerophon (Huber *et al.*, 2004) and manually inspected for quality. After chimera filtering, the remaining sequences were aligned using CLUSTALW (Thompson *et al.*, 1997) with the default settings. Sequence alignment was used to construct a distance matrix with the DNADIST programme from the PHYLIP package (Felsenstein, 1989). The distance matrix was the input for the MOTHUR programme (Schloss *et al.*, 2009), with which sequences were clustered at a distance of 0.03 in order to define the OTUs. Diversity indexes and rarefaction curves of each group of scorpions were evaluated. Additionally, we used SPLITSTREE4 programme (Huson and Bryant, 2006) to calculate the phylogenetic diversity (Faith, 1992).

BLASTN (Altschul *et al.*, 1990) and the Ribosomal Database Project (Maidak *et al.*, 2001) were used to classify the 16S rRNA gene sequences that were deposited in the NCBI GenBank database under accession numbers KM978218–KM978331. An un-rooted tree generated by PHYML (Guindon *et al.*, 2010) from OTUs and a chart for their relative abundance in the six scorpion groups were used as input to generate a heat map in the web page of Interactive Tree of Life, iTol (Letunic and Bork, 2007). A clustering cladogram – using Euclidian distances – of the scorpion groups based on the absence/presence of bacterial OTUs was constructed with HeatMap R package (R Core Team, 2014). Clustering cladogram was merged to the iTol phylogenetic heat map. Weighted and unweighted PCoA were conducted with the UNIFRAC online server (Lozupone and Knight, 2005) using as input a rooted tree file generated by the neighbour programme, which constructs neighbour-joining trees, of the PHYLIP package (Felsenstein, 1989) and an environment file that links each sequence to a clone library.

Phylogenetic analysis

Representative sequences of the denominated SG1, *Mycoplasma* amplified from *V. smithi* (MVs), *Mycoplasma*

amplified from *C. limpidus* (MCI) and OTU4 were used to construct a Tenericutes phylogeny. Additionally, a bacterial domain phylogeny was assembled with the above representative sequences except OTU4. The set of sequences to generate the Tenericutes phylogenetic tree was created from the Ribosomal Database Project and was composed of 171 species. *Bacillus subtilis* and *Streptococcus pneumoniae* 16S rRNA sequences were used as outgroup. The sequences were < 1200 bp, type strains and good quality. The set of sequences to generate the bacterial domain was created in the Greengenes database (DeSantis *et al.*, 2006). The set consisted of 194 bacterial and archeal species, three to five representative sequences from each bacterial phylum and five archeal sequences that were used to root the tree. For both sets, sequences were aligned using CLUSTALW. Tenericutes phylogenetic tree was constructed with PHYML (Guindon *et al.*, 2010) with NNI + SPR searches and 100 bootstrap replicates. Bacterial domain phylogenetic tree was constructed on RAxML (Stamatakis *et al.*, 2008) using maximum likelihood and the GTR + γ model of evolution with 100 bootstrap replicates. Tenericutes tree was visualized and edited with the MEGA6 (Tamura *et al.*, 2013). The bacterial domain tree was visualized in the iTol web page.

Metagenome predictions from 16S rRNA surveys and functional analysis

Metagenomic predictions of each of the three treatment groups of both species were made taking as input the clone abundance of previously predicted OTUs. The dataset that consisted of the total six sets was inputted to QIIME (Caporaso *et al.*, 2010) using a 'closed-reference' OTU picking protocol. Sequence search was done against the Greengenes reference collection at 97% identity. The biom-formatted OTU table obtained from QIIME was used as input to PICRUST (Langille *et al.*, 2013), which estimates the gene families content of bacteria for which no genome sequence is available, using their sequenced relatives as a reference. The OTU table was normalized and used to create the final metagenomic functional predictions. We used QIIME to plot the functional categories collapsing the PICRUST predictions to the hierarchical KEGG levels 2 and 3. Downstream pathway coverage prediction from the KEGG orthologue datasets of the predictions and the shotgun metagenomic sequencing was done with HUMANN version 0.99 (Abubucker *et al.*, 2012) and visualized with GraPhlAn (Asnicar *et al.*, 2015). HUMANN predicted pathways were subjected to a differential abundance analysis with LEfSe (Segata *et al.*, 2011). For the factorial Kruskal–Wallis test among classes and the pairwise Wilcoxon test between subclasses, we used a significance level (alpha) of 0.05. The threshold on the logarithmic LDA score to discriminate significantly abundant pathway categories was 2.0.

Shotgun metagenomic sequencing

Total DNA from 10 pooled guts of food-deprived *V. smithi* and *C. limpidus* was extracted, sheared and sized to produce DNA whole-genome shotgun library according to the manufacturer's protocol (GS FLX Titanium General Library Preparation Kit Roche Applied Science, USA). DNA sequencing

was performed on a 454 GS FLX Titanium platform according to manufacturer's instructions (Roche 454 Life Sciences, USA) by a sequencing provider (Macrogen, Korea). One-eighth plate corresponded to a sample's 3 kb mate pair. The *V. smithi* and *C. limpidus* samples yielded a total of 87 373 and 141 214 reads, respectively. The total pooled DNA extracted from three *V. smithi* and three *C. limpidus* guts were sequenced by means of an Illumina HiSeq 1000 platform (Illumina, USA) 72 × 72 bp paired-end reads in multiplex, one sixth of lane per sample, at the sequencing unit of the National Autonomous University of Mexico. The *V. smithi* sample yielded a total of 15 511 990 reads and the *C. limpidus* sample 12 860 650 reads.

Filtering, assembly, annotation and analysis of shotgun metagenomic sequences

Filtering analysis was performed with PRINSEQ (Schmieder and Edwards, 2011) for 454 and Illumina reads. After filtering the 454 GS FLX datasets, we kept 81 279 and 134 812 reads from *V. smithi* and *C. limpidus*, respectively, and for the Illumina data 12 931 972 reads from *V. smithi* and 10 351 398 from *C. limpidus*.

Metagenomic *de novo* assembly was generated independently for each dataset – Illumina and 454 reads – with IDBA-UD (Peng *et al.*, 2012) with a minimum contig length of 250 bp. Validation of contigs was achieved by mapping the reads against the contigs with Bowtie (Langmead *et al.*, 2009) for Illumina paired-end reads and NEWBLER 2.5 for 454 reads. Gene calling of contigs was performed with FragGeneScan 1.18 (Rho *et al.*, 2010). Taxonomic assignment of the putative coding genes was performed via BLASTP searches against the NCBI protein nr database with an e-value of 1×10^{-1} . Annotation results for BLASTP against the NCBI nr database were loaded into the MEtaGenome ANalyzer 4 (MEGAN 4) software (Huson *et al.*, 2011). We selected only the contigs that contained predicted coding genes, which align with bacterial sequences. A function-based classification was determined using KEGG and COG identifiers, according to the parameters for the lowest common ancestor algorithm (maximum number of match per coding sequence: 10; min support: 5; min score: 50; and top per cent: 10). Bacterial BLASTP results were used as HUMANN input with a contig-reads alignment to predict full pathway abundance in the dataset and compare it with the metagenome predictions. Shotgun bacterial metagenomic sequence data have been deposited at the NCBI under Bioproject PRJNA266890.

Quantification of bacterial loads in food-deprived scorpions

A cultivation-independent method was used to quantify the bacterial loads in the gut tissue of food-deprived scorpions from both species. Filtered samples enriched with bacteria cells and stained with DAPI (4',6-diamidino-2-phenylindole) were screened and counted under a confocal microscope.

Whole gut tissues from three *V. smithi* and three *C. limpidus* scorpions in food-deprived treatment were dissected and homogenized in PBS (as previously described).

The homogenized samples containing scorpion and bacterial cells were centrifuged and then re-suspended in 500 µl of a fixative solution with 4% formaldehyde in PBS for 2 h. Samples were re-suspended in 500 µl of PBS. Fixated samples were passed through a 5 µm filter, enriching the samples with bacterial cells. Of each sample, 15 µl was stained with 2.5 µl of DAPI (300 nM). As a positive control for bacterial DAPI staining, we used a swab from *E. coli* DH5α, and the negative control was the PBS used during the bacterial enrichment steps. We chose randomly four fields from each slide, using a 60× objective mounted onto an FV100 confocal microscope (Olympus, Japan) in the National Laboratory of Advanced Microscopy, UNAM Campus Morelos facility. Total bacterial cells were counted manually in each field using IMAGEJ image analysis programme (Abràmoff *et al.*, 2004). Auto-fluorescence scorpion gut debris (non-uniform particles) was selectively discarded with the support of bright-field images. The average of bacterial cells per field was used to estimate the total bacterial count per sample with the next formula: Bacteria/sample = (bacteria average/field volume) × 60 µl. Bacterial cells per field values for each individual and total sample bacterial counts comparison between food-deprived *V. smithi* and *C. limpidus* were analysed (chi-square goodness-of-fit test) and graphed with R package.

Acknowledgements

Financial support was from CONACyT Basic Science (Grant Number 154453). We are thankful to the Programa de Doctorado en Ciencias Biomédicas from UNAM. L. M. Bolaños had a CONACyT PhD scholarship. We thank Soledad Ochoa-Mendez for technical assistance and M. Dunn for critically reading the manuscript. Authors have no conflict of interest to declare.

References

- Abràmoff, M.D., Magalhães, P.J., and Ram, S.J. (2004) Image processing with ImageJ. *Biophotonics Int* **11**: 36–43.
- Abubucker, S., Segata, N., Goll, J., Schubert, A.M., Izard, J., Cantarel, B.L., and Huttenhower, C. (2012) Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* **8**: e1002358.
- Adams, A.S., Adams, S.M., Currie, C.R., Gillette, N.E., and Raffa, K.F. (2010) Geographic variation in bacterial communities associated with the red turpentine beetle (Coleoptera: Curculionidae). *Environ Entomol* **39**: 406–414.
- Adams, A.S., Boone, C.K., Bohlmann, J., and Raffa, K.F. (2011) Responses of bark beetle-associated bacteria to host monoterpenes and their relationship to insect life histories. *J Chem Ecol* **37**: 808–817.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.
- Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., and Segata, N. (2015) Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**: e1029.
- Bolaños, L.M., Servín-Garcidueñas, L.E., and Martínez-Romero, E. (2014) Arthropod–*Spiroplasma* relationship in the genomic era. *FEMS Microbiol Ecol* **91**: 1–8.

- Brune, A., and Ohkuma, M. (2011) Role of the termite gut microbiota in symbiotic digestion. In *Biology of Termites: A Modern Synthesis*. Bignell, D.E., Roisin, Y., and Lo, N. (eds). Dordrecht, the Netherlands: Springer, pp. 439–476.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., and Knight, R. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.
- Cirimotich, C.M., Dong, Y., Clayton, A.M., Sandiford, S.L., Souza-Neto, J.A., Mulenga, M., and Dimopoulos, G. (2011) Natural microbe-mediated refractoriness to *Plasmodium* infection in *Anopheles gambiae*. *Science* **332**: 855–858.
- Corby-Harris, V., Pontaroli, A.C., Shimkets, L.J., Bennetzen, J.L., Habel, K.E., and Promislow, D.E. (2007) Geographical distribution and diversity of bacteria associated with natural populations of *Drosophila melanogaster*. *Appl Environ Microbiol* **73**: 3470–3479.
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., and Andersen, G.L. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.
- Dillon, R.J., Vennard, C.T., Buckling, A., and Charnley, A.K. (2005) Diversity of locust gut bacteria protects against pathogen invasion. *Ecol Lett* **8**: 1291–1298.
- Dillon, R.J., Webster, G., Weightman, A.J., and Charnley, A.K. (2010) Diversity of gut microbiota increases with aging and starvation in the desert locust. *Antonie Van Leeuwenhoek* **97**: 69–77.
- Dudareva, N., Pichersky, E., and Gershenzon, J. (2004) Biochemistry of plant volatiles. *Plant Physiol* **135**: 1893–1902.
- Dunlop, J.A. (2010) Geological history and phylogeny of Chelicerata. *Arthropod Struct Dev* **39**: 124–142.
- Engel, P., and Moran, N.A. (2013) The gut microbiota of insects—diversity in structure and function. *FEMS Microbiol Rev* **37**: 699–735.
- Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* **61**: 1–10.
- Felsenstein, J. (1989) PHYLIP-phylogeny interference package (version 3.2). *Cladistics* **5**: 164–166.
- Francke, O.F. (2014) Biodiversidad de Arthropoda (Chelicerata: Arachnida ex Acari) en México. *Revista mexicana de biodiversidad* **85**: 408–418.
- Fraune, S., and Zimmer, M. (2008) Host-specificity of environmentally transmitted *Mycoplasma*-like isopod symbionts. *Environ Microbiol* **10**: 2497–2504.
- Geib, S.M., Jimenez-Gasco, M.D.M., Carlson, J.E., Tien, M., and Hoover, K. (2009) Effect of host tree species on cellulase activity and bacterial community composition in the gut of larval Asian longhorned beetle. *Environ Entomol* **38**: 686–699.
- Gonzalez-Ceron, L., Santillan, F., Rodriguez, M.H., Mendez, D., and Hernandez-Avila, J.E. (2003) Bacteria in midguts of field-collected *Anopheles albimanus* block *Plasmodium vivax* sporogonic development. *J Med Entomol* **40**: 371–374.
- González-Santillán, E. (2001) Catálogo de escorpiones de la Colección Nacional de Arácnidos (CNAN). Dissertation, Tesis. DF, Mexico: Facultad de Ciencias, Universidad Nacional Autónoma de México.
- Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307–321.
- Guo, L., Karpac, J., Tran, S.L., and Jasper, H. (2014) PGRP-SC2 promotes gut immune homeostasis to limit commensal dysbiosis and extend lifespan. *Cell* **156**: 109–122.
- Hadley, N.F. (1970) Water relations of the desert scorpion, *Hadrurus arizonensis*. *J Exp Biol* **53**: 547–558.
- Hadley, N.F. (1974) Adaptational biology of desert scorpions. *J Arachnol* **2**: 11–23.
- Huang, X.F., Bakker, M.G., Judd, T.M., Reardon, K.F., and Vivanco, J.M. (2013) Variations in diversity and richness of gut bacterial communities of termites (*Reticulitermes flavipes*) fed with grassy and woody plant substrates. *Microb Ecol* **65**: 531–536.
- Huber, T., Faulkner, G., and Hugenholtz, P. (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**: 2317–2319.
- Huson, D.H., and Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254–267.
- Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N., and Schuster, S.C. (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Res* **21**: 1552–1560.
- Karsch, F. (1879) Scorpionologische Beiträge. I. *Mitt Münch Entomol Ver* **3**: 6–22.
- Kikuchi, Y., Hayatsu, M., Hosokawa, T., Nagayama, A., Tago, K., and Fukatsu, T. (2012) Symbiont-mediated insecticide resistance. *PNAS* **109**: 8618–8622.
- Klepzig, K.D., Adams, A.S., Handelsman, J., and Raffa, K.F. (2009) Symbioses: a key driver of insect physiological processes, ecological interactions, evolutionary diversification, and impacts on humans. *Environ Entomol* **38**: 67–77.
- Koch, H., and Schmid-Hempel, P. (2011) Bacterial communities in central European bumblebees: low diversity and high specificity. *Microb Ecol* **62**: 121–133.
- Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., and Huttenhower, C. (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**: 814–821.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Lee, F.J., Rusch, D.B., Stewart, F.J., Mattila, H.R., and Newton, I.L. (2014) Saccharide breakdown and fermentation by the honeybee gut microbiome. *Environ Microbiol* **17**: 796–815.
- Letunic, I., and Bork, P. (2007) Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127–128.
- Lozupone, C., and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.
- Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker, C.T., Jr, Saxman, P.R., Farris, R.J., and Tiedje, J.M. (2001) The

- RDP-II (Ribosomal Database Project). *Nucleic Acids Res* **29**: 173–174.
- Morales-Jiménez, J., de León, A.V.P., García-Domínguez, A., Martínez-Romero, E., Zúñiga, G., and Hernández-Rodríguez, C. (2013) Nitrogen-fixing and uricolytic bacteria associated with the gut of *Dendroctonus rhizophagus* and *Dendroctonus valens* (Curculionidae: Scolytinae). *Microb Ecol* **66**: 200–210.
- Narasimhan, S., Rajeevan, N., Liu, L., Zhao, Y.O., Heisig, J., Pan, J., and Fikrig, E. (2014) Gut microbiota of the tick vector *Ixodes scapularis* modulate colonization of the Lyme disease spirochete. *Cell Host Microbe* **15**: 58–71.
- Ohkuma, M. (2008) Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trends Microbiol* **16**: 345–352.
- Padilla, A., Govezensky, T., Possani, L.D., and Larralde, C. (2003) Experimental envenoming of mice with venom from the scorpion *Centruroides limpidus limpidus*: differences in mortality and symptoms with and without antibody therapy relating to differences in age, sex and strain of mouse. *Toxicon* **41**: 959–965.
- Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–1428.
- Piskorski, R., Ineichen, S., and Dorn, S. (2011) Ability of the oriental fruit moth *Grapholitha molesta* (Lepidoptera: Tortricidae) to detoxify juglone, the main secondary metabolite of the non-host plant walnut. *J Chem Ecol* **37**: 1110–1116.
- Pocock, R.I. (1902) Arachnida, Scorpiones, Pedipalpi and Solifugae. In *Biologia Centrali-Americana*. Godman, F.D., and Salvin, O. (eds). Londres: Taylor & Francis, p. 71.
- Polis, G.A. (ed.) (1990) *The Biology of Scorpions*. Stanford, CA, USA: Stanford University Press.
- Pontin, M., Bottini, R., Burba, J.L., and Piccoli, P. (2015) *Allium sativum* produces terpenes with fungistatic properties in response to infection with *Sclerotium cepivorum*. *Phytochemistry* **S0031-9422**: 50–53. doi:10.1016/j.phytochem.2015.02.003.
- Prendini, L. (2011) Order Scorpiones CL Koch, 1850. In Zhang, Z.-Q. (ed.) *Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness*. *Zootaxa* **3148**: 115–117.
- R Core Team (2014) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing [WWW document]. URL <http://www.R-project.org/>.
- Rein, J.O. (2012) The scorpion files. Trondheim, Norway: Norwegian University of Science and Technology [WWW document]. URL <http://www.ntnu.no/ub/scorpion-files/>.
- Rho, M., Tang, H., and Ye, Y. (2010) FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* **38**: e191.
- Ridley, E.V., Wong, A.C., Westmiller, S., and Douglas, A.E. (2012) Impact of the resident microbiota on the nutritional phenotype of *Drosophila melanogaster*. *PLoS ONE* **7**: e36765.
- Rodríguez, A., San Andrés, V., Cervera, M., Redondo, A., Alquézar, B., Shimada, T., et al. (2011) The monoterpene limonene in orange peels attracts pests and microorganisms. *Plant Signal Behav* **6**: 1820–1823.
- Rowe, A.H., Xiao, Y., Joseph Scales, J., Klaus, D., Linse, K.D., Rowe, M.P., et al. (2011) Isolation and characterization of CvIV4: a pain inducing α -scorpion toxin. *PLoS ONE* **6**: e23520.
- Schloss, P.D., Delalibera, I., Jr, Handelsman, J.O., and Raffa, K.F. (2006) Bacteria associated with the guts of two wood-boring beetles: *Anoplophora glabripennis* and *Saperda vestita* (Cerambycidae). *Environ Entomol* **35**: 625–629.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., and Weber, C.F. (2009) Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Schmieder, R., and Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**: 863–864.
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011) Metagenomic biomarker discovery and explanation. *Genome Biol* **12**: R60.
- Sharon, G., Segal, D., Ringo, J.M., Hefetz, A., Zilber-Rosenberg, I., and Rosenberg, E. (2010) Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *PNAS* **107**: 20051–20056.
- Shelomi, M., Lo, W.S., Kimsey, L.S., and Kuo, C.H. (2013) Analysis of the gut microbiota of walking sticks (Phasmatodea). *BMC Res Notes* **6**: 368.
- Stamatakis, A., Hoover, P., and Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* **57**: 758–771.
- Tamura, K., Stecher, G., Peterson, D., Filipiński, A., and Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**: 2725–2729.
- Tang, X., Freitak, D., Vogel, H., Ping, L., Shao, Y., Cordero, E.A., and Boland, W. (2012) Complexity and variability of gut commensal microbiota in polyphagous lepidopteran larvae. *PLoS ONE* **7**: e36978.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882.
- Tripathi, A.K., Prajapati, V., Khanuja, S.P.S., and Kumar, S. (2003) Effect of d-limonene on three stored-product beetles. *J Econ Entomol* **96**: 990–995.
- Vasanthakumar, A., Handelsman, J., Schloss, P.D., Bauer, L.S., and Raffa, K.F. (2008) Gut microbiota of an invasive subcortical beetle, *Agrilus planipennis* Fairmaire, across various life stages. *Environ Entomol* **37**: 1344–1353.
- Wang, Y., Stingl, U., Anton-Erxleben, F., Geisler, S., Brune, A., and Zimmer, M. (2004) 'Candidatus hepatoplasma crinochetorum,' a new, stalk-forming lineage of Mollicutes colonizing the midgut glands of a terrestrial isopod. *Appl Environ Microbiol* **70**: 6166–6172.
- Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., and Leadbetter, J.R. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560–565.

- Weisburg, W.G., Barns, S.M., Pelletier, D.A., and Lane, D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173**: 697–703.
- Werren, J.H., Baldo, L., and Clark, M.E. (2008) Wolbachia: master manipulators of invertebrate biology. *Nat Rev Microbiol* **6**: 741–751.
- Wong, A.C., Chaston, J.M., and Douglas, A.E. (2013) The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis. *ISME J* **7**: 1922–1932.
- Wong, C.N.A., Ng, P., and Douglas, A.E. (2011) Low-diversity bacterial community in the gut of the fruitfly *Drosophila melanogaster*. *Environ Microbiol* **13**: 1889–1900.
- Wöll, S., Kim, S.H., Greten, H.J., and Efferth, T. (2013) Animal plant warfare and secondary metabolite evolution. *Nat Prod Bioprospect* **3**: 1–7.
- Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., and 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.
- Ye, Y., and Doak, T.G. (2009) A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput Biol* **5**: e1000465.
- Yun, J.H., Roh, S.W., Whon, T.W., Jung, M.J., Kim, M.S., Park, D.S., and Bae, J.W. (2014) Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Appl Environ Microbiol* **80**: 5254–5264.
- Zaneveld, J.R., Lozupone, C., Gordon, J.I., and Knight, R. (2010) Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives. *Nucleic Acids Res* **38**: 3869–3879.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Rarefaction curves of the 16S rRNA clone libraries from the treatment groups of each scorpion species. Rarefaction analysis was performed under an OTU threshold of 97% identity of the representative clone sequences for each individual library.

Fig. S2. KEGG gene categorization of the predicted profiles. The gene metabolic functions of microbiome predictions show a high consistency across samples, suggesting that they are not affected by diet treatment or scorpion species. No significant differences between the groups could be identified at different categorization levels. (A) Relative abundance of genes clustered in KEGG level 2 categories. (B) Relative abundance of the 20 most abundant gene clusters in KEGG level 3 categories.

Fig. S3. Shotgun metagenomic description of the assembled contigs for food-deprived scorpions. (A) Distribution of bacterial and non-bacterial contigs in each dataset. (B) Taxonomic distribution of the bacterial contigs in *V. smithi* bacterial fraction. *Gamma*- and *Alphaproteobacteria*, accounting for $\approx 95\%$, dominate the dataset. (C) Open reading frames from *V. smithi* shotgun metagenome assigned to KEGG metabolic categories. Carbohydrate and amino acid metabolism comprised the main fraction (24.5%) as in the predicted metagenomes.

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

Table 1. General features of the sequenced *Spiroplasma* genomes.

	<i>S. chrysopicola</i> DF-1	<i>S. syrphydicola</i> EA-1	<i>S. citri</i> Gll3-3X	<i>S. melliferum</i> IPMB4A	<i>S. melliferum</i> KC3	<i>S. diminutum</i> CUAS-1 ^T	<i>S. taiwanense</i> CT-1 ^T	<i>S. apis</i> B31 ^T	<i>S. culicicola</i> AES-1 ^T	<i>S. sabaudiense</i> Ar-1343 ^T
Arthropod host	<i>Chrysops</i> sp.	<i>E. arbustorum</i>	<i>C. haematoceps</i>	<i>Apis-Mellifera</i>	<i>A. Mellifera</i>	<i>Culex-annulus</i>	<i>C. tritaeni-orhynchus</i>	<i>A. Mellifera</i>	<i>Aedes-sollicitans</i>	<i>stricticus/A. vexans</i>
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
Chromosomal 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 leafhopper *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 syrphid 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 mutualistic *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. diminutum* 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. diminutum* 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: Mycooides-Entomoplasmataceae, Apis, Citri-Chrysopicola-Mirum and Ixodetis. *Mycoplasma*, *Mesoplasma* and *Entomoplasma* genera are restricted to Mycooides-Entomoplasmataceae clade. *Spiroplasma* species are distributed in the other three clades.

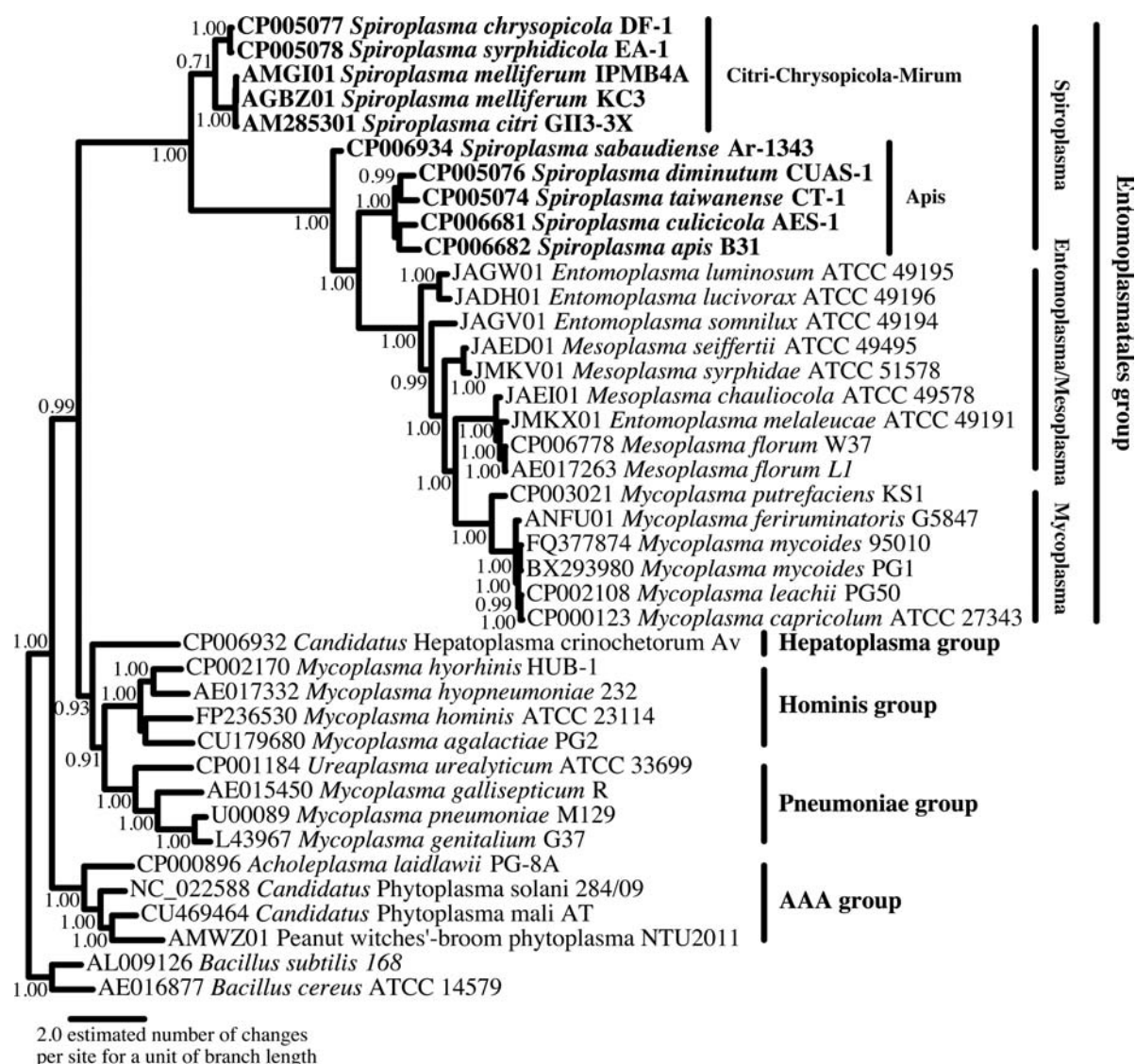


Figure 1. Topology of a phylogenomic analysis showing the predicted evolutionary relationships of sequenced *Spiroplasma* strains within the Mollicutes. The tree was reconstructed with PhyloPhlAn using a multisequence alignment of 388 conserved proteins. PhyloPhlAn performs individual alignments from 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 $94.9\% \pm 0.4$ 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. chrysipicola* 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 *Chrysipicola* and *Apis* clades seems to indicate that susceptibility of viral invasions may be restricted only to the Citri clade. *Spiroplasma chrysipicola* 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. chrysipicola*, 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. syrphidicola*. 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. chrysipicola* 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-Chrysipicola-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 spiralinins 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. chrysipicola* 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 + O₂ to glycerone phosphate + H₂O₂ (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 H₂O₂ 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 deerflies and syrphidflies. 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 pyrophosphate synthesis and the pathway for nucleotide biosynthesis. The non-mevalonate pathway for isopentenyl pyrophosphate (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 afl avin 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 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 twoflower 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. *Spiroplasmafloricola*, 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, Gullet 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 (*Anaerospasma* 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 non-mevalonate 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 lower 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 flies (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 hellicoides* 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 fruitfly. *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, Hjerten 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.