UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

DOCTORADO EN CIENCIAS BIOMÉDICAS

CENTRO DE CIENCIAS GENÓMICAS

IDENTIFICACIÓN MOLECULAR Y FUNCIONAL DE HONGOS ASOCIADOS A LA COCHINILLA DEL NOPAL *Dactylopius* spp.

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS BIOMÉDICAS

PRESENTA:

Q.B.P. ARTURO VERA PONCE DE LEÓN

DIRECTOR DE TESIS

DRA. ESPERANZA MARTÍNEZ ROMERO

CENTRO DE CIENCIAS GENÓMICAS

COMITÉ TUTOR

DR. JESUS AGUIRRE LINARES

INSTITUTO DE FISIOLOGÍA CELULAR

DR. JORGE NIETO SOTELO

INSTITUTO DE BIOLOGÍA

CIUDAD DE MÉXICO, FEBRERO, 2017

Universidad Nacional
Autónoma de México
Biblioteca Central
Dirección General de Bibliotecas de la UNAM

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.

El trabajo de investigación aquí reportado se realizó en el Laboratorio de Ecología Genómica del Centro de Ciencias Genómicas de la UNAM. Con el apoyo de CONACyT 331625.

Agradecimientos

A la Dra. Esperanza Martínez Romero por sus enseñanzas, paciencia, guía y sobre todo porque siempre me dejó hacer lo que más me gusta en la vida "Ciencia", por más extrañas o descabelladas que fueran mis ideas.

A los doctores Jesús Aguirre y Jorge Nieto, ya que sin sus comentarios, críticas e ideas este trabajo nunca se hubiera realizado.

A la Dra. Mónica Rosenblueth porque siempre me ayudó y enseñó a formarme como científico.

Al M en C. Marco Antonío Rogel por que cualquier duda en el laboratorio él lo resolvía.

Al Dr. Alejandro Sánchez Flores por convertirme en bioinformático.

A la Dra. Alejandra Escobar por su ayuda en el trabajo de investigación.

A los miembros de la UUAB de la UNAM (Jerôme, Karel, Verónica, Lety, Luciana y Ricardo).

A los miembros de la UATI-CCG (Romualdo, Víctor y Alfredo).

Al Dr. Pablo Vinuesa y Christian Sohlenkamp.

A la Dra. Angela Douglas por permitirme hacer una estancia en su laboratorio y mostrarme el mundo de la simbiosis.

Al M en C. Victor Higareda por su apoyo, ayuda y amistad a lo largo del doctorado.

A la Dra. Tabita Ramírez que me enseñó a trabajar con el modelo Dactylopius*.*

A mis compañeros y amigos del laboratorio: Tonalli, Rafa, Lorena, Jhony, Yessica, Julio, Jesús, Miguel, Giovani, Violeta, Jazmín, Paco, Tania, Luis Bolaños, Luis Servin, Mauro, Luis "Mampo", Araceli, Martín y todos aquellos que mi mala memoria hace olvidar.

A mis alumnas que siempre me ayudaron a sacar el trabajo: Diana, Ángeles, Paulina, María y Pilar.

A la UNAM y el PDCB que me ha formado.

Por mi raza hablará el espíritu.

Agradecimientos personales

A mi familia por su apoyo incondicional.

A mis padres Sergio Vera y Estrella Ponce de León por las lecciones de vida que me llevaron a ser lo que soy.

A mis hermana Andrea por su cariño y comprensión.

A mi hermano "El Brian" por sus bromas y escucharme cuando lo necesitaba.

A mis abuelos Jorge y Mónica porque siempre me han enseñado que el estudio es lo más importante y por estar ahí cada vez que los necesitaba.

A mis tíos que me guiaron desde pequeño.

A mi abuelo Arturo ya que desde niño me ha mostrado como "arreglar" las cosas.

A mi abuela Pila que aunque no está aquí su energía siempre me ha acompañado.

A mis hermanos no biológicos Javier, Juan y Oscar porque siempre me han apoyado, criticado y ayudado a superar obstáculos.

A los buzos y miembros del staff de Abismo Pablo, Rodrigo "Borre", Ale, Nayelli, Pelletier y Mario que hacían que el estudio fuera más relajado.

A Delia Enriquez y Jaime Gaitan por su apoyo en estos últimos años.

En especial a Meztlli Gaytán por su amor tan grande y por haberme apoyado, seguirme en mi viaje y alentarme a ir más lejos.

Research is to see what ev erybody else has seen, and to think

what nobody else has thought

Albert Szent-Györgi

Índice

Resumen I

Abstract II

1. Introducción 1

1.1 Interacciones microbianas en insectos fitófagos 1

1.2 Relaciones bacteria-insectos. 2

1.3 Fijación de nitrógeno por bacterias diazótrofas simbiontes de insectos 9

1.4 Relaciones hongos-insectos 10

1.5 Reciclaje de nitrógeno por hongos asociados a insectos 12

1.6 La cochinilla del carmín *Dactylopius* spp. (Hemiptera:Dactylopiidae) 14

1.7 Microorganismos asociados a *Dactylopius* spp 16

1.8 La cochinilla de la laca *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae) 18

Capítulo I. 20

Reciclaje del ácido úrico por hongos asociados a la cochinilla del carmín *Dactylopius* spp.

Capitulo II. 26

Hongos asociados a la cochinilla chiapaneca "Nijj" *Llaveia axin axin.*

Capitulo III 30

Fijación de nitrógeno por bacterias simbiontes de Dactylopius

Capitulo IV. 35

Una Gammaproteobacteria asociada al Nijj

Discusión 41

Conclusión 53

Perspectivas 54

Bibliografía 55

Anexos 69

Anexo 1 Arturo Vera-Ponce de León, Alejandro Sanchez-Flores , Mónica Rosenblueth and Esperanza Martínez-Romero (2016). **Fungal community associated with *Dactylopius* (Hemiptera: Coccoidea: Dactylopiidae) and its role in uric acid metabolism.** Front. Microbiol. 7:954. doi: 10.3389/fmicb.2016.00954.

Anexo 2 Arturo Vera-Ponce de León, Tania Rosas-Pérez, Mónica Rosenblueth, Esperanza Martínez-Romero**. Metagenomic analysis of the fungal community associated with the scale insect *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae).**

Anexo 3 Arturo Vera-Ponce de León, Ernesto Ormeño-Orrillo, Shayim T. Ramírez-Puebla, Mónica Rosenblueth, Julio Martinez, Esperanza Martínez-Romero**. *Candidatus* Dactylopiibacterium carminicum a nitrogen-fixing symbiont of the cochineal insect**

Anexo 4 Tania Rosas-Pérez, **Arturo Vera-Ponce de León**, Mónica Rosenblueth, Shamayim T. Ramírez-Puebla, Reiner Rincón-Rosales, Julio Martínez-Romero, Michael F. Dunn, Esperanza Martínez-Romero. (2017)**. The Symbiome of Llaveia Cochineals (Hemiptera: Coccoidea: Monophlebidae) Includes a Gammaproteobacterial Co-symbiont Sodalis TME1 and the Known *Candidatus* Walczuchella monophlebidarum.** Entomology, ISBN 978-953-51-5041-1.

Anexo 5 Shamayim T. Ramírez-Puebla, Luis E. Servín-Garcidueñas, Ernesto Ormeño-Orrillo, **Arturo Vera-Ponce de León**, Mónica Rosenblueth, Luis Delaye, Julio Martínez, Esperanza Martínez-Romero. (2015). **Species in *Wolbachia*? Proposal for the designation of '*Candidatus* Wolbachia bourtzisii', '*Candidatus* Wolbachia onchocercicola',**

'*Candidatus* Wolbachia blaxteri', '*Candidatus* Wolbachia brugii', '*Candidatus* Wolbachia taylori', '*Candidatus* Wolbachia collembolicola' and '*Candidatus* Wolbachia multihospitum' for the different species within *Wolbachia* supergroups. Syst App Microbiol. 2015, 36:390-399.

Anexo 6 Alejandra Escobar-Zepeda, **Arturo Vera-Ponce de León,** Alejandro Sanchez-Flores.(2015). **The road to metagenomics: from microbiology to DNA sequencing technologies and Bioinformatics.** Front. Genet. doi: 10.3389/fgene.2015.00348.

Anexo 7 Shamayim T. Ramírez-Puebla, Ernesto Ormeño-Orrillo, **Arturo Vera-Ponce de León**, Luis Lozano,Alejandro Sanchez-Flores, Mónica Rosenblueth,and Esperanza Martínez-Romero. (2016). **Genomes of *Candidatus* Wolbachia bourtzisii wDacA and *Candidatus* Wolbachia pipientis wDacB from the Cochineal Insect *Dactylopius coccus* (Hemiptera: Dactylopiidae).**G3. 6: 3343−3349

Resumen

Las interacciones microbianas con insectos fitófagos han favorecido la expansión y evolución de estos organismos. En este trabajo se estudiaron dos insectos nativos mexicanos: *Dactylopius* (cochinilla del carmín) y *Llaveia* (cochinilla de la laca) los cuales se asocian con diferentes microorganismos. Particularmente se observó que hongos de los géneros *Cryptococcus, Rhodotorula* y *Penicillium,* aislados del interior de *Dactylopius coccus y Dactylopius opuntiae*, pueden usar el ácido úrico como única fuente de carbono y/o nitrógeno. El estudio del metagenona de *D. coccus* mostró secuencias fúngicas codificantes para enzimas del catabolismo del ácido úrco. Ensayos con antifúngicos resultaron en una disminución significativa en la talla y peso de los insectos tratados en comparación con los controles así como un aumento en la concentración de ácido úrico y disminución de la enzima uricasa. Además, se encontró en el mismo metagenoma de *D. coccus* secuencias para la fijación de nitrógeno por bacterias. La filogenia de los genes *nif*HDK encontrados en el metagenoma correspondieron a una β-proteobacteria emparentada con *Uliginosibacterium gangwonense*. Se reconstruyó un genoma de 3.6 Mb de esta β-proteobacteria y el análisis del core-genómico aunado a la filogenia del 16S rRNA demostró que el genoma pertenecía al simbionte *Dactylopiibacterium carminicum* hallado en todas las especies de *Dactylopius* analizadas previamente en el grupo de trabajo. Ensayos de reducción de acetileno y RT-PCR de genes *nif* de esta bacteria demostraron que es un diazótrofo activo en hemolinfa y ovarios de *D. coccus*.

Estudios metagenómicos de la cochinilla de la laca (*Llaveia axin axin*) revelaron secuencias ribosomales de hongos, principalmente Ascomicetos. En el análisis por cultivo, hongos del género *Aspergillus* fueron aislados de los intestinos de *L. axin axin*. Al igual que en *Dactylopius,* secuencias para el catabolismo del AU fueron halladas en la anotación metagenómica de *L. axin axin*. Previamente en el grupo de trabajo se reportaron secuencias de una γ-proteobacteria en los metagenomas de *L. axin axin*, y en este trabajo se realizaron estudios de genómica comparada del genoma ensamblado de esta bacteria la cual se encuentra filogenéticamente dentro del género *Sodalis*. Con base en los resultados de la identidad promedio de nucleótidos e hibridación DNA-DNA proponemos que el género *Sodalis* puede dividirse en cuatro filogrupos. Además, el genoma de *Sodalis* TME1 (~3.4 Mb) posee la información genética para complementar metabólicamente la vía del ciclo de Krebs y los aminoácidos faltantes sintetizados por el simbionte primario de *Llaveia* (*Walczuchella monophlebidarum*). También se encontraron secuencias para utilizar productos del catabolismo del ácido úrico en el genoma de TME1 y ensayos de expresión diferencial (RNA-Seq) mostraron alta expresión del gen de *Sodalis* TME1 codificante para la alantoinasa. Estos resultados indican que aunque existen diferencias en la comunidad microbiana de las dos cochinillas, al parecer en ambos modelos su microbiota converge para ayudar al reciclaje de ácido úrico aportando una fuente de nitrógeno extra para aliviar las carencias en la dieta de los insectos.

Abstract

Microbial interactions in phytophagous insects are responsible for expansion and evolution of these organisms. In this work two native Mexican insects *Dactylopius* (carmine cochineal) and *Llaveia axin axin* (lacquer cochineal) were studied. Fungi of genera *Rhodotorula*, *Cryptococcus* and *Penicillium* were isolated from *D. coccus* and *D. opuntiae*. These fungi could use uric acid (UA) as sole carbon and nitrogen source. Metagenomic analysis of *D. coccus* showed fungal sequences for UA catabolism. Moreover, significant differences in weight and length were observed in antifungal treated insects. UA was higher and uricases lower in treated insects in comparison with controls. Sequences for nitrogen fixing (*nif*HDK) from β-proteobacteria were detected in same *D. coccus* metagenome and they were phylogenetically related to those of *Uliginosibacterium ganwonense*. A 3.6 Mb genome assembly of all β-proteobacteria sequences was obtained and phylogenomics of 16S rRNA and core-genome sequences showed that this genome belonged to *Dactylopiibacterium carminicum*, a common symbiont of *Dactylopius* genus. Acetylene reduction assays and RT-PCR of *nif* genes showed that *D. carminicum* is an active nitrogen fixer in *Dactylopius*.

In another cochineal metagenome, from the lacquer cochineal, fungal ribosomal sequences of Ascomycetes were found. Furthermore, *Aspergillus* was revealed by a culture-dependent approach in *L. axin axin*. Likewise in *Dactylopius* uric acid catabolism genes were found in *L. axin axin* metagenome. Previous analysis of *L. axin axin* metagenomes has shown sequences of γ-proteobacteria. Here we report a 3.6 Mb genome assembly of this γ-proteobacteria, phylogenomics situated this genome as *Sodalis*. Average Nucleotide Identity and DNA-DNA Hybridization studies suggested that *Sodalis* genus could split into 4 phylo-species. Sodalis TME1 has all the genomic machinery to metabolically complement the biosynthesis of the essential amino acid and the truncated Krebs cycle in *L. axin axin* primary symbionts *Walczuchella*. Genes for purine catabolism were found and RNA-Seq expression analysis showed that allantoinase gene of *Sodalis* TME1 was over expressed. These results allow us to conclude that even though there are differences in microbiota composition between two cochineals here analyzed, metabolic capabilities for uric acid recycling are shared.

1. Introducción

1.1 Interacciones microbianas en insectos fitófagos.

En varios linajes de animales es común una relación estrecha con distintos grupos de microorganismos (bacterias, hongos y protozoarios). Este tipo de interacciones pueden ser benéficas (mutualismo), nulas (comensalismo) o dañinas (parasitismo). Los insectos son un grupo de animales con una alta variabilidad en comparación con otros habitantes de la biósfera ya que se encuentran distribuidos en la mayoría de los ambientes y pueden alimentarse de una extensa variedad de dietas. Esta variabilidad se debe al repertorio genético que poseen, el cual les permite adaptarse rápidamente así como a las asociaciones mutualistas que mantienen con microorganismos simbiontes los cuales les proveen el acceso a nuevas capacidades metabólicas (Janson et al., 2008; Douglas, 2012; Harris et al., 2010). Se considera que entre el 1-10 % de la biomasa de un insecto sano es biomasa microbiana (Douglas, 2015); ya que los insectos interactúan con microorganismos residentes y aunque en algunos casos los artrópodos no son dependientes obligados a su microbiota, hay evidencia de que estos microorganismos moldean diversos rasgos de sus hospederos, siendo una de las más importantes, el aporte de nutrientes extras carentes en la dieta (Douglas, 2014). La hervivoría ha acelerado la diversificación de las especies de insectos, sin embargo, los insectos fitófagos se enfrentan a varios retos a la hora de consumir este tipo de dietas como son: el desbalance nitrógeno-carbono, la presencia de compuestos tóxicos de la planta y la poca digestibilidad de los polímeros de alto peso molecular constituyentes de la pared de las células vegetales (Hansen and Moran, 2014). Es por ello, que en este tipo de insectos han evolucionado estrategias que les permitan adaptarse a este alimento; una de ellas es la asociación con microorganismos (bacterias, hongos o protistas) benéficos que les proveen de metabolitos carentes en el huésped.

1.2 Relaciones bacteria-insectos.

Las bacterias mutualistas son aquellas que se presentan en forma persistente, no invasivas y que proveen de nutrientes y/o capacidades metabólicas benéficas al huésped (Hansen and Moran, 2014). Algunas de estas capacidades son: el uso de dietas deficientes en nutrientes, resistencia a parásitos y patógenos, y la transferencia de genes que intervienen en las vías de síntesis y/o degradación de metabolitos, que favorecen la especiación (Moran, 2006; Harris et al., 2010). El conocimiento de las relaciones bacteria-insecto se rastrea hasta mediados del siglo XX con las observaciones de Buchner en áfidos. En su libro "Endosymbiosis of Animals with Plant Microorganisms", Buchner describe la relación de bacterias habitando distintos tejidos de insectos fitófagos (Buchner, 1965). Se sabe que este tipo de simbiontes alteran funciones celulares de los huéspedes que van desde la transducción de señales hasta la apoptosis (Bentley et al., 2007; Ikeya et al., 2009). Debido a lo recalcitrante que es cultivar este tipo de bacterias, algunas de estas funciones sólo han sido predicciones llevadas a cabo por la reconstrucción genómica de los simbiontes a partir de metagenomas de los huéspedes.

La localización de bacterias en los insectos es variada, pueden habitar la cutícula del insecto, como los actinomicetos de hormigas y abejas (Kaltenpoth, 2009), en las criptas intestinales o el lumen propio del canal alimentario (Heddi and Gross, 2011; Hansen and Moran, 2014) las gónadas y canales reproductores (Matsuura et al., 2012a, 2012b; Watanabe et al., 2014) o bien, forman órganos especializados llamados bacteriomas; dónde células "gigantes" del insecto (bacteriocitos) albergan

bacterias en su citoplasma (Douglas, 1998; Baumann, 2005; Matsuura et al., 2012a). Se sabe que estas bacterias son adquiridas por transferencia directa de la madre a la cría. Este último grado de asociación se ha descrito en al menos 15 % de las especies de insectos y se cree que resultó de una antigua infección que dio como origen una co-cladogénesis entre los dos organismos (Baumann, 2005). En esta relación endosimbiótica la bacteria hospedera es brindada de un hábitat estable que la protege de la competencia con otros organismos, mientras que recibe un flujo constante de nitrógeno y carbono por el insecto. En algunos hospederos estas relaciones son tan estrechas que en los bacteriomas las señales inmunes de defensa contra bacterias son casi nulas (Douglas et al., 2011; Login et al., 2011). Sin embargo en algunos modelos, como lo son los gorgojos, las señales de inmunidad incrementan para controlar y disminuir el número de simbiontes intracelulares cuando éstos ya no son requeridos, o han cumplido con las funciones metabólicas que establece la simbiosis (Login et al., 2011; Vigneron et al., 2014; Masson et al., 2015). Incluso, hay registros de que algunos mutualistas poseen la capacidad de "manipular" la biología del huésped para asegurar su descendencia heredándose a las siguientes generaciones. El ejemplo más común es la infección con algunas cepas de *Wolbachia* o *Spiroplasma* (α-Proteobacteria), las cuales producen incompatibilidad citoplásmica o feminización de la especie (Werren et al., 2008; Harris et al., 2010; Zug and Hammerstein, 2014).

Hasta el momento se han reportado dos maneras en las que los simbiontes de insectos pueden ser adquiridos, la primera como se mencionó es una transmisión

directa de las madres a las crías (transmisión vertical), como es el caso de las bacterias endosimbiotes de los insectos fitófagos. La manera en que la madre transmite las bacterias a la progenie puede ser por la contaminación bacteriana en la superficie de los embriones (recubrimiento de huevecillos), por el excremento de la madre (proctofagía) o bien por la deposición de una cápsula llena de simbiontes ("pelotilla" simbiótica) en los huevecillos (Buchner, 1965; Kikuchi et al., 2005). Por otro lado, se encuentra una trasferencia de manera ambiental, es decir bacterias del medio que al ser ingeridas se establecen en el interior del insecto (e.g *Burkholderia* en chinches de la familia Alydidae; Kikuchi et al., 2005; Kim et al., 2013). Este mecanismo de adquirir bacterias ambientales pudo desencadenar la generación de endosimbiontes, donde bacterias de vida libre (algunas patógenas) fueron engullidas por el huésped y con los años se llegó a un mutualismo. Se infiere que para que este mutualismo suceda se requiere un proceso de comunicación celular por parte de la bacteria hacia el insecto, es por ello que en algunos modelos propuestos como simbiontes de adquisición reciente se mantienen activos sistemas de secreción u otros mecanismos de exportación de moléculas de señalización (Dale et al., 2001, 2005; Husník et al., 2011).

En hemípteros, la clase de bacterias asociadas más comúnmente reportada en las proteobacterias. Dentro de las cuales las α, β y γ proteobacterias se han encontrado como las más representativas en endosimbiontes primarios intracelulares de insectos (Tabla 1) (Heddi and Gross, 2011).Aunque, la presencia de Bacteroidetes, Mollicutes (eg. *Spiroplasma*) y Flavobacteiras (Tabla 1) también han sido reportadas

como endosimbiontes mutualistas en este tipo de insectos (Rosenblueth et al., 2012; Hansen and Moran, 2014; Bolaños et al., 2015). Estas bacterias establecen relaciones nutricionales, en las cuales los endosimbiontes producen aminoácidos esenciales, vitaminas y otros metabolitos que el insecto no puede producir. Todos los metazoarios somos auxótrofos de diez aminoácidos (llamados esenciales) que a continuación se listan, aromáticos: fenilalanina, y triptófano; básicos: histidina, arginina, y lisina; y neutros: leucina, isoleucina, treonina, valina y metionina. Además de ser incapaces de sintetizar vitamina B12 y aparentemente riboflavina. Por lo que la asociación con bacterias productoras de estos metabolitos genera una ventaja evolutiva.

El estudio de las relaciones mutualistas en insectos ha incrementado considerablemente gracias al advenimiento de la genómica que hizo posible el estudio de los endosimbiontes ya que como se ha mencionado son difíciles de cultivar y mantener en laboratorio. El primer genoma de un endosimbionte primario (*Buchnera*) asociado a los áfidos reveló diversos genes faltantes para poder ser una bacteriea de "vida libre" (Shigenobu et al., 2000). Debido a la ausencia de genes claves para la síntesis de pared celular, síntesis de fosfolípidos, el número reducido de transportadores y proteínas de la superficie celular así como un genoma reducido, se cuestionó si las bacterias endosimbiontes se debían tratar como células vivas u organelos de las mismas (Andersson, 2000). Posteriormente se propuso el término "simbionelo" para designar a algunos de los endosimbiontes con genomas extremadamente reducidos (Reyes-Prieto et al.,

2014), de los que se cuestiona su papel en simbiosis. Otras características de los genomas de endosimbiontes primarios son un alto grado de pseudogenización, proliferación de elementos móviles, múltiples re-arreglos genéticos, recambios en el uso de codones, un sesgo en la composición nulceotídica así como una evolución rápida de sus secuencias (McCutcheon and Moran, 2012; Hansen and Moran, 2014). Algunos de estos cambios en la dinámica de los genomas de endosimbiontes pueden deberse a la pérdida de elementos de edición y corrección de las secuencias genéticas. El grado de co-cladogénesisis entre los endosimbiontes y sus hospederos está estrictamente ligado al aporte benéfico que dan las bacterias al huésped. Esta co-cladogénesis se refleja en el mantenimiento de vías para la síntesis de aminoácidos esenciales o vitaminas carentes en las dietas del huésped (Bennett and Moran, 2015). Sin embargo, se han reportado casos en donde el simbionte no es capaz de seguir cumpliendo con la función necesitada por el huésped y este es remplazado por un nuevo simbionte, asegurando con esto la sobrevivencia del insecto y haciendo un "callejón sin salida en la herencia" para simbiontes con alta pérdida de funcionalidad genética (Hansen and Moran, 2014; Bennett and Moran, 2015). Este remplazo de simbiontes generalmente se da por bacterias (endosimbiontes secundarios) presentes previamente en los lugares donde se encuentran los endosmibiontes primarios y en algunas ocasiones el recambio de simbiontes no sólo es de bacterias si no de procariontes por microeucariontes (hongos), los cuales cumplen con las mismas funciones metabólicas que los simbiontes remplazados (Zacchi and Vaughan-Martini, 2003; Sacchi et al., 2008;

Jones et al., 1999). Las funciones remplazadas son: el aporte de vitaminas, síntesis de feromonas o bien participar en el ciclo del nitrógeno.

Tabla 1. Ejemplos de bacterias endosimbiontes primarios presentes en insectos hemípteros.

Orden	Suborden	Familia	Endosimbionte Primario	ClaseTaxonómica bacteriana
Hemiptera	Heteroptera	Cimicidae (Chinche de cama)	*Midichloria*	Alphaproteobacteria
		Pentatomidae (Chinches apestosas)	*Ishikawaella*	Gammaproteobacteria
	Auchenorrhyncha	Fulgoridae (Cigarras)	*Vidania fulgoroideae*	Betaproteobacteria
	Auchenorrhyncha	Cicadellidae (chicharritas)	*Sulcia* *Baumannia*	Gammaproteobacteria
			Hodgkinia	Alphaproteobacteria
	Sternorrhyncha	Aphididae (Pulgones)	*Buchnera*	Gammaprotebacteria
		Psyllidae (Psílidos)	*Carsonella*	Gammaproteobacteria
		Aleyrodidae (Mosquita blanca)	*Portiera*	Gammaproteobacteria
		Pseudococcidae (Cochinillas)	*Tremblaya*	Betaproteobacteria
		Monophlebidae (cochininllas gigantes)	*Walczulchella*	Flavobacteria
		Dactylopiidae (Cochinilla del carmín)	*Dactylopiibacterium carminicum*	Betaproeobacteria

1.3 Fijación de nitrógeno por bacterias diazótrofas simbiontes de insectos.

El nitrógeno es el elemento más abundante en la atmósfera terrestre (79%) y aunque todos los organismos requieren este elemento para la síntesis de biomóleculas sólo los procariontes poseen la maquinaria enzimática para adquirir y biotransformar este elemento directamente del aire (Dixon and Kahn, 2004). Para los insectos fitófagos cuya dieta es pobre en nitrógeno el relacionarse con bacterias diazótrofas confiere una ventaja al colonizar sus plantas hospederas. La fijación biológica del nitrógeno (FBN) se ha observado en distintos linajes de insectos como hormigas (Pinto-Tomás et al., 2009; Sapountzis et al., 2015), termitas (Desai and Brune, 2012), cucarachas (Tai et al., 2016), escarabajos (Morales-Jiménez et al., 2009, 2013) y moscas (Murphy et al., 1988; Behar et al., 2005). No obstante, hasta la fecha no existen reportes de FBN en hemípteros. Esta FBN se ha atribuido principalmente a bacterias simbiontes de las clases γ-proteobacterias y bacteroidetes. Estos simbiontes se han reportado como bacterias intestinales (Behar et al., 2005; Morales-Jiménez et al., 2013) o bien como bacterias asociadas a los jardines fúngicos de las hormigas de la tribu Attini (Pinto-Tomás et al., 2009). Técnicas como la medición de la expresión de los genes codificantes para la nitrogenasa (*nif*), ensayos de reducción de acetileno (ARA) *in vivo* o la cuantificación de N^{15} depositado en los tejidos de los insectos han servido para demostrar el papel de las bacterias simbiontes en el aporte de nitrógeno fijado hacia los insectos (Morales-Jiménez et al., 2013; Tai et al., 2016). No obstante, se desconoce cuantitativamente la contribución neta de nitrógeno que se adquiere por

esta vía. Además, son inciertos los mecanismos por los cuales las bacterias simbiontes exportan el nitrógeno fijado.

1.4 Relaciones hongos-insectos.

Al momento se han estudiado las interacciones de al menos ocho órdenes distintos de insectos fitófagos como las hormigas, termitas, áfidos, cigarras y escarabajos que mantienen algún tipo de interacción con hongos levaduriformes o filamentosos (Vega and Blackwell, 2005; Gibson and Hunter, 2010). Estos hongos pueden encontrarse habitando cavidades dentro del insecto como son el micangio en escarabajos descortezadores (Jones et al., 1999; Klepzig and Six, 2004; Ganter, 2006), o bien en células altamente especializadas llamadas micetocitos, como en las cigarras *Nilaparvata lugens* o en las moscas *Drosophila melanogaster* (Cheng and Hou, 2001; Ebbert et al., 2003). Así mismo, los hongos pueden alojarse recubriendo tejidos del huésped como el canal alimentario, las gónadas, e incluso glándulas productoras de veneno como en la avispa *Comperia merceti* (Gibson and Hunter, 2009; Rivera et al., 2009; Ricci et al., 2011a, 2011b). En años recientes ha habido un auge en el estudio de hongos asociados a artrópodos y se sugiere que la mayor diversidad de hongos unicelulares podría encontrarse en intestinos de insectos, principalmente escarabajos (Suh et al., 2005). La adquisición de los hongos por parte del insecto puede ser, al igual que las bacterias intestinales, ambiental principalmente por micofagia. En esta forma de adquisición, los insectos se alimentan directamente del micelio o de las levaduras presentes en su dieta, como sucede en varias especies de *Drosophila* (Gibson and Hunter, 2010; Becher et al.,

2012). Por otro lado, se ha reportado la transferencia vertical (padres-crías), como es el caso de *N. luggens* y sus simbiontes tipo levadura (YLS) (Cheng and Hou, 2001), los mosquitos A*nopheles stephensi* y la levadura *Wickerhamomyces anomalus* (Ricci et al., 2011a, 2011b) y por último en los escarabajos anóbidos y el hongo *Symbiotaphrina* spp. (Noda and Kodama, 1996). En los últimos ejemplos, las células fúngicas penetran los huevecillos de los huéspedes formando una "pelotilla simbiótica", dejando las nuevas generaciones inoculadas con los simbiontes. Por otro lado, los hongos simbiontes de los insectos mico-cultivadores como las hormigas mirmicinos de la tribu Attini, los escarabajos descortezadores del género *Dendroctonus* (Coleoptera: Scolytidae) y algunas abejas meliponias (*Scaptotrigona depilis*) (D'Ettorre et al., 2002; Rivera et al., 2009; Menezes et al., 2015), aunque también son transmitidos de manera vertical, estos no son depositados directamente en los huevecillos de las crías. A diferencia, estos son llevados de un sitio de crianza a otro por la "reina" de la colonia. El micelio es cortado de los "jardines" fúngicos por el insecto y este es inoculado en las cámaras de crecimiento de las pupas, el cual sirve de alimento a las larvas. En ambos casos la eliminación de los hongos antes de su inoculación causa una disminución en la adecuación de sus hospederos (Sasaki et al., 1996; Menezes et al., 2015).

Algunas de las funciones que desempeñan los hongos mutualistas de insectos son: el aporte de metabolitos nitrogenados y lípidos esenciales carentes en la dieta; la degradación de polímeros biológicos de alta densidad, el reciclaje del ácido úrico, la biotransformación de químicos tóxicos del ambiente y la producción de

feromonas (D'Ettorre et al., 2002; Nasir and Noda, 2003; Vega and Blackwell, 2005; Adams et al., 2011); mientras que los insectos funcionan como vectores y proveen un hábitat favorable para los hongos. No obstante, los efectos netos en la adecuación de los hospederos por parte de los hongos ha sido demostrada en pocos casos concretos como lo son las cigarras (Sasaki et al., 1996) y algunos escarabajos (anóbidos y escolitinos) (Ayres et al., 2000; Nasir and Noda, 2003). Por lo anterior, es importante conocer con más detalle la interacción hongos-insectos.

1.5 Reciclaje del ácido úrico por hongos asociados a insectos.

El ácido úrico (AU) es el principal metabolito de desecho del ciclo del nitrógeno en los insectos terrestres. Se ha estimado que este compuesto constituye cerca del 80 % del producto del catabolismo nitrogenado en estos organismos (Pant, 1988). El AU es movilizado por la hemolinfa a partir del cuerpo graso hasta los túbulos de Malpighi. Estos órganos encargados de la excreción en insectos, se encuentran localizados generalmente entre la unión del intestino medio y el intestino anterior (Fig. 1) (Cochran, 1985; Chapman, 2013). Están distribuidos en prácticamente todos los insectos, a excepción de los áfidos. Los túbulos pueden variar en número desde dos en los insectos escama hasta 150 en las langostas del desierto (Chapman, 2013). Es en estos órganos que el AU y otros desechos son depositados para después ser metabolizados o bien ser excretados en forma de cristales en la orina del insecto (Pant, 1988). Para poder ser metabolizado, el AU tiene que ser bioconvertido a allantoína por la enzima uricasa o urato oxidasa (EC 1.7.3.3; Fig. 2) (Gabison et al., 2008). Aunque la mayoría de los insectos carecen de

los genes para la síntesis de esta enzima en algunos de ellos no ha sido posible detectar la presencia de AU o los productos subsecuentes del metabolismo de las purinas (Fig2; Fig. 3 Anexo 1) en su orina (Mullins and Cochran, 1975; Hongoh et al., 2000; Sabree et al., 2009). Por lo que se ha propuesto que en hemípteros, coleópteros, blátidos e himenópteros el AU es metabolizado por uricasas de microorganismos simbiontes (Potrikus and Breznak, 1980; Sasaki et al., 1996; van Borm et al., 2002; Morales-Jiménez et al., 2013; Patiño-Navarrete et al., 2014). Particularmente, en la cigarra *Nilaparvata lugens* el AU es catabolizado por simbiontes tipo levadura alojados en el cuerpo graso del insecto (Sasaki et al., 1996; Hongoh et al., 2000). Estos hongos, movilizan el AU depositado en los órganos del huésped y lo metabolizan a aminoácidos que luego el insecto aprovecha (Hongoh and Ishikawa, 1997). Hasta el momento éste es el único caso reportado donde los hongos reciclan AU a metabolitos útiles para los insectos.

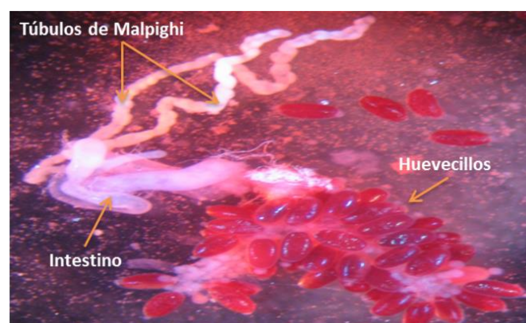


Figura 1. Estructuras internas de *Dactylopius coccus*. Se muestran los túbulos de Malpighi, intestino y ovarios con huevecillos

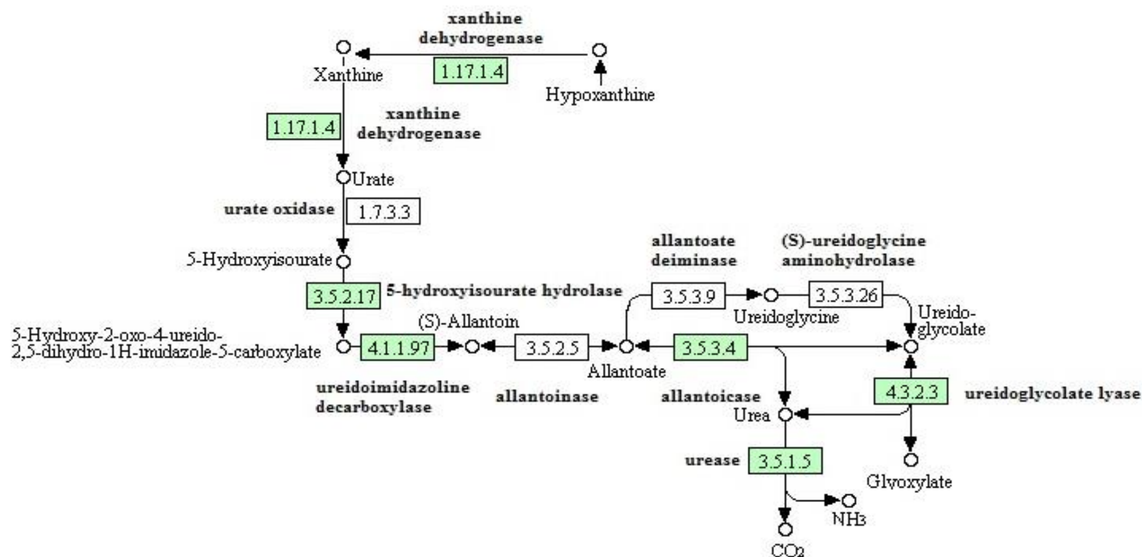


Figura 2. Metabolismo del ácido úrico desde hypoxantina (precursor) hasta amonio y CO_2 (producto final). Se muestran las enzimas participantes en el catabolismo. Imagen tomada del KEGG pathway mapper tool (http://www.kegg.jp/kegg/tool/map_pathway.html)

1.6 La cochinilla del carmín *Dactylopius* spp. (Hemiptera:Dactylopiidae).

El género *Dactylopius* comprende 11 especies *Dactylopius ceylonicus* (Green), *Dactylopius austrinus* (De Lotto), *Dactylopius bassi* (no reconocible), *Dactylopius coccus* (Costa), *Dactylopius confertus* (De Lotto), *Dactylopius confusus* (Cockerell), *Dactylopius gracilipilus* (Van Dam & May), *Dactylopius opuntiae* (Cockerell), *Dactylopius salmianus* (De Lotto), *Dactylopius tomentosus* (Lamark), *Dactylopius zimmermanni* (De Lotto) (Williams and Ben-Dov, 2015). Los individuos dentro de este grupo de hemípteros presentan cuerpos negros de 1−6 mm de longitud. Exhiben un dimorfismo sexual siendo los machos alados y más pequeños que las hembras (Fig. 3). Tienen un ciclo de vida hemimetábolo de aproximadamente 110

días, pasando por los estadios de huevecillo, ninfa I, ninfa II o pupa y adulto (Perez-Guerra and Kosztarab, 1992). Durante los primeros días después de eclosionar las ninfas caminan sobre el cladodio de los cactáceos huéspedes hasta que encuentran un lugar propicio para clavar su aparato bucal (estilete). Las hembras adultas son sésiles y presentan una talla aproximada de 3 mm. A los 58 días después de eclosionar se presenta la cópula, las hembras son fecundadas y 3-4 días después los machos mueren. Las hembras ovopositan alrededor de 200 huevecillos, muriendo pocas semanas después de la ovoposición (Pérez-Guerra, 1991).

Todos los individuos de ambos sexos se encuentran cubiertos por una secreción de cera algodonosa que usan como protección contra depredadores y cambios en el ambiente (Chávez-Moreno et al., 2009).

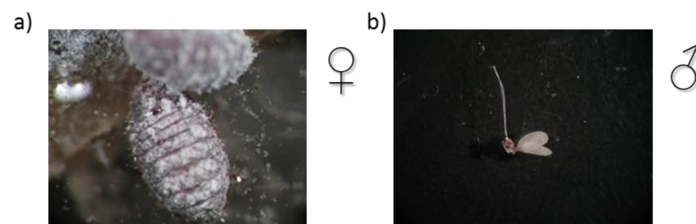


Figura 3. Dimorfismo sexual presente en *Dactylopius*. a) Hembra. b) Macho

Este género de insectos es la fuente de ácido carmínico (carmín o rojo E120); un glúcido antraquinónico (Fig. 4) usado industrialmente como colorante o pigmento en alimentos, cosméticos, fármacos y textiles (Deveoglu et al., 2011), el cual es 50 % del peso húmedo del insecto (Stintzing and Carle, 2005). Aunque todas las especies del género producen ácido carmínico sólo *D. coccus* se ha cultivado desde época

precolombina para la extracción de carmín, debido a su alta concentración y calidad (Rodríguez et al., 2005).

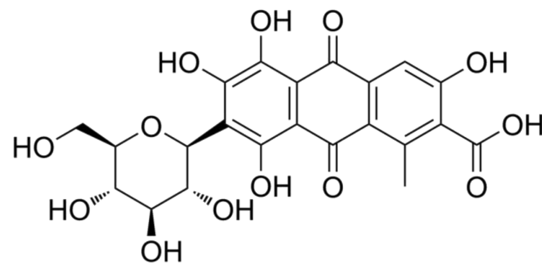


Figura 4. Estructura del ácido carmínico (ácido 7-α-D-Glucopyranosyl-9,10-dihydro-3,5,6,8tetrahydroxy-1-methyl-9,10-dioxoanthracenecarboxílico).

Estos insectos son exclusivamente fitófagos, alimentándose de la savia de cactáceas de los géneros *Opuntiae* y *Nopalea* (Cactaceae: Opuntioidae) (Chávez-Moreno et al., 2009). La sabia está compuesta principalmente de agua (85-95%), carbohidratos (3-7 %) y fibra (1-2%); sin embargo es pobre en proteínas (0.5-1%) y lípidos (0-0.2%) ((Stintzing and Carle, 2005). Se sugiere que evolutivamente, estos insectos han encontrado estrategias que permitan contrarrestar la carencia nutricional de su dieta. Una de estas estrategias es la asociación con microorganismos simbiontes los cuales proporcionen los elementos carentes en la comida.

1.7 Microorganismos asociados a *Dactylopius* spp.

Existen cuatro reportes que exhiben la diversidad de microorganismos asociadas a las cochinillas del género *Dactylopius* (Pankewitz et al., 2007; Ramírez-Puebla et al., 2010, 2015, 2016). Estos trabajos se han centrado en la presencia de bacterias, principalmente las clases α, β y γ proteobacterias (Ramírez-Puebla et al., 2010). Dentro de la clase α-proteobacteria se ha descrito con detalle la presencia de

Wolbachia asociada a *D. coccus*. Gracias a los estudios de estas bacterias el grupo de trabajo logró proponer el uso de especies en el género (más allá de *Wolbachia pipientis*) y no sólo denominarlas como supergrupos (Ramírez-Puebla et al., 2015, 2016). Particularmente se reconoció a *Candidatus* Wolbachia bourtzisii *w*DacA y a *Candidatus* Wolbachia pipientis *w*DacB como simbiontes de *D. coccus* (Ramírez-Puebla et al., 2015, 2016; Anexo 5 y 7). Así mismo, el análisis de clonación de genes ribosomales bacterianos (16S rDNA) a partir de DNA total extraído de varias especies de *Dactylopius* reconoció algunas α−proteobacterias como: *Sphingomonas*, *Mesorhizobium* y *Candidatus* Hepatincola porcellionum. Las cuales se han reportado como endófitas de plantas. Doce ribo-clonas de *D. confusus* se emparentaron con *Acinetobacter* (γ-proteobacteria). Por otro lado, miembros de las β−protobacterias como: *Herbaspirillum* y *Massilia* también fueron reportadas. Particularmente, dentro de esta clase bacteriana se describe a un filogrupo bactriano (END-1) emparentado con *Uliginosibacteriu gangwoense* presente en todas las especies de *Dactylopius* muestreadas y la única en *D. coccus* (Ramírez-Puebla et al., 2010). Sin embargo, todos estos reportes se han limitado al análisis de la diversidad bacteriana y ninguno ha descrito el papel de los microorganismos asociados a las cochinillas del carmín. Aunado a esto, no existen reportes sobre la comunidad fúngica asociada a los insectos del género *Dactylopius* y su papel en la interacción con el huésped.

1.8 La cochinilla de la laca *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae)

Dentro del género *Llaveia* se encuentra una especie de cochinilla llamada *L. axin axin* o niij (Fig 1 Anexo 4) por los nativos mesoamericanos. Su ciclo de vida es anual y presenta hemimetabolismo. Como todos los insectos escama, es exclusivamente fitófago, se alimenta de sabia de plantas como: *Acacia cochliacantha*, *Acaciella angustissima, Jatropha curcas* y *Spondias* sp. entre otras (Rincón-Rosales and Gutiérrez-Miceli, 2008). Este hemíptero es de valor económico en el sur de México, ya que del mismo se extrae una laca usada para recubrir artesanías (Williams and MacVean, 1995). Cabe mencionar que la población de nijj ha disminuido últimamente, principalmente debido a la sobre explotación, deforestación y quema de los bosques donde habita (Rincón-Rosales and Gutiérrez-Miceli, 2008; Rosas-Pérez et al., 2014). La biología de este insecto es poco conocida, sin embargo la microbiota de estos insectos ha empezado a explorarse a partir de muestras metagenómicas. Se ha encontrado a una flavobacteria (nombrada *Candidatus* Walczuchella monophlebidarum) como el simbionte primario de estos insectos (Rosas-Pérez et al., 2014). Por otro lado, del mismo metagenoma se han encontrado bacterias propuestas como endosimbiontes secundarios, dentro de ellas una gamaproteobacteria parecida a *Sodalis* y una alfa protebacteria (Rosas-Pérez et al., 2014). A la fecha no existen otros reportes de hongos simbiontes asociados a este insecto.

En esta tesis analizamos: 1) La relación que mantienen diversos géneros fúngicos asociados a tres especies de *Dactylopius* (*D. coccus*, *D. opuntiae* y *D: confusus*). A partir de reconstrucciones metagenómicas, técnicas cultivables y análisis metabólicos, se demostró que estos hongos participan activamente en el reciclaje de ácido úrico dentro del huésped. 2) De igual manera se estudió la comunidad fúngica asociada a la cochinilla de la laca *L. axin axin* usando técnicas dependientes de cultivo y dos metagenomas, de los cuales se infirió que estos hongos también pueden participar en el reciclaje de ácido úrico en el insecto. 3) Del mismo modo de tres metagenomas obtenidos de *D. coccus* se ensamblo el genoma de *Ca. Dactylopiibacterium carminicum*, una bacteria comúmente reportada en asociación al género *Dactylopius*. Se encontró que esta bacteria posé toda la maquinaria genética para fijar nitrógeno y que puede habitar ambientes con potencial REDOX reducidos. A partir de experimentos de reducción de acetileno *in vivo* y ensayos de expresión de los genes para fijar nitrógeno (*nif*), se propone que esta bacteria puede participar en el aporte de nitrógeno al huésped a partir de la diazotrofía. 4) Por último, haciendo uso de los datos metagenómicos obtenidos del "nijj", se reconstruyó el genoma de una γ-proteobacteria previamente descrita como co-simbionte de *Ca.* Walczuchella monophlebidarum, el simbionte primario de *L. axin axin*. Este genoma, perteneciente al género *Sodalis*, codifica para genes de síntesis de aminoácidos y metabólitos del ciclo de Krebs carentes en *Ca.* Walczuchella monophlebidarum. También se pudo inferir que el género *Sodalis* puede ser dividido en tres filogrupos de acuerdo a análisis de génomica comparada.

Capítulo 1. Reciclaje del ácido úrico por hongos asociados a la cochinilla del carmín *Dactylopius* spp.

Esta sección se basa en el trabajo publicado como producto del proyecto de investigación: **Fungal community associated with *Dactylopius* (Hemiptera: Coccoidea: Dactylopiidae) and its role in uric acid metabolism. Arturo Vera-Ponce de León** , Alejandro Sanchez-Flores , Mónica Rosenblueth and Esperanza Martínez-Romero **Front. Microbiol. 7:954. doi: 10.3389/fmicb.2016.00954**.

- **Planteamiento del problema.**

Las cochinillas del carmín (*Dactylopius*) se enfrentan a carencias nutricionales debido a la deficiencia de metabolitos nitrogenados y lípidos en la savia que consumen. La asociación con microorganismos puede ser una alternativa para combatir dichas deficiencias. Aunque existen reportes de bacterias asociadas a *Dactylopius* ninguno de estos ha explorado la diversidad fúngica asociada a este grupo de insectos. El estudio de los hongos asociados a las cochinillas del carmín y sus capacidades metabólicas podría explicar la adaptación de estos insectos para colonizar nopales.

- **Hipótesis:**

Los hongos asociados a las cochinillas del género *Dactylopius* spp. aportan nutrientes (metabolitos nitrogenados y lípidos) a su hospedero lo cual ayudará a contrarrestar la carencia de dichas moléculas en la dieta del insecto.

- **Objetivo**

Identificar hongos del interior de diversas especies de cochinillas del género *Dactylopius* y proponer algunas funciones en la simbiosis con los insectos.

1. Objetivos Particulares

1. Aislar e identificar hongos por métodos dependientes e independientes de cultivo usando como marcadores moleculares a las regiones del espaciador intergénico "ITS" y el gen 26S rRNA.
2. Identificar la comunidad fúngica del interior de *D. coccus* mediante un enfoque metagenómico "Shotgun metagenomics" y analizar la presencia de genes de hongos relacionados con la vía de la degradación y transformación del ácido úrico.
3. Cuantificar el ácido úrico en hemolinfa, ovarios y túbulos de Malpighi de diferentes especies de *Dactylopius* y medir la actividad enzimática de la uricasa en dichos tejidos.
4. Cuantificar la capacidad uricolítica de diversos hongos aislados y su relación en el reciclaje de metabolitos nitrogenados en el huésped.
5. Localizar a los hongos en *Dactylopius* usando sondas para hibridación fluorescente *in situ* (FISH).

Resultados

A partir de métodos dependientes de cultivo se aislaron 37 hongos del canal alimentario, ovarios y cuerpos completos de *D. coccus*, *D. opuntiae*, y *D. confusus*. Estos se agruparon en 14 filogrupos dentro de las clases Ascomycota y Basidiomycota (Anexo 1: Fig. 1, Fig. supl. 1. Tabla 2). Se observó que la presencia de las distintas unidades taxonómicas operativas (OTUs por sus siglas en inglés) no es uniforme en todas las especies de *Dactylopius* muestreadas (Fig. 5).

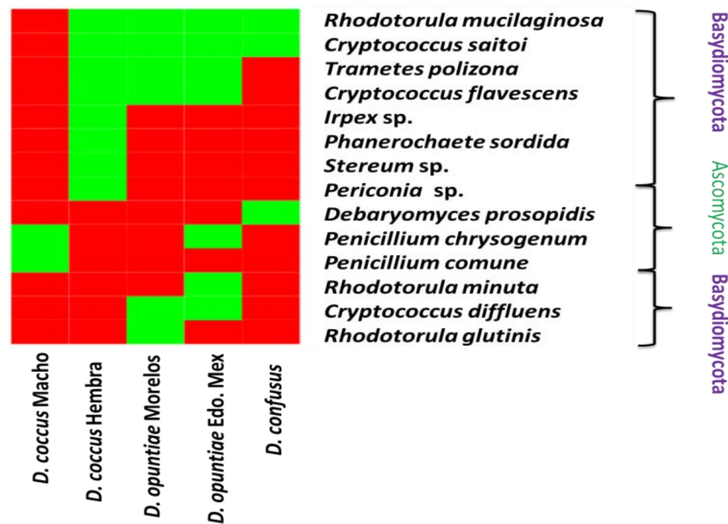


Figura 5. Agrupamiento por presencia-ausencia de OTUs fúngicos recuperadas por cultivo de distintas especies de *Dactylopius*. Rojo=Ausencia; Verde=Presencia.

Rhodotorula mucilaginosa y *Cryptococcus saitoi* fueron los hongos comunes en las tres especies de insectos (*D. coccus*, *D. opuntiae*, *D. confusus*) (Fig. 5). Por otro lado, el análisis de cultivo permitió encontrar miembros de los géneros *Trametes* y *Phanerochaete* los cuales son hongos basidiomicetos caracterizados por su potencial metabólico para la degradación de moléculas de alto peso molecular como

colorantes, lignina y celulosa lo cual sugeriría que estos hongos pudieran aprovechar el carmín como fuente de carbono.

El análisis metagenómico del DNA de intestinos y hemolinfa de *D. coccus* mostró la presencia de secuencias ribosomales (18S rDNA) de las clases de hongos Ascomycota, Basidiomycota, Glomeromycota y Chitridiomycota (Fig. 2 Anexo 1). Se observó diferencia entre las clases encontradas en hemolinfa y canal alimentario, siendo los ascomicetos y basidiomicetos las clases más frecuentemente encontradas en hemolinfa, mientras que en los intestinos los basidiomicetos y secuencias de hongos no-clasificados fueron los más frecuentes (Fig 2 Anexo 1, hoja suplementaria 1 Anexo 1). Así mismo, la reconstrucción *in silico* de los genes ribosomales de hongos apartir de dos metagenomas públicos de Oaxaca México (DCoax) y Peru (DCperu) (Campana et al., 2015), mostró que en la muestra DCoax los hongos no-clasificados (al igual que los intestinos) son más frecuentemente encontrados (Fig. 2, Anexo 1; Hoja suplementaria 1, Anexo 1). Mientras que en la muestra DCperu, el único género de secuencias de hongos encontradas fue *Candida* sp. (Hoja suplementaria 1, Anexo 1).

Por otro lado, en el mismo estudio se logró determinar la cantidad de ácido úrico (AU) y la actividad de uricasas presentes en dos especies de *Dactylopius* (*D. coccus* y *D. opuntiae*). Se observó que existe diferencia significativa en la concentración de AU y uricasa en los distintos estadios del ciclo de vida del insecto (Fig. 4A y 4B; Anexo 1). El análisis metagenómico no reveló ningún gen codificante para uricasas

de insectos, sin embargo se encontraron 20 y 85 genes de hongos que codifican para uricasas en los metagenomas de la hemolinfa y canal alimentario respectivamente (Tabla suplementaria 1 Anexo 1). En los tres metagenomas de *D. coccus* estudiados se encontraron genes fúngicos que codifican para enzimas que participan en la ruta de biotransformación de AU hasta amonio (Fig. 3, Tabla suplementria 1,2 y 3; Fig. suplementaria 3 Anexo 1). Por otro lado, se logró determinar la presencia de AU y actividad enzimática de uricasas en diferentes tejidos de dos especies de *Dactylopius* a lo largo de su ciclo de vida (Fig. 4, Anexo 1). El análisis de las cinéticas de crecimiento usando AU como única fuente de nitrógeno y carbono, mostraron que las levaduras *Cryptococcus saitoi*, *Rhodotorula mucilaginosa*, *Rhodotorula minuta* y *Cryptococcus flavescens* así como el hongo filametoso *Penicillium* son capaces de metabolizar el AU (Fig. 5 Anexo 1). Además, en insectos tratados con antifúngicos (Fludioxonil, Anfotericina B y Cicloheximida), se observó una diferencia significativa en la talla y peso de los insectos tratados en comparación con los controles (sin anti fúngico) (Fig. 6; Fig. 6 Anexo 1 y Figura suplementaria 4; Anexo 1).

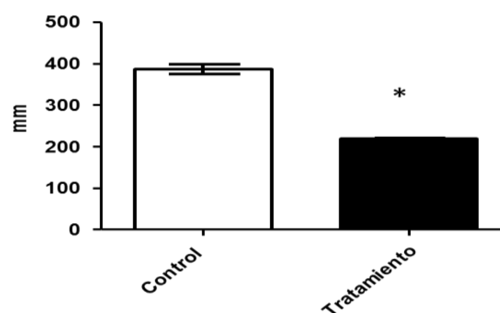


Figura 6. Talla en milímetros de insectos tratados con antifúngico. * Muestra diferencia significativa entre tratamientos (Prueba de t 14 g.l= 4 P=0.0001). Se presenta la media ± SEM de cinco experimentos independientes.

Además la concentración de AU fue significativamente mayor y actividad enzimática de uricasas fueron significativamente menores en los insecto tratados en comparación con los controles (n= 6; t=10.05; P=0.0098 g.l =2 y n= 6; t=3.671; P=0.014; g.l=4; respectivamente Fig. 6 Anexo 1).

Por último, los experimentos de FISH identificaron la presencia de *Cryptococcus* (levadura uricolítica) en los túbulos de Malpighi y huevecillos de *D. coccus* y *D. opuntiae* (Figura 7 Anexo 1). Toda esta evidencia sugiere que los hongos pueden participar en el proceso de reciclaje de ácido úrico en el insecto, proveyendo de fuentes alternativas de nitrógeno para el huésped.

Capitulo II. Hongos asociados a la cochinilla de la laca "nijj" *Llaveia axin axin*.

Esta sección se explora el trabajo en prensa: **Metagenomic analysis of the fungal community associated with the scale insect *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae). Arturo Vera-Ponce de León,** Tania Rosas-Pérez, Mónica Rosenblueth, Esperanza Martínez-Romero.

2. Planteamiento del problema:

La cochinilla de la laca *Llaveia axin axin* ha servido a los artesanos mexicanos como fuente de grasas para el recubrimiento de artesanías. Sin embargo la biología de este insecto ha sido poco estudiada, particularmente se ha investigado la asociación simbiótica con bacterias relacionadas con el nijj sin embargo no se ha estudiado si existe una comunidad fúngica relacionada con el insecto. Este trabajo pretende explorar los hongos asociados al nijj así como estudiar las posibles funciones que estos jueguen en la asociación con los insectos.

3. Hipótesis.

Si existen hongos asociados a la cochinilla de la laca *L. axin axin* entonces estos podrían aportar metabolitos útiles para el insecto.

4. Objetivo

Identificar en distintos tejidos de *L.axin axin* hongos asociados por métodos dependientes de cultivo e inferir posibles roles metabólicos a partir de genes fúngicos encontrados en metagenomas del insecto previamente reportados.

5. Objetivos particulares

1. Aislar e identificar hongos usando como marcadores moleculares a las regiones ITS y el gen 26S rRNA.

2. Identificar la comunidad fúngica asociada a *Llaveia axin axin* a partir de explorar metagenomas de bacteriomas y un macerado abdominal del insecto.

3. Explorar genes fúngicos de estos metagenomas que pudieran participar en el aporte de nutrientes al huésped.

Resultados

El estudio de la comunidad fúngica cultivable de *Llaveia axin axin* nos llevó a aislar 3 hongos del interior de diferentes insectos colectados en distintos años. El análisis filogenético del gen 26S rRNA posicionó a estos aislados dentro del género *Aspergillus* sp. (Fig. 1 Anexo 2). El análisis de las secuencias cortas de los metagenomas logró obtener fragmentos de los genes ribosomales 18S rRNA y 26S rRNA. Estas secuencias permitieron reconocer a los phyla Basidiomycota, Chitridiomycota y algunos Ascomycota en el metagenoma obtenido de los bacteriomas del insecto. Por otro lado, de una muestra obtenida por macerado y fraccionamiento por Percoll (muestra del Percoll de aquí en adelante) de insectos completos, se logró obtener secuencias de los phyla Ascomycota, Basidiomycota, Glomeromycota entre otras (Fig. 2 Anexo 2). A partir de estas lecturas, de la muestra del Percoll, se pudieron reconstruir parcialmente los genes 18S rRNA de una secuencia emparentada con *Cordyceps confragosa* y otro parcial emparentada con *Glomus* sp. (Fig. 7 y Fig. 3 Anexo 2). El análisis funcional de los metagenomas identificó distintos genes fúngicos involucrados en síntesis de triglicéridos y metabolismo de purinas (Fig. 5 y Fig. 6 Anexo 2). Esto sugiere que los hongos dentro de *Llaveia axin axin* pudieran participar en síntesis de lípidos y ayudar a un aporte de nitrógeno a partir del reciclaje de moléculas como el AU y sus derivados al igual que sucede en otros insectos (Morales-Jiménez et al., 2013; Nasir and Noda, 2003; Vera-Ponce de León et al., 2016).

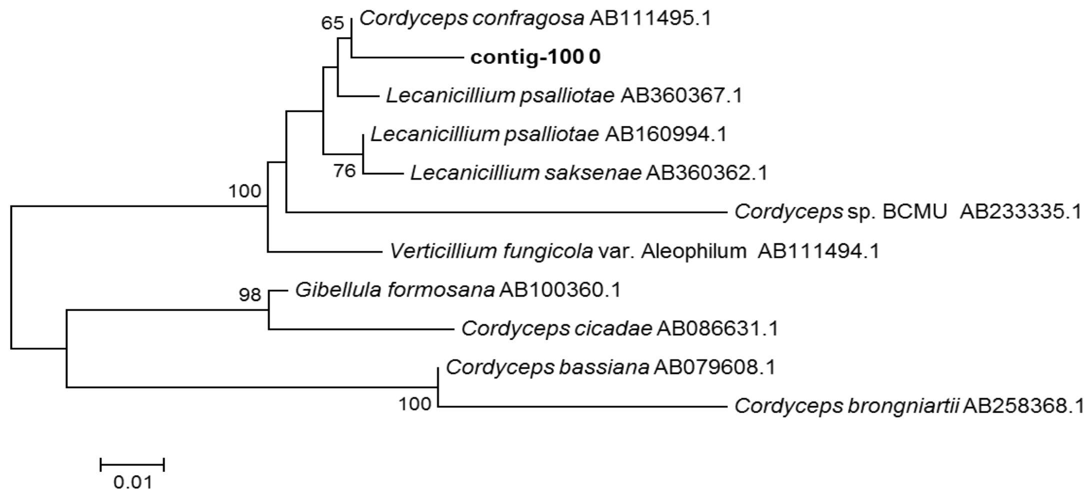


Figura 7. Arból filogenético de Máxima verosimilitud (-ln L=140176) de la secuencia 18S rRNA de *Cordyceps* y otras secuencias del GenBank obtenida del metagenoma de *Llaveia axin axin*. Se muestran los valores Boostrap > 50 %. La barra indica 1 % de divergencia entre los datos.

Particularmente genes codificantes para la enzima uricasa de *Aspergillus* fueron encontrados en el meta-ensamble de la muestra de Percoll y Bacterioma (Tabla suplementaria 2 Anexo 2). También fueron encontrados genes para la degradación de xenobióticos y el rompimiento de celulosa (Tabla suplementaria 2 Anexo 2). Esta evidencia sugiere que los hongos asociados a el Nijj jueguen papeles en el aporte de nitrógeno (recilaje de ácido úrico), degradación de moléculas tóxicas presentes en la dieta y provisión de lípidos.

Capitulo III Fijación de nitrógeno por bacterias simbiontes de *Dactylopius*.

Esta sección se explora el trabajo**: *Candidatus* Dactylopiibacterium carminicum a nitrogen-fixing symbiont of the cochineal insect *Dactylopius coccus* Arturo Vera-Ponce de León**, Ernesto Ormeño-Orrillo, Shayim T. Ramírez-Puebla, Mónica Rosenblueth, Julio Martinez, Esperanza Martínez-Romero.

- **Planteamiento del problema**

Aunque se ha reportado a *Dactylopiibacterium carminicum* como un simbionte presente en todas las especies de *Dactylopius*, no se conoce las posibles funciones ni la localización en el huésped. La reconstrucción (ensamble) y el análisis del genoma de *D. carminiucm* permitirán proponer posibles funciones en el aporte de metabolitos de esta bacteria al huésped. Así mismo el uso de sondas para la búsqueda de la bacteria en el insecto ayudará a conocer su localización.

- **Hipótesis**

Si *Dactylopiibacterium carminicum* se encuentra en todas las especies de *Dactylopius* estudiadas podrá ser considerada como un simbionte primario, y deberá cumplir con los siguientes criterios: tener un genoma reducido, residir dentro de células del insecto y aportar metabolitos útiles al huésped.

- **Objetivo**

Ensamblar y anotar el genoma de *Dactylopibacterium carminicum* a partir de tres metagenomas de *Dactylopius coccus*.

- Objetivos particulares

1. Buscar genes del ciclo del nitrógeno en el genoma de *Dactylopiibacterium carminicum*.
2. Buscar genes para la comunicación con el huésped en el genoma de *Dactylopiibacterium carminicum*
3. Realizar ensayos de reducción de acetileno en diferentes tejidos del insecto.
4. Localizar a *D. carminicum* en el insecto usando sondas diseñadas para hibridación flourescente *in situ* (FISH).

Resultados.

A partir de 3 metagenomas de *D. coccus* se pudo reconstruir un genoma de ca. 3.6 Mb (Tabla 1 Anexo 3) correspondiente a una β-proteobacteria. El gen 16S rRNA localizado en el genoma se emparentó con una identidad del 99-100 % con las secuencias de *Dactylopiibactrium carminicum* END1-C y END1-O reportadas previamente en el grupo de trabajo (Anexo 3). La filogenómica usando 293 genes ortólogos de copia única del core genoma posicionó a esta bacteria dentro de las β-proteobacterias con *Uliginosibacterium ganwonense*, *Aromatoleum aromaticum y Azoarcus* como las especies más cercanas (Fig. 5 Anexo 3). El análisis de la comparación promedio de identidad nucleótidica de estas secuencias ortólogas (gANI) fue de 77.16 % y 74.94 % en la comparación *Dactylopiibacterium-Uliginosibacateriu* y *Dactylopiibacterium-Aromatoleum*, respectivamente. Mientras que la hibridación DNA-DNA *in silico* (DDH) tomó un valor de 12.50 % y 16.80 % en las mismas comparaciones. Siguiendo los estándares de oro para la delimitación de especies (gANI > 95 % DDH > 70 %; (Varghese et al., 2015)) podemos aseverar que *Dactylopiibacterium carminicum* es una nueva especie dentro de la familia Rodocyclaceae.

Dentro de las capacidades metabólicas encontradas en el genoma de este simbionte está la síntesis de vitaminas (cobalamina, riboflavina, piridoxina y tiamina). Aunque no se encontraron transportadores para glucosa (PTS) secuencias para transportadores de xilosa y trealosa fueron hallados en la anotación del genoma.

Esto también se refleja en encontrar vías truncas para la glicólisis, sin embargo la obtención de energía puede ser llevada a cabo a partir del ciclo de las pentosas del cual se encuentran genes para llevar de xilosa a piruvato y acetil Co-A. El ciclo de los ácido tricarboxílicos es completo. Se encontró que este organismo puede llevar a cabo reacciones fermentativas gracias a la presencia de las enzimas piruvato formato-liasa y lactato deshidrogenasa (Fig. 6, Anexo 3). Así mismo, se encontraron genes para la respiración de nitritos. Esta bacteria posee la capacidad de síntesis de todos los aminoácidos esenciales con excepción de asparagina (Fig. 6 Anexo 3).

Es interesante encontrar genes para el sistema de secreción tipo IV, I y II, los cuales podrían participar en la comunicación celular con el huésped. Así mismo, se encontró toda la maquinaria genética para la síntesis y armado de un flagelo (Anexo 3). Por otro lado, también fueron encontradas dos secuencias completas y tres incompletas de profagos insertadas en el ensamble genómico de *D. carminicum* (Anexo 3). El análisis metagenómico de *Dactylopius* localizó secuencias para la fijación biológica del nitrógeno como: Mo-Fe nitrogenasa (*nifDK*), nitrogenasa reductasa (*nifH*), proteínas de regulación y transportadores específicos para cofactores de la nitrogenasa (nifQA y 4Fe-4S ferredoxina) (Fig. 1 Anexo 3). La reconstrucción de los genes *nifHDK* emparentó a estos genes con los reportados en *Uliginosibacterium ganwonense* (Fig. 2, Anexo 3). La búsqueda de estos genes en dos metagenomas públicos reveló la presencia de secuencias en *D. coccus* peruanos y oaxacaqueños (Anexo 3). Por otro lado ensayos de RT-PCR mostraron que existe

expresión del gen *nifH* en hemolinfa, ovarios y embriones pero no en intestinos de *D. coccus* (Fig. 4 Anexo 3). Ensayos de reducción de acetileno (ARA) cuantificaron actividad de fijación de nitrógeno *in vivo* en hemolinfa y ovarios de *D. coccus* (38.00 ± 5.98 y 11.58 ± 1.38 nmol Acetileno h^{-1} g^{-1} tejido respectivamente; Fig. 3 Anexo 3). Aunado a esto, se logró localizar a *D. carminicum* en ovarios de *D.coccus* y *D. opuntiae* por experimentos tipo FISH (Fig. 7 Anexo 3). Toda esta evidencia sugiere que *D. carminicum* es un diazótrofo activo en el interior de *Dactylopius* ayudando con las deficiencias de nitrógeno en la dieta del insecto.

Capitulo IV. Una Gamma proteobacteria asociada al Nijj.

Esta sección se explora el trabajo en prensa**: The Symbiome of *Llaveia* Cochineals (Hemiptera: Coccoidea: Monophlebidae) Includes a Gammaproteobacterial Co-symbiont *Sodalis* TME1 and the Known *Candidatus* Walczuchella monophlebidarum.** Tania Rosas-Pérez, **Arturo Vera-Ponce de León**, Mónica Rosenblueth, Shamayim T. Ramírez-Puebla, Reiner Rincón-Rosales, Julio Martínez-Romero, Michael F. Dunn, Esperanza Martínez-Romero. **Entomology, ISBN 978-953-51-5041-1.**

- **Planteamiento del problema**

Aunque existe el reporte de una gamma proteobacteria asociada a *Llaveia axin axin*, la exploración de las capacidades completas codificadas en el genoma de este organismo así como su posicionamiento filogenético en comparación con otros simbiontes no han sido descritos.

- **Hipótesis**

El genoma de *Sodalis* TME1 asociado a *Llavea axin axin* presentará secuencias codificantes para enzimas que complementen las vías metabólicas truncas del simbionte primario *Wlaczuchella monophlebidarum*.

- **Objetivo**

Explorar el genoma de *Sodalis* TME1 asociado a *L. axin axin* así como hacer un estudio de genómica comparada que ayude a entender la biología de este simbionte en su huésped.

Objetivos particulares

1. Realizar la anotación funcional del genoma de *Sodalis* TME1 y proponer vías metabólicas útiles en la simbiosis.
2. Comparar el genoma de *Sodalis* TME1 con otros *Sodalis* simbiontes de insectos y de vida libre.

Resultados

A partir de 3 metagenomas obtenidos con diferentes plataformas de secuenciación, FLX-454, Illumina GAIIX e Illumina HiSeq200, de *Llaveia axin axin* se logró reconstruir 679 scafolds correspondientes a una γ-Proteobacteria. El anális del gen 16S rRNA de estas secuencias relacionó filogenéticamente al genoma TME1 con el género *Sodalis* (Fig. 8). La filogenómica usando 143 genes ortólogos del core-genoma de *Sodalis*, agrupó a TME1 con *Sodalis praecaptivus* y *Sodalis pierantonius* como especies hermanas (Fig. 3 Anexo 4). El análisis de comparación del promedio de identidad de nucleótidos (gANI) sugiere que el género *Sodalis* puede ser dividido en tres filogrupos putativos (Fig.3 Anexo 4; Tabla 1 Anexo 4).

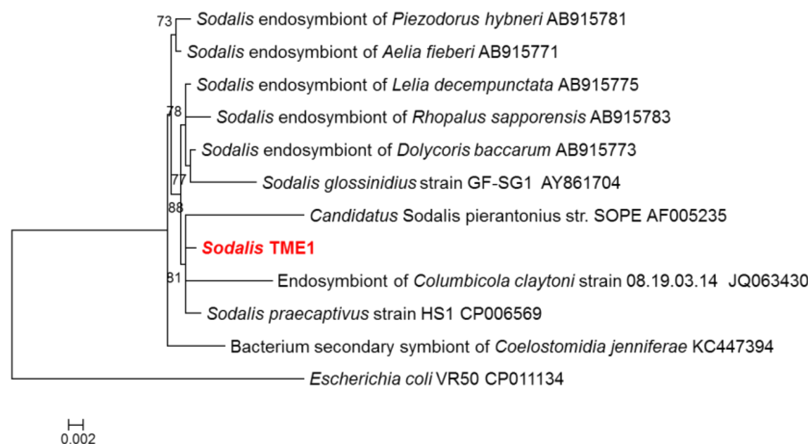


Figura 8. Árbol por máxima verosimilitud (-ln L=-2935.20316) de la secuencia del gen 16S rRNA obtenida del genoma de *Sodalis* TME1 con otras secuencias del GenBank. La barra muestra 2 % de divergencia entre las secuencias. Se muestran valores SH-like ≥ 50. La secuencia del gen 16S rRNA de *Escherichia coli* VR50 es usada como grupo externo.

Tabla 2.Porcentajes de la hibridación DNA-DNA (DDH) entre las cepas de *Sodalis*. Valores en negritas corresponden a DDH > 70 %. Colores corresponden: verde, insectos comedores de plantas; rojo, comedores de sangre; azul, vida libre.

	***Sodalis* str. TME1**	*Sodalis pierantonius* SOPE	*Sodalis praecaptivus* HS1	*Sodalis glossinidius* str. morsitans	*Sodalis*-like str. PSPU	*Sodalis*-like str. SPI-1
***Sodalis* str. TME1**						
Sodalis pierantonius str. SOPE	**70**					
Sodalis praecaptivus str. HS1	**71.7**	**73**				
Sodalis glossinidius str. morsitans	32.9	32.7	33.10			
Sodalis-like str. PSPU	31.1	32.6	33.30	52.50		
Sodalis-like str. SPI-1	33.2	24.4]	30.30	26.30	27.30	

Por otro lado la hibridación DNA-DNA (DDH) reconoce a las especies *Sodalis* TME1, *Sodalis pierantonius* str. SOPE y *Sodalis praecaptivus* str. HS1 como el mismo filogrupo (Fig. 3 Anexo 4, Tabla 2). Además, usando esta herramienta es plausible separar el filogrupo A en dos filoespecies teniendo en cuenta el valor aceptado para la propuesta de separación de especies (≥ 70 % DDH, (Varghese et al., 2015) (Tabla 2).

El genoma de 3.4 Mb codifica para 3,067 genes de los cuales la categorías funcionales se dividen en el ciclo de los ácidos tricarboxílicos completo (TCA); toda

la maquinaría para la síntesis de todos los aminoácidos esenciales así como síntesis, edición y reparación de ácidos nucleicos (Fig. 4 Anexo 4). El genoma reveló que esta bacteria posee elementos para la síntesis de sistemas de secreción tipo III y IV, así como la síntesis de flagelo (Fig. 9).

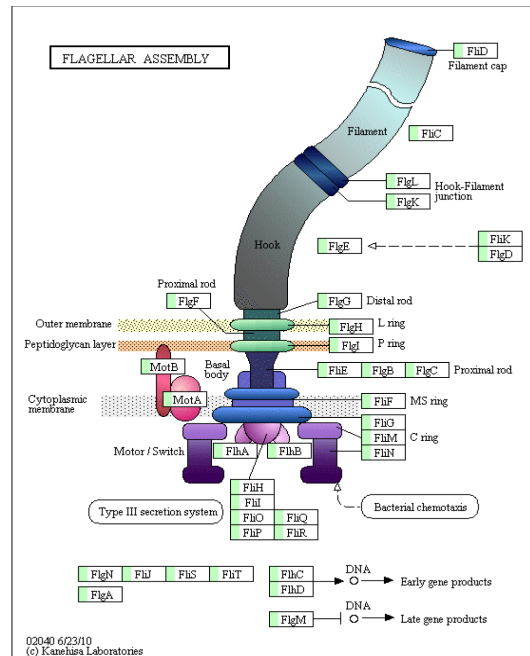


Figura 9. Elementos genéticos para la síntesis de flagelo anotados en *Sodalis* TME1.

Esta bacteria posee la capacidad para respirar nitritos y nitratos, además de usar el ácido úrico y sus subproductos metabólicos debido a la presencia de genes codificantes para uricase (*uaZ*), allantoinasa (*allB*), allantoato deiminasa (*allC*) y ureasa (*ureC* y *ureD*). Esto concuerda con la evidencia experimental de encontrar presencia de ácido úrico y la enzima urato oxidasa en los tejidos del insecto (Anexo 4). Así mismo, un análisis transcriptómico mostró que trasncritos de *Sodalis* TME1 para la síntesis de flagelo y el uso de alantoína se encuentran altamente expresados

en ovarios y bacteriomas de *L. axin axin* (Tabla 4 Anexo 4). Estos resultados sugieren que *Sodalis* TME1 es una bacteria de adquisición reciente en el nijj y que esta pudiera estar participando activamente en el reciclaje de metabolitos nitrogenados, como lo es el AU, en el huésped además de complementar metabólicamente al simbionte primario *Ca.* Walczuchella monophlebidarum.

Discusión

Los insectos fitófagos han sido capaces de ocupar distintos hábitats debido a la interacción con distintas comunidades microbianas. En este trabajo se estudiaron distintos modelos de relación microorganismo-hemíptero. Las dos cochinillas que se usaron son de interés cultural y económico en México. Ambas se clasifican dentro de los insectos escama, y como tal, se alimentan de sabia de plantas, la cual es pobre en nitrógeno. Estos organismos, al igual que las plantas de las que se alimentan, son nativos de México y en ambos modelos se encontraron distintos mutualistas fúngicos y bacterianos (Tabla 3). Se ha propuesto que las relaciones mutualistas entre bacterias y hongos en insectos fitófagos han favorecido la adecuación de los mismos. El aporte de metabolitos nitrogenados se ha considerado una de las maneras en que estos microoganismos influyen en los insectos. En este contexto, esta tesis se centró en el estudio de las interacciones entre hongos y bacterias en dos cochinillas mexicanas (*Dactylopius* y *L. axin axin*), así como su participación en el aporte de nitrógeno a partir del reciclaje de ácido úrico (hongos, Tabla 3) y la fijación biológica del nitrógeno (bacterias, Tabla 3).

- **Hongos asociados a las cochinillas de la laca y el carmín.**

En este trabajo se observó una diferencia entre los filogrupos de hongos encontrados en *Dactylopius* y *Llaveia* (Fig. 4 Anexo 1; Fig. 1 Anexo 2; Tabla 3). En el caso particular de la cochinilla del carmín encontramos que la mayoría de los hongos pertenecen al phylum Basidiomycota, mientras que en la cochinilla de la

laca los más comunes son los Ascomycota y algunos Chytridiomycota (Anexo 2, Fig. 2). Así mismo, las especies de hongos *Cryptococcus saitoi* y *Rhodotorula mucilaginosa* se encontraron presentes en todas las especies de *Dactylopius* muestradas (Fig. 5). Los análisis de hibridación *in situ* mostraron la presencia de C. *saitoi* en los ovarios y túbulos de Malpighi en dos especies de *Dactylopius* (Fig. 7 Anexo 1). Esto sugiere que los hongos se transmiten verticalmente en la cochinilla de carmín. El análisis de función de los metagenomas (anotación metagenómica) de *Llaveia* y *Dactylopius* mostró la presencia de secuencias fúngicas codificantes para genes de la vía del catabolismo del AU (Fig. 3 Anexo 1; Tabla 3 Anexo 2). La participación de hongos mutualistas en el reciclaje de nitrógeno usando el AU excretado por el huésped ya ha sido propuesto en hemípteros (Sasaki et al., 1996; Hongoh et al., 2000). Sin embargo, ni en *Dactylopius* ni en *Llaveia* se ha explorado este tipo de funciones. En este trabajo se observó que ambos insectos producen AU como metabolito y que la enzima uricasa se encuentra presente (Fig. 4 Anexo 1; Anexo 4). Evidencia experimental de que los hongos participan en el reciclaje de AU se observó particularmente en *Dactylopius*. Experimentos con fludoxionil y anfotericina-B (antifúngicos) mostraron que los insectos tratados presentaban una talla y un peso más pequeño con respecto a los controles. Además, la concentración de AU fue mayor y la actividad de uricasas menor en insectos tratados con respecto a los controles (Fig. 6; Fig. 6 Anexo 1; Figura suplementaria 4 Anexo 1). Así mismo, los hongos aislados del *Dactylopius* puedieron usar el AU como única fuente de carbono y/o nitrógeno (Fig. 5 Anexo 1, Tabla 3 Anexo 1). Todo esto

sugiere que los hongos asociados al *Dactylopius* participan en el reciclaje de AU en el insecto participando en el aporte de nitrógeno al huésped. Con esto podríamos inferir que en otros insectos escama como el nijj los hongos jueguen un papel similar en el reciclaje de moléculas nitrogenadas como el AU.

.

Tabla 3. Relaciones mutualistas en la cochinilla del carmín y de la laca.

Insecto hospedero	Simbionte	Principales representantes	Phylum	Posible papel en la simbiosis
Dactylopius	Bacteriano	*Dactylopiibacterium carminicum*	Proteobacteria (β)	Fijación biológica de nitrógeno
		Wolbachia burtzisii wDacA *Wolbachia pipentis* wDacB	Proteobacteria (α)	Desconocida
	Fúngico	*Cryptoccous saitoi* *Rhodotorula mucilaginosa*	Basidiomycota	Reciclaje de ácido úrico
		Penicillium sp.	Ascomycota	
Llaveia axin axin	Bacteriano	*Walzuchella monophlebidarum*	Bacteroidetes (Flavobacteria)	Síntesis de aminoácidos esenciales
		Sodalis TME1	Proteobacteria (γ)	Complementación metabólica de *Walczuchella monophlebidarum* Reciclaje de alantoina
	Fúngico	*Aspergillus* sp.	Ascomycota	Síntesis de lípidos Reciclaje de ácido úrico

- **Los co-simbiontes bacterianos de *Walczuchella* en la cochinilla de la laca.**

En el nijj, usado para la obtención de laca, se encontró un genoma correspondiente a la clase γ-Proteobacteria clasificado como *Sodalis* TME1. Esta bacteria se ha propuesto como un co-simbionte de *Candidatus* Walzuchela monophlebidarum (Tabla 3) una flavobacteria común a la especie de cochinillas *L. axin axin* (Rosas-Pérez et al., 2014). Este co-simbionte, en comparación con otros *Sodalis* se agrupa filogenéticamente con dos especies *S. pieriantonius* SOPE (simbionte del gorgojo del arroz *Sitophilus oryzae*) y un especie no simbionte denominada *Sodalis praecaptivus* str. HS1 (Fig. 3 Anexo 3; Tabla 1 Anexo 3). El valor de gANI fue mayor al 98 % al compararse TME1 con estas dos especies pero menor con otros *Sodalis* reportados previamente como simbiontes de insectos (Fig. 3 Anexo 3). Estos mismos valores de gANI y los análisis filogenéticos del core genoma, usando los cinco genomas de *Sodalis* reportados hasta el momento, nos ayudan a proponer 3 filo-especies de *Sodalis* A-C (Fig. 3 Anexo 3). *Sodalis* TME1 queda posicionado en el grupo C distante de otros *Sodalis* simbiontes de insectos (Fig. 4 Anexo 3). El hecho de proponer especies utilizando gANI ha causado controversia, y por ello los taxónomos recomiendan el uso de técnicas como el cultivo axénico y un sin fin de pruebas fenótipicas (no concluyentes) para proponer nuevos *taxa* bacterianos (Rosselló-Móra and Amann, 2015). Sin embargo, en el caso de simbiontes intracelulares, donde pocas veces han podido cultivarse, se propone el uso de gANI aunado a alineamientos de fracciones de ortólogos (AF) así como la imitación *in silico* de la DDH (DDH) para la delimitación de especies o grupos

taxonómicos (Meier-Kolthoff et al., 2013; Varghese et al., 2015; Rosselló-Móra and Amann, 2015). La posibilidad de separación de filo-grupos en *Sodalis* no ha sido propuesta, sin embargo se han manejado dos linajes, uno considerado simbionte primario y el otro un co-mutualista (Lo et al., 2016). En el trabajo realizado con TME1 se observa que estos grupos pueden ser separados incluso en filo-especies, apoyando la idea de una adquisición reciente por parte del insecto. La ocurrencia de este tipo de adquisiciones recientes de simbiontes se ha reportado en otros modelos. En el caso de los insectos de la familia Monophlebidae (Tabla 1) se sabe que existe una co-simbiosis entre simbiontes primarios de la clase Flavobacteria (P-simbiontes) fechada en ca. 200 millones de años (McCutcheon and Moran, 2010) y otras clases bacterianas denominados generalmente simbiontes secundarios (S-simbiontes) de adquisición reciente (Rosenblueth et al., 2012). Los simbiontes tipo Flavobacteria se han mantenido en distintos linajes de insectos escama (incluyendo *Llaveia axina axin*, Tabla 3) (Rosenblueth et al., 2012) sin embargo los S-simbiontes varían entre las clases α, β o γ protebacterias (Koga and Moran, 2014). En el nijj encontramos a estos S-simbiontes representados por *Sodalis* TME1 (γ-proteobacteria) y una α-proteobacteria cercana al grupo de los ricketsiales (Rosas-Pérez, datos no publicados). El por qué hay una variación de clases bacterianas en los simbiontes secundarios es algo desconocido, no obstante se ha observado que en la mayoría de los casos estos S-simbiontes complementan las deficiencias metabólicas de los P-simbiontes (McCutcheon and Moran, 2007; Moran et al., 2008; McCutcheon et al., 2009; Rosas-Pérez et al., 2014). En el caso particular de *Ca.*

Walczuchella monophlebidarum (Flavobacteria) y *Sodalis* TME1 (γ-proteobacteria) se propone que TME1 proporcione aminoácidos como lisina, leucina, valina, isoleucina e histidina así como corismato; completando con esto las vías anabólicas truncas en *Walczuchella* (Rosas-Pérez et al., 2014). De igual manera podemos especular que *Sodalis* pueda participar en el aporte de intermediarios para el ciclo de los ácidos tricarboxílicos (TCA) y el aporte de energía, ya que *Walczuchella* al igual que otras flavobacterias simbiontes de hemípteros presentan la ruta metabólica del TCA incompleta (Bennett and Moran, 2013; Koga and Moran, 2014). *Sodalis* TME1 al contener toda la maquinaria genética para el ciclo del citrato (Anexo 3) pudiera brindar metabolitos intermedios para la generación de energía a *Wlaczuchela*. Al igual que sucede con *Sodalis*-like PSPU simbionte de la chicharra espumadora (*Philaenus spumarius*) y su Flavobacteria asociada (*Sulcia muelleri* PSPU) (Koga and Moran, 2014). Aunque hasta el momento desconocemos la localización exacta del simbionte, podemos presumir su localización en bacteriocitos. El hecho de que el genoma de esta bacteria presenta secuencias para la síntesis de la nitrato y nitrito reductasa (Anexo 3) sugiere que esta bacteria podría usar el nitrato como aceptor final de electrones en un ambiente reducido como es el citosol celular. En otras enterobacterias como *E. coli* y *Salmonella*, se ha visto que en ambientes reducidos y en presencia de niveles moderados de nitrato los genes codificantes para la nitrato reductasa (*nar*/*nap*) se expresan para poder respirar (Torres et al., 2016). El hecho de poder usar distintos aceptores de electrones daría una ventaja a la hora de competir con el oxígeno usado por la

mitocondria y/o *Wlaczucella* en el interior de las células del Nijj. Por tanto, al ser el amonio el producto final de la reducción del nitrito por la nitrito reductasa este podría ser excretado al medio, siendo este nitrógeno reducido en forma de amonio, una fuente extra de nitrógeno para el huésped. No obstante, esto debería estudiarse en un futuro. La respiración del nitrato ha sido estudiada en simbiontes quimiolitoautótrofos asociados a organismos marinos (Hentschel and Felbeck, 1993, 1995) pero no hay reportes de este tipo de respiración en simbiontes de insectos. En el caso particular del bivalvo *Lucinoma aequizonata* se observó que los simbiontes bacterianos usan nitrato y lo reducen para no competir por el poco oxígeno del medio (Hentschel et al., 1996). Proponemos que *Sodalis* TME1 podría hacer lo mismo en *Llaveia axin axin*. Por otro lado, TME1 se encontró que posee los elementos genéticos para producir poliaminas y que además se tiene un sistema de secreción para una de ellas, la espermidina. Esta se produce de un aminoácido esencial, arginina y tal vez esta también sea otra forma de contribución del endosimbionte al hospedero. Se sabe que diferentes poliaminas pueden tener efectos muy variados en los organismos (Igarashi and Kashiwagi, 2000; Zhou et al., 2007) y hasta ahora no se había descrito la producción de poliaminas por endosimbiontes. Los transportadores de espermidina que encontramos tienen los aminoácidos que se requieren para unir al sustrato y también las hélices transmembranales de estos transportadores (M.Dunn, datos no publicados). En general se ha encontrado que las poliaminas se pegan a RNA y cambian su conformación y estabilidad, lo cual pudiera modificar la expresión genética del

hospedero ayudando al simbionte a establecerse. La manera en que los simbiontes se establecen en el huésped ha sido debatida, en *Sodalis* se sabe que la interacción bacteria-hospedero es llevada acabó por una comunicación vía el sistema de secreción tipo III (SSTT), donde bacterias mutantes en este tipo de sistema son incapaces de llevar a cabo la colonización del huésped (Dale et al., 2001, 2005). *Sodalis* TME1 al igual que los otros *Sodalis* presenta los genes para el ensamblaje del SSTT. Así bien, el análisis transcriptómico mostró alta expresión de los genes de SSTT en el bacterioma del insecto (Taba 4 Anexo 4). Esto sugiere que el mecanismo por el cual se establece la comunicación para la simbiosis nijj-*Sodalis* podría ser similar a la que ocurre en la Mosca Tse-Tse o en el gorgojo del maíz (Dale et al., 2005). Con esto y con la evidencia de que *Sodalis* TME1 es cercano filogéneticamente a una bacteria no simbionte, la similitud en complementación metabólica con el co-simbionte *Walczuchella* y la presencia de los genes para el SSTT apuntan a que *Sodalis* TME1 es una bacteria simbionte de adquisición reciente pero necesaria para el desarrollo y beneficio en la adecuación de *Llavea axin axin*.

- ***Dactylopiibacterium carminicum* y la fijación de nitrógeno en *Dactylopius*.**

Una de las funciones más estudiadas de las bacterias asociadas a eucariontes es el aporte de metabolitos nitrogenados al huésped (Kneip et al., 2007; Douglas, 2009). Anteriormente describimos que los hongos asociados tanto a *Dactylopius* como a

Llaveia pueden participar en el reciclaje de productos del catabolismo del ácido úrico (Capítulo 1; Capítulo 2). No obstante, se ha visto que en distintos linajes de insectos las bacterias asociadas pueden proporcionar nitrógeno aprovechable al huésped por medio de la diazotrofía (Murphy et al., 1988; Pinto-Tomás et al., 2009; Morales-Jiménez et al., 2013).Sin embargo, este proceso no ha sido estudiado en hemípteros, incluyendo los insectos escama. En este trabajo se observó que *Dactylopibacterium carminicum*, posee la maquinaria genética para fijar nitrógeno (Anexo 4). Esta bacteria se describió como un simbionte primario en el género *Dactylopius* (Ramírez-Puebla et al., 2010) pero sus capacidades metabólicas y participación en la simbiosis con el insecto no habían sido estudiadas. Al contrario con la simbiosis entre otras betaproteobacterias con insectos (*Tremblaya princeps, Nasuia deltacephanicola, Zinderia insecticola* y *Profftella armatura*) donde se observan genomas reducidos y complementariedad metabólica (McCutcheon and Von Dohlen, 2011; Bennett and Moran, 2013; Nakabachi et al., 2013) el genoma de *Dactylopiibacterium* es casi 10 veces más grande que los genomas de los endosimbiontes mencionados anteriormente (Tabla 2, Anexo 4). Por otro lado, se ha mencionado que la relación que han llevado estos simbiontes tipo β-proteobacteria es muy antigua (~260 millones de años) en el caso de los insectos del suborden Auchenorrhynca complementando en la mayoría de los casos el metabolismo del simbionte primario *Sulcia* (Noda et al., 2012; Bennett and Moran, 2013). Aunque, la relación de *Dactylopibacterium* con su huésped parece ser de reciente adquisición. El genoma de esta bacteria mantiene toda la maquinaria para

la síntesis de pared, conserva los genes para la síntesis de flagelo y es filogenéticamente cercano con bacterias diazótrofas de vida libre *Uliginosibacterium gangwonense* y *Azoarcus* sp. (Fig 5 y Fig. 6 Anexo 4). Sin embargo *Dactylopibacterium* no ha podido ser cultivado *in vitro* en laboratorio. Como se ha mencionado la posible función de esta β-proteobacteria es el aporte de nitrógeno gracias a su capacidad fijadora, en este trabajo se demostró que los genes canónicos para la nitrogenasa (*nif*HDK) de *Dactylopibacterium carminicum* NFE1 están filogenéticamente cercanos a los de *Uliginosibacterium ganwonense*, *Azovibrio restrictus* y *Dechloromonas* (Fig. 1 y Fig. 2 Anexo 4), bacterias descritas como diazótrofas (Reinhold-hurek and Hurek, 2006). Ensayos de reducción de acetileno en distintos tejidos de *D. coccus* y *D. opuntiae* demostraron que hay actividad diazotrófica en el huésped, al igual que sucede en otros insectos (Murphy et al., 1988; Morales-Jiménez et al., 2013). Por último la expresión de los genes *nif*H y *nif*D de NFE1 en ovarios y hemolinfa de *D. coccus* comprueba que la diazotrófia está ocurriendo en el insecto (Fig 3 Anexo 4). La cercanía filogenética con bacterias diazótroficas-endófitas de plantas podría sugerir que *Dactylopibacterium* NFE1 pudiera haberse adquirido a partir de la cactácea hospedera de *Dactylopius*. La relación de bacterias fijadoras de nitrógeno con plantas es muy exitosa porque hay un intercambio de N_2 por carbono, sin embargo este intercambio solo se ha reportado con los rizobios y en cianobacterias (Ohlendieck et al., 2000). Se ha observado que cuando las tasas de fijación de N_2 son muy elevadas y superan las necesidades de nitrógeno de las bacterias entonces se excreta N_2 al medio o a otro

(micro)organismo. Existe la evidencia de que en cianobacterias hay una transferencia del 97% del N_2 fijado a una diatomea mutialista, por lo que se ha pensado que debe de haber un mecanismo por el cual la transferencia de N_2 sea estimulada por parte de la diatomea (Foster et al., 2011). En insectos, aunque se ha estudiado que el nitrógeno fijado por simbiontes bacterianos es incorporado a las proteínas del huésped a partir de experimentos con N^{15} (Pinto-Tomás et al., 2009; Tai et al., 2016), se desconoce si este nitrógeno es liberado en forma de metabolito secundario o bien las bacterias son recicladas en forma de proteína como es el caso de *Sodalis* en los escarabajos (Masson et al., 2015). Queda por investigar si NFE1 tiene un mecanismo de liberación de nitrógeno a su hospedero. Planeamos realizar experimentos con N^{15} para probar esto. En rizobios, se publicó que además de amonio, el aminoácido alanina era uno de los productos de la fijación que se exportaba a las plantas (Waters et al 1998), pero después este hallazgo se cuestionó (Kumar et al., 2005). En cianobacterias se ha considerado que glutámico y glutamina pudieran exportarse de los heterocistos que son células fijadoras de nitrógeno (Thomas et al., 1977). Mientras que el hecho de encontrar siete genes que codifican para transportadores de urea en NFE1 y tres en *U. ganwonese* nos ha llevado a especular que tal vez la urea sea el producto de la fijación de N_2 excretado por la bacteria fijadora de nitrógeno NFE1 a su hospedero y esta pudiera ser una característica del grupo bacteriano.

Conclusión

En contexto las relaciones endosimbioticas con insectos escama han sido poco estudiadas. En México el uso de *Dactylopius* y *Llaveia* de manera comercial ha sido explotado desde épocas prehispánicas, sin embargo hasta ahora no se ha ahondado en la biología de estos insectos. Este trabajo pretende ampliar el conocimiento de la microbiota asociada a ambas cochinillas. Los estudios permiten concluir que hongos dentro de las cochinillas mexicanas del carmín y de la laca pueden participar en el aporte de nitrógeno a partir del reciclaje de ácido úrico. Por otro lado se descubrió que el género *Sodalis* puede ser dividido en tres filogrupos y que *Sodalis* TME1 puede complementar metabólicamente a *Walczuchela* tanto en la síntesis de aminoácidos como en el ciclo de Krebs. Finalmente describimos que el P-simbionte de *Dactylopius*, *Dactylopiibacterium carminicum*, es un diazótrofo activo en el insecto y puede participar en el ciclo del nitrógeno en el huésped.

Perspectivas

- Secuenciar y analizar el metagenoma de *D. opuntiae* y *D. tomentosus* para hacer una comparación entre las comunidades microbianas de una especie domesticada *D. coccus* y dos silvestres y determinar la divergencia de sus simbiontes fúngicos y bacterianos como son: *D. carminicum* y *Spiroplasma* sp.

- Secuenciar y analizar el metatranscriptoma de la especie domesticada D. cocus y del silvestre D. opuntiae para identificar genes expresados diferencialmente de las comunidades microbianas entre las dos especies.

- Realizar ensayos con N15 de tejidos completos de Datylopius para determinar si el nitrógeno fijado por D. carminicum es aprovechado directamente por el huésped.

- Realizar ensayos de tratamientos con antibióticos para generar insectos axénicos y corroborar que las bacterias mutualistas de Dactylopius favorecen la adecuación del huésped.

- Proponer un "draft" del genoma de D. coccus usando los datos metagenómicos y metatranscriptómicos obtenidos en este trabajo, para buscar rutas de síntesis como es el ácido carmínico

.

Bibliografía

- 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, 808–17. doi:10.1007/s10886-011-9992-6.
- Andersson, J. O. (2000). Evolutionary genomics: Is *Buchnera* a bacterium or an organelle? Curr. Biol. 10, 866–868. doi:10.1016/S0960-9822(00)00816-2.
- Ayres, M. P., Wilkens, R. T., Ruel, J. J., Lombardero, M. J., Vallery, E., and Aug, N. (2000). Nitrogen Budgets of Phloem-Feeding Bark Beetles with and without Symbiotic Fungi. Ecology. 81, 2198–2210.
- Baumann, P. (2005). Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects. Annu. Rev. Microbiol. 59, 155–189. doi:10.1146/annurev.micro.59.030804.121041.
- Becher, P. G., Flick, G., Rozpędowska, E., Schmidt, A., Hagman, A., Lebreton, S., Larsson, M. C., Hansson, B. S., Piškur, J., Witzgall, P., et al. (2012). Yeast, not fruit volatiles mediate *Drosophila melanogaster* attraction, oviposition and development. Funct. Ecol. 26, 822–828. doi:10.1111/j.1365-2435.2012.02006.x.
- Behar, a, Yuval, B., and Jurkevitch, E. (2005). Enterobacteria-mediated nitrogen fixation in natural populations of the fruit fly *Ceratitis capitata*. Mol. Ecol. 14, 2637–43. doi:10.1111/j.1365-294X.2005.02615.x.
- Bennett, G. M., and Moran, N. a. (2015). Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. Proc. Natl. Acad. Sci. 2015, 201421388. doi:10.1073/pnas.1421388112.
- Bennett, G. M., and Moran, N. A. (2013). Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect. Genome Biol. Evol. 5, 1675–1688. doi:10.1093/gbe/evt118.
- Bentley, J. K., Veneti, Z., Heraty, J., and Hurst, G. D. (2007). The pathology of embryo death caused by the male-killing *Spiroplasma bacterium* in *Drosophila nebulosa*. BMC Biol. 5, 9. doi:10.1186/1741-7007-5-9.

- Bolaños, L. M., Servín-Garcidueñas, L. E., and Martínez-Romero, E. (2015). Arthropod-*Spiroplasma* relationship in the genomic era. FEMS Microbiol. Ecol. 91, 1–8. doi:10.1093/femsec/fiu008.
- van Borm, S., Buschinger, A., Boomsma, J. J., and Billen, J. (2002). *Tetraponera* ants have gut symbionts related to nitrogen-fixing root-nodule bacteria. Proc. Biol. Sci. 269, 2023–7. doi:10.1098/rspb.2002.2101.
- Buchner, P. (1965). Endosymbiosis of animals with plant microorganisms. New York: Jhon Wiley & Sons, Inc.
- Campana, M. G., Robles García, N. M., and Tuross, N. (2015). America's red gold: multiple lineages of cultivated cochineal in Mexico. Ecol. Evol. 5, 607–17. doi:10.1002/ece3.1398.
- Chapman, R. (2013). The Insects Structure and Function. 5th ed. , eds. A. E. Douglas and S. J. Simpson Cambridge University Press.
- Chávez-Moreno, C. K., Tecante, a., and Casas, a. (2009). The *Opuntia* (Cactaceae) and *Dactylopius* (Hemiptera: Dactylopiidae) in Mexico: a historical perspective of use, interaction and distribution. Biodivers. Conserv. 18, 3337–3355. doi:10.1007/s10531-009-9647-x.
- Cheng, D. J., and Hou, R. F. (2001). Histological observations on transovarial transmission of a yeast-like symbiont in *Nilaparvata lugens* Stal (Homoptera, Delphacidae). Tissue Cell 33, 273–279. doi:10.1054/tice.2001.0173.
- Cochran, D. G. (1985). Nitrogen Excretion in Cockroaches. Annu. Rev. Entomol. 30, 29–49. doi:10.1146/annurev.en.30.010185.000333.
- D'Ettorre, P., Mora, P., Dibangou, V., Rouland, C., and Errard, C. (2002). The role of the symbiotic fungus in the digestive metabolism of two species of fungus-growing ants. J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 172, 169–176. doi:10.1007/s00360-001-0241-0.
- Dale, C., Jones, T., and Pontes, M. (2005). Degenerative evolution and functional diversification of type-III secretion systems in the insect endosymbiont S*odalis glossinidius*. Mol. Biol. Evol. 22, 758–766. doi:10.1093/molbev/msi061.
- Dale, C., Young, S. A., Haydon, D. T., and Welburn, S. C. (2001). The insect endosymbiont Sodalis *glossinidius* utilizes a type III secretion system for cell

invasion. Proc. Natl. Acad. Sci. U. S. A. 98, 1883–8. doi:10.1073/pnas.021450998.

- Desai, M. S., and Brune, A. (2012). Bacteroidales ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites. ISME J. 6, 1302–13. doi:10.1038/ismej.2011.194.
- Deveoglu, O., Karadag, R., and Yurdun, T. (2011). Qualitative HPLC determination of main anthraquinone and lake pigment contents from *Dactylopius coccus* dye insect. Chem. Nat. Compd. 47, 103–104. doi:10.1007/s10600-011-9842-3.
- Dixon, R., and Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. Nat. Rev. Microbiol. 2, 621–31. doi:10.1038/nrmicro954.
- Douglas, a. E. (2009). The microbial dimension in insect nutritional ecology. Funct. Ecol. 23, 38–47. doi:10.1111/j.1365-2435.2008.01442.x.
- Douglas, A. E. (2012). "Can't Live without You:" Essential Animal-Bacterial Relationships. Microbe Mag. 7, 273–277. doi:10.1128/microbe.7.273.1.
- Douglas, A. E. (2015). Multiorganismal Insects: Diversity and Function of Resident Microorganisms. Annu. Rev. Entomol. 60, 17–34. doi:10.1146/annurev-ento-010814-020822.
- Douglas, A. E. (1998). Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria *Buchnera*. Annu. Rev. Entomol. 43, 17–37. doi:10.1146/annurev.ento.43.1.17.
- Douglas, A. E. (2014). The molecular basis of bacterial-insect symbiosis. J. Mol. Biol. 426, 3830–3837. doi:10.1016/j.jmb.2014.04.005.
- Douglas, A. E., Bouvaine, S., and Russell, R. R. (2011). How the insect immune system interacts with an obligate symbiotic bacterium. Proc. R. Soc. B Biol. Sci. 278, 333–338. doi:10.1098/rspb.2010.1563.
- Ebbert, M. a, Marlowe, J. L., and Burkholder, J. J. (2003). Protozoan and intracellular fungal gut endosymbionts in *Drosophila*: prevalence and fitness effects of single and dual infections. J. Invertebr. Pathol. 83, 37–45. doi:10.1016/S0022-2011(03)00033-8.

- Foster, R. a, Kuypers, M. M. M., Vagner, T., Paerl, R. W., Musat, N., and Zehr, J. P. (2011). Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses. ISME J. 5, 1484–1493. doi:10.1038/ismej.2011.26.
- Gabison, L., Prangé, T., Colloc'h, N., El Hajji, M., Castro, B., and Chiadmi, M. (2008). Structural analysis of urate oxidase in complex with its natural substrate inhibited by cyanide: mechanistic implications. BMC Struct. Biol. 8, 32. doi:10.1186/1472-6807-8-32.
- Ganter, P. (2006). "Yeast and Invertebrate Associations," in Biodiversity and ecophysiology of yeasts The Yeast Handbook., eds. G. Péter and C. Rosa (Berlin/Heidelberg: Springer-Verlag), 303–370. doi:10.1007/3-540-30985-3.
- Gibson, C. M., and Hunter, M. S. (2010). Extraordinarily widespread and fantastically complex: comparative biology of endosymbiotic bacterial and fungal mutualists of insects. Ecol. Lett. 13, 223–34. doi:10.1111/j.1461-0248.2009.01416.x.
- Gibson, C. M., and Hunter, M. S. (2009). Negative fitness consequences and transmission dynamics of a heritable fungal symbiont of a parasitic wasp. Appl. Environ. Microbiol. 75, 3115–3119. doi:10.1128/AEM.00361-09.
- Hansen, A. K., and Moran, N. a (2014). The impact of microbial symbionts on host plant utilization by herbivorous insects. Mol. Ecol. 23, 1473–96. doi:10.1111/mec.12421.
- Harris, H. L., Brennan, L. J., Keddie, B. A., and Braig, H. R. (2010). Bacterial symbionts in insects: balancing life and death. Symbiosis 51, 37–53. doi:10.1007/s13199-010-0065-3.
- Heddi, A., and Gross, R. (2011). "Proteobacteria as primary endosymbionts of arthropods," in Manipulative Tenants: Bacteria Associated with Arthropods, eds. E. Zchori-Fein and K. Bourtzis (Boca Raton, Fl: CRC Press), 1–27. Available at: http://www.crcnetbase.com/isbn/9781439827505.
- Hentschel, U., and Felbeck, H. (1995). Nitrate respiration in chemoautotrophic symbionts of the bivalve *Lucinoma aequizonata* is not regulated by oxygen. Appl. Environ. Microbiol. 61, 1630–1633.
- Hentschel, U., and Felbeck, H. (1993). Nitrate respiration in the hydrothermal vent tubeworm *Riftia pachyptila*. Nature 366, 338–340. doi:10.1038/366338a0.

❖ Hentschel, U., Hand, S. C., and Felbeck, H. (1996). The contribution of nitrate respiration to the energy budget of the symbiont-containing clam *Lucinoma aequizonata*: A calorimetric study. J. Exp. Biol. 199, 427–433. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9318064.

❖ Hongoh, Y., and Ishikawa, H. (1997). Uric Acid as a Nitrogen Resource for the Brown Planthopper, *Nilaparvata lugens*: Studies with Synthetic Diets and Aposymbiotic Insects. Zoolog. Sci. 14, 581–586. doi:10.2108/zsj.14.581.

❖ Hongoh, Y., Sasaki, T., and Ishikawa, H. (2000). Cloning, sequence analysis and expression in *Escherichia coli* of the gene encoding a uricase from the yeast-like symbiont of the brown planthopper, *Nilaparvata lugens*. Insect Biochem. Mol. Biol. 30, 173–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10696593.

❖ Husník, F., Chrudimský, T., and Hypša, V. (2011). Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches. BMC Biol. 9, 87. doi:10.1186/1741-7007-9-87.

❖ Igarashi, K., and Kashiwagi, K. (2000). Polyamines: mysterious modulators of cellular functions. Biochem Biophys Res Commun 271, 559–564. doi:10.1006/bbrc.2000.2601.

❖ Ikeya, T., Broughton, S., Alic, N., Grandison, R., and Partridge, L. (2009). The endosymbiont Wolbachia increases insulin/IGF-like signalling in *Drosophila*. Proc. Biol. Sci. 276, 3799–3807. doi:10.1098/rspb.2009.0778.

❖ Janson, E. M., Stireman, J. O., Singer, M. S., and Abbot, P. (2008). Phytophagous insect-microbe mutualisms and adaptive evolutionary diversification. Evolution 62, 997–1012. doi:10.1111/j.1558-5646.2008.00348.x.

❖ Jones, K. G., Dowd, P. F., and Blackwell, M. (1999). Polyphyletic origins of yeast-like endocytobionts from anobiid and cerambycid beetles. Mycol. Res. 103, 542–546. Available at: http://journals.cambridge.org/abstract_S0953756298007308

❖ Kaltenpoth, M. (2009). Actinobacteria as mutualists: general healthcare for insects? Trends Microbiol. 17, 529–35. doi:10.1016/j.tim.2009.09.006.

❖ Kikuchi, Y., Meng, X., and Fukatsu, T. (2005). Gut symbiotic bacteria of the genus *Burkholderia* in the broad-headed bugs *Riptortus clavatus* and *Leptocorisa chinensis* (Heteroptera: Alydidae). Appl. Environ. 71, 4035–4043. doi:10.1128/AEM.71.7.4035.

❖ Kim, J. K., Kim, N. H., Jang, H. A., Kikuchi, Y., Kim, C.-H., Fukatsu, T., and Lee, B. L. (2013). Specific midgut region controlling the symbiont population in an insect-microbe gut symbiotic association. Appl. Environ. Microbiol. 79, 7229–33. doi:10.1128/AEM.02152-13.

❖ Klepzig, K., and Six, D. (2004). Bark beetle-fungal symbiosis: Context dependency in complex associations. Symbiosis 37, 189–205.

❖ Kneip, C., Lockhart, P., Voss, C., and Maier, U.-G. (2007). Nitrogen fixation in eukaryotes-new models for symbiosis. BMC Evol. Biol. 7, 55. doi:10.1186/1471-2148-7-55.

❖ Koga, R., and Moran, N. A. (2014). Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. ISME J. 8, 1237–46. doi:10.1038/ismej.2013.235.

❖ Kumar, S., Bourdès, A., and Poole, P. (2005). De novo alanine synthesis by bacteroids of *Mesorhizobium loti* is not required for nitrogen transfer in the determinate nodules of *Lotus corniculatus*. J. Bacteriol. 187, 5493–5495. doi:10.1128/JB.187.15.5493-5495.2005.

❖ Lo, W.-S., Huang, Y.-Y., and Kuo, C.-H. (2016). Winding paths to simplicity: genome evolution in facultative insect symbionts. FEMS Microbiol Rev 2, 158–161. doi:10.1093/femsre/fuw028.

❖ Login, F. H., Balmand, S., Vallier, A., Vincent-Monégat, C., Vigneron, A., Weiss-Gayet, M., Rochat, D., and Heddi, A. (2011). Antimicrobial peptides keep insect endosymbionts under control. Science 334, 362–5. doi:10.1126/science.1209728.

❖ Masson, F., Moné, Y., Vigneron, A., Vallier, A., Parisot, N., Vincent-Monégat, C., Balmand, S., Carpentier, M.-C., Zaidman-Rémy, A., and Heddi, A. (2015). Weevil endosymbiont dynamics is associated with a clamping of immunity. BMC Genomics 16, 819. doi:10.1186/s12864-015-2048-5.

- Matsuura, Y., Kikuchi, Y., Hosokawa, T., Koga, R., Meng, X.-Y., Kamagata, Y., Nikoh, N., and Fukatsu, T. (2012a). Evolution of symbiotic organs and endosymbionts in lygaeid stinkbugs. ISME J. 6, 397–409. doi:10.1038/ismej.2011.103.

- Matsuura, Y., Kikuchi, Y., Meng, X. Y., Koga, R., and Fukatsu, T. (2012b). Novel clade of alphaproteobacterial endosymbionts associated with stinkbugs and other arthropods. Appl. Environ. Microbiol. 78, 4149–4156. doi:10.1128/AEM.00673-12.

- McCutcheon, J. P., and Von Dohlen, C. D. (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr. Biol. 21, 1366–1372. doi:10.1016/j.cub.2011.06.051.

- McCutcheon, J. P., McDonald, B. R., and Moran, N. A. (2009). Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc. Natl. Acad. Sci. U. S. A. 106, 15394–9. doi:10.1073/pnas.0906424106.

- McCutcheon, J. P., and Moran, N. a (2012). Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. 10, 13–26. doi:10.1038/nrmicro2670.

- McCutcheon, J. P., and Moran, N. A. (2010). Functional convergence in reduced genomes of bacterial symbionts spanning 200 my of evolution. Genome Biol. Evol. 2, 708–718. doi:10.1093/gbe/evq055.

- McCutcheon, J. P., and Moran, N. A. (2007). Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc. Natl. Acad. Sci. U. S. A. 104, 19392–19397. doi:10.1073/pnas.0708855104.

- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14, 60. doi:10.1186/1471-2105-14-60.

- Menezes, C., Vollet-Neto, A., Marsaioli, A. J., Zampieri, D., Fontoura, I. C., Luchessi, A. D., and Imperatriz-Fonseca, V. L. (2015). A Brazilian Social Bee Must Cultivate Fungus to Survive. Curr. Biol. 25, 2851–2855. doi:10.1016/j.cub.2015.09.028.

- Morales-Jiménez, J., Vera-Ponce de León, A., 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. doi:10.1007/s00248-013-0206-3.

- Morales-Jiménez, J., Zúñiga, G., Villa-Tanaca, L., and Hernández-Rodríguez, C. (2009). Bacterial community and nitrogen fixation in the red turpentine beetle, *Dendroctonus valens* LeConte (Coleoptera: Curculionidae: Scolytinae). Microb. Ecol. 58, 879–91. doi:10.1007/s00248-009-9548-2.

- Moran, N. a (2006). Symbiosis. Curr. Biol. 16, R866–71. doi:10.1016/j.cub.2006.09.019.

- Moran, N. a, McCutcheon, J. P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. 42, 165–90. doi:10.1146/annurev.genet.41.110306.130119.

- Mullins, D. E., and Cochran, D. G. (1975). Nitrogen metabolism in the american cockroach—I. An examination of positive nitrogen balance with respect to uric acid stores. Comp. Biochem. Physiol. Part A Physiol. 50, 489–500. doi:10.1016/0300-9629(75)90306-0.

- Murphy, K. M., Rae, I. C. Mac, and Teakle, D. S. (1988). Nitrogenase Activity in the Queensland Fruit Fly , *Dacus tryoni*. 447–452.

- Nakabachi, A., Ueoka, R., Oshima, K., Teta, R., Mangoni, A., Gurgui, M., Oldham, N. J., Van Echten-Deckert, G., Okamura, K., Yamamoto, K., et al. (2013). Defensive bacteriome symbiont with a drastically reduced genome. Curr. Biol. 23, 1478–1484. doi:10.1016/j.cub.2013.06.027.

- Nasir, H., and Noda, H. (2003). Yeast-like symbiotes as a sterol source in anobiid beetles (Coleoptera, Anobiidae): possible metabolic pathways from fungal sterols to 7-dehydrocholesterol. Arch. Insect Biochem. Physiol. 52, 175–82. doi:10.1002/arch.10079.

- Noda, H., and Kodama, K. (1996). Phylogenetic position of yeastlike endosymbionts of anobiid beetles. Appl. Environ. Microbiol. 62, 162–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=167783&tool=pmcentrez&rendertype=abstract.

- Noda, H., Watanabe, K., Kawai, S., Yukuhiro, F., Miyoshi, T., Tomizawa, M., Koizumi, Y., Nikoh, N., and Fukatsu, T. (2012). Bacteriome-associated endosymbionts of the green rice leafhopper *Nephotettix cincticeps*

(Hemiptera: Cicadellidae). Appl. Entomol. Zool. 47, 217–225. doi:10.1007/s13355-012-0110-1.

❖ Ohlendieck, U., Stuhr, A., and Siegmund, H. (2000). Nitrogen fixation by diazotrophic cyanobacteria in the Baltic Sea and transfer of the newly fixed nitrogen to picoplankton organisms. J. Mar. Syst. 25, 213–219. doi:10.1016/S0924-7963(00)00016-6.

❖ Pankewitz, F., Zöllmer, A., Hilker, M., and Gräser, Y. (2007). Presence of *Wolbachia* in insect eggs containing antimicrobially active anthraquinones. Microb. Ecol. 54, 713–21. doi:10.1007/s00248-007-9230-5.

❖ Pant, R. (1988). Nitrogen excretion in insects. Proc. Anim. Sci. 97, 379–415. doi:10.1007/BF03179946.

❖ Patiño-Navarrete, R., Piulachs, M.-D., Belles, X., Moya, A., Latorre, A., and Peretó, J. (2014). The cockroach *Blattella germanica* obtains nitrogen from uric acid through a metabolic pathway shared with its bacterial endosymbiont. Biol. Lett. 10, 7–10. doi:10.1098/rsbl.2014.0407.

❖ Perez-Guerra, G., and Kosztarab, M. (1992). Biosystematics of the family Dactylopiidae (Homoptera: Coccinea) with emphasis on the life cycle of *Dactylopius coccus* Costa. Bull. Virginia Agric. Exp. Stn. 92, 1–90.

❖ Pérez-Guerra, G. M. del (1991). Biosystematics of the family Dactylopiidae (Homoptera:Coccinea) whit emphasis on the life cycle of *Dactylopius coccus* Costa.

❖ Pinto-Tomás, A. A., Anderson, M. A., Suen, G., Stevenson, D. M., Chu, F. S. T., Cleland, W. W., Weimer, P. J., and Currie, C. R. (2009). Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. Science 326, 1120–3. doi:10.1126/science.1173036.

❖ Potrikus, C. J., and Breznak, J. a (1980). Anaerobic degradation of uric Acid by gut bacteria of termites. Appl. Environ. Microbiol. 40, 125–32. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=291535&tool=pmcentrez&rendertype=abstract.

❖ Ramírez-Puebla, S. T., Ormeño-Orrillo, E., León, A. V.-P. de, Lozano, L., Sánchez, A., Rosenblueth, M., and Martínez-Romero, E. (2016). Genomes of *Candidatus* Wolbachia Bourtzisii wDacA and *Candidatus* Wolbachia Pipientis

wDacB from the Cochineal Insect *Dactylopius coccus* (Hemiptera: Dactylopiidae). G3 Genes|Genomes|Genetics 6, g3.116.031237. doi:10.1534/G3.116.031237.

- Ramírez-Puebla, S. T., Rosenblueth, M., Chávez-Moreno, C. K., Catanho Pereira de Lyra, M. C., Tecante, A., and Martínez-Romero, E. (2010). Molecular phylogeny of the genus *Dactylopius* (Hemiptera: Dactylopiidae) and identification of the symbiotic bacteria. Environ. Entomol. 39, 1178–83. doi:10.1603/EN10037.

- Ramírez-Puebla, S. T., Servín-Garcidueñas, L. E., Ormeño-Orrillo, E., Vera-Ponce de León, A., Rosenblueth, M., Delaye, L., Martínez, J., and Martínez-Romero, E. (2015). Species in *Wolbachia*? Proposal for the designation of "*Candidatus* Wolbachia bourtzisii", "*Candidatus* Wolbachia onchocercicola", "*Candidatus* Wolbachia blaxteri", "*Candidatus* Wolbachia brugii", "*Candidatus* Wolbachia taylori", '*Candidatus* Wolbachia collembol. Syst. Appl. Microbiol. 38, 390–9. doi:10.1016/j.syapm.2015.05.005.

- Reinhold-hurek, B., and Hurek, T. (2006). The Genera *Azoarcus* , Azovibrio , Azospira and Azonexus. The Prokaryotes 5 (Proteob, 873–891. doi:10.1007/0-387-30745-1.

- Reyes-Prieto, M., Latorre, A., and Moya, A. (2014). Scanty microbes, the "symbionelle" concept. Environ. Microbiol. 16, 335–338. doi:10.1111/1462-2920.12220.

- Ricci, I., Damiani, C., Scuppa, P., Mosca, M., Crotti, E., Rossi, P., Rizzi, A., Capone, A., Gonella, E., Ballarini, P., et al. (2011a). The yeast *Wickerhamomyces anomalus* (*Pichia anomala*) inhabits the midgut and reproductive system of the Asian malaria vector *Anopheles stephensi*. Environ. Microbiol. 13, 911–21. doi:10.1111/j.1462-2920.2010.02395.x.

- Ricci, I., Mosca, M., Valzano, M., Damiani, C., Scuppa, P., Rossi, P., Crotti, E., Cappelli, A., Ulissi, U., Capone, A., et al. (2011b). Different mosquito species host *Wickerhamomyces anomalus* (*Pichia anomala*): perspectives on vector-borne diseases symbiotic control. Antonie Van Leeuwenhoek 99, 43–50. doi:10.1007/s10482-010-9532-3.

- Rincón-Rosales, R., and Gutiérrez-Miceli, F. (2008). Características biológicas de *Acaciella angustissima* (Mill.) Britton & Rose en su habitat natural y evaluación de su potencial cortical en Chiapas, México. Agrociencia, 129–

137. Available at: http://www.scielo.org.mx/scielo.php?pid=S1405-31952008000100013&script=sci_arttext [Accessed January 11, 2016].

- Rivera, F. N., González, E., Gómez, Z., López, N., Hernández-Rodríguez, C., Berkov, A., and Zúñiga, G. (2009). Gut-associated yeast in bark beetles of the genus *Dendroctonus* Erichson (Coleoptera: Curculionidae: Scolytinae). Biol. J. Linn. Soc. 98, 325–342. doi:10.1111/j.1095-8312.2009.01289.x.

- Rodríguez, L. C., Faúndez, E., Seymour, J., Escobar, C. A., Espinoza, L., Petroutsa, M., Ayres, A., and Niemeyer, H. M. (2005). Factores Bióticos y Concentración de Ácido Carmínico en la Cochinilla (*Dactylopius coccus* Costa) (Homoptera: Dactylopiidae). Agric. Técnica 65, 323–329. doi:10.4067/S0365-28072005000300011.

- Rosas-Pérez, T., Rosenblueth, M., Rincón-Rosales, R., Mora, J., and Martínez-Romero, E. (2014). Genome sequence of "*Candidatus* Walczuchella monophlebidarum" the flavobacterial endosymbiont of *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae). Genome Biol. Evol. 6, 714–26. doi:10.1093/gbe/evu049.

- Rosenblueth, M., Sayavedra, L., Sámano-Sánchez, H., Roth, A., and Martínez-Romero, E. (2012). Evolutionary relationships of flavobacterial and enterobacterial endosymbionts with their scale insect hosts (Hemiptera: Coccoidea). J. Evol. Biol. 25, 2357–2368. doi:10.1111/j.1420-9101.2012.02611.x.

- Rosselló-Móra, R., and Amann, R. (2015). Past and future species definitions for Bacteria and Archaea. Syst. Appl. Microbiol. 38, 209–216. doi:10.1016/j.syapm.2015.02.001.

- Sabree, Z. L., Kambhampati, S., and Moran, N. a (2009). Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. Proc. Natl. Acad. Sci. U. S. A. 106, 19521–6. doi:10.1073/pnas.0907504106.

- Sacchi, L., Genchi, M., Clementi, E., Bigliardi, E., Avanzati, a M., Pajoro, M., Negri, I., Marzorati, M., Gonella, E., Alma, a, et al. (2008). Multiple symbiosis in the leafhopper *Scaphoideus titanus* (Hemiptera: Cicadellidae):

details of transovarial transmission of *Cardinium* sp. and yeast-like endosymbionts. Tissue Cell 40, 231–42. doi:10.1016/j.tice.2007.12.005.

- Sapountzis, P., Zhukova, M., Hansen, L. H., Sørensen, S. J., Schiøtt, M., and Boomsma, J. J. (2015). *Acromyrmex* leaf-cutting ants have simple gut microbiota with nitrogen-fixing potential. Appl. Environ. Microbiol. 81, 5527–5537. doi:10.1128/AEM.00961-15.

- Sasaki, T., Kawamura, M., and Ishikawa, H. (1996). Nitrogen recycling in the brown planthopper, *Nilaparvata lugens*: Involvement of yeast-like endosymbionts in uric acid metabolism. J. Insect Physiol. 42, 125–129.

- Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., and Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407, 81–86. doi:10.1038/35024074.

- Stintzing, F. C., and Carle, R. (2005). Cactus stems (*Opuntia* spp.): a review on their chemistry, technology, and uses. Mol. Nutr. Food Res. 49, 175–94. doi:10.1002/mnfr.200400071.

- Suh, S.-O., McHugh, J. V., Pollock, D. D., and Blackwell, M. (2005). The beetle gut: a hyperdiverse source of novel yeasts. Mycol. Res. 109, 261–265. doi:10.1017/S0953756205002388.

- Tai, V., Carpenter, K. J., Weber, P. K., Nalepa, C. a, Perlman, S. J., and Keeling, P. J. (2016). Genome evolution and nitrogen-fixation in bacterial ectosymbionts of a protist inhabiting wood-feeding cockroaches. Appl. Environ. Microbiol. doi:10.1128/AEM.00611-16.

- Thomas, J., Meeks, J. C., Wolk, C. P., Shaffer, P. W., Austin, S. A. M. M., and W.-S. CHIEN (1977). Formation of glutamine from [^{13}N] ammonia, [^{13}N]dinitrogen , and [^{14}C] glutamate by heterocysts isolated from *Anabaena cylindrica*. J. Bacteriol. 129, 1545–1555.

- Torres, M. J., Simon, J., Rowley, G., Bedmar, E. J., and Richardson, D. J. (2016). Nitrous Oxide Metabolism in Nitrate-Reducing Bacteria : Physiology and Regulatory Mechanisms. 1st ed. Elsevier Ltd. doi:10.1016/bs.ampbs.2016.02.007.

- Varghese, N. J., Mukherjee, S., Ivanova, N., Konstantinidis, K. T., Mavrommatis, K., Kyrpides, N. C., and Pati, A. (2015). Microbial species

delineation using whole genome sequences. Nucleic Acids Res. 43, gkv657–. doi:10.1093/nar/gkv657.

- Vega, F., and Blackwell, M. (2005). Insect-fungal associations: ecology and evolution. , eds. F. E. Vega and M. Blackwell New York: Oxford University Press Available at: http://fmedicine.ajums.ac.ir/_fmedicine/documents/Insect fungal associations ecology and evolution_20130423_101512.pdf [Accessed August 15, 2014].

- Vera-Ponce de León, A., Sanchez-Flores, A., Rosenblueth, M., and Martínez-Romero, E. (2016). Fungal Community Associated with *Dactylopius* (Hemiptera: Coccoidea: Dactylopiidae) and Its Role in Uric Acid Metabolism. Front. Microbiol. 7, 1–15. doi:10.3389/fmicb.2016.00954.

- Vigneron, A., Masson, F., Vallier, A., Balmand, S., Rey, M., Vincent-Monégat, C., Aksoy, E., Aubailly-Giraud, E., Zaidman-Rémy, A., and Heddi, A. (2014). Insects recycle endosymbionts when the benefit is over. Curr. Biol. 24, 2267–2273. doi:10.1016/j.cub.2014.07.065.

- Watanabe, K., Yukuhiro, F., Matsuura, Y., Fukatsu, T., and Noda, H. (2014). Intrasperm vertical symbiont transmission. Proc. Natl. Acad. Sci. U. S. A. 111, 7433–7. doi:10.1073/pnas.1402476111.

- Werren, J. H., Baldo, L., and Clark, M. E. (2008). *Wolbachia*: master manipulators of invertebrate biology. Nat. Rev. Microbiol. 6, 741–751.

- Williams, D. J., and Ben-Dov, Y. (2015). Scale insect species names that have been combined with the genus name *Dactylopius Costa* (Hemiptera: Sternorrhyncha: Coccomorpha). Zootaxa 4006, 161–70. doi:http://dx.doi.org/10.11646/zootaxa.4006.1.8.

- Williams, M., and MacVean, C. (1995). Ethnococcidology: use of the giant margarodids, by indigenous peoples of mesoamerica in their culture , medicine and arts. Isr. Jounal Entomol. XXIX, 147–148.

- Zacchi, L., and Vaughan-Martini, A. (2003). Distribution of three yeast and yeast-like species within a population of soft scale insects (*Saissetia oleae*) as a function of developmental age. Ann. Microbiol. 53, 43–46. Available at: <Go to ISI>://000182397000005.

- Zhou, L., Wang, J., and Zhang, L. H. (2007). Modulation of bacterial type III secretion system by a spermidine transporter dependent signaling pathway. PLoS One 2. doi:10.1371/journal.pone.0001291.

- Zug, R., and Hammerstein, P. (2014). Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. Biol. Rev. 90, 89–111. doi:10.1111/brv.12098.

Anexos

frontiers in Microbiology

ORIGINAL RESEARCH
published: 23 June 2016
doi: 10.3389/fmicb.2016.00954

Fungal Community Associated with *Dactylopius* (Hemiptera: Coccoidea: Dactylopiidae) and Its Role in Uric Acid Metabolism

Arturo Vera-Ponce de León[1]*, *Alejandro Sanchez-Flores*[2], *Mónica Rosenblueth*[1] *and Esperanza Martínez-Romero*[1]

[1] Programa de Ecología Genómica, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavca, Mexico, [2] Unidad de Secuenciación Masiva y Bioinformática, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavca, Mexico

We studied fungal species associated with the carmine cochineal *Dactylopius coccus* and other non-domesticated *Dactylopius* species using culture-dependent and -independent methods. Thirty seven fungi were isolated in various culture media from insect males and females from different developmental stages and *Dactylopius* species. 26S rRNA genes and ITS sequences, from cultured fungal isolates revealed different species of *Cryptococcus, Rhodotorula, Debaryomyces, Trametes,* and *Penicillium*, which are genera newly associated with *Dactylopius*. Uric acid (UA) and uricase activity were detected in tissues extracts from different insect developmental stages. However, accumulation of high UA levels and low uricase activities were found only after antifungal treatments, suggesting an important role of fungal species in its metabolism. Additionally, uricolytic fungal isolates were identified and characterized that presumably are involved in nitrogen recycling metabolism. After metagenomic analyses from *D. coccus* gut and hemolymph DNA and from two published data sets, we confirmed the presence of fungal genes involved in UA catabolism, suggesting that fungi help in the nitrogen recycling process in *Dactylopius* by uricolysis. All these results show the importance of fungal communities in scale insects such as *Dactylopius*.

Keywords: fungal-metagenomics, *Cryptococcus*, scale insects, *Rhodotorula*, ITS region, purine metabolism, carmine cochineal

OPEN ACCESS

Edited by:
Vijai Kumar Gupta, National University of Ireland Galway, Ireland

Reviewed by:
Guilherme Lanzi Sassaki, Universidade Federal do Paraná, Brazil
Alejandro De Las Penas, Instituto Potosino de Investigación Científica y Tecnológica, Mexico

***Correspondence:**
Arturo Vera-Ponce de León
avera@ccg.unam.mx

Specialty section:
This article was submitted to Fungi and Their Interactions, a section of the journal Frontiers in Microbiology

Received: 14 March 2016
Accepted: 02 June 2016
Published: 23 June 2016

Citation:
Vera-Ponce de León A, Sanchez-Flores A, Rosenblueth M and Martínez-Romero E (2016) Fungal Community Associated with Dactylopius (Hemiptera: Coccoidea: Dactylopiidae) and Its Role in Uric Acid Metabolism. Front. Microbiol. 7:954. doi: 10.3389/fmicb.2016.00954

INTRODUCTION

Insects are the most diverse arthropods in the biosphere and dwell in almost all environments. They can feed on a wide variety of nutrients, probably due to their associated microorganisms, including fungal species (Douglas, 2009). There is evidence that many arthropods harbor yeast-like microorganisms inside their bodies (Buchner, 1965), and at least eight orders of insects, including 143 species, have been reported to be associated with fungi (Vega and Blackwell, 2005; Gibson and Hunter, 2010). Fungi are located either inside the insect body in highly specialized cells called mycetocytes, as in *Nilaparvata lugens* and *Drosophila melanogaster*, which harbor yeasts (Chen et al., 1981; Ebbert et al., 2003), or in cavities named mycangia as in bark beetles (Jones et al., 1999; Klepzig and Six, 2004; Ganter, 2006). Fungi have also been found in the insect gut, as well as in their

reproductive organs and fat tissues (Buchner, 1965; Gibson and Hunter, 2009; Rivera et al., 2009; Ricci et al., 2011). Moreover, studies on fungi-insect symbioses show that fungi play important roles in insect development and fitness (Gibson and Hunter, 2010). Fungi are capable of providing nitrogen compounds that are limited in the diets of some insects, or can degrade high molecular weight molecules and produce pheromones for mating and communication (Brand et al., 1976; Sasaki et al., 1996; Nasir and Noda, 2003; Gibson and Hunter, 2010). In some insects like cockroaches, termites, shield bugs, planthoppers, and bark beetles uric acid (UA), the major product of nitrogen excretion, is recycled by bacterial or fungal symbionts (Mullins and Cochran, 1975; Potrikus and Breznak, 1981; Pant, 1988; Kashima et al., 2006; Morales-Jiménez et al., 2013; Patiño-Navarrete et al., 2014). However, to our knowledge, there are no reports on the UA content or catabolism in scale insects.

The Dactylopiidae family includes only one genus, *Dactylopius* (Costa), commonly called "cactus cochineals" or "cochineal scale insects." They are obligate phytophagous hemipterous from the scale insects family (Coccoidea). Ten species have been described as belonging to this genus and six of them, *D. ceylonicus*, *D. confusus*, *D. opuntiae*, *D. coccus*, *D. bassi,* and *D. tomentosus,* inhabit Mexico (Ben-Dov and Marotta, 2001; Chávez-Moreno et al., 2009). These insects are the main source of carminic acid, a glycoside-anthraquinone molecule used in the textile, cosmetic, pharmaceutical, and food industries as a dye or pigment (Deveoglu et al., 2011). All of these *Dactylopius* species produce carminic acid, but only *D. coccus* is cultivated and used for commercial purposes due to the higher amount and quality of its pigment (Rodríguez et al., 2005). Moreover, since non-cultivated *Dactylopius* are considered a cactus plague, in some countries they are used as biological control for these plants (Zimmermann and Moran, 1991; Spodek et al., 2013; Pérez-Ramirez et al., 2014; da Silva Santos et al., 2015).

Dactylopius cochineals spend their life feeding on *Opuntia* and *Nopalea* cactus sap (Chávez-Moreno et al., 2009), which is mainly composed of water (88–95% wet weight) and has low protein concentration (0.5–1% wet weight; Stintzing and Carle, 2005). Thus, we supposed that nitrogen deficiencies may be supplied by associated symbiotic microorganisms. The diversity of microbial symbionts in *Dactylopius* has been scarcely described. There are a few reports of the bacterial communities in *Dactylopius* species (Pankewitz et al., 2007; Ramírez-Puebla et al., 2010, 2015). However, there are no reports on the fungal community and their possible roles in association with this cochineal insect. The aim of this work was to identify and describe fungi from diverse stages and tissues of *Dactylopius* species, as well as to determine their role in uric acid catabolism in these insects.

TABLE 1 | Collection sites of *Dactylopius* species.

Location	Location code	Latitude/Longitude	Insect species
Campo Carmín, Xochitepec, Morelos state	CC	18°44′46.7″N 99°11′17.8″W	*D. coccus*
Teotihuacán, Mexico state	TEM	19°40′47.3″N 98°50′59.4″W	*D. opuntiae*
Ecatepec, Mexico state	ECM	19°35′27.3″N 98°59′57.5″W	*D. opuntiae*
Jiutepec, Morelos state	JM	18°53′52.5″N 99°10′56.8″W	*D. opuntiae*
Coyoacán, Federal district	CDF	19°19′18.9″N 99°11′09.8″W	*D. confusus*
Milpalta, Federal district	MADF	19°12′26.7″N 99°1′28.8″W	*D. confusus*

MATERIALS AND METHODS

Insect Sampling and Identification

Dactylopius coccus samples were obtained from Campo Carmín Company (**Table 1**). Wild species of *Dactylopius* (*D. opuntiae* and *D. confusus*) were collected from three states in Mexico (**Table 1**). Insects were obtained from *Opuntia* spp. cactus and were transported together with their host plants to the laboratory. For species identification, ten female adults from the different locations were preserved in fixation buffer (chloroform: ethanol: glacial acetic acid 4:3:1). The superficial wax was removed by placing the insects in 10% KOH for 10 min at 60°C. Body contents were removed by cutting a slit in the body margin and expelling the contents with a spatula. Cleaned specimens were transferred into 70% alcohol for 10 min. Then, all specimens were transferred and kept in a staining solution (2% aqueous solution of acid fuchsin) overnight. Specimens were washed in 70% alcohol for 10 min and dehydrated in 100% alcohol for 10 min. Each specimen was placed face down on a slide with a drop of Canada balsam and covered with a slip. Microscopic observations with the keys described by Perez-Guerra and Kosztarab allowed the morphological identification of *Dactylopius* species (Perez-Guerra and Kosztarab, 1992). Specimens were deposited in the collection of Héctor González-Hernández from COLPOS, Mexico.

Fungal Isolation, DNA Extraction, and PCR Amplification

Insects from 1st instar nymph, 2nd instar nymph and adult stages of *D. coccus* and of wild *Dactylopius* (*D. opuntiae* and *D. confusus*) were detached from their host plant, submerged in 100% ethanol and the wax cover was removed with forceps under a stereoscope. They were then surface disinfected with 70% ethanol and rinsed twice with sterile water. A pool of five washed and disinfected insects from each developmental stage mentioned above of *D. coccus, D. opuntiae, D. confusus* and a pool of 20 *D. coccus* adult males were totally macerated (hereafter named as whole body samples) with a sterile Eppendorf® pestle in a 1.5 microtube with 500 µl of 0.85% NaCl. Additionally, two

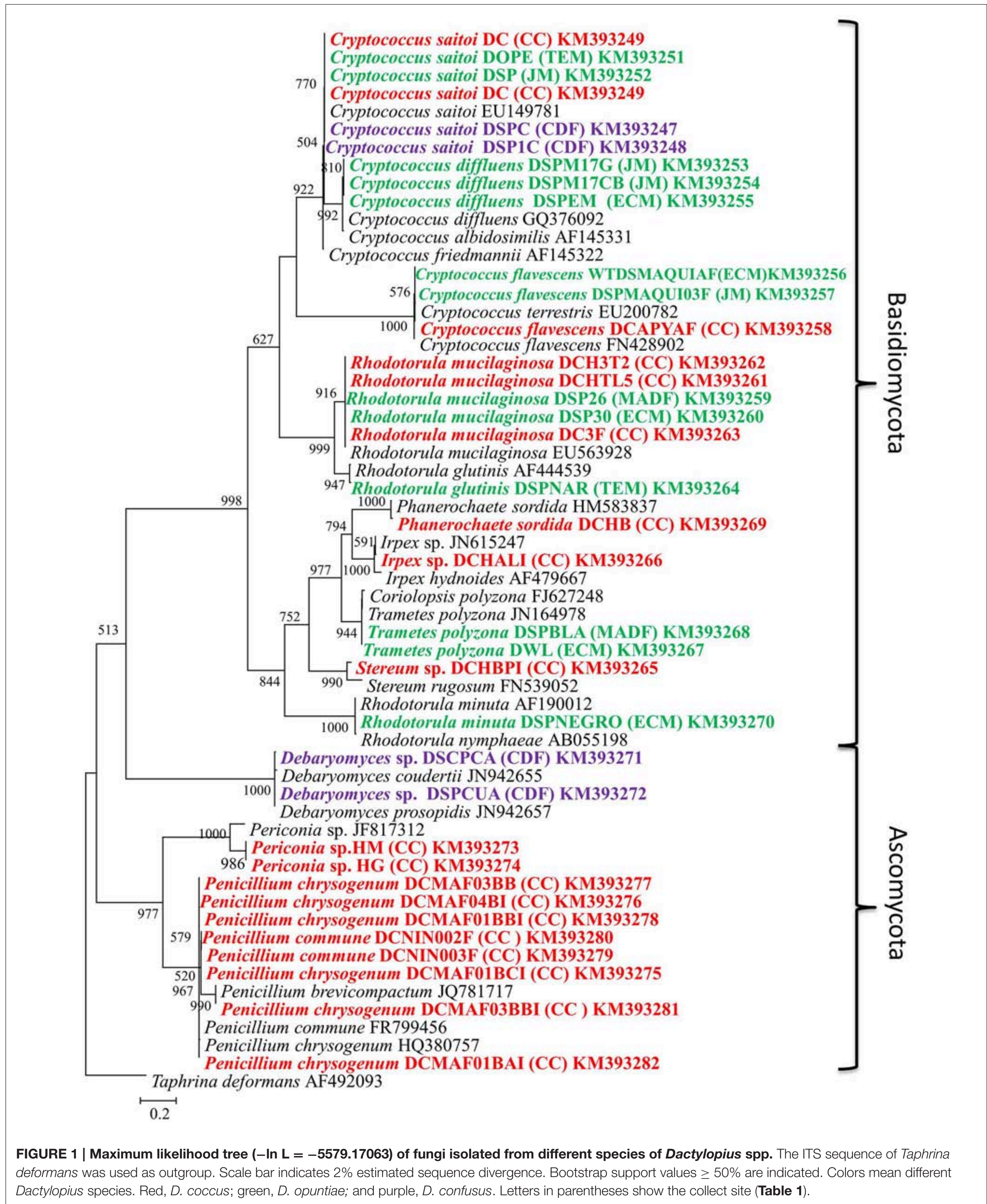


FIGURE 1 | Maximum likelihood tree (−ln L = −5579.17063) of fungi isolated from different species of *Dactylopius* spp. The ITS sequence of *Taphrina deformans* was used as outgroup. Scale bar indicates 2% estimated sequence divergence. Bootstrap support values ≥ 50% are indicated. Colors mean different *Dactylopius* species. Red, *D. coccus*; green, *D. opuntiae;* and purple, *D. confusus*. Letters in parentheses show the collect site (**Table 1**).

TABLE 2 | Fungi associated with different *Dactylopius* species in culture-dependent analysis.

Insect host	Isolate name	Most related fungi ITS sequence from GenBank (identity %)	Most related fungi 26S rRNA sequence from GenBank (identity %)	OTU Number	Morphology Yeast (Y) Mold (M)	Isolated from: Ovary-eggs (O) Gut (G) Whole body (W)	Insect host stage
Dactylopius coccus	DCHTL5	*Rhodotorula mucilaginosa* EU56392 (100)	*Rhodotorula mucilaginosa* DQ832198 (100)	1	Y	G	Adult female
	DC3F				Y	O	Egg
	DCH3T2				Y	W	Adult female
	DC	*Cryptococcus saitoi* EU149781 (100)	*Cryptococcus saitoi* JX188127 (100)	4	Y	W	Adult female
	DCAPYAF	*Cryptococcus flavescens* FN428902 (99.76)	*Cryptococcus flavescens* FJ743610 (98.5)	5	Y	W	Adult female
	DCHBPI	*Stereum* sp. GQ999353 (77.58)	*Phlebiopsis flavidoalba* EU118662 (97.8)	9	M	W	Adult female
	DCALI	*Irpex* sp. JN615247 (99.78)	*Irpex lacteus* JN710547 (99.8)	8	M	G	Adult female
	DCHBP	*Trametes polyzona* JN164978 (99.77)	*Trametes polyzona* JN164790 (100)	7	M	G	Adult female
	HG	*Periconia* sp. JN164978 (88.85)	*Periconia macrospinosa* JN859484 (93.74)	11	M	O	Egg
	HM				M	O	Egg
	DCHB	*Phanerochaete sordida* HM583837 (98.60)	*Phanerochaete sordida* HM595608 (97.8)	10	M	G	Adult female
	DCNin003F	*Penicillium commune* FR799456 (99.06)	*Penicillium nalgiovense* JQ434685 (100)	13	M	W	1st instar
	DCNin002F				M	G	1st instar
	DCNIN01F	*Penicillium chrysogenum* HQ380757 (99.76)	*Penicillium cavernicola* JQ434692 (100)	14	M	W	1st instar
	DCMAF01BCI				M	W	Adult male
	DCMAF04BI				M	W	Adult male
	DCMAF01BAI				M	W	Adult male
	DCMAF01BBI				M	W	Adult male
	DCMAF03BB				M	W	Adult male
Dactylopius confusus	DSPC	*Cryptococcus saitoi* EU149781 (100)	*Cryptococcus saitoi* JX188127 (100)	4	Y	W	Adult female
	DSCP1C				Y	G	Adult female
	DSP26	*Rhodotorula mucilaginosa* EU56392 (100)	*Rhodotorula mucilaginosa* DQ832198 (100)	1	Y	G	2nd instar
	DSPCUA	*Debaryomyces prosopidis* JN942657 (100)	*Debaryomyces hansenii* AB470569 (100)	12	Y	G	Adult female
	DSPA				Y	G	Adult female
Dactylopius opuntiae	DSPNAR	*Rhodotorula glutinis* AF444539 (100)	*Rhodotorula glutinis* KC494740 (100)	2	Y	G	Adult female
	DSP30	*Rhodotorula mucilaginosa* EU56392 (100)	*Rhodotorula mucilaginosa* DQ832198 (100)	1	Y	G	2nd instar
	DSPNEGRO	*Rhodotorula minuta* AF190012 (100)	*Rhodotorula minuta* EU583491 (99.8)	3	Y	G	Adult female
	DWL	*Trametes polyzona* JN164978 (99.77)	*Trametes polyzona* JN164790 (100)	7	Y	W	Adult female
	DSPMGT17CB	*Cryptococcus diffluens* GQ376092 (99.58)	*Cryptococcus diffluens* AF335981 (100)	6	Y	G	Adult female

(Continued)

TABLE 2 | Continued

Insect host	Isolate name	Most related fungi ITS sequence from GenBank (identity %)	Most related fungi 26S rRNA sequence from GenBank (identity %)	OTU Number	Morphology Yeast (Y) Mold (M)	Isolated from: Ovary-eggs (O) Gut (G) Whole body (W)	Insect host stage
	DSPEM				Y	G	2nd instar
	DSPM17G				Y	G	Adult female
	DOP	*Cryptococcus saitoi* EU149781 (100)	*Cryptococcus saitoi* JX188127 (100)	4	Y	W	Adult female
	DOPE				Y	O	Egg
	DSP				Y	W	1st instar
	WTDSMAQUIAF	*Cryptococcus flavescens* FN428902 (100)	*Cryptococcus flavescens* FJ743610 (98.5)	5	Y	G	Adult female
	DSPMAQUI03F				Y	G	Adult female
	DSPBLA	*Trametes polyzona* JN164978 (99.77)	*Trametes polyzona* JN164790 (100)	7	M	G	Adult female

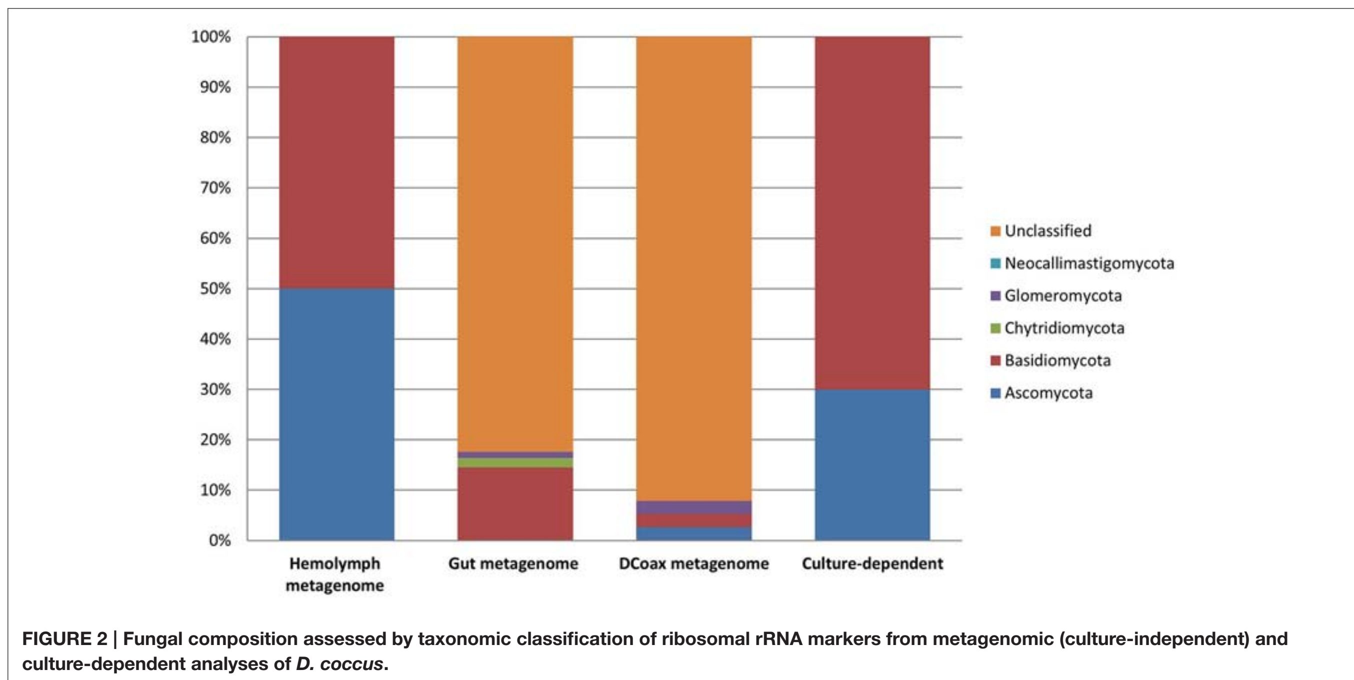


FIGURE 2 | Fungal composition assessed by taxonomic classification of ribosomal rRNA markers from metagenomic (culture-independent) and culture-dependent analyses of *D. coccus*.

individuals of 2nd instar nymphs and adult females from *D. coccus*, *D. opuntiae,* and *D. confusus* were dissected under sterile conditions to obtain guts (gut samples) and ovary-eggs (ovary samples). Dissections were performed by making a transverse cut in the cuticle and removing the organs with fine sterile forceps. These organs were submerged in 600 µl of sterile 0.85% NaCl and macerated using sterile pestles. After maceration, all samples were indirectly sonicated for 30 s in a Bransonic® Ultrasonic MH Cleaning Bath. One hundred microliters of this suspension were inoculated in 50 ml of YPD media (1% w/v yeast extract, 2% w/v peptone, and 2% w/v dextrose), malt extract media (Difco) and two minimal media: MMT [NH_4Cl 3 g l^{-1}; K_2HPO_4 1 g l^{-1}; $MgSO_4$ 0.025 g l^{-1}; $CaCl_2$ 0.25 g l^{-1}; KCl 0.025 g l^{-1}; $FeSO_4$ 0.02 g l^{-1}; yeast extract (Difco) 0.02 g l^{-1}; trehalose 0.01 g l^{-1}; glucose 10 g l^{-1}; and sucrose 5 g l^{-1}] and MMTC [NH_4Cl3 g l^{-1}; K_2HPO_4 1 g l^{-1}; $MgSO_4$ 0.025 g l^{-1}; $CaCl_2$ 0.25 g l^{-1}; KCl 0.025 g l^{-1}; $FeSO_4$ 0.02 g l^{-1}; $CuSO_4$ 0.02 g l^{-1}; yeast extract (Difco) 0.02 g l^{-1}; Carmine dye 0.01 g l^{-1} (Merck microscopy grade)] and were incubated at 25 ± 2°C at 180 rpm for 72 h. After the incubation period, 100 µl of the liquid medium was spread on the corresponding solid medium for selection of yeast and filamentous isolates. To test the best conditions for growing fungi, 100 µl of the initial macerate suspension was also spread directly on solid media MMTC and MMT and incubated in CO_2 generation Gaspack™ EZ CampyPuch™ System at room temperature for 1 week. Pure cultures were obtained and stored at −70°C in 25% glycerol for further analysis.

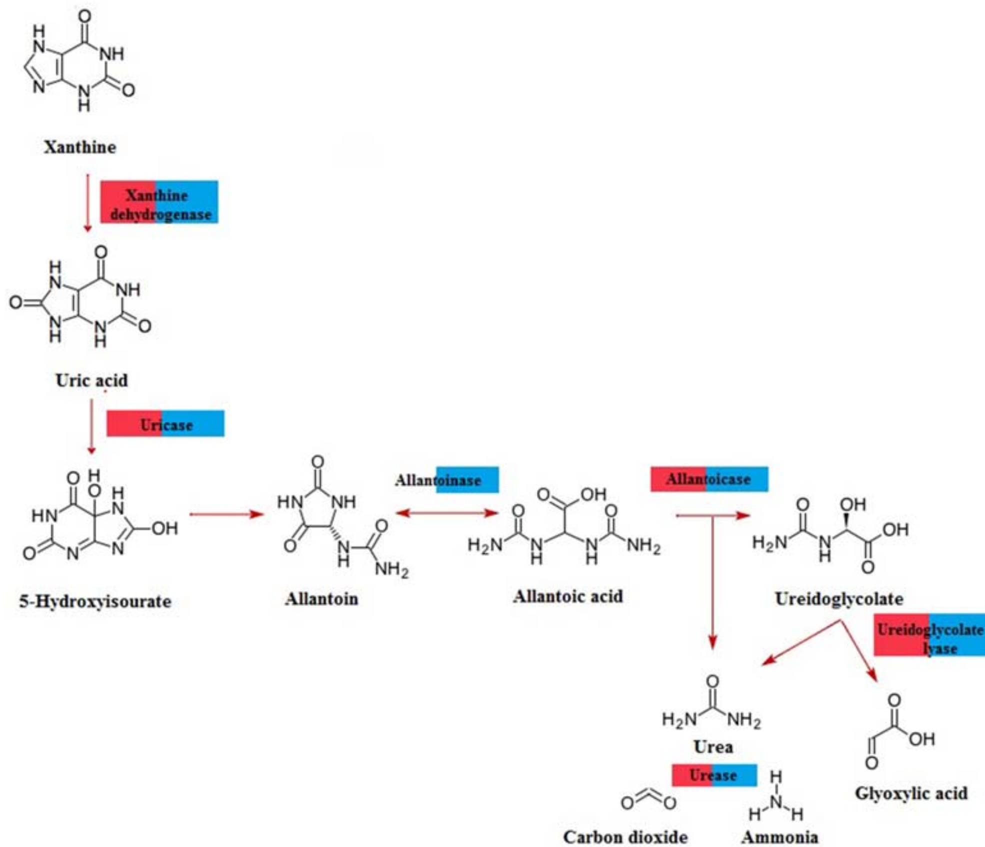


FIGURE 3 | Pathway for purine catabolism of fungal genes predicted from *D. coccus* gut (blue) and hemolymph (red) metagenomes.

DNA from fungal isolates was extracted following the protocols described by Hoffman and Winston (1987). ITS regions were amplified using primers ITS1 (5′ TCCGTAGGTGAACCT GCGG 3′) and ITS2 (5′TCCTCCGCTTATTGATATGC 3′) that we designed for this study. D1-D2 26S rRNA gene region from fungal isolates were amplified using primers 26S-A1 (5′ CAT ATCAATAAGCGGAGCAAAAG 3′) and 26S-A2 (5′ ìCAGTTC TGCTTACCAAAAATGG 3′; Scorzetti et al., 2002). The final concentration for 50 µl PCR reactions was as follows: 10 ng of total DNA, 0.8 pmol of each primer, 0.2 mM dNTPs, 2.5 mM MgCl, 0.5 U *Taq* polymerase and 1x *Taq* polymerase buffer (Invitrogen Life Technologies, Sao Paulo, Brazil). The reaction conditions were 94°C for 5 min; 35 cycles of 60 s at 94°C, 60 s at 57°C, and 90 s at 72°C; and a final extension at 72°C for 10 min. PCR products were purified using the High Pure PCR Product Purification Kit (Roche) and sequenced by Macrogen Inc. (Seoul, Korea) by Sanger technology.

Insect DNA Extraction

For shotgun metagenomic analysis, 30 adult females of *D. coccus* were externally disinfected and dissected as described above. All 30 guts (including the Malpighian tubules) were placed in 200 µl of lysis buffer solution (Tris-HCl 10 mM, pH. 8; EDTA 1 mM; NaCl 10 mM; SDS 1%; Triton X-100 2%). For DNA extraction, samples were macerated with sterile pestles, additionally 0.3 g of sterile glass beads and 200 µl of phenol-chloroform-isoamyl alcohol (25:24:1) were added to the macerate. The samples were mixed by vortexing, warmed at 65°C for 1 h, followed by centrifugation at 15996 × g and the aqueous phase was recovered. Nucleic acids were precipitated with 1 ml of absolute ethanol for 20 min at −20°C, washed twice with 70% ethanol then dried in a vacuum concentrator, resuspended in 50 µl of deionized water and cleaned with DNeasy Blood and Tissue Kit (QIAGEN) columns (this sample is hereafter called as gut metagenome). Additionally, hemolymph from another 30 individuals of *D. coccus* adult females was obtained by dissection. Insect debris was separated by centrifugation in a Percoll (Sigma) gradient, and hemolymph cells were resuspended into 200 µl of PBS and macerated using sterile plastic pestles (Eppendorf). DNA extraction and purification from this sample (hereafter called as hemolymph metagenome) was performed with DNeasy Blood and Tissue Kit (QIAGEN) following manufacturer's instructions.

DNA Sequencing

For gut metagenome DNA Illumina sequencing libraries were prepared using a fragment size of 400 bases and sequenced

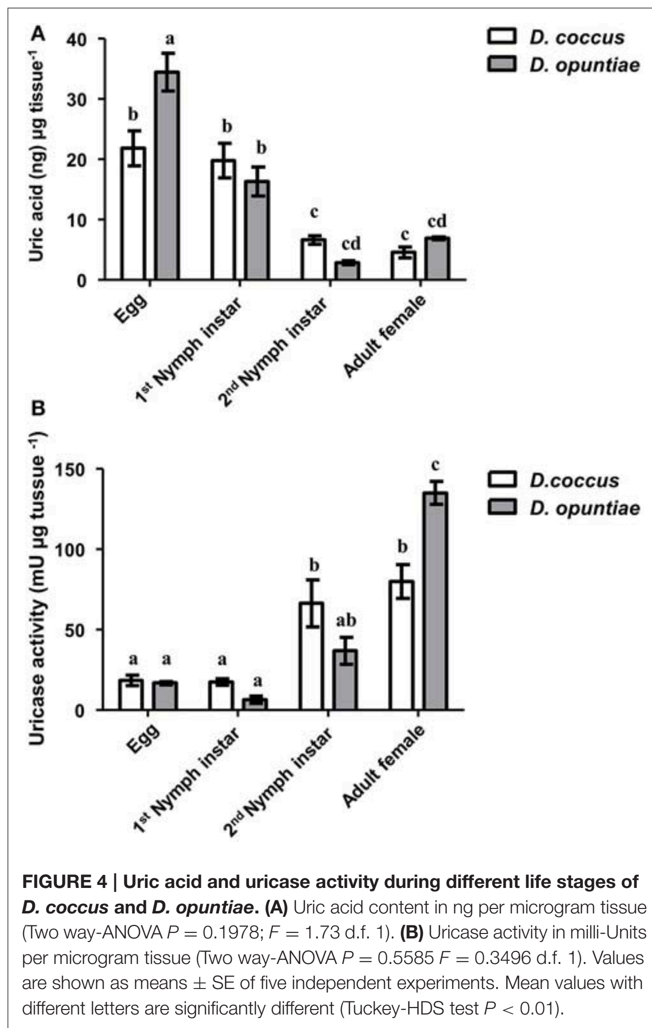


FIGURE 4 | Uric acid and uricase activity during different life stages of *D. coccus* and *D. opuntiae*. (A) Uric acid content in ng per microgram tissue (Two way-ANOVA $P = 0.1978$; $F = 1.73$ d.f. 1). **(B)** Uricase activity in milli-Units per microgram tissue (Two way-ANOVA $P = 0.5585$ $F = 0.3496$ d.f. 1). Values are shown as means ± SE of five independent experiments. Mean values with different letters are significantly different (Tuckey-HDS test $P < 0.01$).

by Illumina HiSeq2000 platform using a configuration of 200 cycles to obtain pair-end reads of 100 base length. Both library preparation and sequencing were performed at Macrogen Inc. (Korea). The sample yielded a total of 58,146,564 reads. Additionally, DNA from hemolymph metagenome was sequenced using the 454 GS-FLX platform yielding 811,305 single reads.

Metagenomic Fungal Ribosomal Gene *In silico* Reconstruction and Characterization

Ribosomal genes from all metagenomic reads were obtained using Parallel-meta 2.4 (Su et al., 2014) algorithm. Eukaryotic ribosomal sequences were recovered using -E option against the SILVA database within an *e*-value of 1×10^{-10} cutoff. Fungal 18S rRNA sequences were retrieved from parsing Parallel-meta result tables. Fungal hits were visualized in Krona graphs (Ondov et al., 2011). 18S rRNA gene sequences were recovered from long reads of the hemolymph metagenome (>200 nt), compared to taxonomically related sequences from NCBI using BLASTn 2.2.30+ (Camacho et al., 2009) and used for maximum likelihood phylogenetic analysis. MODELTEST 3.06 was used to select appropriate models of sequence evolution by the AIC model. Model TrN was the best model (A = 0.25409; C = 0.14918; G = 0.20597; T = 0.39076). The ribosomal sequence retrieved was deposited in the GenBank database under the accession number KT351777.

Gene Annotation and Purine Pathway Reconstruction

To eliminate bacterial sequences, all metagenomic reads were mapped to *Wolbachia w*DacA and *w*DacB genomes previously obtained from *D. coccus* metagenome (Ramírez-Puebla et al., 2015) using Bowtie2 2.2.4 (Langmead and Salzberg, 2012). Un-mapped reads were retrieved by Samtools 1.2 (Li et al., 2009). High-quality shotgun unmapped reads longer than 100 nucleotides were used directly for gene prediction and annotation. Gene prediction was performed using FragGeneScan 1.20 (Rho et al., 2010) with –w 0 –p 16 –t illumina_5 (gut, DCoax and DCperu metagenomes) and –t 454_5 (hemolymph-metagenome) parameters. Metabolic annotation was obtained from all putative coding gene predicted using GhostKoala tool from KEGG (Kanehisa et al., 2015). Fungal annotation was obtained by parsing the annotation result table using KEGGREST Bioconductor library (http://bioconductor.org/packages/release/bioc/html/KEGGREST.html). A metabolic pathway of uric acid catabolism was constructed using KEEG Mapper–Search & color Pathway tool (http://www.genome.jp/kegg/tool/map_pathway2.html) from fungal annotation results. All metagenomics reads from gut and hemolymph metagenomes were deposited in GenBank under SRA accession study SRP074499.

Additionally, to extend our metagenomic results we analyzed the two available *Dactylopius* metagenomes from the whole body (here after called DCoax and DCperu metagenome) deposited in GenBank under BioProject PRJNA244295 (Campana et al., 2015). For this, we performed a fungal ribosomal gene *in silico* reconstruction and the annotation of fungal reads related to uric acid catabolism as was described above.

Phylogenetic Analysis

Nucleotide sequences were compared against non-redundant GeneBank library by BLASTn 2.2.30+ (Camacho et al., 2009) and taxonomically related sequences were collected from NCBI. Cultured fungi were identified by ITS and 26S rRNA phylogenies obtained by Maximum likelihood. MODELTEST 3.06 was used to select appropriate models of sequence evolution by the AIC model (Posada, 2008). GTR+I+G ($\alpha = 1.772$ for gamma distribution; $A = 0.25778$; $C = 0.23041$; $G = 0.22501$; $T = 0.28681$) was the best model for the ITS gene, while GTR + I ($\alpha = 0.383$ for gamma distribution; $A = 0.25061$; $C = 0.20735$; $G = 0.29982$; $T = 0.24222$) was the best model for 26S rRNA gene. A p-distance among sequences was calculated using DNAdist algorithm from Phylip 3.6 software (Felsenstein, 1989). Limits for genus and species were established at 95 and 97%, respectively. To compare the sequences and quantify the number of fungi operational taxonomical units (OTUs) related with *Dactylopius* spp., a cluster analysis was performed using MOTHUR (Schloss et al., 2009) and ribosomal sequences were clustered at 0.03%

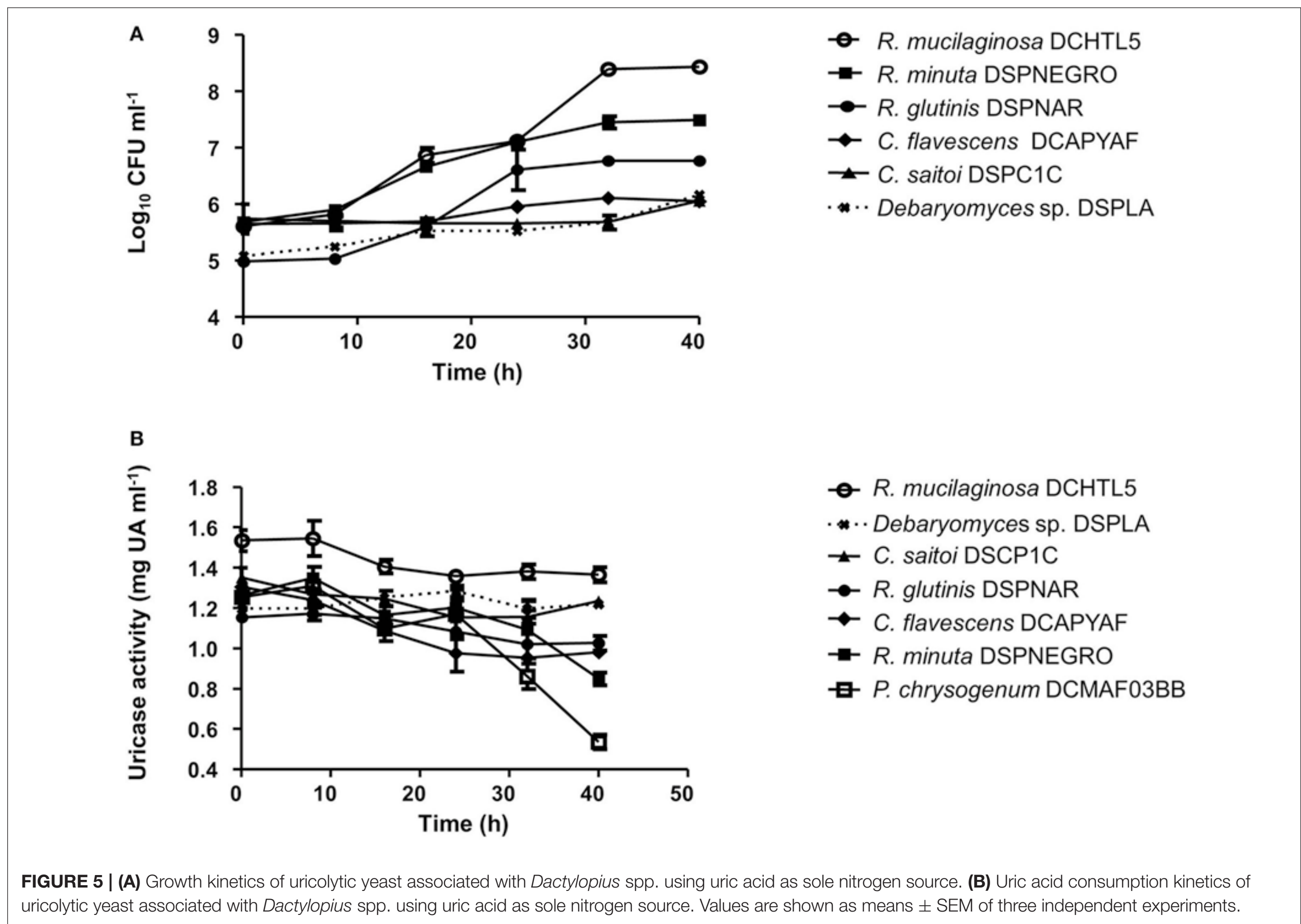


FIGURE 5 | (A) Growth kinetics of uricolytic yeast associated with *Dactylopius* spp. using uric acid as sole nitrogen source. **(B)** Uric acid consumption kinetics of uricolytic yeast associated with *Dactylopius* spp. using uric acid as sole nitrogen source. Values are shown as means ± SEM of three independent experiments.

distance. All sequences generated from ITS and 26S rRNA of cultured fungi were deposited in the GenBank database under the accessions numbers KM393247 to KM393282 and KT351741 to KT351776, respectively.

Determination of Uric Acid and Uricase Activity in *Dactylopius* spp.

Three guts from *D. coccus* and *D. opuntiae* in 1st instar nymph, 2nd instar nymph and adults, as well as eggs from both species, were dissected as mentioned above. Additionally male bodies were resuspended in 200 µl AmplexRed buffer solution. Also, 10 µl of honeydew from *D. coccus* and *D. opuntiae* were resuspended in 100 µl of the same buffer solution. UA and uricase activity were determined using the Amplex® Red Uric Acid/Uricase Assay Kit (Life Technologies Eugene, OR) following the manufacturer's instructions. Means of the UA content as well as uricase activity were compared using two-way ANOVA, and a Tukey-HSD *post-hoc* test was applied for pairwise comparisons between insects. Furthermore, to compare differences in UA content between honeydew and adult female guts a *t*-test was performed. All statistics test were performed using R version 3.1.

Fungal Uricolytic Activity

Individual guts and Malpighian tubules, from adults of *D. opuntiae* and *D. coccus* were placed separately in microtubes and macerated with sterile pestles in 200 µl of sterile PBS. Serial 10-fold dilutions from 10^{-1} to 10^{-3} were spread on duplicate plates of MU media (K_2HPO_4 2.5 g l^{-1}; KH_2PO_4 5 g l^{-1}; $MgSO_4$•7 H_2O 0.2 g l^{-1}; $MnSO_4$ 0.02 g l^{-1}; $CaCl_2$ 0.05 g l^{-1}; $FeSO_4$ 0.05 g l^{-1}; uric acid (Sigma) 1.5 g l^{-1}; glucose 10 g l^{-1} and agar 15 g l^{-1}). Plates were incubated at 28°C in CO_2 atmosphere generated by BD GasPak EZ Pouch Systems™ for 7 days. Colonies with yeast-like macro and microscopic morphology surrounded with a clear halo (suggestive of uric acid utilization) were counted and colony forming units (CFU) per gut were obtained. All isolates were stored at −70°C. Additionally, uricolytic activity of 37 isolated fungi from *Dactylopius* spp. was tested measuring a degradation halo in YPU (Yeast extract 10 g l^{-1}; Peptone 10 g l^{-1}, UA 7 g l^{-1}) medium. Enzyme activity was determined as described by Morales-Jiménez et al. (2013). To find out if UA was used by fungi isolates as sole nitrogen source, growth and UA consumption kinetics were performed. Microbial growth was measured quantifying the CFU ml^{-1} for yeast and by weighing the final biomass for molds grown in liquid MU media. UA consumption was quantified by measuring the decrease in

TABLE 3 | Uric acid consumed as sole nitrogen source by fungi isolated from *Dactylopius*.

Isolate	Uric acid consumed (μg ml^{-1})	Sperman correlation *R*-value	*P*-value
Rhodotorula glutinis DSPNAR	127.6 ± 42.54	−0.922	0.0045
Cryptococcus saitoi DSPC1C	119.9 ± 62.16	−0.725	0.0515
Rhodotorula minuta DSPNEGRO	414.8 ± 66.43	−0.897	0.0128
Rhodotorula mucilaginosa DCHTL5	170.5 ± 89.54	−0.867	0.0127
Cryptococcus flavescens DCAPYAF	323.5 ± 37.34	−0.925	0.0041
Debaryomyces sp. DSPA	0.0 ± 0.0	0.221	0.3372
Penicillium sp. DCM03BB	717.9 ± 27.05	−0.892	0.0085

absorbance at 295 nm. These results were compared against a standard curve of UA. A Sperman correlation was performed to assess a negative correlation and differences in UA consumption in relation to time.

Antifungal Treatment

A group of 15 first instar nymphs of *D. opuntiae* was fed on a prickly pear pad of *Opuntia ficus-indica* injected with 5 ml of 20 μg ml^{-1} antifungal cocktail of Ketoconazol (Sigma), Anfotericine B (Sigma), and Fludioxonil (Sigma). Fleshy leaves were injected weekly for 4 weeks and then female insects were removed. *O. ficus-indica* leaves without antifungal were similarly infested and used as negative controls. After treatment, a pool of six individuals of each leaf was used to measure differences in dry weight, UA content and uricase activity. Five replicates of this experiment were performed. UA content, uricase activity and dry weight data were compared between controls and treatments using a *t*-test.

Fluorescent *In situ* Hybridization (FISH)

FISH was performed as previously described by Koga et al. (2009) with slight modifications. Ninety-day old *D. coccus* and *D. opuntiae* were collected. Malpighian tubules, as well as ovaries and embryos (25 from *D. coccus* and 20 from *D. opuntiae*) were dissected as described above. These organs were embedded in 3% agarose and treated as described by Rosas-Pérez et al. (2014). The oligonucleotide probe used was Cy5-Cry851 (5′-TGATGCGA GTTTCTGCTATC-3′), which targets 26S rRNA of *Cryptococcus saitoi* (designed for this work). After washing with PBS the samples were stained with 2.4 μg ml^{-1} of DAPI and mounted with citifluor antifade solution. To confirm probe specificity, control experiments were performed with no probe and RNAse digestion. The samples were observed under an Olympus FV100 Multi-photonic confocal microscopy. Images were processed using FIJI 2.0.0 software (Schindelin et al., 2012).

RESULTS

Culture-Dependent and Culture-Independent Analyses of Fungal Communities

A total of 37 fungal isolates were cultured from *D. coccus*, *D. opuntiae,* and *D. confusus*. Isolates were obtained from guts, whole bodies and ovary samples (**Table 2**). Nucleotide sequences of 26S rRNA genes and ITS regions from different morphotypes corresponded to 14 OTUs. 26S rRNA and ITS phylogenetic analyses showed sequences belonging to Ascomycota and Basidiomycota with *Rhodotorula*, *Cryptococcus* and *Penicillium* as the most frequent genera (**Figure 1**; Supplementary Figure 1). Fungal species like *Rhodotorula mucilaginosa* and *Cryptococcus saitoi* were present in the three *Dactylopius* species sampled, whereas *Trametes polizona* was present in *D. coccus* and *D. opuntiae* (**Table 2**). Three filamentous fungi had an ITS sequence identity of 77.6 and 88.9% to *Stereum* sp. and *Periconia* sp. (DCHG and DCHM) respectively (**Figure 1**; **Table 2**). In 26S rRNA phylogenies, the closest related sequences of these novel fungi were *Phlebiopsis flavidoalba* (DCHBPI) with 97.8% identity and *Periconia macrospinosa* (DCHG and DCHM) with 93.74% identity (**Table 2**, Supplementary Figure 1). Likewise, from *D. coccus* we could isolate the mold *Penicillium* from 1st instar nymphs ($n = 3$) and males ($n = 5$) but not from adult females (**Figure 1**; **Table 2**; Supplementary Figure 1).

From the metagenomic data of the hemolymph and gut metagenomes, fungal 18S rRNA gene sequences were detected. Hemolymph metagenome sequences were assigned particularly to *Sebacina vermifera*, *Bullera ninhbinhensis* (Basidiomycetes), and *Candida lignicola* (Ascomycetes; **Figure 2**; Supplementary Data Sheet 1). In congruence, a phylogenetic reconstruction of 18S rRNA (∼200 nt) from this sample showed the presence of *Pichia anomala* (100% identity) in *Dactylopius* hemolymph (Supplementary Figure 2). In gut metagenome, we found sequences related to Basidiomycota, particularly to the Sebacinaceae family (*Craterocolla* sp. and *Sebacina* sp.) and Ustilaginaceae family (*Rhodosporidium* sp.), as well as sequences related to Chytridiomycota and Glomeromycota phyla (**Figure 2**; Supplementary Data Sheet 1). Remarkably, most of the fungal sequences obtained by the metagenomic analysis were associated with uncultured and unclassified fungi (**Figure 2**; Supplementary Data Sheet 1). Analysis of DCoax metagenome showed sequences related to Basidiomycota (*Agaricus bisporus* and *Thanatephorus cucumeris*), Ascomycota (*Blastobostrys adeninivorans* and *Candida* sp.), Glomeromycota and some unclassified fungi (**Figure 2**; Supplementary Data Sheet 1). From DCperu metagenome the only fungal species detected was *Candida* sp.

Metagenomic Annotation of Fungal Genes Involved in Uric Acid Catabolism

A total of 518,258 open reading frames (ORFs) were predicted from the hemolymph metagenome and 20,136,058 ORFs from the gut metagenome. From those, only 2,874 and 66,502 corresponded to fungal ORFs, respectively. Metabolic annotation

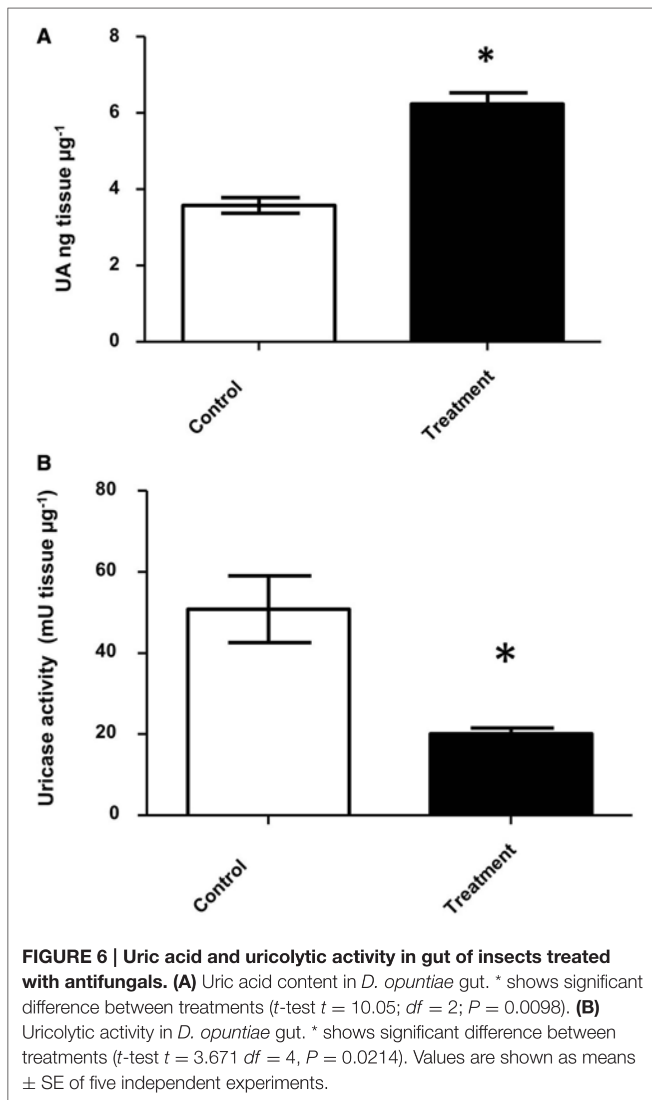


FIGURE 6 | Uric acid and uricolytic activity in gut of insects treated with antifungals. (A) Uric acid content in *D. opuntiae* gut. * shows significant difference between treatments (t-test $t = 10.05$; $df = 2$; $P = 0.0098$). **(B)** Uricolytic activity in *D. opuntiae* gut. * shows significant difference between treatments (t-test $t = 3.671$ $df = 4$, $P = 0.0214$). Values are shown as means ± SE of five independent experiments.

of these fungal ORFs revealed genes related to UA metabolism (**Figure 3**). Particularly, we detected the presence of 20 and 85 fungal genes involved in UA catabolism from hemolymph and gut metagenome, respectively (Supplementary Table 1). All coding genes for xanthine degradation to urea were present in gut metagenome whereas in hemolymph metagenome we did not find any allantoinase fungal genes (**Figure 3**). From DCoax metagenome a total of 8,911,722 ORFs were estimated and 8,901,672 were properly annotated by Ghost-KOALA, from which 262,623 corresponded to fungal sequences. We found 128 putative genes involved in uric acid catabolism in this metagenome (Supplementary Table 2). From the DCperu metagenome, 8,619,769 ORFs were predicted; 8,611,041 had a functional annotation and 226,810 belonged to fungal sequences. A total of 101 putative genes of uric acid catabolism were present in this sample (Supplementary Table 3). As in gut metagenome, all genes for xanthine catabolism to urea were found in DCoax and DCperu metagenomes (Supplementary Figure 3).

UA and Uricase Activity in *Dactylopius* spp. Guts

UA and uricase activities were detected in *D. opuntiae* and *D. coccus* extracts where the changes in UA concentration depended on the insect developmental stage (**Figure 4A**). The highest amount of UA was present in eggs of both species (21.87 ± 2.91 and 34.49 ± 3.11 ng μg^{-1} tissue, respectively; Supplementary Table 4) whereas the lowest was in *D. coccus* adult male, *D. coccus* female and in *D. opuntiae* 2nd instar nymph (4.49 ± 0.38; 4.61 ± 0.91 and 2.91 ± 0.32 ng μg^{-1} tissue respectively; Supplementary Table 4).

Post-hoc comparison using Tukey-HSD test showed significant differences in UA content among eggs, 1st instar nymph, and adults in both species, although no significant difference was seen between 2nd nymph instar and adult (**Figure 4A**).

Urate oxidase or uricase (EC 1.7.3.3 or UOX) is a homotetramer that catalyzes the conversion of UA and molecular oxygen to 5-hydroxyurate and hydrogen peroxide (Gabison et al., 2008). In our results, this enzyme showed high activity in adult females of both *Dactylopius* species (80 mU μg^{-1} tissue for *D. coccus* and 135 mU μg^{-1} tissue for *D. opuntiae*; **Figure 4B**; Supplementary Table 4). *Post-hoc* test showed significant differences in uricolytic activity in all stages (**Figure 4B**). The content of uric acid in adult's honeydew in both scale species was low, 0.18 ± 0.05 and 0.58 ± 0.05 ng μl^{-1} in *D. coccus* and *D. opuntiae,* respectively. A t-test showed a significant difference between UA content in honey dew and adults gut (*D. coccus* $P = 0.0006$; $t = 4.856$; $df = 8$; *D. opuntiae* $P < 0.0001$; $t = 26.85$; $df = 8$), moreover no urate oxidase activity was detected in these samples. This supports the idea that UA is metabolized inside the insect.

Uricolytic Fungi Associated with *Dactylopius*

The number of uricolytic yeast CFUs in MU from *D. opuntiae* gut was estimated in $4.1 \times 10^2 \pm 0.74 \times 10^2$ CFU gut^{-1}. The isolates *C. flavescens* DCPYAF01, *R. mucilaginosa* DCHTL5, *R. minuta* DSPNEGRO, *R. glutinis* DSPNAR, *C. saitoi* DSPCUB, and the mold *Penicillium* sp. DCFM03BB (**Figure 1**; **Table 2**), were capable of growth and consumption of UA as sole nitrogen source (**Figures 5A,B**; **Table 3**). The maximum consumption rates were with *Penicillium* sp. DCMAF03BB and *R. minuta* DSPNEGRO (717.9 ± 27.05 and 414.8 ± 66.43 µg of UA respectively; **Table 3**). *Debaryomyces* sp. DSPA showed no significant growth and there was no evidence for UA uptake by this strain (**Figures 5A,B**; **Table 3**).

Antifungal Effects on UA Concentration and Uricase Activity in *D. opuntiae*

After four weeks with antifungal treatment *D. opuntiae* weight was significantly lower in comparison to the controls (2.50 ± 0.15 and 0.58 ± 0.12 mg respectively; $t = 6.954$; $df = 4$; $P = 0.0201$; Supplementary Figure 4). Uric acid concentration was significantly higher in fungicide treated insects vs. controls (6.25 ± 0.28 and 3.58 ± 0.21 UA ng μg^{-1} tissue [1] respectively;

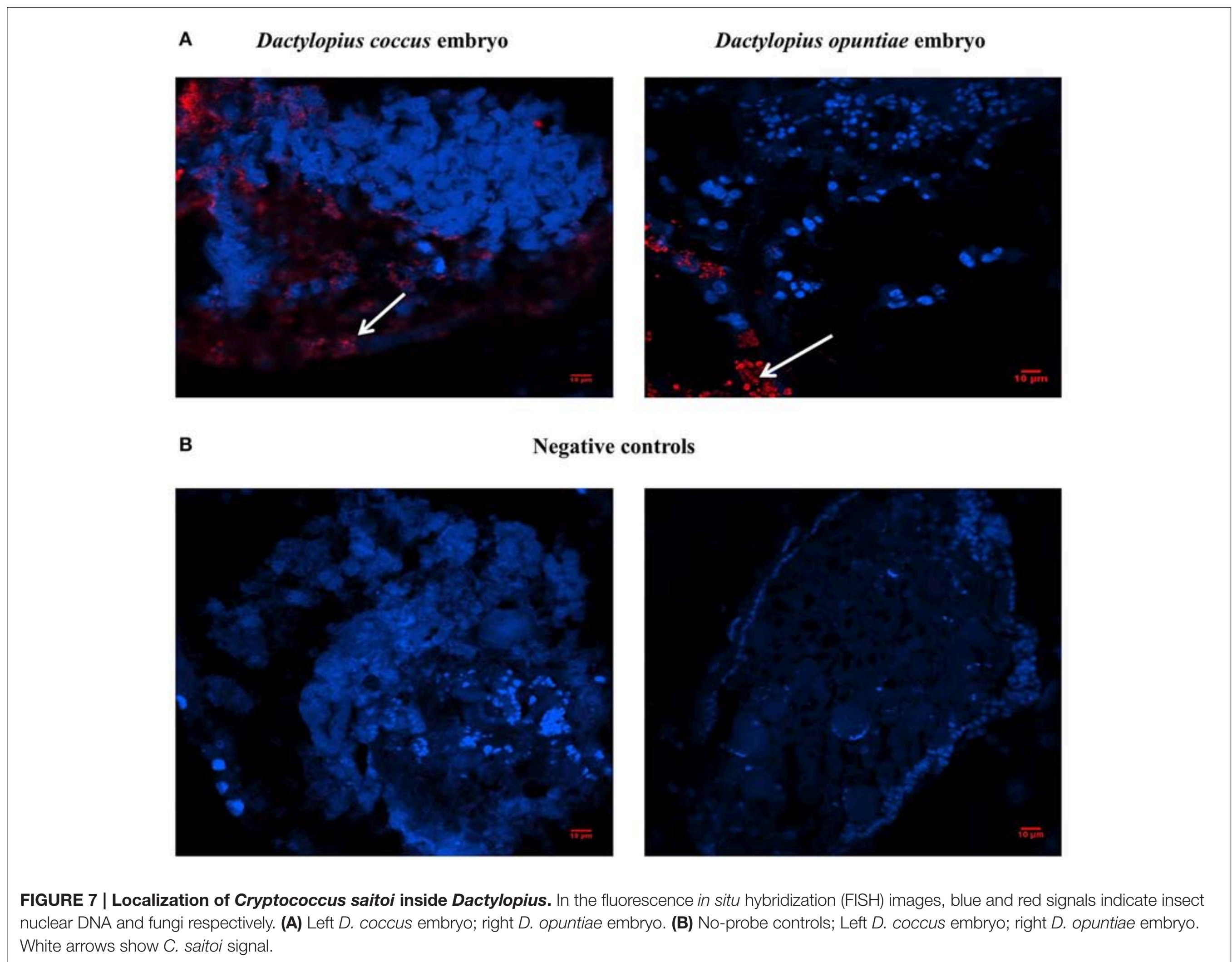


FIGURE 7 | Localization of *Cryptococcus saitoi* inside *Dactylopius*. In the fluorescence *in situ* hybridization (FISH) images, blue and red signals indicate insect nuclear DNA and fungi respectively. **(A)** Left *D. coccus* embryo; right *D. opuntiae* embryo. **(B)** No-probe controls; Left *D. coccus* embryo; right *D. opuntiae* embryo. White arrows show *C. saitoi* signal.

Figure 6A). Additionally, uricase activity was significant lower in antifungal treatments than in controls (20.20 $\pm$ 1.35 and 50.91 $\pm$ 8.26 mU tissue μg^{-1}, respectively; **Figure 6B**).

Cryptococcus saitoi Localization in *Dactylopius*

Fluorescent *in situ* hybridization of *D. coccus* and *D. opuntiae* showed the presence of *C. saitoi* in embryos of both species (**Figures 7A,B**). Of 25 embryos of *D. coccus* and 20 of *D. opuntiae,* 17 (68%) and 14 (70%) contained the fluorescent signal. FISH analysis showed that *C. saitoi* fungi were on the egg surface. Additionally, *C. saitoi* was observed by FISH in a distal part of the Malpighian tubules in *D. coccus* (Supplementary Figure 5).

DISCUSSION

A comprehensive study of the fungal community associated with *Dactylopius* is presented here, where different species in four fungal phyla were found by culture and culture-independent analyses. *C. saitoi* and *R. mucilaginosa* were found in most female samples (**Figure 2**) while *Penicillum* was the only fungus found in males (**Figure 1**; **Table 2**; Supplementary Figure 1). *Penicillium* has been associated with other insects such as bees, beetles, termites, and as well as in *Triatoma* sp. guts (Batra et al., 1973; Lage-Moraes et al., 2001; Pérez et al., 2003). The cultured fungi obtained belonging to *Rhodotorula*, *Cryptococcus*, *Trametes*, *Penicillium*, and *Debaryomyces* (**Figure 1**; Supplementary Figure 1) were previously found in other phytophagous insects (Jones et al., 1999; Guevara et al., 2000; Suh et al., 2001; Ganter, 2006; Kobayashi et al., 2007). Particularly in the scale insect *Saissetia oleae*, *Cryptococcus,* and *Rhodotorula* yeasts were isolated from the gut and reproductive organs (Zacchi and Vaughan-Martini, 2003). Similarly, in the reproductive tissues and guts from *D. coccus* and *D. opuntiae,* we found *Cryptococcus* and *Rhodotorula* by a culture dependent approach and by FISH (**Figure 7**; **Table 2**; Supplementary Figure 5). In culture we also found *P. flavidoalba* (DCHBPI), *Periconia macrospinosa* (DCHG and DCHM) and *Irpex lacteus* (DCALI) which, to our knowledge, have not been previously isolated from insect's inner

tissues. In this work ITS and 26S rDNA markers were used for culture-fungi identification and in few cases genus or species assignment differed depending on the marker used (**Table 2**), indicating that single gene phylogenetic stories are not fully reliable and a better sample of the genome is needed in novel groups.

Most of the fungal ribosomal sequences from the female metagenomic analyses belonged to uncultured or non-classified fungi. With ∼100–300 base pair reads an accurate classification may be difficult. Additionally, fungal sequences are underrepresented in metagenomics because of limited information in databases used for the analysis and problems in fungal DNA extraction from different samples (Lindahl and Kuske, 2013; Escobar-Zepeda et al., 2015). However, members of Chytridiomycota and Glomeromycota phyla (**Figure 2**; Supplementary Data Sheet 1) were recovered form *D. coccus* metagenomes. There are reports of entomopathogenic Chytridiomycota associated with elm bark beetles, blackflies, and aquatic dipteran larvae (Humber et al., 1990; Powell, 1993), but not in scale insects. Glomeromycota is a phylum of asexual fungi from arbuscular mycorrhiza of plants, they are obligate endosymbionts and cannot be grown in pure culture in the absence of their plant host (Hempel et al., 2007; Gianinazzi-Pearson and Van Tuinen, 2012). Interestingly, there are no reports of this fungal phylum associated with insects, although some sequences related to mycorrhizal fungi have been found in other habitats like the human oral cavity (Ghannoum et al., 2010; Cui et al., 2013). In *Dactylopius* we found sequences of Glomeromycota in gut and whole body (Supplementary Data Sheet 1). It is tempting to speculate that its presence could mediate a close interaction between insects and their host plant. This is the first report of Glomeromycota in insects.

Sequences of *Candida,* which we did not recover in cultures (**Figure 1**; **Table 2**; Supplementary Figure 1), were found in all female *Dactylopius* metagenomes (Supplementary Data Sheet 1; Supplementary Figure 2). Species of *Candida* have been isolated from insect guts as well as in mycetocytes of other hemipterans (Gibson and Hunter, 2005; Vega and Blackwell, 2005; Suh et al., 2008; Hughes et al., 2011).

Additionally, we report here the presence of uricolityc fungi associated with *Dactylopius* spp. Nitrogen content in *O. ficus-indica* cladodes is around 0.5–1% of wet weight (Stintzing and Carle, 2005). Meanwhile in *Dactylopius* this element constitutes about 32% of wet weight (Gómez-Hernández, 2006). This means that *Dactylopius* has to accumulate 30 times the nitrogen present in the cactus. It is known that N_2 recycling by UA catabolism provides nitrogen to plant feeding insects (Potrikus and Breznak, 1981; Sasaki et al., 1996; Morales-Jiménez et al., 2013; Patiño-Navarrete et al., 2014). However, bacteria are often mentioned as major recyclers in these scenarios and only in the brown plant hopper (*Nilaparvata luggens*) it has been shown that many unicellular fungi symbionts called yeast-like symbionts (YLS) are involved in insect UA metabolism (Sasaki et al., 1996). Plant hoppers produce and store UA when fed nitrogen-rich diets, but when nitrogen is limited their YLS mobilize the stored UA using the enzyme uricase (EC:1.7.3.3). This process may turn UA into amino acids for insects. Yeast isolates from *D. coccus* and *D. opuntiae* females as well as the mold *Penicillium* from *D. coccus* males were capable of metabolizing UA as sole nitrogen source (**Figures 5A,B**; Supplementary Table 4) There are reports for UA catabolism in *Cryptococcus* and *Penicillium* (Allam and Elzainy, 1969; Lee et al., 2013) but to our knowledge there are no reports for uricolytic *Rhodotorula* (Middelhoven et al., 1985). In termites (*Reticulotermes flavipes*) and in bark beetles (*Dendroctonus valens* and *Dendroctonus rhizophagus*) uricolytic microorganisms have been isolated from their guts (Potrikus and Breznak, 1980; Morales-Jiménez et al., 2013), in agreement most of the *Dactylopius* uricolytic fungi come from the alimentary canal (**Figures 5A,B**; **Table 2**). FISH analysis showed the presence of *Cryptoccocus* (uricolytic yeast) in Malpighian tubules of *D. coccus* (Supplementary Figure 5). Additionally, metagenomic analysis of guts and hemolymph of *D. coccus* and whole body of other *D. coccus* revealed the presence of fungal genes involved in UA catabolism (**Figure 3**; Supplementary Figure 3; Supplementary Tables 2, 3). Uricase catalyzes the first step in UA catabolism (Gabison et al., 2008). Even though putative genes for uricase were present in all metagenomes analyzed, there was only one ORF codifying for this enzyme in hemolymph metagenome; meanwhile in the gut metagenome 18 of these genes were found (Supplementary Table 2). This supports the idea that UA could be metabolized by fungi in *Dactylopius* gut, as in other insects, rather than directly in hemolymph. Besides, putative fungal genes for allantoinase, allantoicases, and ureases were also found. This suggests that UA can be catabolized to urea and ammonia by fungi (**Figure 3**; Supplementary Figure 3). It is known that in silkworm *Bombix mori* and in the larvae of the bruchid beetle *Caryedes brasiliensis* urea can be incorporated into insect proteins as an alternative nitrogen source (Hirayama et al., 1999). In *Dactylopius* uric acid could be metabolized into urea by their associated fungi and then used as nitrogen by its insect host.

Different levels of UA during life stages have been detected in other Hemiptera. Particularly in *Parastrachia japonensis*, UA is higher before copulation and during ovarian development and lower in nymph stages (Kashima et al., 2006). In contrast, in *Dactylopius* we found that UA is higher in nymphs as compared to adults (**Figure 4A**; Supplementary Table 4). Uricase activity was detected in *Dactylopius* guts in all life stages, in contrast this enzyme is absent in the majority of insects (Pant, 1988). However, some insect symbionts present uricase activity (Potrikus and Breznak, 1981; Hongoh and Ishikawa, 2000). In the shield bug *P. japonensis* treatment with antibiotics produce a reduction in uricolytic activity and in amino acid concentration in hemolymph (Kashima et al., 2006). In *Dactylopius*, antifungal treatment showed a similar significant decrease of uricase activity (**Figure 6B**), additionally UA concentration was higher in those insects treated (**Figure 6A**). As mentioned, the metagenomic approach revealed fungal uricase genes (**Figure 3**; Supplementary Figure 3; Supplementary Tables 1–3), that in addition to the experimental evidence of UA accumulation and lower uricolytic activity in antifungal treated insects (**Figures 6A,B**), suggest that the uricase detected in the enzymatic assay on *Dactylopius* (**Figure 4B**; **Table 3**) may come from their

associated fungi. In conclusion fungi associated to *Dactylopius* could recycle nitrogen in order to supply deficiencies in their diet.

AUTHOR CONTRIBUTIONS

The experiments were conceived and designed by AV, AS, MR, and EM, and were conducted and analyzed by AV and AS. All authors contributed to interpreting the results and writing the article.

FUNDING

This work was supported by Consejo Nacional de Ciencia y Tecnologia grant 154453 and graduate student (AVPL) grant 331625.

ACKNOWLEDGMENTS

To CONACyT Basic Science grant 154453. AVPL is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received the 331625 fellowship from Consejo Nacional de Ciencia y Tecnología (CONACyT). To Michael Dunn, Angela E. Douglas and Meztlli Gaytán for discussions and valuable comments on the manuscript. To J. Martínez-Romero for technical help. To Héctor González-Hernández for his help to identify the wild *Dactylopius* specimens. To Jesús A. Chávez-Vargas and Campo Carmín Company for *D. coccus* provided. To Alejandra Escobar-Zepeda, Jérôme Verleyen, Leticia Vega, Karel Estrada, and Veronica Jiménez-Jacinto for their help in bioinformatics support. All bioinformatics were performed in USMB-UNAM servers. We thank the "Unidad de Secuenciación Masiva y Bioinformática" of the "Laboratorio Nacional de Apoyo Tecnológico a las Ciencias Genómicas," CONACyT #260481, at the Instituto de Biotecnología/UNAM.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.00954

REFERENCES

Allam, A., and Elzainy, T. (1969). Degradation of xanthine by *Penicillium chrysogenum*. *J. Gen. Microbiol.* 56, 293–300.

Batra, L. R., Batra, S. W. T., and Bohart, G. E. (1973). The mycoflora of domesticated and wild bees (Apoidea). *Mycopathol. Mycol. Appl.* 49, 13–44.

Ben-Dov, Y., and Marotta, S. (2001). Taxonomy and family placement of *Coccus bassi* Targioni Tozzetti, 1867 (Hemiptera: Coccoidea). *Phytoparasitica* 29, 169–170. doi: 10.1007/BF02983961

Brand, J. M., Bracke, J., Briton, L., Markovetz, A., and Barras, S. (1976). Bark beetle pheromones: production of verbenone by a mycangial fungus of *Dendroctonus frontalis*. *J. Chem. Ecol.* 2, 195–199.

Buchner, P. (1965). *Endosymbiosis of Animals with Plant Microorganisms*. New York, NY: John Wiley & Sons, Inc.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Campana, M. G., Robles García, N. M., and Tuross, N. (2015). America's red gold: multiple lineages of cultivated cochineal in Mexico. *Ecol. Evol.* 5, 607–617. doi: 10.1002/ece3.1398

Chávez-Moreno, C. K., Tecante, A., and Casas, A. (2009). The *Opuntia* (Cactaceae) and *Dactylopius* (Hemiptera: Dactylopiidae) in Mexico: a historical perspective of use, interaction and distribution. *Biodivers. Conserv.* 18, 3337–3355. doi: 10.1007/s10531-009-9647-x

Chen, C., Cheng, L., and Hou, R. (1981). Studies on the intracellular yeast—like symbiote in the brown planthopper, *Nilaparvata lugens* Stal. *Zeitschrift Angew* 92, 440–449.

Cui, L., Morris, A., and Ghedin, E. (2013). The human mycobiome in health and disease. *Genome Med.* 5, 63. doi: 10.1186/gm467

da Silva Santos, A. C., Oliveira, R. L. S., da Costa, A. F., Tiago, P. V., and de Oliveira, N. T. (2015). Controlling *Dactylopius opuntiae* with *Fusarium incarnatum–equiseti* species complex and extracts of *Ricinus communis* and *Poincianella pyramidalis*. *J. Pest Sci.* 89, 539–547. doi: 10.1007/s10340-015-0689-4

Deveoglu, O., Karadag, R., and Yurdun, T. (2011). Qualitative HPLC determination of main anthraquinone and lake pigment contents from *Dactylopius coccus* dye insect. *Chem. Nat. Comp.* 47, 103–104. doi: 10.1007/s10600-011-9842-3

Douglas, A. E. (2009). The microbial dimension in insect nutritional ecology. *Funct. Ecol.* 23, 38–47. doi: 10.1111/j.1365-2435.2008.01442.x

Ebbert, M., A, Marlowe, J. L., and Burkholder, J. J. (2003). Protozoan and intracellular fungal gut endosymbionts in *Drosophila*: prevalence and fitness effects of single and dual infections. *J. Invertebr. Pathol.* 83, 37–45. doi: 10.1016/S0022-2011(03)00033-8

Escobar-Zepeda, A., Vera-Ponce de León, A., and Sanchez-Flores, A. (2015). The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. *Front. Genet.* 6:348. doi: 10.3389/fgene.2015.00348

Felsenstein, J. (1989). Phylip: phylogeny inference package (version 3.2). *Cladistics* 5, 164–166.

Gabison, L., Prangé, T., Colloc'h, N., El Hajji, M., Castro, B., and Chiadmi, M. (2008). Structural analysis of urate oxidase in complex with its natural substrate inhibited by cyanide: mechanistic implications. *BMC Struct. Biol.* 8:32. doi: 10.1186/1472-6807-8-32

Ganter, P. (2006). "Yeast and invertebrate associations," in *Biodiversity and ecophysiology of yeasts The Yeast Handbook*, eds G. Péter and C. Rosa (Berlin/Heidelberg: Springer-Verlag), 303–370.

Ghannoum, M. A, Jurevic, R. J., Mukherjee, P. K., Cui, F., Sikaroodi, M., Naqvi, A., et al. (2010). Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog.* 6:e1000713. doi: 10.1371/journal.ppat.1000713

Gianinazzi-Pearson, V., and Van Tuinen, D. (2012). "Exploring the genome of glomeromycotan fungi," in *Fungal Associations*, ed B. Hock (Heidelberg: Springer-Verlag), 1–21.

Gibson, C. M., and Hunter, M. S. (2005). Reconsideration of the role of yeasts associated with Chrysoperla green lacewings. *Biol. Control* 32, 57–64. doi: 10.1016/j.biocontrol.2004.06.006

Gibson, C. M., and Hunter, M. S. (2009). Negative fitness consequences and transmission dynamics of a heritable fungal symbiont of a parasitic wasp. *Appl. Environ. Microbiol.* 75, 3115–3119. doi: 10.1128/AEM.00361-09

Gibson, C. M., and Hunter, M. S. (2010). Extraordinarily widespread and fantastically complex: comparative biology of endosymbiotic bacterial and fungal mutualists of insects. *Ecol. Lett.* 13, 223–234. doi: 10.1111/j.1461-0248.2009.01416.x

Gómez-Hernández, B. (2006). *Extracción y Acetilación de los Componentes de la Grana Cochinilla* (*Dactylopius coccus* Costa), Master dissertation, Universidad Autonoma del Estado de Hídalgo, Mexico. Available online at: http://repository.uaeh.edu.mx/bitstream/bitstream/handle/123456789/11155/Extracci%c3%b3n%20y%20acetilaci%c3%b3n%20de%20los%20componentes%20de%20la%20grana%20cochinilla%20%28Dactylopius%20coccus%20COSTA29.pdf?sequence=1&isAllowed=y

Guevara, R., Rayner, A. D. M., and Reynolds, S. E. (2000). Effects of fungivory by two specialist ciid beetles (*Octotemnus glabriculus* and *Cis boleti*) on the reproductive fitness of their host fungus, Coriolus versicolor. *New Phytol.* 145, 137–144. doi: 10.1046/j.1469-8137.2000.00552.x

Hempel, S., Renker, C., and Buscot, F. (2007). Differences in the species composition of arbuscular mycorrhizal fungi in spore, root and soil communities in a grassland ecosystem. *Environ. Microbiol.* 9, 1930–1938. doi: 10.1111/j.1462-2920.2007.01309.x

Hirayama, C., Sugimura, M., and Shinbo, H. (1999). Recycling of urea associated with the host plant urease in the silkworm larvae, *Bombyx mori. J. Insect Physiol.* 45, 15–20.

Hoffman, C. S., and Winston, F. (1987). A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli. Gene* 57, 267–272.

Hongoh, Y., and Ishikawa, H. (2000). Evolutionary studies on uricases of fungal endosymbionts of aphids and planthoppers. *J. Mol. Evol.* 51, 265–277. doi: 10.1007/s002390010088

Hughes, G. L., Allsopp, P. G., Webb, R. I., Yamada, R., Iturbe-Ormaetxe, I., Brumbley, S. M., et al. (2011). Identification of yeast associated with the planthopper, *Perkinsiella saccharicida*: potential applications for Fiji leaf gall control. *Curr. Microbiol.* 63, 392–401. doi: 10.1007/s00284-011-9990-5

Humber, R. A., Samson, R. A., Evans, H. C., and Latge, J.-P. (1990). Atlas of entomopathogenic fungi. *Mycologia* 82, 148.

Jones, K. G., Dowd, P. F., and Blackwell, M. (1999). Polyphyletic origins of yeast-like endocytobionts from anobiid and cerambycid beetles. *Mycol. Res.* 103, 542–546.

Kanehisa, M., Sato, Y., and Morishima, K. (2015). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 4, 726–731. doi: 10.1016/j.jmb.2015.11.006

Kashima, T., Nakamura, T., and Tojo, S. (2006). Uric acid recycling in the shield bug, *Parastrachia japonensis* (Hemiptera: Parastrachiidae), during diapause. *J. Insect Physiol.* 52, 816–825. doi: 10.1016/j.jinsphys.2006.05.003

Klepzig, K., and Six, D. (2004). Bark beetle-fungal symbiosis: context dependency in complex associations. *Symbiosis* 37, 189–205. Available online at: http://www.srs.fs.usda.gov/pubs/ja/ja_klepzig003.pdf

Kobayashi, C., Fukasawa, Y., Hirose, D., and Kato, M. (2007). Contribution of symbiotic mycangial fungi to larval nutrition of a leaf-rolling weevil. *Evol. Ecol.* 22, 711–722. doi: 10.1007/s10682-007-9196-2

Koga, R., Tsuchida, T., and Fukatsu, T. (2009). Quenching autofluorescence of insect tissues for *in situ* detection of endosymbionts. *Appl. Entomol. Zool.* 44, 281–291. doi: 10.1303/aez.2009.281

Lage-Moraes, A. M., Reis-de-Figueiredo, A., Vieira-Junqueira, A. C., Lara-da-Costa, G., Aguiar, R. K., and Cunha-de-Oliveira, P. (2001). Fungal flora of the digestive tract of *Panstrongylus megistus* (Reduviidae) used for experimental xenodiagnosis of Trypanosoma (Schizotripanum) cruzi Chagas, 1909. *Rev. Iberoam. Micol.* 18, 79–82.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lee, I. R., Yang, L., Sebetso, G., Allen, R., Doan, T. H. N., Blundell, R., et al. (2013). Characterization of the complete uric acid degradation pathway in the fungal pathogen *Cryptococcus neoformans. PLoS ONE* 8:e64292. doi: 10.1371/journal.pone.0064292

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Lindahl, B. D., and Kuske, C. R. (2013). "Metagenomics for study of fungal ecology," in *The Ecological Genomics of Fungi,* ed F. Martin (Hoboken, NJ: John Wiley & Sons, Inc.), 279–303.

Middelhoven, W. J., de Kievit, H., and Biesbroek, A. L. (1985). Yeast species utilizing uric acid, adenine, n-alkylamines or diamines as sole source of carbon and energy. *Antonie Van Leeuwenhoek* 51, 289–301.

Morales-Jiménez, J., Vera-Ponce de León, A., 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. doi: 10.1007/s00248-013-0206-3

Mullins, D. E., and Cochran, D. G. (1975). Nitrogen metabolism in the american cockroach-I. An examination of positive nitrogen balance with respect to uric acid stores. *Comp. Biochem. Physiol. A Physiol.* 50, 489–500.

Nasir, H., and Noda, H. (2003). Yeast-like symbiotes as a sterol source in anobiid beetles (Coleoptera, Anobiidae): possible metabolic pathways from fungal sterols to 7-dehydrocholesterol. *Arch. Insect Biochem. Physiol.* 52, 175–182. doi: 10.1002/arch.10079

Ondov, B. D., Bergman, N. H., and Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12:385. doi: 10.1186/1471-2105-12-385

Pankewitz, F., Zöllmer, A., Hilker, M., Gräser, Y., Anja, Z., and Gräser, Y. (2007). Presence of *Wolbachia* in insect eggs containing antimicrobially active anthraquinones. *Microb. Ecol.* 54, 713–721. doi: 10.1007/s00248-007-9230-5

Pant, R. (1988). Nitrogen excretion in insects. *Proc. Anim. Sci.* 97, 379–415.

Patiño-Navarrete, R., Piulachs, M.-D., Belles, X., Moya, A., Latorre, A., and Peretó, J. (2014). The cockroach *Blattella germanica* obtains nitrogen from uric acid through a metabolic pathway shared with its bacterial endosymbiont. *Biol. Lett.* 10, 7–10. doi: 10.1098/rsbl.2014.0407

Pérez, J., Infante, F., Vega, F. E., Holguín, F., Macías, J., Valle, J., et al. (2003). Mycobiota associated with the coffee berry borer (*Hypothenemus hampei*) in Mexico. *Mycol. Res.* 107, 879–887. doi: 10.1017/S0953756203007986

Perez-Guerra, G., and Kosztarab, M. (1992). Biosystematics of the family dactylopiidae (Homoptera: Coccinea) with emphasis on the life cycle of *Dactylopius coccus* Costa. *Bull. Virginia Agric. Exp. Stn.* 92, 1–90.

Pérez-Ramirez, A., Castrejón-Ayala, F., and Jiménez-Pérez, A. (2014). Potential of terpenoids and mealybug extract to determine the establishment of *Dactylopius opuntiae* (Hemiptera: Dactylopiidae) crawlers on Opuntia ficus-indica. *Florida Entomol.* 97, 269–271.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256. doi: 10.1093/molbev/msn083

Potrikus, C. J., and Breznak, J. A. (1980). Uric acid-degrading bacteria in guts of termites [*Reticulitermes flavipes* (Kollar)]. *Appl. Environ. Microbiol.* 40, 117–124.

Potrikus, C. J., and Breznak, J. A. (1981). Gut bacteria recycle uric acid nitrogen in termites: a strategy for nutrient conservation. *Proc. Natl. Acad. Sci. U.S.A.* 78, 4601–4605.

Powell, M. J. (1993). Looking at mycology with a Janus Face: a glimpse at Chytridiomycetes active in the environment. *Mycologia* 85, 1–20.

Ramírez-Puebla, S. T., Rosenblueth, M., Chávez-Moreno, C. K., Catanho Pereira de Lyra, M. C., Tecante, A., and Martínez-Romero, E. (2010). Molecular phylogeny of the genus *Dactylopius* (Hemiptera: Dactylopiidae) and identification of the symbiotic bacteria. *Environ. Entomol.* 39, 1178–1183. doi: 10.1016/j.syapm.2015.05.005

Ramírez-Puebla, S. T., Servín-Garcidueñas, L. E., Ormeño-Orrillo, E., Vera-Ponce de León, A., Rosenblueth, M., Delaye, L., et al. (2015). Species in *Wolbachia*? Proposal for the designation of "*Candidatus* Wolbachia bourtzisii", "*Candidatus* Wolbachia onchocercicola", "*Candidatus* Wolbachia blaxteri", "*Candidatus* Wolbachia brugii", "*Candidatus* Wolbachia taylori", "*Candidatus* Wolbachia collembolicola". *Syst. Appl. Microbiol.* 38, 390–399. doi: 10.1016/j.syapm.2015.05.005

Rho, M., Tang, H., and Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* 38, e191. doi: 10.1093/nar/gkq747

Ricci, I., Mosca, M., Valzano, M., Damiani, C., Scuppa, P., Rossi, P., et al. (2011). Different mosquito species host *Wickerhamomyces anomalus* (*Pichia anomala*): perspectives on vector-borne diseases symbiotic control. *Antonie Van Leeuwenhoek* 99, 43–50. doi: 10.1007/s10482-010-9532-3

Rivera, F. N., González, E., Gómez, Z., López, N., Hernández-Rodríguez, C., Berkov, A., et al. (2009). Gut-associated yeast in bark beetles of the genus *Dendroctonus* Erichson (Coleoptera: Curculionidae: Scolytinae). *Biol. J. Linn. Soc.* 98, 325–342. doi: 10.1111/j.1095-8312.2009.01289.x

Rodríguez, L. C., Faúndez, E., Seymour, J., Escobar, C. A., Espinoza, L., Petroutsa, M., et al. (2005). Factores bióticos y concentración de ácido carmínico en la cochinilla (*Dactylopius coccus* Costa) (Homoptera: Dactylopiidae). *Agric. Técnica* 65, 323–329. doi: 10.4067/s0365-28072005000300011

Rosas-Pérez, T., Rosenblueth, M., Rincón-Rosales, R., Mora, J., and Martínez-Romero, E. (2014). Genome sequence of "*Candidatus* Walczuchella monophlebidarum" the flavobacterial endosymbiont of *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae). *Genome Biol. Evol.* 6, 714–726. doi: 10.1093/gbe/evu049

Sasaki, T., Kawamura, M., and Ishikawa, H. (1996). Nitrogen recycling in the brown planthopper, *Nilaparvata lugens*: involvement of yeast-like endosymbionts in uric acid metabolism. *J. Insect Physiol.* 42, 125–129.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Scorzetti, G., Fell, J. W., Fonseca, A., and Statzell-Tallman, A. (2002). Systematics of basidiomycetous yeasts: a comparison of large subunit D1/D2 and internal transcribed spacer rDNA regions. *FEMS Yeast Res.* 2, 495–517. doi: 10.1111/j.1567-1364.2002.tb00117.x

Spodek, M., Ben-Dov, Y., Protasov, A., Carvalho, C. J., and Mendel, Z. (2013). First record of *Dactylopius opuntiae* (Cockerell) (Hemiptera: Coccoidea: Dactylopiidae) from Israel. *Phytoparasitica* 42, 377–379. doi: 10.1007/s12600-013-0373-2

Stintzing, F. C., and Carle, R. (2005). Cactus stems (*Opuntia* spp.): a review on their chemistry, technology, and uses. *Mol. Nutr. Food Res.* 49, 175–194. doi: 10.1002/mnfr.200400071

Su, X., Pan, W., Song, B., Xu, J., and Ning, K. (2014). Parallel-META 2.0: enhanced metagenomic data analysis with functional annotation, high performance computing and advanced visualization. *PLoS ONE* 9:e89323. doi: 10.1371/journal.pone.0089323

Suh, S.-O., Nguyen, N. H., and Blackwell, M. (2008). Yeasts isolated from plant-associated beetles and other insects: seven novel *Candida* species near *Candida albicans*. *FEMS Yeast Res.* 8, 88–102. doi: 10.1093/oxfordjournals.molbev.a003901

Suh, S. O., Noda, H., and Blackwell, M. (2001). Insect symbiosis: derivation of yeast-like endosymbionts within an entomopathogenic filamentous lineage. *Mol. Biol. Evol.* 18, 995–1000. doi: 10.1093/oxfordjournals.molbev.a003901

Vega, F., and Blackwell, M. (2005). *Insect-Fungal Associations: Ecology and Evolution*. New York, NY: Oxford University Press. Available at: http://fmedicine.ajums.ac.ir/_fmedicine/documents/Insect fungal associations ecology and evolution_20130423_101512.pdf (Accessed August 15, 2014).

Zacchi, L., and Vaughan-Martini, A. (2003). Distribution of three yeast and yeast-like species within a population of soft scale insects (*Saissetia oleae*) as a function of developmental age. *Ann. Microbiol.* 53, 43–46.

Zimmermann, H. G., and Moran, V. C. (1991). Biological control of prickly pear, *Opuntia ficus-indica* (Cactaceae), in South Africa. *Agric. Ecosyst. Environ.* 37, 29–35. doi: 10.1016/0167-8809(91)90137-M

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Vera-Ponce de León, Sanchez-Flores, Rosenblueth and Martínez-Romero. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Elsevier Editorial System(tm) for Fungal Biology
Manuscript Draft

Manuscript Number:

Title: Metagenomic analysis of the fungal community associated with the scale insect Llaveia axin axin (Hemiptera: Coccoidea: Monophlebidae).

Article Type: Original Research

Keywords: Fungal metagenomics
26S rRNA
Uric acid
Aspergillus
Glomeromycota

Corresponding Author: Mr. Arturo Vera-Ponce de Leon,

Corresponding Author's Institution: Center for Genomics Sciences

First Author: Arturo Vera-Ponce de Leon

Order of Authors: Arturo Vera-Ponce de Leon; Tania Rosas-Pérez, Ph.D; Mónica Rosenblueth, Ph.D; Esperanza Martínez-Romero, Ph.D.

Abstract: Association between fungi and insects are widespread in nature. Recent studies are taking advantage of the new DNA sequencing technologies to explore the microbial communities inside insects. Here we report the presence of fungi in the scale insect Llaveia axin axin using a metagenomic approach. L. axin axin feeds on plant sap and contains bacterial endosymbionts, however there are no reports of fungi associated with this insect. The most abundant fungal sequences found corresponded to the Ascomycota, Basidiomycota, Glomeromycota and Chytridiomycota phyla. Using a culture-dependent approach, three different fungal morphotypes were isolated from L. axin axin, and sequences of their 26S rRNA genes showed 99 % similarity to Aspergillus. Metagenomic analyses showed the presence of fungal genes for sterol and glycerol biosynthesis, as well as for uric acid and xenobiotic catabolism. This evidence suggests that associated fungi may help insects in recycling nitrogen, synthesizing lipids or detoxifying toxic chemicals from the plant.

Suggested Reviewers: Angela Douglas
Cornell University
aes326@cornell.edu

Laila Partida-Martínez
CINVESTAV
laila.partida@ira.cinvestav.mx

Amparo Latorre
Instituto Cavanilles de Biodiversidad y Biología Evolutiva Universidad de Valencia
amparo.latorre@uv.es

Cover Letter

January 25th, 2016

Editor in Chief

Fungal Biology

Please find enclosing a submission of manuscript titled "Metagenomic analysis of the fungal community associated with the scale insect *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae)". The manuscript is not under consideration for publication in another journal or book. All supporting material files have been uploaded with the original manuscript. Manuscript submission for publication has been approved by all the authors and institutions; and all persons entitled to authorship have been so named. All authors have seen and agreed to the submitted version of the manuscript

Thank you for your time and consideration.

Sincerely,

Arturo-Vera Ponce de León
Centro de Ciencias Genómicas, UNAM
Programa de Ecología Genómica
Cuernavaca, Mexico
email: avera@ccg.unam.mx

Esperanza Martínez-Romero
Centro de Ciencias Genómicas, UNAM
Programa de Ecología Genómica
Cuernavaca, Mexico
email: emartine@ccg.unam.mx

Manuscript
Click here to download Manuscript: ManuscriptAVPL_250115.docx Click here to view linked References

1 # Metagenomic analysis of the fungal community associated with
2 # the scale insect *Llaveia axin axin* (Hemiptera: Coccoidea:
3 # Monophlebidae).

4 Arturo Vera-Ponce de León[1], Tania Rosas-Pérez[1], Mónica Rosenblueth[1], Esperanza
5 Martínez-Romero[1]

6 [1]Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Ap. P. 565-
7 A, Cuernavaca, Morelos 62251, Mexico, telephone number.: +52 777 3131697, fax
8 number: +52 777 3175581, e-mail: avera@ccg.unam.mx

9 **Abstract**

10 Associations between fungi and insects are widespread in nature. Recent studies are taking
11 advantage of the new DNA sequencing technologies to explore the microbial communities
12 inside insects. Here we report the presence of fungi in the scale insect *Llaveia axin axin*
13 using a metagenomic approach. *L. axin axin* feeds on plant sap and contains bacterial
14 endosymbionts, however there are no reports of fungi associated with this insect. The most
15 abundant fungal sequences found corresponded to the Ascomycota, Basidiomycota,
16 Glomeromycota and Chytridiomycota phyla. Using a culture-dependent approach, three
17 different fungal morphotypes were isolated from *L. axin axin*, and sequences of their 26S
18 rRNA genes showed 99 % similarity to *Aspergillus*. Metagenomic analyses showed the
19 presence of fungal genes involved in sterol and glycerol biosynthesis, as well as in uric acid
20 and xenobiotic catabolism. This evidence suggests that associated fungi may help insects in
21 recycling nitrogen, synthesizing lipids or detoxifying toxic chemicals from the plant.

Introduction

Relationships between insects and fungi are common in nature. In plant-feeding insects fungal symbionts play important roles such as degradation of high molecular weight compounds, detoxification, lipid production, and nitrogen recycling (Sasaki, Kawamura, and Ishikawa 1996; D'Ettorre et al. 2002; Nasir and Noda 2003; Adams et al. 2008; Bouvaine et al. 2012). Many tools have been used to explore fungal associations with insects. These range from basic culture techniques to molecular non-culture-based approaches like restriction fragment length polymorphism (RFLP), oligonucleotide fingerprint of rRNA genes (ORFG), denaturing gradient gel electrophoresis (DGGE), and *in-situ* hybridization (Cui, Morris, and Ghedin 2013). Next generation sequence technologies (NGS), however, have revolutionized our knowledge of microbial communities. Nowadays microbial ecologists can analyze all (or almost all) the microorganisms found in an environment (Sharpton 2014). Cost efficiency and high-throughput data make NGS the gold standard for "mycobiome studies". Shotgun metagenomics (SM) can reveal genes with ecological relevance such as those for nitrogen recycling (uricases), carbon utilization (cellulases) (Scully et al. 2013) or chemical biotransformation (monooxygenases) (Adams et al. 2011). However, fungal genome analyses by SM face challenges due to the large size of fungal genomes (≥ 20 Mb) (Raffaele and Kamoun 2012), the presence of multiple nuclear stages, introns and alternative splicing, as well as large intergenic non coding regions that may be over represented in shotgun sequencing (Lindahl and Kuske 2013). Other alternatives, such as metatranscriptomics, can help to increase metagenomic data mining from fungi inside communities (Qi et al. 2011).

45 *Llaveia axin axin* also called “niij” by native Mexican people (Hemiptera:
46 Coccoidea: Monophlebidae), is a “giant” scale insect which feeds exclusively on plant sap.
47 It is a parasite of some plants like *Acacia cochliacantha*, *Acaciella angustissima, Jatropha*
48 *curcas*, and *Spondias* sp. (Rincón-Rosales and Gutiérrez-Miceli 2008; Suazo-Ortuño, Del
49 Val-De Gortari, and Benitez-Malvido 2013). Adult females are covered by a powdery waxy
50 coating for body protection. Males are winged and only live a few days for mating. This
51 insect has a single generation per year, hatching in the rainy season and ending their life
52 cycle in the middle of fall. Native people use niij for lacquer production. Mexican
53 craftsman collect adult females to obtain a yellow fat that is used to coat traditional
54 handcrafts like gourds, wood and ceramics (Williams and MacVean 1995). Microbial
55 communities associated with *L. axin axin* have been little studied and have focused on
56 bacterial endosymbionts, reporting the presence of Flavobacteria and Enterobacteria
57 (Rosas-Pérez et al. 2014). However, no reports exist on fungi associated with these insect
58 species. Here we present a culture-dependent and shotgun metagenome study of the *L. axin*
59 *axin* fungal community.

60 **Material and Methods**

61 **Insect collection**

62 *L. axin axin* adult females were collected from *Jatropha curcas* plants at Ejido Flores
63 Magón, Mpo. Venustiano Carranza, and at Chiapa de Corzo in Chiapas, Mexico.

64 **Fungal isolation and DNA extraction**

65 Three adult females of *L. axin axin* were removed from their host plant and submerged in
66 100 % ethanol for wax removal. They were then surface-disinfected with 70 % ethanol and

67 rinsed twice with sterile water. Dissection was performed by making a longitudinal cut in
68 the abdomen and extracting the organs with fine sterile forceps. Gut and fat body were
69 separately submerged in 500 µl of sterile 0.85 % NaCl and homogenized using sterile
70 pestles. Fifty microliters of each homogenate was inoculated in 50 ml of YPD liquid
71 medium (1 % w/v yeast extract, 2 % w/v peptone and 2 % w/v dextrose) and malt extract
72 broth medium (Difco). Additionally, some plates with the same media but with 100 µl of
73 the initial homogenate-were incubated in Gaspack ™ EZ CampyPuch™ (in
74 microaerobiosis). All plates were incubated at room temperature for one week. Pure
75 cultures were obtained and stored at -70 °C for further analysis. For DNA extraction, a
76 small ball of mycelia from a liquid-culture grown for 72 hr was placed in 200 µL of lysis
77 buffer solution (Tris-HCl 10 mM, pH. 8; EDTA 1 mM; NaCl 10 mM; SDS 1%; Triton X-
78 100 2 %). Around 0.3 g of sterile glass beads and 200 µl of phenol-chloroform-isoamyl
79 alcohol (25:24:1) were added to the suspension. Then samples were mixed by vortexing
80 and warmed at 65 °C for 1 hr. Samples were then centrifuged at 15,996 x g and the aqueous
81 phase was recovered. DNA was precipitated by adding 1 ml of absolute ethanol and stored
82 for 20 min at -20 °C. Then, samples were centrifuged at 15,996 x g and washed twice with
83 70 % ethanol. DNA was dried in a vacuum concentrator and resuspended in 50 µl of
84 deionized water.

85 **PCR and isolate identification**

86 The 26S rRNA gene was amplified with primers 26S-1 (5'-CATATCAATAAGCGG
87 AGCAAAAG-3') and 26S-2 (5'-CAGTTCTGCTTACCAAAAATGG-3') (González-
88 Quijano et al. 2014). Final concentrations for 50 µl PCR reactions were as follows: 10 ng of
89 total DNA, 0.8 pmol of each primer, 0.2 mM dNTPs, 2.5 mM MgCl, 0.5 U Taq polymerase

90 and 1x Taq polymerase buffer (Invitrogen Life Technologies, Sao Paulo, Brazil). The
91 reaction conditions were 94 °C for 5 min; 35 cycles of 60 s at 94 °C, 60 s at 57 °C, and 90 s
92 at 72 °C; and a final extension at 72 °C for 10 min. PCR products were purified with the
93 High Pure PCR product purification Kit (Roche) and sequenced by Macrogen Inc. (Seoul,
94 Korea) with Sanger technology.

95 **Shotgun metagenomics sequencing and assembly**

96 The whole bodies of fifteen fresh adult females were surface-disinfected and homogenized
97 in PBS with mortar and pestle and sequentially filtered through 100, 20 and 11 μm pore
98 membranes. To remove insect tissue debris, the sample was subjected to a 40% Percoll
99 gradient and centrifuged at 12,000 x *g* for 45 minutes. Layers of Percoll were analyzed
100 under microscope and the one where microorganisms could be observed was centrifuged at
101 16,000 x *g* for 10 minutes. The pellet was used for DNA extraction and purification with
102 the Qiagen DNeasy Blood and Tissue Kit following the manufacturer's instructions. Five
103 micrograms of the DNA were used for metagenomic sequencing with Illumina GAIIx
104 sequencing (hereafter called the Percoll-fraction). In addition to the Percoll-fraction
105 metagenomic data described above, another metagenomic read set from Rosas-Pérez et al.
106 (2014) was analyzed. This sample was obtained from bacteriomes of two adult females
107 sequenced by Illumina HiSeq2000 (hereafter called the Bacteriome sample) (Rosas-Pérez et
108 al., 2014). A total of 61,593,058 100-bp-reads from the insect's bacteriome, and 22,448,562
109 of 36-bp-long from Percoll-fraction were yielded. Illumina sequences were assembled
110 using IDBA 1.1.1 software (Peng et al. 2012) with default parameters.

111 **Fungal community identification**

112 Ribosomal genes from metagenomic reads and assembled contigs were obtained using
113 Parallel-meta 2.0 (Su et al. 2014) algorithm. Eukaryotic ribosomal sequences were
114 recovered using –E option against the SILVA database within an e-value of 1×10^{-10}
115 cutoff. Fungal 18S rRNA sequences from data bases were retrieved using custom-made
116 scripts. Moreover, to recover other fungal sequences not present in the GenBank database,
117 all reads and contigs were aligned to 26S rRNA sequences from cultured niij's fungus
118 using BLAST+ v 2.2.8 (Altschul 1997). Reads with an identity higher than 90 % and length
119 coverage > 80 % were retained. Additionally, a recursive search against of large subunit
120 rRNA (LSU) and small subunit rRNA (SSU) genes was performed using all metagenomics
121 reads by METAXA2 pipeline (Bengtsson-Palme et al. 2015). Tables with fungi hits were
122 visualized in Krona graphs (Ondov, Bergman, and Phillippy 2011).

123 **Metagenomic annotation of fungal communities**

124 Open reading frames (ORFs) coming from all the contigs were obtained using
125 FragGeneScan 1.20 software (Rho, Tang, and Ye 2010). All the ORFs were annotated
126 against the KEGG data base using GhostKOALA 2.0 package
127 (http://www.kegg.jp/ghostkoala/). Fungi metabolic annotations were obtained by custom
128 Perl and R scripts using KEGGREST Bioconductor library
129 (http://bioconductor.org/packages/release/bioc/html/KEGGREST.html).

130 **Phylogenetic analysis**

131 Sequences from 26S rRNA genes, as well as contigs longer than 300 bp identified as fungal
132 LSU or SSU were compared against a non-redundant GenBank library by BLAST

133 (Altschul 1997), and taxonomically related sequences were collected from NCBI. We used
134 jModeltest (Posada 2008) to identify the best substitution model . TrN + G (α= 0.022 for
135 gamma distribution; A= 0.24192, C= 0.23157, G= 0.31386, T= 0.21266) was the best
136 model for 26S rRNA gene. On the other hand, TIM2 + G (α= 0.022 for gamma distribution;
137 A= 0.29073, C= 0.16060, G= 0.22999, T= 0.31867) was the best model for 18S rRNA gene
138 sequence analysis. Sequences were deposited in GenBank data base under the accession
139 numbers KP725294 to KP725297.

140 **Results and discussion**

141 **Culture dependent and culture-independent analysis of nijj's fungal community**

142 Using the culture-dependent approach, three isolates were obtained from each of three
143 independent insects. Isolates were recovered in plates maintained under microaerophilic
144 conditions. 26S rRNA gene sequences from three morphologically distinct isolates (FD01,
145 FD02 and FD03) placed them in the Ascomycota phylum, in a single group near
146 *Aspergillus* in a phylogenetic tree (Fig. 1). Around 3% of the reads belonging to
147 Ascomycota from the Percoll-fraction metagenomic analysis were related to *Aspergillus*
148 *versicolor* (Fig. S1). Similarly, we recovered around 40 reads (≥ 95 nt length) from the
149 metagenomic analysis from the bacteriome fraction with identity greater than 90% to the
150 FD02 and FD01 26S rRNA sequences (data not shown). *Aspergillus* has been described as
151 an opportunistic pathogenic fungus of insects (Rohlfs and Churchill 2011). However, a
152 beneficial interaction has been reported between *A. versicolor* and the granary weevil
153 *Sitophilus granaries*, where the mold prevents parasitoid attacks against weevil larvae

154 (Steiner et al. 2007). Up to now there has not been any report of *Aspergillus* associated with
155 scale insects.

156 Additionally, from the metagenome, 205 and 113 fungal 18S rRNA gene sequences were
157 recovered from the Percoll-fraction and bacteriome samples respectively. 18S rRNA partial
158 gene sequences were classified as belonging to Ascomycota, Basidiomycota,
159 Chytridiomycota, and Glomeromycota phyla. Additionally some *incertis sedis* fungi were
160 revealed (Fig S1). Ascomycota were more abundant in Percoll than Bacteriome samples
161 (Figure 2 and Fig S1) and we found sequences related to fungi that were previously
162 reported as being associated with other insects, such as *Pichia*, *Kazaschtania*, *Capronia* and
163 *Aspergillus*. (Vega and Blackwell 2005; S.-O. Suh and Zhou 2011; Voglmayr et al. 2011).
164 In the Bacteriome sample, Chytridiomycota was the most abundant phyla (Fig 2 and Fig
165 S1). There are reports of entomopathogenic Chytridiomycota associated with Blackfiles
166 and aquatic dipteran larvae (Humber et al. 1990). In Basidiomycota all sequences
167 corresponded to Agaricomycetes and Agaricomycotina classes, which have also been
168 reported in insects, especially those associated with ants (Little and Currie 2007).

169 On the other hand, using LSU ribosomal markers analysis from Percoll-fraction we
170 detected sequences from the order Hypocreales, particularly Ophicordicipitacea (n=8; i.e.
171 *Ophiocordyceps sinensis*); Schizososaccharomycetacea (n=3; i.e. *Zygossacharomyces*
172 *rouxii*) and Phaeosphaeriaceae (n=2; i.e. *Phaeosphaeria nodorum*) families (Fig. S2).
173 Members of Hypocreles have been described as pathogens, however it is known that yeast
174 like symbionts of aphids and planthoppers belong to this order (S. O. Suh, Noda, and
175 Blackwell 2001).

176 In addition, from the assembled contigs a 18S rRNA gene related to Glomeromycetes was
177 reconstituted (Figure 3). Phylogenetic analysis clustered the niij-derived sequence with
178 *Glomus aggregatum* (Glomeromycota) with 55% similarity. Glomeromycota fungi are
179 normally found associated with arbuscular mycorrhiza of plants (Hempel, Renker, and
180 Buscot 2007). Moreover some sequences related to mycorrhiza fungi have been found in
181 human oral cavity samples (Ghannoum et al. 2010; Cui, Morris, and Ghedin 2013). This is
182 the first report that associates Glomeromycota with insects.

183 **Functional annotation of fungal genes from metagenomic analysis**

184 A total of 125,071 ORFs could be reconstructed from meta-assembly. 9,528 ORFs were
185 assigned taxonomically to fungi (GHOSTX score ≥ 25) from which 3,342 corresponded to
186 Basidiomycetes and 6,053 to Ascomycetes. Fungal genes related to sterol and glycerol
187 metabolism were found (Fig. 4 and 5. Tables 1 and 2). It is known that insects are not
188 capable of synthetizing sterols so they have to take them from their diet and in other cases
189 from their symbionts as occurs in beetles and planthoppers (Wetzel et al. 1992; Nasir and
190 Noda 2003). In this work, 29 fungal genes from Ascomycetes (n=16), Basidiomycetes
191 (n=11) and Microsporidia (n=2) for glycerol synthesis were found in the Bacteriome
192 sample annotation (Fig 5, Supplementary Table 1). Genes related to triacylglycerol
193 biosynthesis such as the ones encoding phospholipid: diacylglycerol acyltransferase (n=5)
194 and diacylglycerol O-acyltransferase 1 (n=1) are also present in this sample (Fig. 5, Table
195 S1). Triacylglycerol and diacylglycerol are the main source of reserve energy in insects.
196 These molecules, named as polyunsatured fatty acids (PUFAs), are not synthetized *de novo*
197 by insects and have to be up taken from the diet or other sources (Chapman 2013). The
198 presence of genes encoding fungal triacylglyecerol and diacilglycerol pathway enzymes

199 suggests that fungi could be a source of PUFAs in *L. axin axin*. Furthermore, glycerolipids
200 have been identified as molecules involved in insect-fungi interactions. In the rice blast
201 fungus *Metarhizium anisopliae* appressoria turgor mechanism for insect colonization is
202 mediated by the lipid-breakdown product glycerol (Wang and St Leger 2007), and a
203 *Metharizium robustant* glycerol-acyl-transferase mutant is not able to colonize insects (Gao
204 et al. 2013). On the other hand, in *Drosophila* it is known that glycerol is the signal for
205 flies feed on fungi, particularly yeast (Koseki et al. 2005). These evidences suggest that
206 glycerol produced by fungi inside nijj would be important for the interaction.

207 On the other hand, fungal genes related to purine degradation were found in the
208 niij's metagenome (Fig 6), including urate oxidase, 5-hydroxyisourate hydrolase and
209 allantoinase from Ascomycetes and Basidiomycetes (Table 3, Fig. 6). Uric acid is the major
210 nitrogen waste from terrestrial insects (Chapman 2013) and there are reports of fungi that
211 can recycle this molecule into amino acids useful for their host. For example, in the
212 planthopper *Nilaparvata lugens* yeast like symbionts metabolize uric acid to amino acids
213 useful for the insect (Sasaki, Kawamura, and Ishikawa 1996; Hongoh and Ishikawa 2000).
214 The presence of fungal uricolytic genes was noted in *L. axin axin*, however only one
215 arthropod gene for purine degradation (*all*B, encoding allantoinase) could be found (Fig 6).
216 Evidence of fungal uricase encoding genes and all other genes for uric acid catabolism to
217 urea may suggest that fungi could recycle nitrogen inside the insect by uricolysis.

218 Genes corresponding to *Aspergillus* were obtained from the metagenomic analysis.
219 Particularly, in the Bacteriome we found 184 genes and in the Percoll-fraction we found 2.
220 These genes were related to uric acid catabolism, cellulose breakdown and xenobiotic
221 degradation (Supplementary Table 2). It is known that insects have to contend with toxic

222 chemicals produced by plants. For example, in bark beetles fungi play an important role in
223 α-pinene detoxification (Hunt and Borden 1990), mediated mainly by monooxygenases.
224 *Jatropha curcas* sap contains phorbol esters that are toxic to many animals (Makkar,
225 Aderibigbe, and Becker 1998; Rakshit et al. 2008) . We suggest that *Aspergillus*
226 monoxygenases may participate in degradation of xenobiotic compounds present in their
227 diet. All this evidence suggests-that *Aspergillus* could participate in uric acid recycling,
228 cellulose uptake and xenobiotic protection inside *L. axin axin*.

229 **Acknowledgements**

230 To CONACyT Basic Science grant 154453. AVPL is a doctoral student from Programa de
231 Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM)
232 and received the 331625 fellowship from Consejo Nacional de Ciencia y Tecnología
233 (CONACyT). To Julio Martínez for technical help and to Michael Dunn and Mariana
234 Mateos for reading the manuscript. To Reiner Rincón-Rosales for collecting the niijs. To
235 Alejandra Escobar-Zepeda and Leticia Vega for their help in script developing. All
236 bioinformatics analyses were performed in Unidad de Secuenciación Masiva y
237 Bioinfomática-UNAM servers.

238 **References**

239 Adams, Aaron S, Celia K Boone, Jörg Bohlmann, and Kenneth F Raffa. 2011. Responses
240 of bark beetle-associated bacteria to host monoterpenes and their relationship to insect life
241 histories. Journal of Chemical Ecology 37 (8): 808–17.

242 Adams, Aaron S, Diana L Six, Sandye M Adams, and William E Holben. 2008. In vitro
243 interactions between yeasts and bacteria and the fungal symbionts of the mountain pine
244 beetle (*Dendroctonus ponderosae*). Microbial Ecology 56 (3): 460–66.

245 Altschul, S. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database
246 search programs. Nucleic Acids Research 25 (17): 3389–3402.

247 Bengtsson-Palme, Johan, Martin Hartmann, Karl Martin Eriksson, Chandan Pal, Kaisa
248 Thorell, Dan Göran Joakim Larsson, and Rolf Henrik Nilsson. 2015. metaxa2: improved
249 identification and taxonomic classification of small and large subunit rRNA in
250 metagenomic data. Molecular Ecology Resources 15 (6): 1403–14.

251 Bouvaine, Sophie, Spencer T Behmer, George G Lin, Marie-Line Faure, Robert J
252 Grebenok, and Angela E Douglas. 2012. The physiology of sterol nutrition in the pea aphid
253 A*cyrthosiphon pisum*. Journal of Insect Physiology 58 (11). Elsevier Ltd: 1383–89.

254 Chapman, RF. 2013. The insects structure and function. Edited by Angela E. Douglas and
255 Stephen J. Simpson. 5th ed. Cambridge University Press.

256 Cui, Lijia, Alison Morris, and Elodie Ghedin. 2013. The human mycobiome in health and
257 disease. Genome Medicine 5 (7): 63.

258 D'Ettorre, Patrizia, P. Mora, V. Dibangou, C. Rouland, and C. Errard. 2002. The role of the
259 symbiotic fungus in the digestive metabolism of two species of fungus-growing ants.
260 Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental
261 Physiology 172 (2): 169–76.

262 Gao, Qiang, Yanfang Shang, Wei Huang, and Chengshu Wang. 2013. glycerol-3-phosphate
263 acyltransferase contributes to triacylglycerol biosynthesis, lipid droplet formation, and host
264 invasion in *Metarhizium robertsii*. Applied and Environmental Microbiology 79 (24):
265 7646–53.

266 Ghannoum, Mahmoud a, Richard J Jurevic, Pranab K Mukherjee, Fan Cui, Masoumeh
267 Sikaroodi, Ammar Naqvi, and Patrick M Gillevet. 2010. Characterization of the oral fungal
268 microbiome (mycobiome) in healthy individuals. PLoS Pathogens 6 (1): e1000713.

269 González-Quijano, Génesis Karendash, Lidia Dorantes-Alvarez, Humberto Hernández-
270 Sánchez, María Eugenia Jaramillo-Flores, María de Jesús Perea-Flores, Arturo Vera-Ponce
271 de León, and César Hernández-Rodríguez. 2014. Halotolerance and survival kinetics of
272 lactic acid bacteria isolated from jalapeño pepper (*Capsicum annuum* l.) Fermentation.
273 Journal of Food Science 79 (8): M1545–53.

274 Hempel, Stefan, Carsten Renker, and François Buscot. 2007. Differences in the species
275 composition of arbuscular mycorrhizal fungi in spore, root and soil communities in a
276 grassland ecosystem. Environmental Microbiology 9 (8): 1930–38.

277 Hongoh, Y, and H Ishikawa. 2000. Evolutionary studies on uricases of fungal
278 endosymbionts of aphids and planthoppers. Journal of Molecular Evolution 51 (3): 265–77.

279 Humber, Richard a., R. a. Samson, H. C. Evans, and J.-P. Latge. 1990. Atlas of
280 entomopathogenic fungi. Mycologia 82 (1): 148.

281 Hunt, DWA, and JH Borden. 1990. Conversion of verbenols to verbenone by yeasts
282 isolated from *Dendroctonus ponderosae* (coleoptera: scolytidae). Journal of Chemical
283 Ecology 16 (4): 1385–97.

284 Koseki, Takaya, Masayuki Koganezawa, Akira Furuyama, Kunio Isono, and Ichiro
285 Shimada. 2005. Stereospecificity of the receptor site for glycerol, a new sweetener, in a
286 labellar sugar receptor cell of *Drosophila*. Chemical Senses 30 Suppl 1 (suppl 1): i277–78.

287 Lindahl, Björn D, and Cheryl R Kuske. 2013. Metagenomics for study of fungal ecology.
288 In The Ecological Genomics of Fungi, edited by Francis Martin. Hoboken, NJ: John Wiley
289 & Sons, Inc.

290 Little, Ainslie E F, and Cameron R Currie. 2007. Symbiotic complexity: discovery of a
291 fifth symbiont in the attine ant-microbe symbiosis. Biology Letters 3 (5): 501–4.

292 Makkar, H.P.S., a.O. Aderibigbe, and K. Becker. 1998. Comparative evaluation of non-
293 toxic and toxic varieties of *Jatropha curcas* for chemical composition, digestibility, protein
294 degradability and toxic factors. Food Chemistry 62 (2): 207–15.

295 Nasir, Habib, and Hiroaki Noda. 2003. Yeast-like symbiotes as a sterol source in anobiid
296 beetles (Coleoptera, Anobiidae): possible metabolic pathways from fungal sterols to 7-
297 Dehydrocholesterol. Archives of Insect Biochemistry and Physiology 52 (4): 175–82.

298 Ondov, Brian D, Nicholas H Bergman, and Adam M Phillippy. 2011. Interactive
299 metagenomic visualization in a web browser. BMC Bioinformatics 12 (1). BioMed Central
300 Ltd: 385.

301 Peng, Yu, Henry C M Leung, S M Yiu, and Francis Y L Chin. 2012. IDBA-UD: a *de novo*
302 assembler for single-cell and metagenomic sequencing data with highly uneven depth.
303 Bioinformatics. 28 (11): 1420–28.

304 Posada, David. 2008. jModelTest: Phylogenetic model averaging. Molecular Biology and
305 Evolution 25 (7): 1253–56.

306 Qi, Meng, Pan Wang, Nicholas O'Toole, Perry S Barboza, Emilio Ungerfeld, Mary Beth
307 Leigh, L Brent Selinger, et al. 2011. Snapshot of the eukaryotic gene expression in
308 muskoxen rumen--a metatranscriptomic approach. PloS One 6 (5): e20521.

309 Raffaele, Sylvain, and Sophien Kamoun. 2012. Genome evolution in filamentous plant
310 pathogens: why bigger can be better. Nature Reviews. Microbiology 10 (6): 417–30.

311 Rakshit, K D, J Darukeshwara, K Rathina Raj, K Narasimhamurthy, P Saibaba, and S
312 Bhagya. 2008. Toxicity studies of detoxified *Jatropha* meal (*Jatropha curcas*) in rats. Food
313 and Chemical Toxicology 46 (12): 3621–25.

314 Rho, Mina, Haixu Tang, and Yuzhen Ye. 2010. FragGeneScan: predicting genes in short
315 and error-prone reads. Nucleic Acids Research 38 (20): e191..

316 Rincón-rosales, r, and fa gutiérrez-miceli. 2008. Características biológicas de *Acaciella*
317 *angustissima* (Mill.) Britton & Rose en su habitat natural y evaluación de su potencial
318 cortical en Chiapas, México. Agrociencia, 129–37.

319 Rohlfs, Marko, and Alice C.L. Churchill. 2011. Fungal secondary metabolites as
320 modulators of interactions with insects and other arthropods. Fungal Genetics and Biology
321 48 (1): 23–34.

322 Rosas-Pérez, Tania, Mónica Rosenblueth, Reiner Rincón-Rosales, Jaime Mora, and
323 Esperanza Martínez-Romero. 2014. Genome sequence of '*Candidatus* Walczuchella
324 monophlebidarum' the flavobacterial endosymbiont of *Llaveia axin axin* (Hemiptera:
325 Coccoidea: Monophlebidae). Genome Biology and Evolution 6 (3): 714–26.

326 Sasaki, T., M. Kawamura, and H. Ishikawa. 1996. Nitrogen recycling in the brown
327 planthopper, *Nilaparvata lugens*: involvement of yeast-like endosymbionts in uric acid
328 metabolism. Journal of Insect Physiology 42 (2): 125–29.

329 Scully, Erin D, Scott M Geib, Kelli Hoover, Ming Tien, Susannah G Tringe, Kerrie W
330 Barry, Tijana Glavina del Rio, Mansi Chovatia, Joshua R Herr, and John E Carlson. 2013.
331 Metagenomic profiling reveals lignocellulose degrading system in a microbial community
332 associated with a wood-feeding beetle. PloS One 8 (9): e73827.

333 Sharpton, Thomas J. 2014. An introduction to the analysis of shotgun metagenomic data.
334 Frontiers in Plant Science 5 (June): 209.

335 Steiner, Sven, Daniel Erdmann, Johannes L M Steidle, and Joachim Ruther. 2007. Host
336 habitat assessment by a parasitoid using fungal volatiles. Frontiers in Zoology 4 (January):
337 3.

338 Su, Xiaoquan, Weihua Pan, Baoxing Song, Jian Xu, and Kang Ning. 2014. Parallel-meta
339 2.0: enhanced metagenomic data analysis with functional annotation, high performance
340 computing and advanced visualization. PloS One 9 (3): e89323.

341 Suazo-Ortuño, Ireri, Ek Del Val-De Gortari, and Julieta Benitez-Malvido. 2013.
342 Rediscovering an extraordinary vanishing bug: L*laveia axin axin*. Revista Mexicana de
343 Biodiversidad 84 (1): 338–46.

344 Suh, S O, H Noda, and M Blackwell. 2001. Insect symbiosis: derivation of yeast-like
345 endosymbionts within an entomopathogenic filamentous lineage. Molecular Biology and
346 Evolution 18 (6): 995–1000.

347 Suh, Sung-Oui, and Jianlong J Zhou. 2011. *Kazachstania intestinalis* sp. Nov., an
348 ascosporogenous yeast from the gut of passalid beetle *Odontotaenius disjunctus*. Antonie
349 van Leeuwenhoek 100 (1): 109–15.

350 Vega, FE, and Meredith Blackwell. 2005. Insect-fungal associations: ecology and
351 evolution. Edited by Fernando E. Vega and Meredith Blackwell. New York: Oxford
352 University Press..

353 Voglmayr, Hermann, Veronika Mayer, Ulrich Maschwitz, Joachim Moog, Champlain
354 Djieto-Lordon, and Rumsaïs Blatrix. 2011. The diversity of ant-associated black yeasts:
355 insights into a newly discovered world of symbiotic interactions. Fungal Biology 115 (10):
356 1077–91.

357 Wang, Chengshu, and Raymond J St Leger. 2007. The *Metarhizium anisopliae* perilipin
358 homolog mpl1 regulates lipid metabolism, appressorial turgor pressure, and virulence. The
359 Journal of Biological Chemistry 282 (29): 21110–15.

360 Wetzel, J M, M Ohnishi, T Fujita, K Nakanishi, Y Naya, H Noda, and M Sugiura. 1992.
361 Diversity in steroidogenesis of symbiotic microorganisms from planthoppers. Journal of
362 Chemical Ecology 18 (11): 2083–94.

363 Williams, M, and CM MacVean. 1995. Ethnococcidology: use of the giant margarodids, by
364 indigenous peoples of mesoamerica in their culture , medicine and arts. Israel Jounal of
365 Entomology XXIX (29): 147–48.

366

367

368

369 **Figure legends**

370 Figure 1. Maximum likelihood tree (-ln L=-2193.57258) of fungi isolated from *L. axin*
371 *axin*. The 26S rRNA sequence of *Thermoascus crustaceus* was used as outgroup. Scale bar
372 indicates 0.005 % estimated sequence divergence. Boostrap support values ≥ 50 % are
373 indicated.

374 Figure 2. Fungus community composition (%) found in *L. axin axin* in the two
375 metagenomes analyzed and in pure culture.

376 Figure 3. Maximum likelihood tree (-ln L=- 2967.29607) of Glomeromycota from fungal
377 18S rRNA sequences from the *L. axin axin* metagenome. The 18S rRNA sequence of
378 *Rhodotorula lactose* was used as outgroup. Scale bar indicates 2 % estimated sequence
379 divergence. Bootstrap support values ≥ 50 % are indicated.

380 Figure 4. Genes involved in steroid biosynthesis from nijj's metagenome. Colors show
381 Green: Arthropod, Blue: Ascomycetes and Pink: Basidiomycete.

382 Figure 5. Genes involved in glycerol metabolism from nijj's metagenome. Colors show
383 Green: Arthropod, Blue: Ascomycetes and Pink: Basidiomycetes.

384 Figure 6. Genes involved in purine metabolism annotated from nijj's metagenome. Colors
385 show Green: Arthropod, Blue: Ascomycetes and Pink: Basidiomycetes.

386

387

388

389 **Supplementary material**

390 Fig S1. Krona map from 18S rRNA fungal sequences of *L. axin axin* metagenomes.

391 Fig S2 Krona map from LSU fungal sequences of Percoll-fraction metagenome.

392 Table S1. Annotation of fungal genes involved in glycerol biosynthesis from *L. axin axin*

393 Bacteriome.

394 Table S2. Genes annotated of *Aspergillus* sequences from *L. axin axin* metagenome.

Table1

Table 1. Annotation of fungal gene products involved in sterol biosynthesis from the *L. axin axin* metagenome.

Fungal phylum	Enzyme	EC number	KO id
Ascomycetes	C-22 sterol desaturase	EC: 1.14.14.-	K09831
	Delta24(24(1))-sterol reductase	EC:1.3.1.71	K00223
	Sterol 14-demethylase	EC:1.14.13.70	K05917
	Sterol-4alpha-carboxylate 3-dehydrogenase (decarboxylating)	EC:1.1.1.170	K07748
	Sterol 24-C-methyltransferase	EC:2.1.1.41	K00559
	Delta14-sterol reductase	EC:1.3.1.70	K00222
	Lathosterol oxidase	EC:1.14.21.6	K00227
	Sterol 14-demethylase	EC:1.14.13.70	K05917
	7-dehydrocholesterol reductase	EC:1.3.1.21	K00213
	Sterol-4alpha-carboxylate 3-dehydrogenase (decarboxylating)	EC:1.1.1.170	K07748
Basidiomycetes	C-22 sterol desaturase	EC:1.14.14.-	K09831
	Sterol 24-C-methyltransferase	EC:2.1.1.41	K00559
	Lanosterol synthase	EC:5.4.99.7	K01852
	Lathosterol oxidase	EC:1.14.21.6	K00227
	Delta14-sterol reductase	EC:1.3.1.70	K00222
	Sterol 24-C-methyltransferase	EC:2.1.1.41	K00559

Table2

Table 2. Annotation of fungal gene products involved in glycerol biosynthesis fromthe *L. axin axin* metagenome.

Fungal phylum	Enzyme	EC number	KO id
Ascomycetes	TAG lipase / steryl ester hydrolase / phospholipase A2 / LPA acyltransferase	EC:3.1.1.3/3.1.1.13/3.1.1.4/2.3.1.51	K14674
	Glycerol-3-phosphate O-acyltransferase / dihydroxyacetone phosphate acyltransferase	EC:2.3.1.15/2.3.1.42	K13507
Basidiomycetes	Glycerol-3-phosphate O-acyltransferase / dihydroxyacetone phosphate acyltransferase	EC:2.3.1.15/2.3.1.42	K13507
	Lysophospholipid acyltransferase	EC:2.3.1.51/2.3.1.23/2.3.1	K13519

Table3

Table 3. Annotation of fungal gene products involved in purine catabolism from the *L. axin axin* metagenome.

Fungal phylum	Enzyme	EC Number	KO ID
Ascomycetes	Xanthine dehydrogenase/oxidase	EC:1.17.1.4/ 1.17.3.2	K00106
	Urate oxidase	EC:1.7.3.3	K00365
	Allantoicase	EC:3.5.3.4	K01477
	Urate oxidase	EC:1.7.3.3	K00365
	Xanthine dehydrogenase/oxidase	EC:1.17.1.4/ 1.17.3.2	K00106
	Urate oxidase	EC:1.7.3.3	K00365
	Xanthine dehydrogenase/oxidase	EC:1.17.1.4 1.17.3.2	K00106
	Xanthine dehydrogenase/oxidase	EC:1.17.1.4 1.17.3.2	K00106
	Allantoicase	EC:3.5.3.4	K01477
	Ureidoglycolate lyase	EC:4.3.2.3	K01483
	Urease	EC:3.5.1.5	K01427
Basidiomycetes	Allantoinase	EC:3.5.2.5	K01466
	Allantoicase	EC:3.5.3.4	K01477
	5-hydroxyisourate hydrolase	EC:3.5.2.17	K07127
	Allantoinase	EC:3.5.2.5	K01466

Figure1

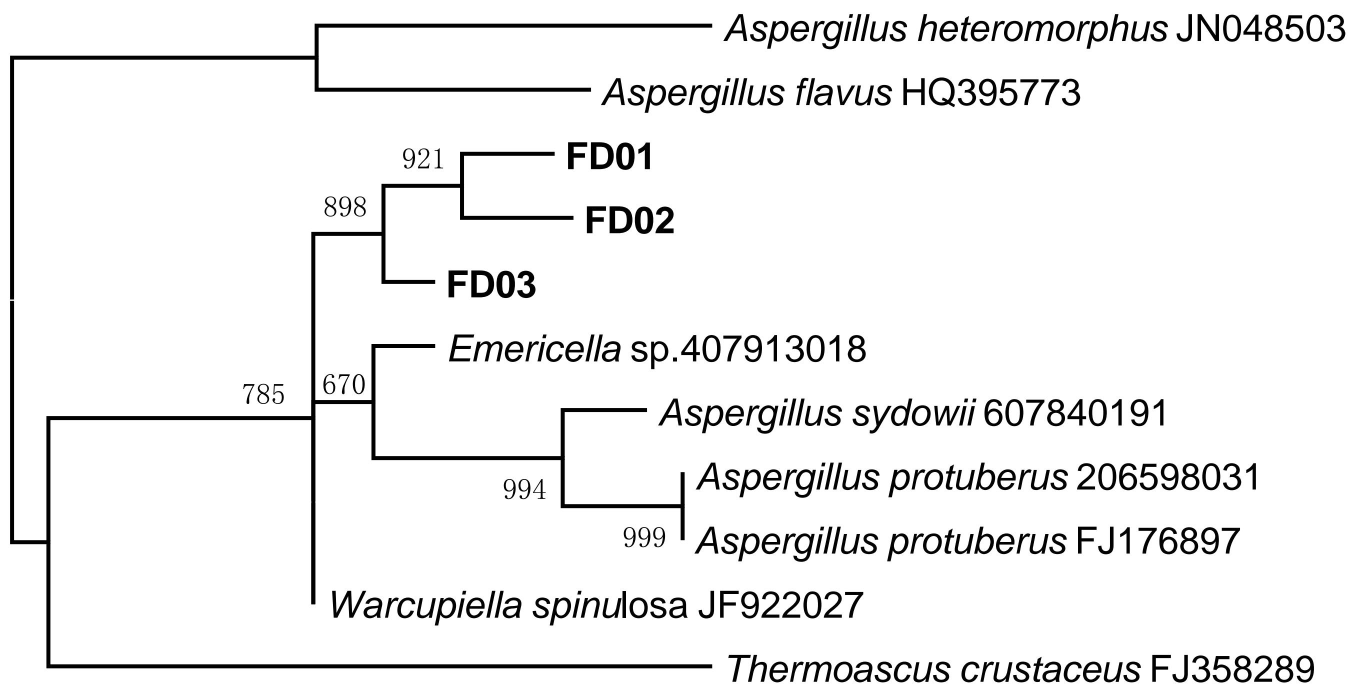


Figure2

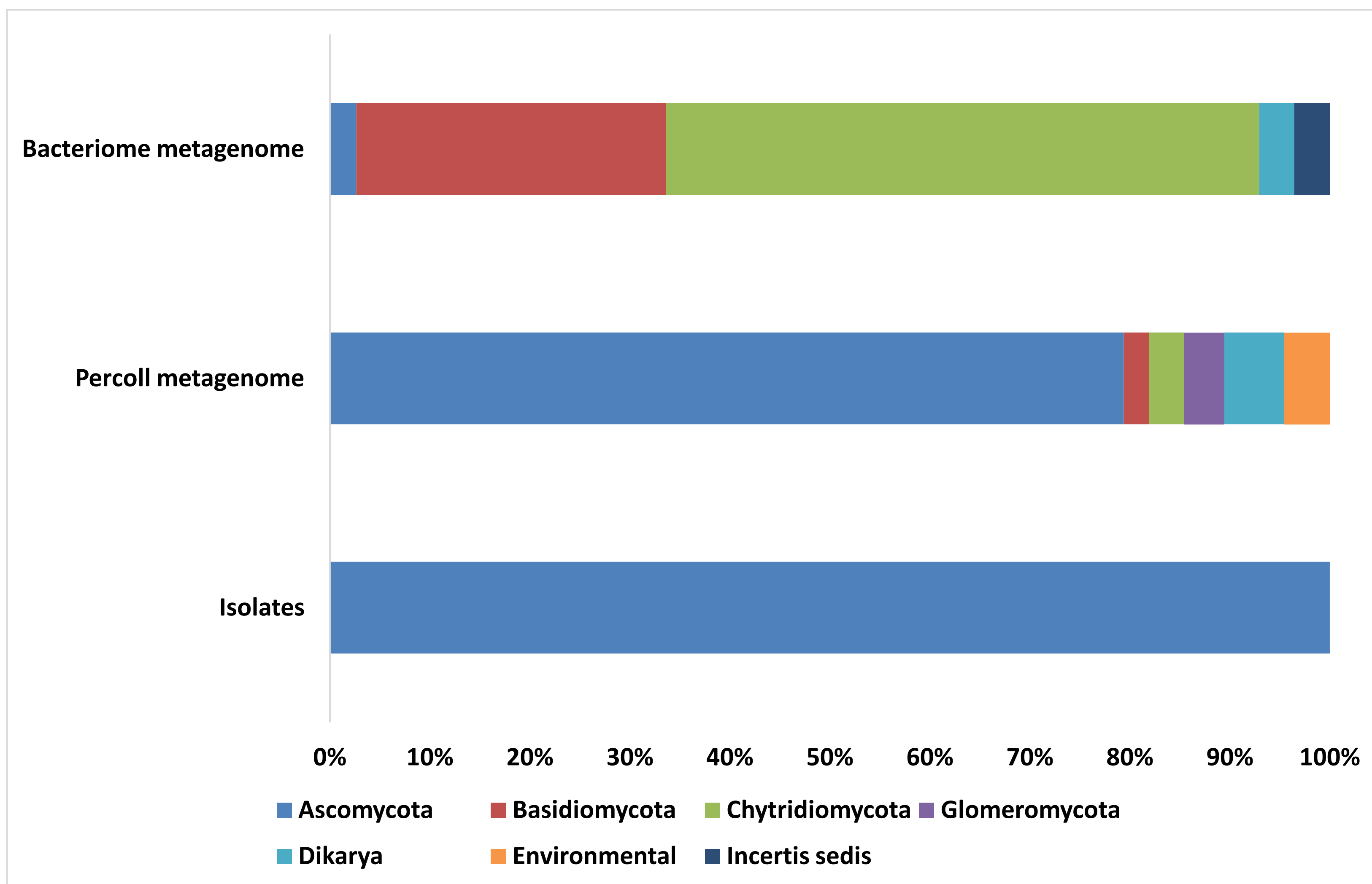


Figure3

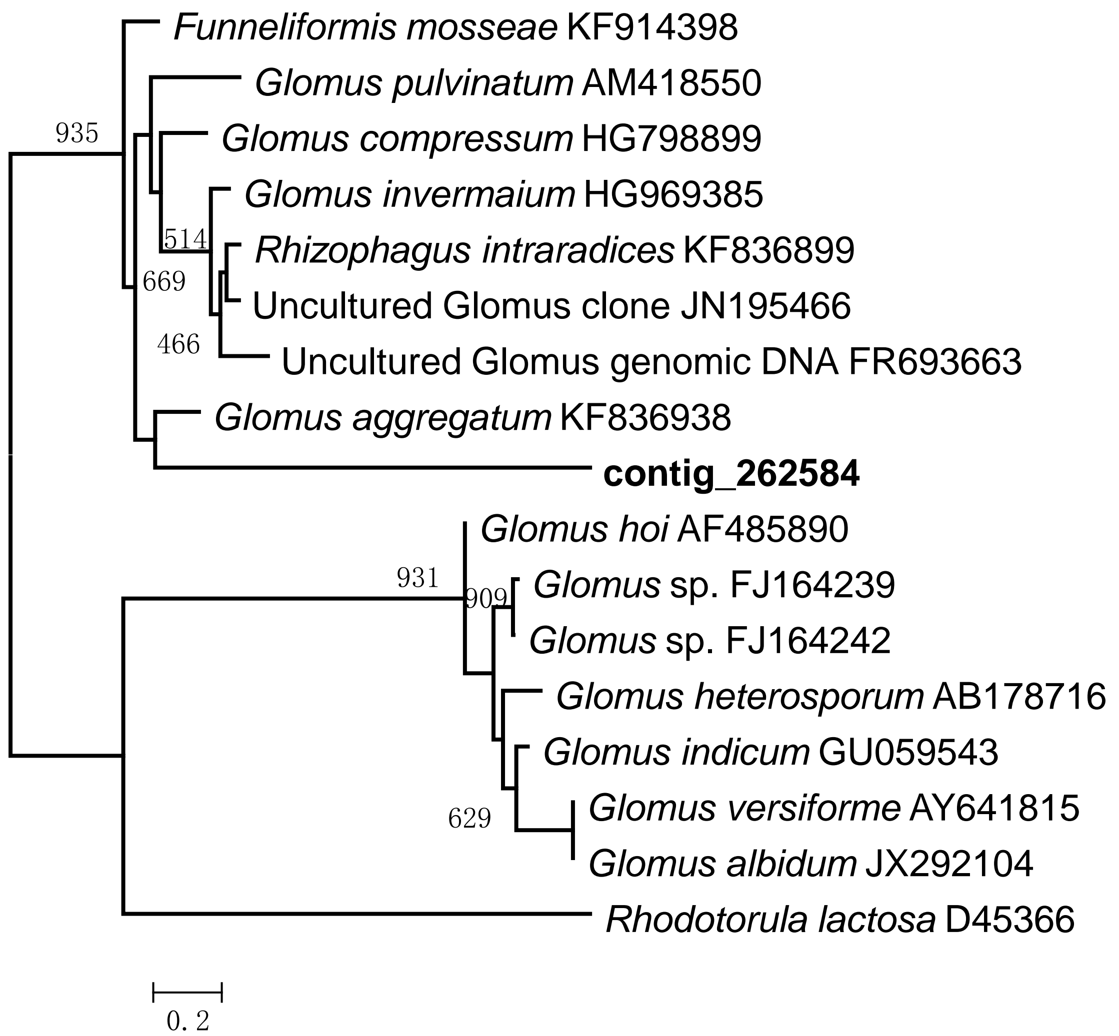


Figure4

STEROID BIOSYNTHESIS

Terpenoid backbone biosynthesis
Farnesyl-PP
2.5.1.21
Presqualene-PP
2.5.1.21
Squalene
1.14.13.132
(S)-Squalene-2,3-epoxide
5.4.99.7
5.4.99.8
24,25-Dihydro-lanosterol
1.3.1.72
Lanosterol
1.14.13.70
4,4-Dimethyl-cholesta-8,14,24-trienol
1.3.1.70
4,4-Dimethyl-cholesta-8-en-3β-ol
1.3.1.72
14-Demethyl-lanosterol
1.14.13.72
4-Methylzymosterol-carboxylate
1.1.1.170
3-Keto-4-methyl-zymosterol
1.1.1.270
4-Methylzymosterol
Cholesta-8-en-3β-ol
1.3.1.72
Zymosterol
ERG6
Fecosterol
5.3.3.5
5.3.3.5
ERG2
Lathosterol
1.3.1.72
Cholesta-7,24-dien-3β-ol
Episterol
1.14.21.6
ERG3
Vitamin D3
7-Dehydrocholesterol
hv
1.3.1.72
7-Dehydro-desmosterol
5,7,24(28)-Ergosta-trienol
1.14.13.159
1.3.1.21
ERG5
Cholesterol ester
3.1.1.13
1.3.1.72
Desmosterol
5,7,22,24(28)-Ergosta-tetraenol
Calcidiol
2.3.1.26
Cholesterol
ERG4
1.14.13.13
2.3.1.73
Steroid hormone biosynthesis
Ergosterol
Calcitriol
Steroid degradation
hv
1.14.13.126
1.14.13.126
Primary bile acid biosynthesis
Vitamin D2
Calcitetrol
Secalciferol

Cycloartenol
SMT1
24-Methylene-cycloartanol
SMO1
Cycloeucalenol
5.5.1.9
Obtusifoliol
CYP51G1
δ 8,14-Sterol
FK
4α-Methylfecosterol
HYD1
24-Methylene-lophenol
SMT2
24-Ethylidene-lophenol
SMO2
SMO2
Episterol
δ 7-Avenasterol
STE1
STE1
5-Dehydroepisterol
5-Dehydroavenasterol
DWF5
DWF5
24-Methylene-cholesterol
Isofucosterol
DWF1
DWF1
Campesterol
Sitosterol
CYP710A
Brassinosteroid biosynthesis
Stigmasterol
Phytosterol

00100 11/29/13
(c) Kanehisa Laboratories

Figure5

Click here to download high resolution image

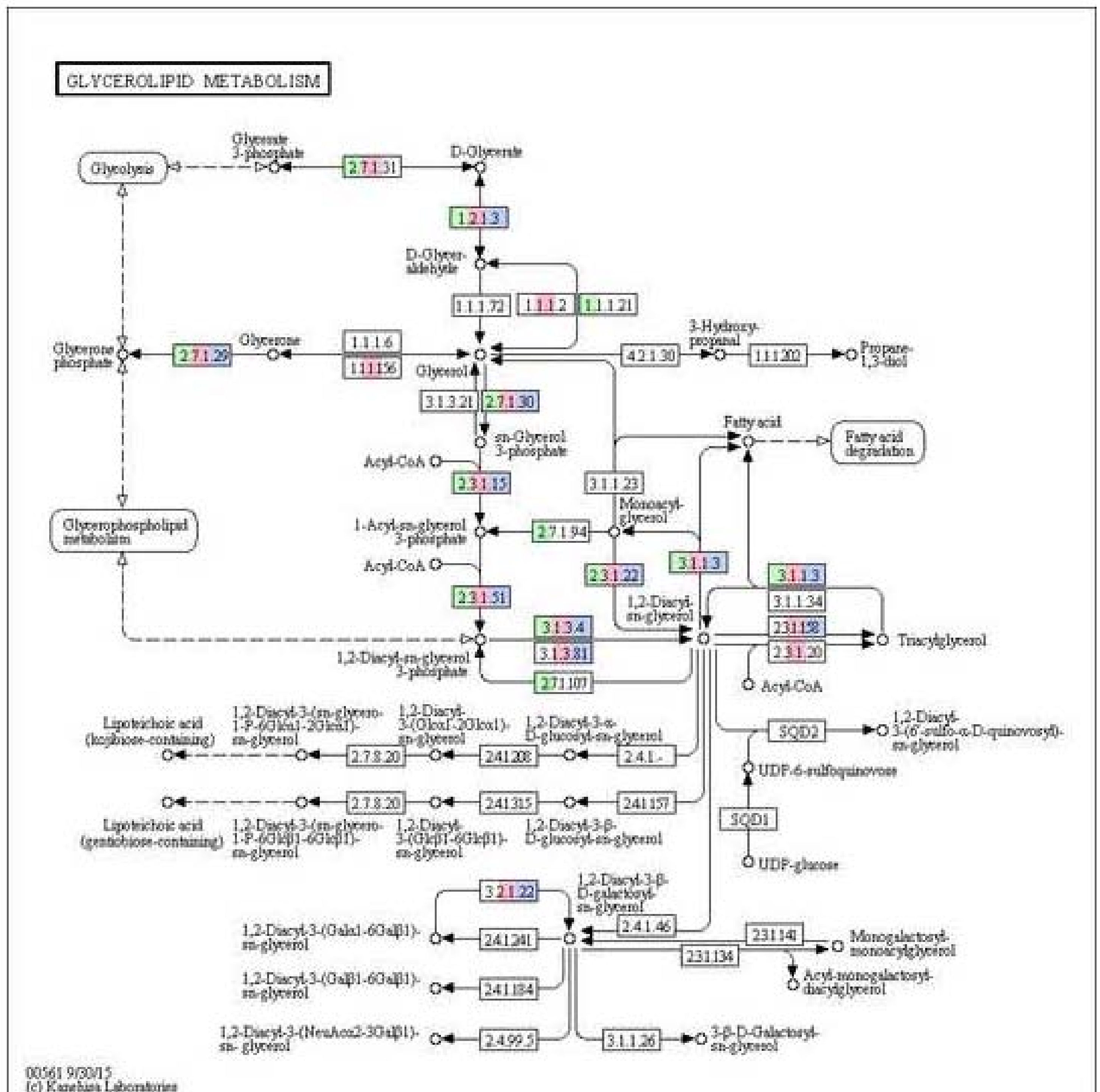
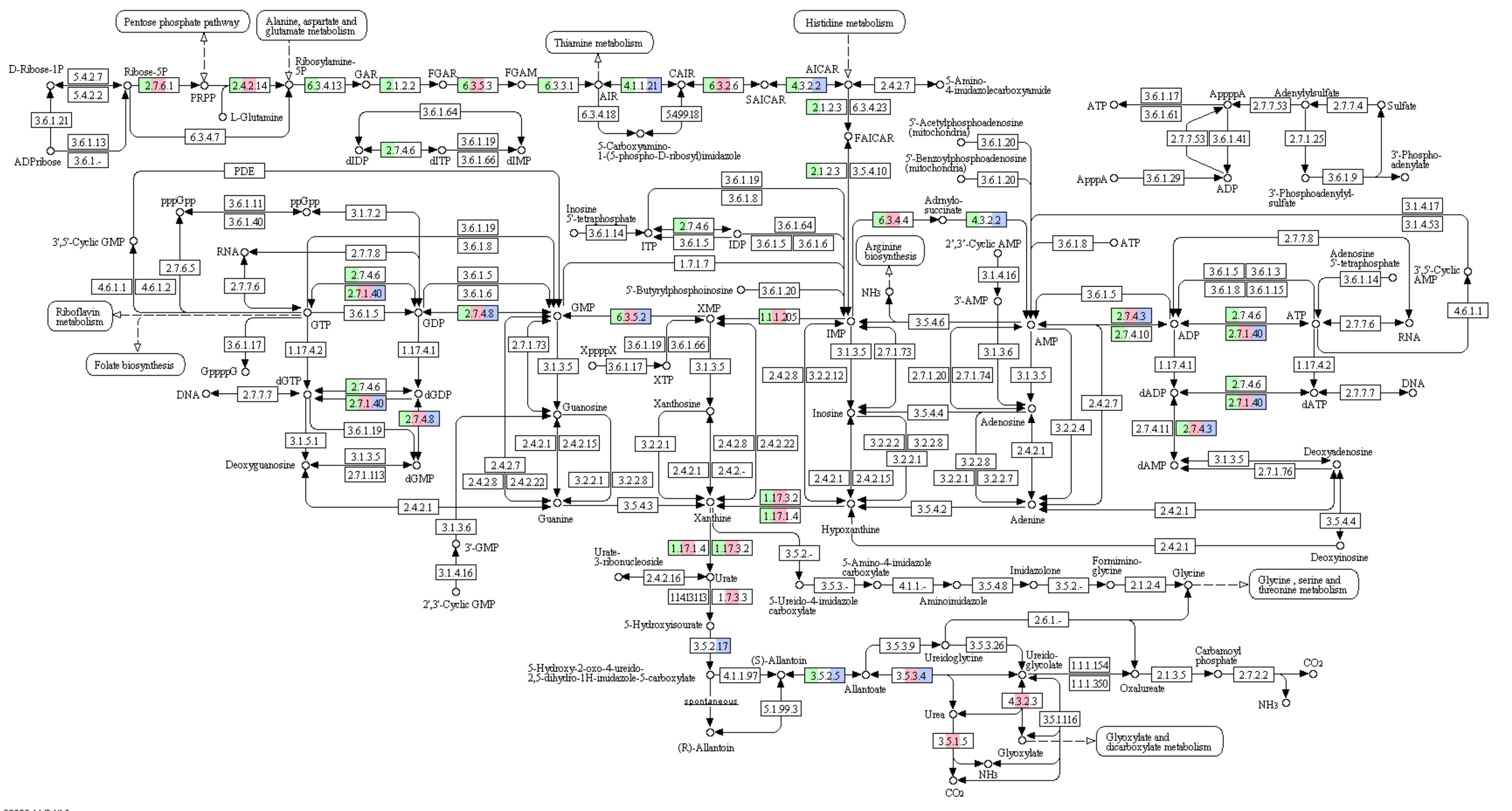


Figure6



figure_s1.html

Click here to download Supplementary File: Figure_S1.html

figure_s2.html
Click here to download Supplementary File: Figure_S2.html

Supplementary Table 1

Click here to download Supplementary File: Table_S1.xlsx

Supplementary Table 2

Click here to download Supplementary File: Table_S2.xlsx

OXFORD UNIVERSITY PRESS | Genome Biology and Evolution

Candidatus Dactylopiibacterium carminicum a nitrogen-fixing symbiont of the cochineal insect Dactylopius coccus (Hemiptera: Dactylopiidae)

Journal:	*Genome Biology and Evolution*
Manuscript ID	GBE-161232
Manuscript Type:	Research Article
Date Submitted by the Author:	15-Dec-2016

Complete List of Authors:	Vera-Ponce de León, Arturo; Universidad Nacional Autónoma De México, Centro de Ciencias Genomicas Ormeño-Orrillo, Ernesto; Universidad Nacional Agraria La Molina; Universidad Nacional Autónoma De México, Centro de Ciencias Genomicas Ramírez-Puebla, Shamayim; Universidad Nacional Autónoma De México, Centro de Ciencias Genomicas Rosenblueth, Monica; Centro de Ciencias Genomicas, Departamento de Ecologia Genomica Degli Esposti, Mauro; Italian Institute of Technology, Nanotoxicology Martinez-Romero, Julio; Universidad Nacional Autónoma De México, Centro de Ciencias Genomicas Martinez-Romero, Esperanza; Universidad Nacional Autónoma De México, Centro de Ciencias Genomicas

Keywords:	nitrogen fixation, scale insect, metagenomics

SCHOLARONE™ Manuscripts

Candidatus Dactylopiibacterium carminicum a nitrogen-fixing symbiont of the cochineal insect *Dactylopius coccus* (Hemiptera: Dactylopiidae)

Arturo Vera-Ponce de León[1a], Ernesto Ormeño-Orrillo[1,2a], Shamayim T. Ramírez-Puebla[1a] Mónica Rosenblueth[1], Mauro Degli Esposti[1], Julio Martinez[1], Esperanza Martinez-Romero[1*]

[1] Programa de Ecología Genómica, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, México,

[2] Universidad Nacional Agraria La Molina, Lima, Perú.

*Corresponding author **e mail** emartine@ccg.unam.mx

[a] authors made equal contributions to the manuscript

For Peer Review

ABSTRACT

Acetylene reduction activity was detected in the tissues of *Dactylopius coccus*, the cochineal insect that produces carmine, the natural red pigment. The metagenome analysis of *Dactylopius coccus* allowed the assembly of a 3.6 Mb genome that contained nitrogen fixation genes. The genome corresponded to the β−Proteobacterial symbiont *Candidatus* Dactylopiibacterium carminicum (*D. carminicum*), which was previously found in five different *Dactylopius* species. The genome size and its close relatedness to others from free-living bacteria (such as *Uliginosibacterium gangwonense* and rice and grass nitrogen-fixing endophytes) suggest that *D. carminicum* is a recently acquired symbiont in the carmine cochineal seemingly substituting the flavobacterial and enterobacterial endosymbionts that are prevalent in this insect group. *D. carminicum* may compensate for nitrogen deficiencies in the cactus sap from which the insects feed, since its nitrogenase genes were expressed in *D. coccus* hemolymph and ovaries and were the only nitrogenase genes found in all metagenomes analyzed.

INTRODUCTION

The ecological success of insects may be related to their association with different microorganisms. It is estimated that insect microbiota constitute around 10% of host biomass (Douglas, 2015). Bacteria may provide nitrogen to the hosts in the form of amino acids, as well as other compounds such as pheromones, lipids and vitamins (Adams et al., 2009; Behmer et al., 2011; Douglas, 2012). In other cases, microorganisms confer protection against parasites (Engel and Moran, 2013; Xie et al. 2014). Some insects have specialized cells called bacteriocytes where they harbor bacterial endosymbionts which are maternally inherited. Only rarely can endosymbionts be cultured in the laboratory (Dale

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

49 and Maudlin, 1999; Dale et al. 2006). Consequently, the advent of genomics paved the way
50 to their detailed study. The first genome from an endosymbiont (that of *Buchnera* from
51 aphids) was published in 2000 (Shigenobu et al. 2000) and now many various genomes are
52 available from endosymbionts belonging to different bacterial phyla (reviewed in
53 McCutcheon and Moran 2012; Martinez-Cano et al. 2015). Typically, they show a
54 reduction in gene content, genome erosion and high AT content, as well as over-
55 representation of genes encoding essential functions for their insect host (McCutcheon and
56 Moran, 2012).

57 *Dactylopius coccus* is the cochineal producer of carmine pigment. These insects
58 were domesticated by pre-Hispanic Americans and still have a large economic value for
59 their production of natural pigments that are used for dyeing cosmetics, textiles and food
60 (Perez-Guerra and Kosztarab, 1992; Williams and Ben-Dov, 2015; Campana et al. 2015).
61 *D. coccus* spends its entire life on cactus plants and, like other sap-feeding insects, has
62 nutritional deficiencies that may be compensated for by its microbial symbionts. Different
63 bacteria and fungi have been detected in the microbiota of *Dactylopius* species (Ramírez-
64 Puebla et al. 2010, 2016a; Vera-Ponce de León et al. 2016), including the Proteobacteria
65 *Herbaspirillum*, *Mesorhizobium* and *Sphingomonas*, species that are commonly associated
66 with plants (James et al. 1997; Kim et al. 1998; López-López et al. 2013). In addition, a
67 novel β-Proteobacterium named *Candidatus* Dactylopiibacterium carminicum (hereafter
68 abbreviated as *D. carminicum*) was identified in 16S rRNA libraries from all *Dactylopius*
69 species sampled (Ramírez-Puebla et al 2010). Furthermore, two *Wolbachia* strains
70 corresponding to supergroups A and B (now called *Candidatus* Wolbachia bourtzisii and
71 *Candidatus* Wolbachia pipientis, respectively (Ramírez-Puebla et al. 2015)) were identified

For Peer Review

in the *D. coccus* metagenome. Different fungal species have also been identified in the microbiota of *D. coccus, D. ceylonicus* and *D. opuntiae* (Vera-Ponce de León et al. 2016).

Nitrogen fixation by insect-associated bacteria alleviates dietary nitrogen deficiencies (Engel and Moran, 2013; Douglas, 2015). For example, it has been estimated that 10-40 Kg N ha^{-1} per year are fixed by the symbiotic bacteria of insects (Nardi et al. 2002). Diverse diazotrophs have been identified in termites (Ohkuma et al. 1999), ants (van Borm et al. 2002), fruit flies (Behar et al. 2005), cockroaches (Tai et al. 2016) and some beetles (Morales-Jiménez et al. 2009, 2013), but not in hemipterans such as *Dactylopius.* Diazotrophic symbionts of termites include spirochaetes that use acetate as carbon source (Desai and Brune, 2012). Conversely, ants, bark beetles and fruit flies have *Klebsiella, Kluyvera* and *Roultella* (γ−Proteobacteria) in their guts as nitrogen-fixing associated bacteria (van Borm et al. 2002; Behar et al. 2005; Pinto-Tomás et al. 2009; Morales-Jiménez et al. 2009, 2013).

On the other hand, nitrogen recycling of uric acid (UA) by fungi seems to occur in Malpighian tubules of *D. coccus* and *D. opuntiae* (Vera-Ponce de León et al. 2016). UA and other nitrogen waste molecules may be used by bacteria for amino acid biosynthesis that allows nitrogen to be recycled in insects (Hirayama et al. 1999). The cactus sap on which the cochineal insects feed is nitrogen poor (Stintzing and Carle, 2005), and therefore it is possible that diazotrophic and UA recycling bacteria are associated with *Dactylopius* to compensate for the nutritional deficiency in nitrogen. Indeed, the genome of the *D. carminicum* endosymbiont that we present here contains all the genomic machinery needed for nitrogen fixation. This and other metabolic features of the novel symbiont will be discussed.

MATERIALS AND METHODS

For Peer Review

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

96 **Genome sequencing and annotation**

97 Extraction and sequencing of the metagenomic DNA from *D. coccus* as well as assembly
98 and binning were previously described (Ramírez-Puebla et al. 2016b). The RAST server
99 was used for gene prediction and annotation (Meyer et al. 2008). Manual curation of
100 relevant genes was performed after comparisons with sequences deposited in the following
101 databases: nr and Refseq via BLASTX (Camacho et al. 2009), the Conserved Domain
102 Database at GenBank (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), the Protein
103 families (PFAM) database (Finn et al. 2016) and the Transport Classification Database
104 (Saier et al. 2016). Additionally, a pathway profiling of metabolic annotation was obtained
105 from all putative coding genes predicted using GhostKoala tool from KEGG (Kanehisa et
106 al. 2016). Manual annotation was further refined using SmartBLAST of the deduced
107 proteins and their hydropathy profile analyzed with the online service of Swissprot TMpred
108 (http://www.ch.embnet.org/cgi-bin/TMPRED_form_parser) as previously described (Degli
109 Esposti et al. 2014). Genome completeness was assessed by the presence of single-copy
110 widespread orthologs with BUSCO (Simão et al. 2015). Prophage sequences were
111 identified, annotated and visualized by PAHAST and PHASTER tools (Zhou et al. 2011;
112 Arndt et al. 2016). The *D. carminicum* genome sequence has been deposited at
113 DDBJ/ENA/GenBank under the accession MQNN00000000. The version described in this
114 paper is version MQNN01000000.

115 **Phylogeny, comparative phylogenomics and phylogenetic reconstruction of nitrogen**
116 **fixation genes**

117 16S rRNA sequences were identified from genome contigs using Metaxa2 pipeline
118 (Bengtsson-Palme et al. 2015). Sequences were aligned using ClustalX 2.0 (Larkin et al.
119 2007) with other 16S rRNA sequences from GenBank. A phylogenetic tree was generated

For Peer Review

1
2
3 120 from a maximum likelihood analysis using PhyML 3.1 (Guindon et al. 2010). JModelTest
4
5 121 2.1.10 (Posada, 2008) was used to select appropriate models of sequence evolution by the
6
7 122 AIC. Model GTR+I+G (α =0.2810 for gamma distribution; A = 0.2448; C = 0.2275; G =
8
9 123 0.3209; T=0.2069, p-inv=0.4320) was the best model for the 16S rRNA genes.
10
11 124 For a comparative phylogenomic analysis of *D. carminicum*, 32 genomes of the
12
13 125 Rhodocyclaceae family were collected from GenBank. Open reading frames (ORF) and
14
15 126 coding sequences of all genomes were retrieved using GeneMark predictor (Besemer and
16
17 127 Borodovsky, 2005). The pangenome and core genome of all strains were obtained by
18
19 128 GETHOMOLOGUES version 2.0 software (Contreras-Moreira and Vinuesa, 2013) with -A
20
21 129 -c -t 0 -M -n 35 and -A -c -t 0 -G -n 35 parameters. A set of 293 single copy orthologous
22
23 130 genes from the core genome-matrix were aligned
24
25 131 A set of 293 single copy orthologous genes from the core genome-matrix was
26
27 132 aligned with Clustal Omega version 1.2.1 (Sievers et al. 2011) and Prottest3 version 3.4.2
28
29 133 (Darriba et al. 2011) was used to select the best amino acid substitution model. A maximum
30
31 134 likelihood phylogeny was constructed using PhyML 3.1 (Guindon et al. 2010) with the
32
33 135 LG+I+G+F model (p-inv=0.261, α=0.786 for gamma distribution) and the Shimodaira–
34
35 136 Hasegawa like procedure for internal branch supporting (Shimodaira and Hasegawa, 1999).
36
37 137 Additionally, for comparison with other β−Proteobacteria reported as insect
38
39 138 endosymbionts, a set of eight genomes were retrieved from NCBI-Genome database, core
40
41 139 and pangenome matrix and phylogeny from a set of eight single copy orthologous genes
42
43 140 were obtained as mentioned above using the LG+I+G model (p-inv= 0.09; α= 0.767 for
44
45 141 gamma distribution). To show the syntenic blocks of genes between *D. carminicum* and
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

142 other Rhodocyclaceae, eight genomes were aligned with Nucmer and plotted with Mummer
143 plot (Delcher et al. 2002).

144 To generate a *nifHDK* phylogeny, *D. carminicum* CDS predicted and annotated as
145 nitrogenase-iron protein, nitrogenase molybdenum-iron protein alpha chain and nitrogenase
146 molybdenum-iron protein subunit beta were aligned using Clustal Omega (Sievers et al.
147 2011) with amino acid sequences of the corresponding genes from other bacteria deposited
148 in GenBank. Final alignment concatenated matrix of the three genes was generated in
149 BioEdit version 7.2.5 (Hall, 1999). Maximum likelihood phylogenetic tree was obtained as
150 above using LG+I+G+F model (α=0.836 for gamma distribution; p-inv=0.225).
151 Additionally, to extend the metagenomic results we analyzed the presence of nitrogenase
152 genes in two available *Dactylopius* metagenomes from the whole body (here after called
153 DCoax and DCperu metagenome) deposited in GenBank under BioProject PRJNA244295
154 (Campana et al. 2015). For this, all reads greater than 100 bp were filtered from GenBank
155 fastq records, amino acid translations of these sequences were obtained using
156 FragGeneScan v 1.20 (Rho et al. 2010). BLASTp searches were performed against all
157 nitrogenases annotated in *D. carminicum*. Blast results were parsed using custom scripts.
158 *nifH* and *nifD* gene phylogenies were obtained as mentioned above.

159 **Nitrogen fixation**

160 Acetylene reduction assays were performed from different tissues of *D. coccus*. For organ
161 dissection, five adult females (~90 days old) were detached from the host plant, and rinsed
162 in 100 % ethanol for wax removal. Insects were washed twice with deionized water and
163 placed in PBS. One µl of hemolymph from each individual was obtained by making a fine
164 puncture in the third segment of the insect, fluid was collected with a micropipette and
165 placed in 10 ml vessels with 5 ml of semisolid nitrogen free medium (NFM) (KH_2P0_4 1.1 g

For Peer Review

166 l^{-1}; K_2HPO_4 2.2 g l^{-1}; $MgSO_4$ g l^{-1}; NaCl 0.1 g l^{-1} ; $CaCl_2$ g l^{-1}.; $FeSO_4$ g l^{-1} 0.05; Na_2MoO_4
167 0.002 g l^{-1}; agar 3 g l^{-1}; fructose 0.25 g l^{-1} sucrose 0.15 g l^{-1} and trehalose 0.1 g l^{-1} and 1.0
168 ml of growth factor solution: 10 mg of biotin and 20 mg of pyridoxin hydrochloride in 100
169 ml of distilled water, sterilized separately). Gut and ovaries from the same individual were
170 dissected as described (Vera-Ponce de León et al. 2016) and were submerged in the media.
171 Samples were incubated at 28 °C for 48 h and acetylene-enriched atmospheres were
172 injected to achieve 10% final concentration. Media with autoclaved organs (121 °C 15 min
173 15 psi) were used as negative controls, and vials inoculated with *Klebsiella variicola* 6A3
174 were used as positive controls. The ethylene formed by nitrogenase was measured as
175 described by Morales-Jiménez et al. (2013). Differences between the ARA activity of
176 negative controls and samples were analyzed by one-way ANOVA and a Tukey-HSD post-
177 hoc test for pairwise comparisons between samples.

178 ***D. carminicum nifH* gene expression in different tissues of *D. coccus*.**

179 For RNA extraction, 10 adult females (~90 days old) of *D. coccus* were externally
180 disinfected with aqueous 70 % ethanol (v/v). Hemolymph, guts and ovaries were dissected
181 as described (Vera-Ponce de León et al. 2016) and placed separately in 200 µl of RNAlater
182 ® Solution (Life Technologies, Carlsbad, CA, USA). Samples were centrifuged 1 min at
183 9300 g^{-1}. RNAlater was removed and RNA was isolated with TRIzol reagent (Invitrogen)
184 following the manufacture's instructions and cleaned with RNeasy column kit (Qiagen,
185 Germantown, MD, USA) adding DNAse before kit column treatment. RNA purity and
186 integrity was checked by electrophoresis and the absorbance ratios (260/280 and 260/230)
187 were measured. cDNA was synthesized for RT-PCR from purified RNA using the High
188 Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, California
189 USA) following the manufacturer's instructions. cDNA was used for PCR amplification

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

190 with 0.8 pmol of primers NiFHBeta-9 (5'AAC GTC AAT GCG CAA TTT ACG-3') and
191 NifHBeta-805 (5'ATG ATG CCG AAC TCC ATC AG-3') (this study). The final
192 concentration for 50 µl PCR reaction was 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 0.5 U Taq
193 polymerase and 1x Taq polymerase buffer (Invitrogen Life Technologies, Sao Paulo,
194 Brazil). The cycling conditions were 94 ºC for 5 min; 35 cycles of 60 s at 94 ºC, 60 s at 58
195 ºC, and 90 s at 72 ºC and final extension at 72 ºC for 10 min. PCR products were purified
196 using the High Pure PCR Product Purification Kit (Roche) and sequenced by Macrogen
197 Inc. (Seoul, Korea). Concentrated RNAs were used as negative controls and genomic DNA
198 of *D. coccus* was used as positive control.

199 **Fluorescent *in situ* hybridization (FISH)**

200 For localization of *D. carminicum* in the host, FISH was performed as described by Koga et
201 al. (2009) with modifications. Adult *D. coccus* females were detached from the host plant
202 and guts and ovaries were dissected as described (Vera-Ponce de León et al. 2016). Insect
203 tissues were fixed with Carnoy's solution and autofluorescence was quenched with 6%
204 H_20_2 for seven days. The oligonucleotide fluorescent probe used was Cy5-END1-1081 (5′-
205 Cy5-CTT GCG TAG CAA CTA ATG ATA AGG -3′) (this study) targeted to the 16S
206 rRNA gene of *D. carminicum*. FISH was performed adding the probe directly to dissected
207 guts and ovaries; slides were incubated overnight at 28 °C. After washing with PBS, the
208 samples were stained with 24 µg l^{-1} of DAPI and mounted with citifluor antifade solution
209 (Ted Pella, Inc.). Control experiments were performed without probe and with RNAse
210 digestion. The samples were observed under an Olympus FV100 Multi-photonic confocal
211 microscope. Images were processed using FIJI 2.0.0 software (Schindelin et al. 2012).

212

213

For Peer Review

214 **RESULTS**

215 **Nitrogen fixation in carmine cochineals**

216 From the cochineal metagenome analysis, we detected nitrogenase genes that were
217 phylogenetically related to those of β−Proteobacteria; they were assembled into four
218 contigs in the *D. carminicum* genome (Figure 1). We found genes encoding Mo-Fe
219 nitrogenase (*nifD-nifK*), nitrogenase reductase (*nifH*), nitrogen regulatory proteins and
220 transporters specific for nitrogenase cofactors (*nifQ-nifA* and 4Fe-4S ferredoxin) *nifHDK*
221 genes of *D. carminicum* clustered with those of *Uliginosibacterium gangwonense* (Figure
222 2), in agreement with the phylogenies obtained using 16S rRNA and 293 concatenated
223 orthologous genes (see later). Analysis of DCoax and DCperu metagenomes showed the
224 presence of *D. carminicum nifD* and *nifH* genes in these samples (Supplementary figure 1).
225 No other *nif* gene sequences were found in neither our samples nor in DCoax-DCperu
226 metagenomes.

227 Acetylene reduction activity, the standard assay for nitrogen fixation, was detected
228 in the hemolymph and ovaries (38 ± 5.98 and 11.58 ± 1.38 nmol ethylene h^{-1} tissue g^{-1},
229 respectively) of *D. coccus*. Post-hoc Tukey test showed significant differences in
230 nitrogenase activities between hemolymph and ovaries (P=0.0000979). However, no
231 significant nitrogenase activity was detected in sterile organs used as negative controls
232 (Figure 3). Additionally, RT-PCR showed the expression of the *nifH* gene in *D.*
233 *carminicum* hemolymph, ovaries and eggs, but not in guts (Supplementary Figure 2).

234 Xylose (a sugar derived from plant hemicellulose) and trehalose (the insect
235 hemolymph sugar), but not glucose, best sustained nitrogen fixation as carbon sources
236 added to the soft agar for the hemolymph acetylene reduction assays (data not shown).

237 **Genome characteristics of *D. carminicum***

For Peer Review

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

238 A bin composed of 122 contigs summing ca. 3.6 Mb was obtained from the metagenome
239 sequence of the cochineal insect *D. coccus* (Table 1). BLAST searches against the nr
240 database revealed that those contigs have β−Proteobacterial sequences as their closest
241 matches. All ribosomal genes were contained in a single contig whose 16S rRNA gene was
242 99-100% identical to our previously reported sequences from the β−Proteobacterium *D.*
243 *carminicum* (Supplementary Figure 3) (Ramírez-Puebla et al. 2010). The recovered contigs
244 represented 97% of the genome as judged by the program BUSCO (Simão et al. 2015). The
245 genome encoded 3633 CDS genes of which 79% could be assigned a putative function
246 (Table 1 and Table 2).

247 **Phylogenomics**

248 Phylogenetic analysis of 293 concatenated orthologous genes showed that *D. carminicum* is
249 closely related to *U. gangwonense* (Weon et al. 2008) (Figure 4). Both species belong to the
250 order Rhodocyclales and have *Azoarcus* and *Thauera* as close genera. *U. gangwonense* has
251 a large plasmid that was not found in the *D. carminicum* genome.

252 A phylogeny of eight orthologous genes from the core genome showed that *D.*
253 *carminicum* was not related to other β−proteobacterial endosymbionts of insects but rather
254 to free-living Rhodocyclacles (Supplementary figure 4). However, synteny was not
255 conserved between the genome of *D. carminicum* and that of other Rhodocyclaceae
256 (Supplementary figure 5).

257 **Metabolic capabilities**

258 The genome of *D. carminicum* encoded complete metabolic pathways for the biosynthesis
259 of biotin, thiamine, ubiquinone, cobalamine, heme, coenzyme B12, riboflavin, pyridoxin,
260 NAD(P), folate, lipoic acid and coenzyme A (Figure 5). This bacterium also possessed

For Peer Review

1
2
3 261 complete pathways for the biosynthesis of most amino acids, with the sole exception of
4
5 262 asparagine (Figure 5). However, the genome showed the ability to produce charged
6
7 263 asparaginyl-tRNA from aspartyl-tRNA because it codes the enzyme Aspartyl-tRNA(Asn)
8
9 264 synthetase. Furthermore, purine and pyrimidine nucleotide biosynthesis pathways were also
10
11 265 present in the genome. Key enzymes for glycolysis and gluconeogenesis appeared to be
12
13 266 absent but the pentose phosphate and Entner-Doudoroff pathways, which are fully
14
15 267 represented in *D. carminicum* genome, can provide alternative routes for carbon
16
17 268 metabolism (Figure 5). All genes of the tricarboxylic acid cycle were found.
18
19 269 **Anaerobic metabolism in *D. carminicum***
20
21 270 Pyruvate formate-lyase and lactate dehydrogenase were present in *D. carminicum*
22
23 271 genome, suggesting that the endosymbiont could perform fermentation reactions (Figure 5).
24
25 272 Indeed, the *D. carminicum* genome codes for various enzymes associated with anaerobic
26
27 273 metabolism. These include genes for nitrite respiration to nitrous oxide (Figure 5), as well
28
29 274 as two terminal oxidases which are active under very low concentrations of oxygen: the bd-
30
31 275 type ubiquinol oxidase and the cbb3-type cytochrome *c* oxidase (Degli Esposti et al. 2014).
32
33 276 Both of these oxidases are over-expressed in proteobacteria that fix N_2 in symbiosis with
34
35 277 plants (Degli Esposti and Martinez Romero, 2016). Conversely, only a few, non-catalytic
36
37 278 genes for the subunits of COX operons encoding cytochrome *c* oxidase (Degli Esposti and
38
39 279 Martinez Romero, 2016) were found in the genome of *D. carminicum*, thereby producing
40
41 280 an alternative respiratory chain equivalent to that of the Rickettsiales parasite *Midichloria*
42
43 281 *mitochondrii* (Degli Esposti et al, 2014). Intriguingly, the same respiratory chain is present
44
45 282 in the close free-living relative *U. gangwonense*, but not in *Azoarcus* and other members of
46
47 283 the family Rhodocyclaceae (results not shown). The genome of *D. carminicum* also
48
49 284 includes the entire operon of fumarate reductase, an enzyme that recycles the reduced
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

285 membrane quinones that are produced by complex I and other dehydrogenases and is
286 typically present in facultatively anaerobic bacteria such as *Escherichia coli*. The operon
287 for this enzyme appears to be present in other Rhodocyclaceae with versatile metabolic
288 capacity such as *Azoarcus*. Fumarate reductase normally uses low potential quinones such
289 as menaquinone, and several rhodocyclales have been found to have menaquinone or to
290 possess enzymes for its biosynthesis. Detailed analysis of membrane and unknown proteins
291 encoded in the genome of *D. carminicum* has indicated that key enzymes for menaquinone
292 biosynthesis are apparently missing, therefore suggesting the possibility that this
293 endosymbiont may produce rhodoquinone under micro-oxic and anaerobic conditions to
294 recycle electrons via fumarate reductase as in metazoans adapted to anaerobiosis, e.g.
295 *Ascaris* (Müller et al. 2012). The enzymes required for rhodoquinone biosynthesis are not
296 yet known (Müller et al, 2012), but we note that at least one member of the Rhodocyclaceae
297 family, *Zooglea*, has been reported to have this quinone (Hiraishi et al. 1992).

298 **Cell surface**

299 Inspection of genes related to cell envelope biosynthesis indicated that *D. carminicum*
300 could produce membranes with phosphatidylglycerol, phosphatidylserine,
301 phosphatidylethanolamine and cardiolipine but not phosphatidylcholine. Additionally,
302 genes for synthesis of lipopolysaccharide including the core and O-antigen, a normal
303 peptidoglycan layer and some form of exo- or capsular polysaccharide were found in its
304 genome. *D. carminicum* was revealed as a potentially motile bacteria possessing two sets of
305 genes for the flagellar machinery (Figure 5, Supplementary Figure 6) and one gene set for
306 chemotaxis. Additionally, genes for type IV pilus biogenesis were found.

307 **Secretion systems**

For Peer Review

D. carminicum also possessed tra/trb genes for the mating-pair formation apparatus of a type IV secretion system (T4SS) that is most related to those found in Integrative Conjugative Elements (ICEs) inserted in the chromosomes of β- and γ-proteobacteria. ICEs are composed of modules required for integration, excision, stabilization and conjugative transfer, as well as modules that differ between ICEs and genes that encode for various accessory functions like symbiosis, antibiotic resistance, and xenobiotic degradation (Burrus and Waldor, 2004). Besides the T4SS conjugative module, we found the integration module but not the excision module or stabilization modules. Due to the draft assembly, only a gene for a TonB-dependent siderophore receptor could be associated with an accessory function encoded within the ICE. The bacterium also possessed genes for type I and II secretion systems.

Prophage

Analysis by PHASTER identified at least five prophage regions (contigs) in the *D. carminicum* genome (Supplementary Figure 7a). Two of them were intact phage sequences (Supplementary Figure 7b). Prophages showed >90 % similarity to *Burkholderia* phage BcepMu and *Xylella* phage Xfas53. Other incomplete prophages were similar to *Bacillus* phage G, *Burkholderia* phage KL1 and *Pseudomonas* phage LPB1.

Nitrogen recycling

UA is the main waste compound of nitrogen catabolism in *D. coccus* (Vera-Ponce de León et al. 2016). The genome of *D. carminicum* has genes for nitrogen recycling of purine waste products and UA degradation to urea. All genes for urea metabolism and transport of its metabolites were predicted and annotated (Figure 5).

Localization of *D. carminicum* inside the insect

For Peer Review

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

331 Fluorescent *in situ* hybridization (FISH) with specific probes targeted to *D. carminicum*
332 showed its presence in ovaries and embryos of both *D. coccus* and *D. opuntiae* samples
333 (Figure 6 and Supplementary figure 8). More experiments are needed to define if *D.*
334 *carminicum* is intracellular as well.
335

336 **DISCUSSION**

337 *D. carminicum* was considered as the primary *Dactylopius* symbiont as it was not only
338 found in the domesticated *D. coccus* but in other wild *Dactylopius* species such as *D.*
339 *ceylonicus, D. confusus, D. opuntiae* and *D. tomentosus*, as well as in *D. cocus* eggs,
340 suggesting maternal inheritance (Ramirez-Puebla et al. 2010). In agreement, FISH
341 experiments showed the presence of *D. carminicum* in ovaries of *D. coccus* and *D.*
342 *opuntiae*, as well as in embryos of *D. coccus* (Figure 6, Supplementary figure 8).
343 The genome of the betaproteobacterium obtained from the *D. coccus* metagenome had 16S
344 rRNA gene sequences identical to those reported previously for *D. carminicum* (Ramírez-
345 Puebla et al. 2010) (supplementary Figure 3). Thus, we presumably have recovered its
346 genome, which unexpectedly turned out to be larger than those of other endosymbionts
347 (ranging from 0.1-2Mb), but was similar in size to those of recently acquired symbionts
348 such as *Sodalis* or *Arsenophonus* (Gil et al. 2008; Oakeson et al. 2014).

349 In most studied families of scale insects (Coccoidea), which feed on plant sap,
350 namely Monophlebidae, Ortheziidae, Coelostomidiidae, Coccidae, Diaspididae,
351 Lecanodiaspididae, Pseudococcidae and some Eriococcidae, the primary endosymbiont is a
352 flavobacterium, in most cases in co-symbiosis with a γ−Proteobacteria (Dhami et al. 2012;
353 Gruwell et al. 2012; Rosenblueth et al. 2012). We suggest that *D. carminicum* substituted

For Peer Review

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

354 the flavobacterium and γ−Proteobacteria symbionts conferring the beneficial role of
355 nitrogen fixation in the carmine cochineal insect (Figure 1, Figure 5). The flavobacterial
356 and γ−Proteobacteria endosymbionts commonly provide essential amino acids to other
357 scale insects, so *D. carminicum* should also fulfill this role. Ammonium assimilation has
358 been described in other diazotrophic symbionts such as *Candidatus* Azobacteroides
359 pseudotrichonymphae from termites (Hongoh et al. 2008; Desai and Brune, 2012). In
360 contrast, other nitrogen-fixing symbionts like rhizobia normally excrete fixed nitrogen as
361 ammonium and not as essential amino acids into to the host plants, while the enzymes for
362 their nitrogen assimilation are repressed (Prell and Poole, 2006). *nif* genes of *D.*
363 *carminicum* are closely related to those of *Azoarcus* strains (Figure 2), which are very
364 efficient nitrogen-fixing endophytes of rice and grass plants (Reinhold-Hurek and Hurek,
365 2006).

366 In the rice weevil, in addition to *Sodalis pierantonius* SOPE (γ−Proteobacteria), *Wolbachia*
367 endosymbionts are also found (Heddi et al. 1999). Thus, it is not unusual to find a diversity
368 of symbiont species in scale insects, as in the case of *D. coccus*, which contains *Wolbachi*a
369 (Ramírez-Puebla et al. 2015, 2016a), *Spiroplasma* (unpublished results), fungi (Vera-Ponce
370 de León et al. 2016) and *D. carminicum*.

371 Endosymbionts belonging to β−Proteobacteria have been found in mealybugs (López-
372 Madrigal et al. 2011), spittlebugs (McCutcheon and Moran, 2012), leafhoppers (Bennett et
373 al. 2016) and in the crypts of stinkbugs (Kikuchi et al. 2012), as well as in fungi (Lackner et
374 al. 2009). None of them are closely related to *D. carminicum* (Supplementary Figure 4).

375 *D. carminicum* has a smaller genome than its closest free-living relative
376 *Uliginosibacterium*, which is ~ 5Mb in size (Table 2). Similarly, in acetic acid bacteria, the

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

377 genome of the *Commensalibacter* symbiont present in the guts of Monarch butterfly shows
378 an AT bias and a slightly smaller size (2.3 Mb) than those of closely related free-living
379 bacteria, indicating a host adaptation process (Servín-Garcidueñas et al. 2014). While
380 *Commensalibacter* could be cultured in special laboratory media, we were unable to do the
381 same with *D. carminicum*. New efforts for culturing *D. carminicum* will be guided by a
382 deep analysis of its genome.

383 The finding of enzymes normally associated with anaerobic metabolism such as lactate
384 dehydrogenase (Figure 5) and fumarate reductase already suggests that low oxygen
385 conditions may facilitate the culture of *D. carminicum*. Nonoxygenic respiration may
386 represent an evolutionary advantage for endosymbionts of metazoans. Endosymbionts of
387 marine bivalves, for instance, use nitrate respiration to avoid oxygen competition with the
388 host (Hentschel and Felbeck, 1995). Moreover, *Sodalis* TME1, a γ−Proteobacterial
389 symbiont of the scale insect *Llaveia axin axin*, also possesses the enzymes for nitrate and
390 nitrite respiration (Rosas-Pérez, in press). Partial nitrite respiration may also occur in the
391 *Dactylopius-D. carminicum* symbiosis, but specific experiments are needed to corroborate
392 this metabolic trait. The conditions needed for nitrogen fixation may be provided by bd-
393 type and cbb3-type terminal oxidases that act as oxygen scavengers. Nitrogenase activity is
394 sensitive to even low concentrations of oxygen. Perhaps the endosymbionts thrive in
395 environments locally deprived of oxygen within host cells or extracellular tissues; once
396 established in such environments, they up-regulate oxygen scavenging systems to facilitate
397 nitrogen fixation. Given the nutritional requirements that the symbiont satisfies for the host,
398 the above possibility would suggest a host-derived manipulation of *D. carminicum*, similar
399 to that occurring in nitrogen-fixing symbionts within the nodules of leguminous plants
400 (Degli Esposti and Martinez Romero, 2016).

For Peer Review

While motility is a rare trait in most insect endosymbionts, *D. carminicum* had all the machinery for complete assembly of flagella and the corresponding genes were found expressed in the cochineal metatranscriptome (Figure 6, Supplementary figure 6). In the β-Proteobacteria symbiont of earthworms *Verminephrobacter eiseniae*, the flagella and type IV pili are also essential for host embryo colonization (Dulla et al. 2012). Likewise, *Vibrio fischeri* requires bacterial flagellar motility for establishing symbiosis in the light organ of the squid *Euprymna scolopes* (Millikan and Ruby, 2003; Brennan et al. 2013) and *Sodalis glossinidius* flagella may be used to colonize the tsetse fly progenie (Wang et al 2013). Hence, we surmise that flagellar motility may be used by *D. carminicum* to colonize *Dactylopius* tissues or offspring.

The *D. carminicum* genome encodes genes for uric acid, allantoin and urea utilization and transport (Figure 5). Thus, the urea cycle and its metabolites may serve as additional nitrogen sources for the host, besides the metabolic input derived from direct nitrogen fixation.

ACKNOWLEDGEMENTS

To CONACyT Basic Science grant 154453. AVPL was a Ph. D. student in the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) with fellowship number 331625 from Consejo Nacional de Ciencia y Tecnología (CONACyT). To José Luis Marquina for providing *D. coccus*. All bioinformatics were performed in CCG-UNAM servers.

REFERENCES

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

423 Adams AS, Currie CR, Cardoza Y, Klepzig KD, Raffa KF. 2009. Effects of symbiotic
424 bacteria and tree chemistry on the growth and reproduction of bark beetle fungal symbionts.
425 1147:1133–1147.
426 Arndt D et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool.
427 Nucleic Acids Res. 44:1–6.
428 Behar A, Yuval B, Jurkevitch E. 2005. Enterobacteria-mediated nitrogen fixation in natural
429 populations of the fruit fly *Ceratitis capitata*. Mol. Ecol. 14:2637–2643.
430 Behmer ST, Grebenok RJ, Douglas AE. 2011. Plant sterols and host plant suitability for a
431 phloem-feeding insect. Funct. Ecol. 25:484–491.
432 Bengtsson-Palme J et al. 2015. Metaxa2: improved identification and taxonomic
433 classification of small and large subunit rRNA in metagenomic data. Mol. Ecol. Resour.
434 15:1403–1414.
435 Bennett GM, Abbà S, Kube M, Marzachì C. 2016. Complete Genome Sequences of the
436 Obligate Symbionts '*Candidatus* Sulcia muelleri' and '*Ca*. Nasuia deltocephalinicola' from
437 the Pestiferous Leafhopper *Macrosteles quadripunctulatus* (Hemiptera: Cicadellidae).
438 Genome Announc. 4:e01604–15.
439 Besemer J, Borodovsky M. 2005. GeneMark: Web software for gene finding in
440 prokaryotes, eukaryotes and viruses. Nucleic Acids Res. 33:451–454.
441 Brennan CA, Mandel MJ, Gyllborg MC, Thomasgard KA, Ruby EG. 2013. Genetic
442 determinants of swimming motility in the squid light-organ symbiont *Vibrio fischeri*.
443 Microbiologyopen. 2:576–594.
444 Burrus V, Waldor MK. 2004. Shaping bacterial genomes with integrative and conjugative
445 elements. Res. Microbiol. 155:376–386.

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

446 Camacho C et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics.
447 10:421.
448 Campana MG, Robles García NM, Tuross N. 2015. America's red gold: multiple lineages
449 of cultivated cochineal in Mexico. Ecol. Evol. 5:607–617.
450 Contreras-Moreira B, Vinuesa P. 2013. GET_HOMOLOGUES, a versatile software
451 package for scalable and robust microbial pangenome analysis. Appl. Environ. Microbiol.
452 79:7696–7701.
453 Dale C, Beeton M, Harbison C, Jones T, Pontes M,. 2006. Isolation , Pure Culture , and
454 Characterization of " *Candidatus* Arsenophonus arthropodicus ," an Intracellular Secondary
455 Endosymbiont from the Hippoboscid Louse Fly *Pseudolynchia canariensis* 72:2997–3004.
456 Dale C, Maudlin I. 1999. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a
457 microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*.
458 Culture. 267–275.
459 Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit
460 models of protein evolution. Bioinformatics. 27:1164–1165.
461 Degli Esposti M et al. 2014. Evolution of mitochondria reconstructed from the energy
462 metabolism of living bacteria. PLoS One. 9:e96566.
463 Degli Esposti M, Martinez Romero E. 2016. A survey of the energy metabolism of
464 nodulating symbionts reveals a new form of respiratory complex I. FEMS Microbiol. Ecol.
465 92:fiw084.
466 Delcher AL, Phillippy A, Carlton J, Salzberg SL. 2002. Fast algorithms for large-scale
467 genome alignment and comparison. Nucleic Acids Res. 30:2478–2483.
468 Desai MS, Brune A. 2012. Bacteroidales ectosymbionts of gut flagellates shape the
469 nitrogen-fixing community in dry-wood termites. ISME J. 6:1302–1313.

For Peer Review

470 Dhami MK, Turner AP, Deines P, Beggs JR, Taylor MW. 2012. Ultrastructural and
471 molecular characterization of a bacterial symbiosis in the ecologically important scale
472 insect family Coelostomidiidae. FEMS Microbiol. Ecol. 81:537–546.
473 Douglas AE. 2012. 'Can't Live without You:' Essential animal-bacterial relationships.
474 Microbe Mag. 7:273–277.
475 Douglas AE. 2015. Multiorganismal insects: Diversity and function of resident
476 microorganisms. Annu. Rev. Entomol. 60:17–34.
477 Dulla GFJ, Go RA, Stahl DA, Davidson SK. 2012. *Verminephrobacter eiseniae* type IV
478 pili and flagella are required to colonize earthworm nephridia. ISME J. 6:1166–1175.
479 Engel P, Moran NA. 2013. The gut microbiota of insects - diversity in structure and
480 function. FEMS Microbiol. Rev. 37:699–735.
481 Finn RD et al. 2016. The Pfam protein families database: Towards a more sustainable
482 future. Nucleic Acids Res. 44:D279–D285.
483 Gil R, et al. 2008. Massive presence of insertion sequences in the genome of SOPE, the
484 primary endosymbiont of the rice weevil *Sitophilus oryzae*. Int. Microbiol. 11:41–48.
485 Gruwell ME, Morse GE, Normark BB. 2012. Phylogenetic congruence of armored scale
486 insects (Hemiptera: Diaspididae) and their primary endosymbionts from the phylum
487 Bacteroidetes. Mol. Phylogen. Evol. 44:267-280.
488 Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood
489 phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.
490 Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis
491 program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41:95–98.

For Peer Review

492 Heddi A, et al. 1999. Four intracellular genomes direct weevil biology: nuclear,
493 mitochondrial, principal endosymbiont, and *Wolbachia*. Proc. Natl. Acad. Sci. U. S. A.
494 96:6814–6819.

495 Hentschel U, Felbeck H. 1995. Nitrate respiration in chemoautotrophic symbionts of the
496 bivalve *Lucinoma aequizonata* is not regulated by oxygen. Appl. Environ. Microbiol.
497 61:1630–1633.

498 Hiraishi A, Shin YK, Sugiyama J, Komagata K. 1992. Isoprenoid quinones and fatty acids
499 of *Zoogloea*. Antonie Van Leeuwenhoek. 61:231–236.

500 Hirayama C, Sugimura M, Shinbo H. 1999. Recycling of urea associated with the host
501 plant urease in the silkworm larvae, *Bombyx mori*. J. Insect Physiol. 45:15–20.

502 Hongoh Y, et al. 2008. Genome of an endosymbiont coupling N_2 fixation to cellulolysis
503 within protist cells in termite gut. Science 322:1108–1109.

504 James EK, Olivares FL, Baldani JI, Döbereiner J. 1997. *Herbaspirillum*, an endophytic
505 diazotroph colonizing vascular tissue *Sorghum bicolor* L. Moench. J. Exp. Bot. 48:785–
506 798.

507 Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools
508 for Functional Characterization of Genome and Metagenome Sequences. J. Mol. Biol.
509 428:726–731.

510 Kikuchi Y, et al. 2012. Symbiont-mediated insecticide resistance. Proc. Natl. Acad. Sci. U.
511 S. A. 109:8618–8622.

512 Kim H, et al. 1998. High population of *Sphingomonas* species on plant surface. J. Appl.
513 Microbiol. 85:731–736.

514 Koga R, Tsuchida T, Fukatsu T. 2009. Quenching autofluorescence of insect tissues for in
515 situ detection of endosymbionts. Appl. Entomol. Zool. 44:281–291.

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

516 Lackner G, Partida-Martinez LP, Hertweck C. 2009. Endofungal bacteria as producers of
517 mycotoxins. Trends Microbiol. 17:570–576.
518 Larkin MA et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23:2947–
519 2948. doi:
520 López-López A, Negrete-Yankelevich S, Rogel MA, Ormeño-Orrillo E, Martínez J,
521 Martínez-Romero E. 2013. Native bradyrhizobia from Los Tuxtlas in Mexico are
522 symbionts of *Phaseolus lunatus* (Lima bean). Syst. Appl. Microbiol. 36:33–38.
523 López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R. 2011. Complete genome sequence
524 of '*Candidatus* Tremblaya princeps' strain PCVAL, an intriguing translational machine
525 below the living-cell status. J. Bacteriol. 193:5587–5588.
526 Martínez-Cano DJ et al. 2015. Evolution of small prokaryotic genomes. Front. Microbiol.
527 5:1–23
528 McCutcheon JP, Moran N. 2012. Extreme genome reduction in symbiotic bacteria. Nat.
529 Rev. Microbiol. 10:13–26.
530 Meyer F, et al. 2008. The metagenomics RAST server - a public resource for the automatic
531 phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 9:386.
532 Millikan DS, Ruby EG. 2003. FlrA, a σ^{54}-Dependent transcriptional activator in *Vibrio*
533 *fischeri*, is required for motility and symbiotic light-organ colonization. J. Bacteriol.
534 185:3547–3557.
535 Morales-Jiménez J, Zúñiga G, Villa-Tanaca L, Hernández-Rodríguez C. 2009. Bacterial
536 community and nitrogen fixation in the red turpentine beetle, Dendroctonus valens LeConte
537 (Coleoptera: Curculionidae: Scolytinae). Microb. Ecol. 58:879–891.
538 Morales-Jiménez J, Vera-Ponce de León A, García-Domínguez A, Martínez-Romero E,
539 Zúñiga G, Hernández-Rodríguez C. 2013. Nitrogen-fixing and uricolytic bacteria

For Peer Review

associated with the gut of *Dendroctonus rhizophagus* and *Dendroctonus valens* (Curculionidae: Scolytinae). Microb. Ecol. 66:200–210.

Müller, M., Mentel, M., van Hellemond, J. J., Henze, K., Woehle, C., Gould, S. B., Yu, R.-Y., van der Giezen, M., Tielens, A. G. M., and Martin, W. F. 2012. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiol. Mol. Biol. Rev. 76:444–495.

Nardi JB, Mackie RI, Dawson JO. 2002. Could microbial symbionts of arthropod guts contribute significantly to nitrogen fixation in terrestrial ecosystems? J. Insect Physiol. 48:751–763.

Oakeson KF, et al. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. Genome Biol. Evol. 6:76–93.

Ohkuma M, Noda S, Kudo T. 1999. Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites phylogenetic 65:4926–4934.

Perez-Guerra G, Kosztarab M. 1992. Biosystematics of the family Dactylopiidae (Homoptera: Coccinea) with emphasis on the life cycle of *Dactylopius coccus* Costa. Bull. Virginia Agric. Exp. Stn. 92:1–90.

Pinto-Tomás AA, Anderson MA, Suen G, Stevenson DM, Chu FST, Cleland WW, Weimer PJ, Currie CR. 2009. Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. Science 326:1120–1123.

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

Prell J, Poole P. 2006. Metabolic changes of rhizobia in legume nodules. Trends Microbiol. 14:161-168.

For Peer Review

563 Ramírez-Puebla ST, Ormeño-Orrillo E, Vera-Ponce de León A, Lozano L, Sanchez A,
564 Rosenblueth M, Martínez-Romero E. 2016a. Genomes of *Candidatus* Wolbachia Bourtzisii
565 *w*DacA and *Candidatus* Wolbachia Pipientis *w*DacB from the Cochineal Insect *Dactylopius*
566 *coccus* (Hemiptera: Dactylopiidae). G3. 6:3343-3349.
567 Ramírez-Puebla ST, Rosenblueth M, Chávez-Moreno CK, Catanho Pereira de Lyra MC,
568 Tecante A, Martínez-Romero E. 2010. Molecular phylogeny of the genus *Dactylopius*
569 (Hemiptera: Dactylopiidae) and identification of the symbiotic bacteria. Environ. Entomol.
570 39:1178–1183.
571 Ramírez-Puebla, S. T., Servín-Garcidueñas, L. E., Ormeño-Orrillo, E., Vera-Ponce de
572 León, A., Rosenblueth, M., Delaye, L., Martínez, J., and Martínez-Romero, E. 2016b. A
573 response to Lindsey et al. "*Wolbachia pipientis* should not be split into multiple species: A
574 response to Ramírez-Puebla et al.". Syst. Appl. Microbiol. 39:223–225.
575 Ramírez-Puebla ST, Servín-Garcidueñas LE, Ormeño-Orrillo E, Vera-Ponce de León A,
576 Rosenblueth M, Delaye L, Martínez J, and Martínez-Romero E. 2015. Species in
577 *Wolbachia*? Proposal for the designation of "*Candidatus* Wolbachia bourtzisii",
578 "*Candidatus* Wolbachia onchocercicola", "*Candidatus* Wolbachia blaxteri", "*Candidatus*
579 Wolbachia brugii", "*Candidatus* Wolbachia taylori", '*Candidatus* Wolbachia collembol.
580 Syst. Appl. Microbiol. 38, 390–9.
581 Reinhold-Hurek B, Hurek T. 2006. The Genera *Azoarcus*, *Azovibrio*, *Azospira* and
582 *Azonexus*. The Prokaryotes. 5:873–891.
583 Rho M, Tang H, Ye Y. 2010. FragGeneScan: predicting genes in short and error-prone
584 reads. Nucleic Acids Res. 38:e191.
585 Rosas-Pérez T, Rosenblueth M, Rincón-Rosales R, Mora J, Martínez-Romero E. 2014.
586 Genome sequence of '*Candidatus* Walczuchella monophlebidarum' the flavobacterial

For Peer Review

587 endosymbiont of *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae). Genome Biol.
588 Evol. 6:714–726.
589 Rosenblueth M, Sayavedra L, Sámano-Sánchez H, Roth A, Martínez-Romero E. 2012.
590 Evolutionary relationships of flavobacterial and enterobacterial endosymbionts with their
591 scale insect hosts (Hemiptera: Coccoidea). J. Evol. Biol. 25:2357–2368.
592 Saier MH, et al. 2016. The Transporter Classification Database (TCDB): Recent advances.
593 Nucleic Acids Res. 44:D372–D379.
594 Schindelin J, et al. 2012. Fiji: an open-source platform for biological-image analysis. Nat.
595 Methods. 9:676–682.
596 Servín-Garcidueñas LE, Sánchez-Quinto A, Martínez-Romero E. 2014. Draft genome
597 sequence of *Commensalibacter papalotli* MX01, a symbiont identified from the guts of
598 overwintering monarch butterflies. Genome Announc. 2:1–2.
599 Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of
600 the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature. 407:81–86.
601 Shimodaira H, Hasegawa M. 1999. Letter to the Editor Multiple Comparisons of Log-
602 Likelihoods with Applications to Phylogenetic Inference. Mol. Biol. Evol. 16:1114–1116.
603 Sievers F et al. 2011. Fast, scalable generation of high-quality protein multiple sequence
604 alignments using Clustal Omega. Mol. Syst. Biol. 7:539.
605 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO:
606 Assessing genome assembly and annotation completeness with single-copy orthologs.
607 Bioinformatics. 31:3210–3212.
608 Stintzing FC, Carle R. 2005. Cactus stems (*Opuntia* spp.): a review on their chemistry,
609 technology, and uses. Mol. Nutr. Food Res. 49:175–194.

For Peer Review

Tai V, Carpenter KJ, Weber PK, Nalepa CA, Perlman SJ, Keeling PJ. 2016. Genome evolution and nitrogen-fixation in bacterial ectosymbionts of a protist inhabiting wood-feeding cockroaches. Appl. Environ. Microbiol. doi:10.1128/AEM.00611-16.

van Borm S, Buschinger A, Boomsma JJ, Billen J. 2002. Tetraponera ants have gut symbionts related to nitrogen-fixing root-nodule bacteria. Proc. Biol. Sci. 269:2023–2027.

Vera-Ponce de León A, Sanchez-Flores A, Rosenblueth M, Martínez-Romero E. 2016. Fungal community associated with *Dactylopiu*s (Hemiptera: Coccoidea: Dactylopiidae) and its role in uric acid metabolism. Front. Microbiol. 7:1–15.

Wang J, Weiss BL, Aksoy S. 2013. Tsetse fly microbiota: form and function. Front. Cell. Infect. Microbiol. 3:69.

Weon HY, Kim BY, Yoo SH, Kwon SW, Go SJ, Stackebrandt E. 2008. *Uliginosibacterium gangwonense* gen. nov., sp. nov. isolated from a wetland, Yongneup, in Korea. Int. J. Syst. Evol. Microbiol. 58:131–135.

Williams DJ, Ben-Dov Y. 2015. Scale insect species names that have been combined with the genus name *Dactylopiu*s Costa (Hemiptera: Sternorrhyncha: Coccomorpha). Zootaxa. 4006:161–170.

Xie J, Butler S, Sanchez G, Mateos M. 2014. Male killing *Spiroplasm*a protects *Drosophila* melanogaster against two parasitoid wasps. Heredity112:399–408.

Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: A Fast Phage Search Tool. Nucleic Acids Res. 39:1–6.

For Peer Review

Figure legends

Figure 1. Gene structure of *D. carminicum* contigs carrying nitrogenase related CDS. Arrows show transcription direction. Scales show position in the assembly.

Figure 2. Maximum likelihood tree (−ln L= -16717.43208) of concatenated *nifHDK* genes of *D. carminicum* and other related sequences from GeneBank. Scale bar indicates 0.5% estimated sequence divergence. SH-like support values ≥ 50% are indicated.

Figure 3. Acetylene reduction activity detected in *D. coccus* tissues. Bar indicates median ± SD. One-Way ANOVA ($P < 0.01$; $F= 34.59$; d.f 3). Plots with different letters indicate significant mean differences (Tuckey-HDS test $P < 0.05$).

Figure 4. Maximum likelihood tree (−ln L=-1403085.372840.) of 293 orthologous of *D. carminicum* and other Rhodocyclaceae genomes. The genome of *Paraburkholderia rhizoxinica* was used as out-group. Scale bar indicates 10 % estimated sequence divergence. SH-like support values ≥ 50% are indicated.

Figure 5. Scheme of metabolic pathways predicted from the *D. carminicum* genome. Gray arrow indicate incomplete pathways. Dotted arrows indicate pathways where an imported or synthesized product is likely to be used.

Figure 6. Fluorescent *in situ* hybridization of *D. carminicum* in *Dactylopius coccus.* Left: ovaries; right: embryo. a) Treatment with probe targeted to *D. carminicum*. b) No-probe controls. Blue signals indicate insect nucleus and red indicates *D. carminicum* signal.

Supplementary

Supplementary figure 1. Maximun likelihood trees of a) *nif*H gene sequence of DC peru metagenomic related to *D. carminicum* and other bacteria from GenBank (FLU used as Aminoacidic substitution model; -ln L=-612.4); b) *nif*D gene sequence of DC_oax

For Peer Review

1
2
3 657 metagenomic related to *D. carminicum* and other bacteria from GenBank (WAG used as
4
5 658 Aminoacidic substitution model; -ln L=-228.2).
6
7 659 Supplementary figure 2. Amplicons of RT-PCR from *D. carminicum nif*H transcript. *D.*
8
9 660 *coccus* total DNA as Positive control and no-DNA template as Negative control.
10
11 661 Supplementary figure 3. Maximum likelihood tree (− ln L = -6164.71130) of 16S rRNA
12
13 662 sequences of *Dactylopiibacterium carminicum* and other bacteria from GenBank. The 16S
14
15 663 rRNA sequence of was used as outgroup. Scale bar indicates 5% estimated sequence
16
17 664 divergence. SH-like support values ≥ 50% are indicated
18
19 665 Supplementary figure 4. Maximum likelihood tree of 8 orthologous from *D. carminicum*
20
21 666 and other β-proteobacterial endosymbionts. The genome of *Paraburkholderia rhizoxinica*
22
23 667 was used as out-group. Scale bar indicates 10 % estimated sequence divergence. SH-like
24
25 668 support values ≥ 50% are indicated.
26
27 669 Supplementary figure 5. Genomic alignments *D. carminicum* (y-axis) and other members
28
29 670 of Rhodocyclaceae (x-axis). Forward matches shown in red. Reverse matches shown in
30
31 671 blue.
32
33 672 Supplementary figure 6. *D. carminicum* annotated genes for flagella assembly.
34
35 673 Supplementary figure 7. Prophages present in *D. coccus* metagenome a) Five regions
36
37 674 (contigs) identified as prophage by Blast. b) Intact prophage sequences of region 2 and 4
38
39 675 Supplementary figure 8. Localization of *D. carminicum* in *Dactylopius opuntiae* ovaries.
40
41 676 Blue signals indicate insect nucleus and red indicates *D. carminicum* signal.
42
43 677
44
45 678
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. Genome features of *Candidatus* Dactylopiibacterium carminicum from *Dactylopius coccus*

Feature	**NFE1**
Number of Contigs	122
N50 (bp)	50,206
Estimated genome size (bp)	3,589,384
G+C content (%)	62.74
CDS genes	3633
With function	2831
Hypothetical	802
rRNA	3
tRNA	46
Genome Completeness	97 %

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Table 2. Comparision of *Dactylopiibacterium carminicum* genome features with related β−Proteobacteria.

Bacterial genome	***Dactylopiibacterium carminicum***	*Uliginosibacterium ganwonese*	*Azoarcus* sp. BH72	*Ca.* Tremblaya princeps	*Ca.* Zinderia insecticola	*Paraburkholderia rhizoxinica*
Accession number	**MQNN00000000**	GCA_000373965	GCF_000061505	GCF_000219195	GCA_000147015	GCA_000198775
Size (bp)	**3,589,384**	5,021,710	3,908,237	138,927	208,564	3,750,139
G+C (%)	**62.74**	56.57	67.92	58.83	13.54	60.7
CDSs	**3,633**	4,717	3,992	140	202	3,878
Host	**Carmine cochineal (*Dactylopius* spp.)**	None	Kallar grass (*Leptochloa fusca*)	Mealybug (*Planococcus citri*)	Spittlebug (*Clastoptera arizonana*)	Zygomycete (*Rhizopus microsporus*)
Life style	**Insect endosymbiont**	Free living	Grass endophyte	Insect endosymbiont	Insect endosymbiont	Fungi endosymbiont
Diazotrophic	**Yes**	Yes	Yes	No	No	No

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

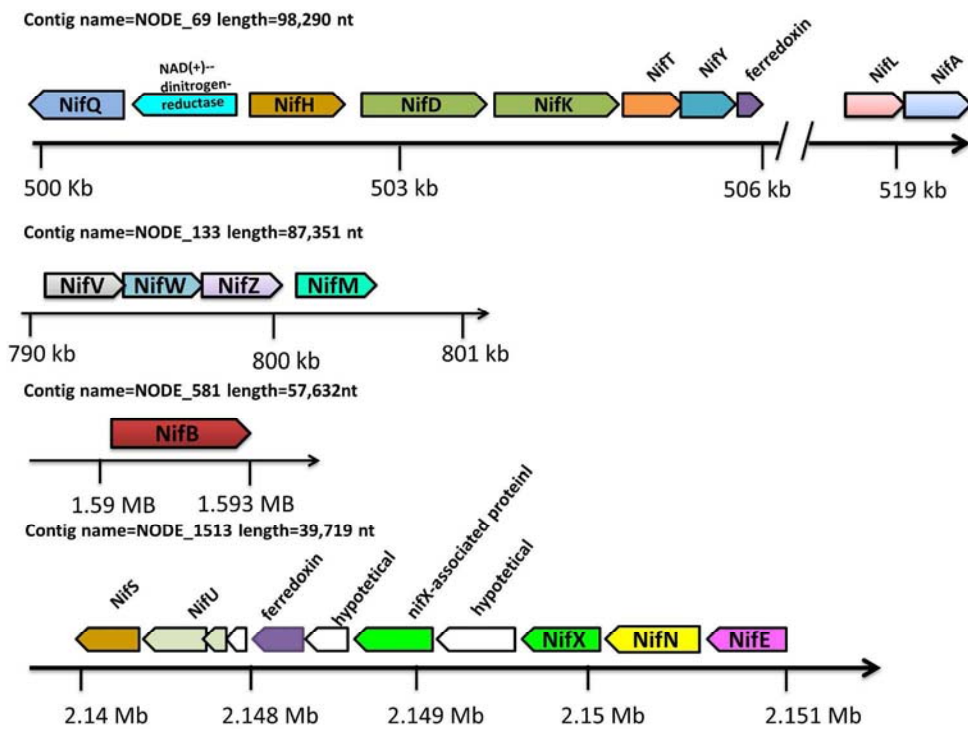


Figure 1. Gene structure of D. carminicum contigs carrying nitrogenase related CDS. Arrows show transcription direction. Scales show position in the assembly

254x190mm (300 x 300 DPI)

review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

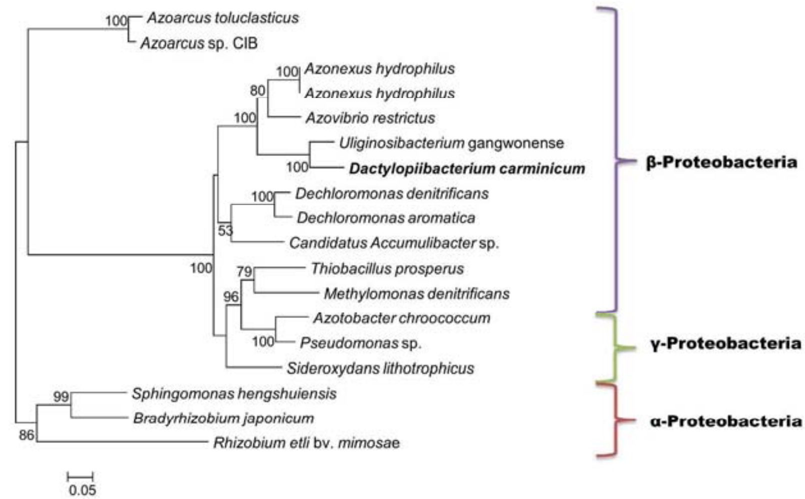


Figure 2. Maximum likelihood tree (−ln L= -16717.43208) of concatenated nifHDK genes of D. carminicum and other related sequences from GeneBank. Scale bar indicates 0.5% estimated sequence divergence. SH-like support values ≥ 50% are indicated

254x190mm (300 x 300 DPI)

view

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

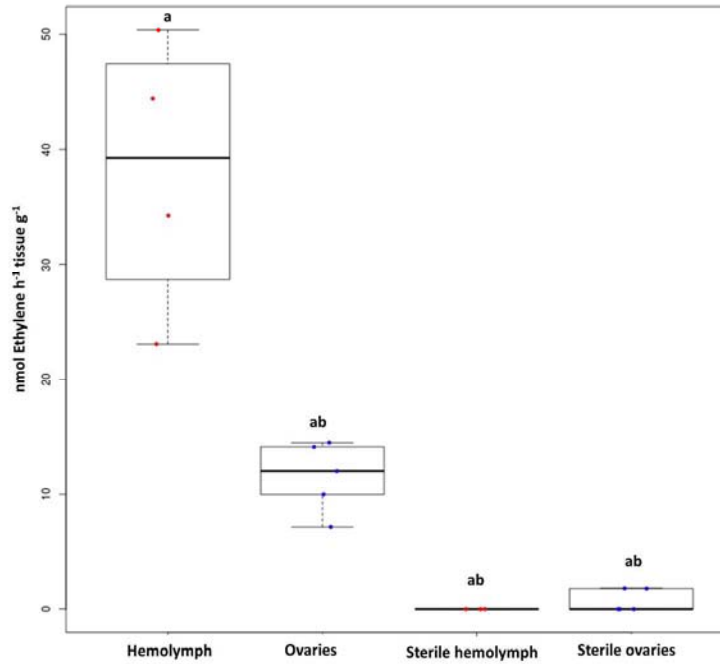


Figure 3. Acetylene reduction activity detected in D. coccus tissues. Bar indicates median ± SD. One-Way ANOVA (P <0.01; F= 34.59; d.f 3). Plots with different letters indicate significant mean differences (Tuckey-HDS test P < 0.05)

254x190mm (300 x 300 DPI)

view

Figure 4. Maximum likelihood tree (−ln L=-1403085.372840.) of 293 orthologous of D. carminicum and other Rhodocyclaceae genomes. The genome of Paraburkholderia rhizoxinica was used as out-group. Scale bar indicates 10 % estimated sequence divergence. SH-like support values ≥ 50% are indicated

260x253mm (299 x 299 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

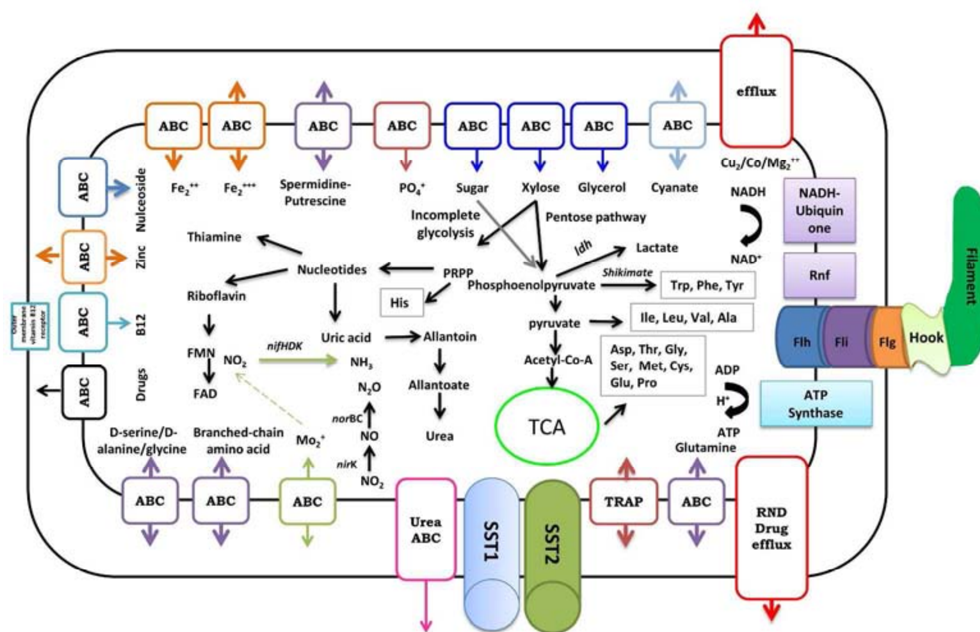


Figure 5. Scheme of metabolic pathways predicted from the D. carminicum genome. Gray arrow indicate incomplete pathways. Dotted arrows indicate pathways where an imported or synthesized product is likely to be used.

265x190mm (293 x 293 DPI)

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

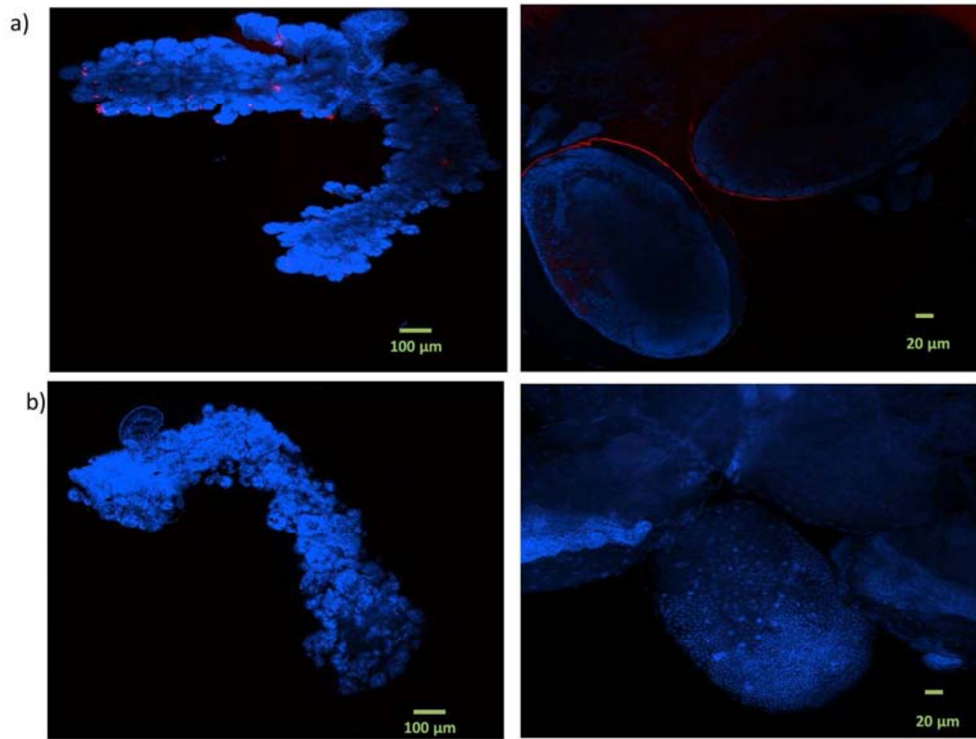


Figure 6. Fluorescent in situ hybridization of D. carminicum in Dactylopius coccus. Left: ovaries; right: embryo. a) Treatment with probe targeted to D. carminicum. b) No-probe controls. Blue signals indicate insect nucleus and red indicates D. carminicum signal

254x190mm (300 x 300 DPI)

view

Provisional chapter

The Symbiome of Llaveia Cochineals (Hemiptera: Coccoidea: Monophlebidae) Includes a Gammaproteobacterial Cosymbiont Sodalis TME1 and the Known *Candidatus* Walczuchella Monophlebidarum

Tania Rosas-Pérez, Arturo Vera-Ponce de León, Mónica Rosenblueth, Shamayim T. Ramírez-Puebla, Reiner Rincón-Rosales, Julio Martínez-Romero, Michael F. Dunn and Esperanza Martínez-Romero

Additional information is available at the end of the chapter

Provisional chapter

The Symbiome of Llaveia Cochineals (Hemiptera: Coccoidea: Monophlebidae) Includes a Gammaproteobacterial Cosymbiont Sodalis TME1 and the Known *Candidatus* Walczuchella Monophlebidarum

Tania Rosas-Pérez, Arturo Vera-Ponce de León, Mónica Rosenblueth, Shamayim T. Ramírez-Puebla, Reiner Rincón-Rosales, Julio Martínez-Romero, Michael F. Dunn and Esperanza Martínez-Romero

Additional information is available at the end of the chapter

Abstract

The genome and transcriptome of the endosymbiotic flavobacterium *Candidatus* Walczuchella monophlebidarum revealed its role in the synthesis of essential amino acids for its host, the wax cochineal *Llaveia axin axin*. There were, however, missing genes in the endosymbiont for some biosynthetic pathways. Here, we characterized TME1, another cochineal symbiont that may metabolically complement *Walczuchella*. TME1 was ascribed to the gammaproteobacterial genus *Sodalis* on a phylogenomic basis using core genome sequences and the core average nucleotide identity (ANI) confirmed its position. Additionally, we describe *Sodalis* as a coherent genus. TME1 genome is around 3.4 Mb and has complete gene sequences for the biosynthesis of 10 essential amino acids, for polyamines, flagella, nitrate respiration, and detoxification among many others. Transcripts from ovaries and bacteriomes allowed the identification of differentially transcribed genes from the endosymbionts and host. Highly transcribed genes were identified in TME1 and transcripts involved in amino acid biosynthesis were found. We review here that cosymbionts that derived from different bacterial classes and genera seem to be advantageous for insects that have Flavobacteria as the primary endosymbionts.

Keywords: endosymbionts, scale insect, gammaproteobacteria, *Sodalis*-like, Alphaproteobacteria, fungi

INTECH open science | open minds

© 2016 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTECH open science | open minds

© 2016 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

All organisms are inhabited by microbes that exert different effects on their hosts. In insects, there are many examples of beneficial associations with symbiotic microbes that have been linked to the insect ecological success. Symbionts that are vertically transmitted from mother to offspring and with an intrinsic interdependence with the insect host are considered as primary endosymbionts and they have reduced genomes [1, 2]; they do not grow on standard laboratory media. In theory, endosymbionts evolved from gut bacteria [3] that are largely more complex and may be determined by the diet and the environment. Primary endosymbionts may reside inside insect cells called bacteriocytes that may be found in specialized host structures called bacteriomes. Bacteriomes may be equivalent to plant-root nodules considering that they are host structures harboring particular bacterial species with specific roles [4]. But even in plants, cosymbionts have been encountered; for example, the slow-growing actinobacteria *Micromonospora* is found in nodules formed by *Bradyhizobium, Rhizobium*, or *Frankia* in several legumes or actinorhizal roots, although Micromonospora is unable to form nodules [5]. *Micromonospora* has been reported to enhance nodulation and promote plant growth, may enhance plant defense responses, or inhibit pathogens [6].

In insects, cosymbiosis is not uncommon and there are cases in which two symbionts are found in the bacteriome [7, 8]. Additionally, other microbes including fungi may be found in the hemolymph or in different insect tissues [9–11]. Fungal symbionts may be found as well in specialized insect structures known as mycangia [12] or inside insect cells called mycetocytes [13].

In insects, primary bacterial endosymbionts synthesize essential amino acids or vitamins for their hosts and reside intracellularly in bacteriomes. In some cases, complementation of metabolic pathways seems to occur among different insect symbionts [14–17]. Additionally, cosymbionts may have different roles, and some have been implicated in defense [18–21], tolerance to stress [22], resistance to high temperatures [23–25], to virus [26–28], or may manipulate sex differentiation [29]. There is an example in which a secondary endosymbiont substituted a lost primary *Buchnera* symbiont in an aphid [30]. Among others, gamma and betaproteobacteria have been found as cosymbionts; for example, the primary endosymbiont *Candidatus* Sulcia muelleri ("*Sulcia*" from here on) (phylum Bacteroidetes, class Flavobacteria) with a highly reduced genome has betaproteobacteria as cosymbionts that are found in green rice leafhoppers [7], stinkbugs [31], and spittlebugs [32, 33]. In leafhoppers, the symbionts occupy different types of bacteriocytes that constitute the outer or inner regions of the bacteriome [7]. The *Sulcia* cosymbionts are *Hodgkinia, Zinderia, Nasuia* [34, 35] with very small genomes, and the gammaproteobacteria *Baummania, Arsenophonus*, or *Sodalis*, the latter considered as a new acquisition. Surprisingly a gammaproteobacterium may be found inside *Sulcia* cells and be transmitted to the next generation [36].

Scale insects (Hemiptera: Coccoidea) feed on plant sap, which is a nutritionally poor diet that lacks most of the essential amino acids. Therefore, these insects have built up symbiotic associations with bacteria that can synthesize them. Most of the families that have been analyzed from scale insects (Hemiptera: Coccoidea) harbor flavobacteria as primary symbionts and enterobacteria as secondary symbionts (Monophlebidae, Coelostomidiidae, Orthezidae, Phenacoccinae from Pseudococcidae, Coccidae, Lecanodiaspididae, Diaspididae, and a clade

of Eriococcidae) [37–39]. It has been reported that the families from scale insects Dactylopiidae, some Eriococcidae, and Pseudococcinae from Pseudococcidae harbor different endosymbionts, which could indicate that they lost their flavobacteria and enterobacteria and acquired other endosymbionts [39]. Flavobacteria seem to be very ancient symbionts, perhaps starting symbiosis before the divergence of scale insects [39] (150–250 mya [40]). Although it has been suggested that Flavobacteria have cospeciated only within Monophlebidae, Coelostomidiidae, Ortheziidae, and Diaspididae [38–41], and host switches seem to have occurred in the other families [39]. Otherwise, enterobacteria have undergone more evolutionary events (losses, duplications, and host switches). Some scale insects have enterobacteria closely related to *Sodalis* endosymbionts (*Sodalis*-like). But others may have symbionts closely related to *Pantoea* and *Klebsiella* [39].

Sodalis cosymbionts have been identified mainly by their 16S rRNA but also by other gene sequences. They have been found within various insect orders including Diptera, Coleoptera, Phthiraptera, and Hemiptera [42–45]. The first described was *S. glossinidius*, the secondary symbiont of tsetse flies [46]. Later, bacteria with related gene sequences were referred as *Sodalis*-like [47] or *Sodalis*-affiliated but more recently several "*Sodalis*-like" bacteria and SOPE [48] are classified as *Sodalis*, others have been assigned to different genera. Still, scientists are in the process of making correct adscriptions for some of these bacteria [49].

The flavobacteria endosymbiont *Candidatus* Walczuchella monophlebidarum ("*Walczuchella*" from here on) was sequenced from the giant wax cochineal *Llaveia axin axin* (Llave) (Coccoidea: Monophlebidae) [50]. This insect has been used to obtain a lacquer to coat traditional art crafts by native people in Mexico and Guatemala since pre-Hispanic times [51]. The flavobacterial genome revealed that the endosymbiont's major role is to synthesize and provide amino acids to the insect host [50]. The Flavobacteria genome was obtained from the analysis of a metagenome of *L. axin axin*. From this metagenome, we could also ensemble sequences from other microorganisms. Here, we present the draft genome of another cosymbiont of *Walczuchella*, a *Sodalis*-like bacteria that is designated here as *Sodalis* TME1. We also present a comparison to the genomes of five other *Sodalis*, as well as preliminary data of a metatranscriptome performed in the bacteriome *of L. axin axin* adults and in the ovaries of senescent adults.

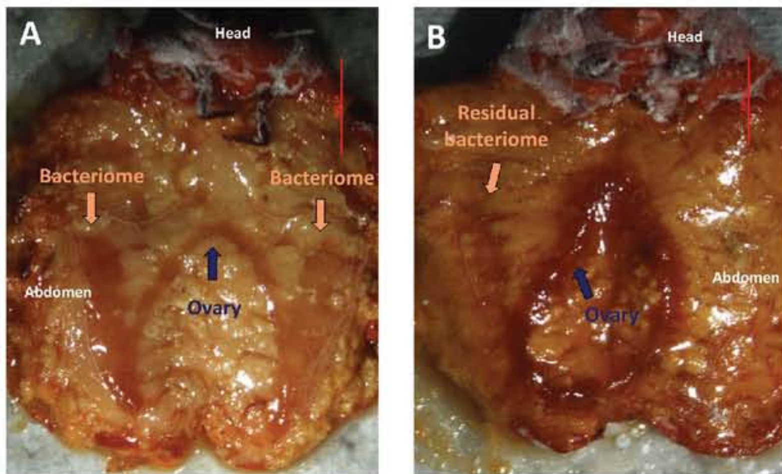
2. Materials and methods

DNA, sequencing, and assembly were performed from bacteriomes (Illumina HiSeq 2000) and from the homogenized of female adults (pyrosequencing) of *L. axin axin* collected in the state of Chiapas, Mexico, as described [50]. A photograph from *L. axin axin* female adults is shown in **Figure 1**. RAST and GosthKOALA from KEGG [52] were used for genomic and metabolic pathway annotation of the metagenomic data that was previously reported when we obtained the *Walczuchella* genome [50].

Comparative phylogenomic analysis was performed with 20 genomes of gammaproteobacteria from GeneBank. Gene calling of all genomes was performed using GeneMark version 2.5 [53]. The pangenome and core genome from orthologous genes of all strains were

01

02 **Figure 1.** *L. axin axin* adult females on a *Jatropha curcas* plant.



01 **Figure 2.** Dissected *L. axin axin* adult females used for the metatranscriptome analysis. (A) early stage and (B) late stage or senescent adults.

02 obtained by GET_HOMOLOGUES version 2.0 software [54] with -A -c -t 0 -M -n 35 and
03 -A -c -t 0 -G -n 35 parameters. We selected a set of 143 unique single-copy orthologous
04 genes from core genome. Translated coding sequences of each gene were concatenated
05 using BioEdit Version 7.2.5 and aligned with Clustal Omega version 1.2.1 [55]. Prottest3
06 version 3.4.2 [56] was used to select the best amino acid substitution model using the AICc
07 correction. The edited alignment contained 47,803 amino acid positions. Maximum likeli-
08 hood phylogeny was performed by PhyML software version 3.1 [57] using the CpREV
09 model with the Shimodaira–Hasegawa-like procedure for internal branch support [58].
10 The genome of *Escherichia coli* K-12 MG1655 was used as outgroup.

11 Comparative genomics was carried out with the following *Sodalis* genomes: *S. glossin-*
12 *idius* morsitans from tsetse fly, *Sodalis*-like endosymbiont from the blood-feeding lice
13 *Proechinophthirus fluctus* (an obligate ectoparasite of fur seals), *S. pierantonus* SOPE from
14 rice weevils *Sitophilus oryzae*, the free-living *S. praecaptivus*, and *Sodalis*-like symbiont of the
15 meadow spittlebug *Philaenus spumarius*. Orthologous genes and the core genomes were
16 obtained by GET_HOMOLOGUES as described above. Core genome matrix was parsed from
17 GET_HOMOLOGUES result, using the parsing_pangenome_matrix.pl script. Shared genes
18 between *Sodalis*-like TME1 and all other strains were retrieved by parsing the core matrix
19 using custom perl scripts. Annotation of each gene cluster was carried out by BLASTp 2.2.30+
20 [59] searches against Uniref100 database. Furthermore, average nucleotide identity (ANI)
21 was determined for all *Sodalis* genomes described above using the ANIcalculator software
22 described by Varghese et al. [60] with the default parameters.

RNA was extracted from the bacteriome of *L. axin axin* female adults and from the ovaries of senescent female adults that do not possess the structure of the bacteriomes (bacteriomes degrade in senescent adults) (**Figure 2**). Sequencing of cDNA was performed by SOLID technology. The sequences were mapped to the genomes of *Walczuchella, Sodalis*-like TME1, and two insect reference genomes, *Drosophila melanogaster* and to the aphid *Acyrthosiphon pisum*. Differentially expressed genes were identified by comparing expression values between samples and using Kal's *Z*-test of proportions [61]. Genes with a change in the expression more than twofold and a *p*-value of <0.01 in the *Z*-test were considered as differentially expressed genes.

To determine the uric acid and uricase activity, *L. axin axin* adult females were individually dissected under sterile conditions. Guts including the Malpighian tubules were extracted and metabolic activities were detected as described [62].

3. Results

We found gene sequences of an enterobacterium (gammaproteobacterium) related to *Sodalis* in the metagenome of the wax cochineal *L. axin axin* [50]. The phylogeny with a set of 143 conserved genes shows that the enterobacterium of *L. axin axin* is closely related to other *Sodalis*-like endosymbionts, especially close to the free-living *S. praecaptivus* [63] (**Figure 3**). The small branches in the *Sodalis* group may indicate that they have recently diverged while the large differences found in genome sizes among these endosymbionts indicate that evolution may be occurring mainly by genome reduction when compared to the larger genome of the free-living *Sodalis* (**Figure 3**).

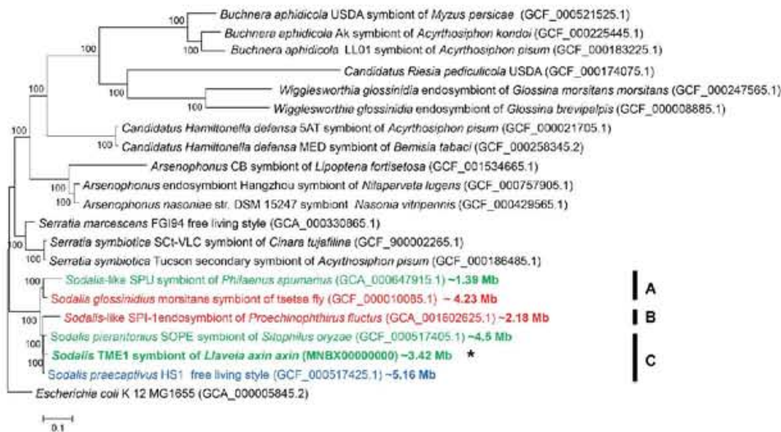


Figure 3. Maximum likelihood phylogeny of sequenced enterobacterial endosymbionts performed with 143 conserved genes. *Sodalis* endosymbionts are shown in green, plant-feeding host; red, blood-feeding host; and blue, free-living style, *Sodalis* TME1 used in this study.

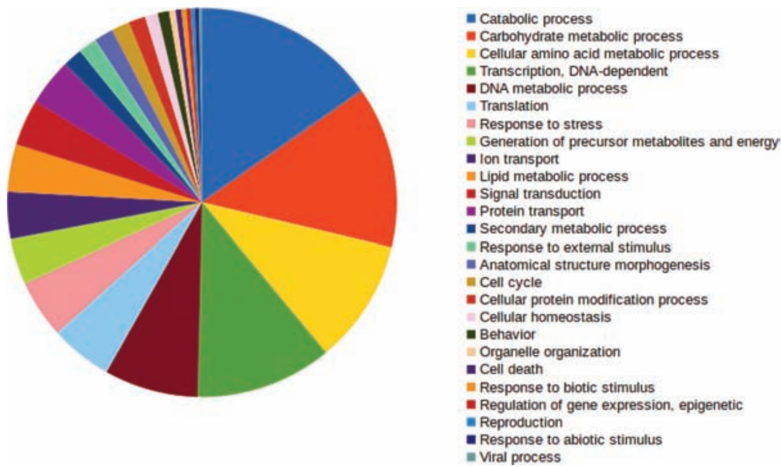
TME1 was compared with the ANI (average nucleotide identity) metric to other *Sodalis* using the same core genome used in the phylogenomic analysis. TME1 showed ANIs well over 95% that is used to delineate species with *S. pierantonius* SOPE and *S. praecaptivus* HS1, but lower than 95% with *S. glossinidius* morsitans, *Sodalis*-like SPU, and *Sodalis*-like SPI-1 (**Table 1**). There was a good correlation of the ANI values obtained and phylogenetic positions that allowed the identification of three groups within *Sodalis* (**Figure 3** and **Table 1**).

The draft assembly of the enterobacterial endosymbiont *Sodalis* TME1 genome consisted of 679 scaffolds with an N50 of 7713 and an average G + C content of 55.6%. The scaffolds sum 3.4 Mb [50]. A total of 3067 genes were identified to which a functional annotation was assigned. The functional categories more represented by the annotated genes were carbohydrates catabolism, amino acids metabolism, transcription, translation, DNA metabolism, and stress response (mainly for oxidative and osmotic stress) (**Figure 4**). Interestingly, many phage-related sequences were found as well as genes for different multidrug efflux pumps and type III and IV secretion systems. TME1 has genes for polyamine biosynthesis and excretion as well as Ankyrin repeat domains and for a lactoyl-glutathione lyase that is a detoxifying enzyme [64]. Among the conserved genes in the core genome of *Sodalis* TME1, *S. pierantonius* str. SOPE and *S. praecaptivus* str. HS1 are genes for the synthesis of flagella and for nitrate reduction (nitrate reductase narV and *nar*G). Maybe nitrate serves in *Sodalis* as an electron acceptor in anaerobiosis as occurs in bacterial symbionts of marine bivalves *Lucinoma aequizonata* [65]. *Sodalis* TME1 genome has genes for uric acid utilization such as uricase (*ua*Z), allantoinase (*all*B), allantoate deiminase (*all*C), and urease (*ure*C and *ure*D). Comparative genomics with all *Sodalis* strains show that *all*C and the alpha subunit for urease gene (*ure*C) orthologous were only present in *Sodalis* TME1. Experimentally, uric acid and uricase activity were quantified in *L. axin axin* female adults. We detected 5.86 ± 0.77 ng of uric acid per tissue μg^{-1} and 32.87 ± 5. 25 mU of uricase per tissue μg^{-1} in female cochineals.

We obtained 11,042,037 and 11,042,428 reads from the cDNA sequence of the bacteriome and the ovaries, respectively. These two organs for studying the differentially expressed genes were selected because endosymbionts are transferred from bacteriomes to the ovaries for vertical transmission to their offspring. It was expected to find genes related to the migration of

	Sodalis str. TME1	*Sodalis pierantonius* SOPE	*Sodalis praecaptivus* HS1	*Sodalis glossinidius* str. morsitans	*Sodalis*-like str. PSPU	*Sodalis*-like str. SPI-1
Sodalis str. TME1						
Sodalis pierantonius str. SOPE	**98.45**					
Sodalis praecaptivus str. HS1	**98.54**	**98.46**				
Sodalis glossinidius str. morsitans	91.24	91.04	91.27			
Sodalis-like str. PSPU	89.81	89.67	89.87	**95.48**		
Sodalis-like str. SPI-1	92.25	91.94	92.15	89.91	85.86	

Table 1. Average nucleotide identity (ANI) percentage among *Sodalis* strains. Values in bold are >95%. Colors correspond to green, plant-feeding host; red, blood-feeding host; blue, free-living style..



01 **Figure 4.** Gene functional categories of *Sodalis* TME1.

02 the endosymbionts from the bacteriome and the colonization of the ovaries. Reads mapped
03 to the reference genomes are shown in **Table 2**. The number of genes that were statistically
04 differentially expressed is shown in **Table 3**.

05 *Walczuchella* in the bacteriome tissue showed only two genes that exhibited differential expres-
06 sion, a putative hydrolase and the chaperone GroEL. Other genes showed a change in expres-
07 sion less than twofold compared to their expression in the ovary. The chaperonin GroES is
08 almost at the limit for differential expression with 1.86-fold (**Table 4**).

09 From the ovary tissue, we found differential expression of *Walczuchella* genes that code for
10 some ATP synthase subunits (some of them annotated previously as pseudogenes), cyto-
11 chrome c oxidase, also some genes of protein translocation systems, tryptophan, histidine
12 and chorismate biosynthesis, one gene related to oxidative stress, and a gene that encodes a
13 possible component of an ABC transporter (**Table 4**).

Reference genome	Bacteriome	Ovaries
Drosophila melanogaster (exons)	2,019,585	2,008,381
Acyrthosiphon pisum (mRNA refseq)	3,082,319	2,912,207
Walczuchella	1,052,077	87,502
Sodalis TME1	*4.1.1.* 409,128	483,601

14

Table 2. Number of reads mapped to the reference genomes.

Reference genome	Bacteriome	Ovaries
Drosophila melanogaster (exons)	494	680
Acyrthosiphon pisum (mRNA refseq)	244	280
Walczuchella	2	89
Sodalis TME1	66	50

Table 3. Number of genes differentially expressed according to *Z*-test ($p < 0.01$).

	Walczuchella	Sodalis TME1	Insect
Bacteriome (differential expression) (high RPKM)	Putative hydrolase	T3SS-secreted effector	Chaperon Hsp70
	Chaperones GroEL, GroES	Allantoinase	ABC transporters
	Hypothetical proteins	Hypothetical proteins	Antiparasitic-like peptide
	ATP synthase B subunit	NAD biosynthesis	Asparaginase
	Amino acids biosynthesis genes	FtsE cell division gene	Unknown genes
		Transcriptional regulation	Extracellular glutamate receptor channel
		Flagellum synthesis	Phospholipids synthesis
			Transcriptional regulation
Ovary (differential expression) (high RPKM)	ATP synthase B and A subunits (pseudogenes)	Hypothetical proteins	ATPase subunit
	AhpC oxidative stress gene	NAD biosynthesis	Transmembrane transporters of sugars and amino acids
	Glycoprotease	Flagellum synthesis	Peptidoglycan-binding protein
	Amino acids biosynthesis genes	FtsE cell division gene	Lysozyme
	Cytochrome c oxydase	Glycolysis	Unknown genes
	SecY translocase	Phage lysozyme	Transcriptional regulation
	Hypothetical proteins	Transcriptional regulation	Phospholipids synthesis

Table 4. Highly expressed and differentially expressed genes in the bacteriome and the ovaries in the endosymbionts *Walczuchella* and *Sodalis* TME1 and the host *L. axin axin*.

In the bacteriome, the enterobacterium TME1 showed very strong overexpression of a gene that codes an effector protein possibly secreted by the type III secretion system (TTSS), expressed 66.8-fold compared to its expression in the ovaries. Also, a gene that codes an allantoinase that participates in uric acid metabolism is highly overexpressed in bacteriome, showing a 50-fold change. Other genes with overexpression in the bacteriome are four ABC transporters, a peroxidase, the heme synthase, two genes related to nucleotides biosynthesis, two genes related to lipid A biosynthesis, and two genes of the type III secretion system (**Table 4**).

In the ovary, TME1-overexpressed genes were related to NAD synthesis, carbohydrate metabolism, stress response, and some transporters and transcriptional regulators (**Table 4**).

Among the insect differentially expressed genes in the bacteriome there were 19 putative transporters (for amino acids, carbohydrates, vitamins, drugs, or unknown substrates), five genes related to defense systems including an antiparasitic peptide with identity to Drosomycin, three from *D. melanogaster*, two genes related to heat-shock response, an oxidative stress response gene, seven genes related to amino acid metabolism, and some genes related to lipid, carbohydrate, and vitamin metabolism (**Table 4**).

On the other hand, we found that in the insect, in the ovaries there was overexpression of 15 transporters, 17 immune response genes, some genes related to heat shock, desiccation, oxidative stress, and hypoxia response, and genes related to lipids, vitamins, carbohydrates, nucleotides, amino acids, and chitin synthesis and metabolism (**Table 4**).

4. Discussion

Due to the annual cycle of the wax cochineal, we are only able to collect insects once a year during the rainy season. It is worth mentioning that in 2015 and 2016, we did not find cochineals in many of the places where we had collected previously. Considering the menace of mosquitoes transmitting Zika, or Chikungunya, extensive fumigations with chemical insecticides have been carried out in many places in Mexico, especially in Chiapas. The relation to the diminished populations of cochineals remains to be established.

A previous survey of symbiotic bacteria from scale insects in Mexico revealed the prevalence of Flavobacteria and Gammaproteobacteria [39]. Some of the Gammaproteobacteria had 16S ribosomal gene sequences closely related to those of TME1, and thus they may be considered as *Sodalis* as well. They were obtained from different scale insects such as *Insignorthezia* sp. and *I. insignia, Icerya purchasi, Cripticerya* sp., and *Pseudococcus longispinus* that together with *Llaveia* would be hosts for *Sodalis*.

While Flavobacteria and insects showed a co-divergent pattern of evolution, the phylogenetic relationships of the Gammaproteobacteria and insects were not parallel, indicating multiple enterobacterial transfers among the different hosts, and a more recent and less dependent symbiosis. In agreement, the genome size of the gammaproteobacterium TME1 is much larger than that from the primary endosymbiont from wax cochineals, the Flavobacteria *Walczuchella*, and also larger than those from other cosymbionts as the Betaproteobacteria that accompany the bacteroidete *Sulcia* found in some insects.

The genome from the gammaproteobacterium TME1 (3.4 Mb) is within the range of those from other *Sodalis* (1.4–4.7 Mb, **Figure 3**). There are very few genomes available from *Sodalis*, namely those from *Sodalis* found in blood-sucking insects as in lice [42] and tsetse flies [66], in plant-feeding insects as the rice weevils [44], in spittlebugs [45], and from a free-living bacterium [67]. The TME1 genome is the first one from an endosymbiont from a sap-feeding insect. The average nucleotide identity (ANI [68] being used for global genomic comparisons and considered now as a gold standard in prokaryote taxonomy [69]) was estimated for the *Sodalis* with available genomes. ANI values and the phylogenomic analysis

performed showed *Sodalis* as a defined and coherent genus with three groups A–C. These groups could represent three different species according to the global standards [69]. Two of these groups were identified as different lineages by Lo et al. [49]. The groups that we described here have a 100 SH-like value support, group A contains *S. glossinidius* from tsetse flies and *Sodalis* from the meadow spittlebug *P. spumarius*, group B is constituted by *Sodalis* from the fur seal *P. fluctus*, and group C contains the closely related TME1, the free-living *S. praecaptivus* and *S. pierantonius* SOPE. The nucleotide sequence conservation among the group A symbiotic and free-living *Sodalis* may reflect that the former were recent acquisitions in insects without enough time for sequence divergence in their hosts. The presence of very similar *Sodalis* in distinct insect isolates reinforces the reports that indicate that they may frequently be transferred among hosts [39, 47].

TME1 has biosynthetic pathways for all essential amino acids and may supply the needs of the wax cochineal and of *Walczuchella* that does not have complete pathways for the biosynthesis of all essential amino acids. It is worth noting that the flavobacterium *Candidatus* Uzinura diaspidicola, an endosymbiont from the armored scale insect *Aphytis melinus* that feed on parenchyma which may provide more nutrients than sap, supplies its host with all nutrients without the need of a cosymbiont [70]. Other armored scale insects have been reported to have a *Sodalis*-like endosymbiont [39].

In *S. glossinidius* that is a secondary symbiont of tsetse flies, a type III secretion system was found implicated in cell invasion and maybe required for colonizing the insect bacteriocytes [71]. Genes encoding for a similar system were found in TME1. Notably, genes that code for the type III secretion system (TTSS) as well as a gene coding for an effector protein that may be secreted by this system were among the most highly induced in the bacteriome of TME1. In *Salmonella enterica*, polyamines are required for full expression of TTSS and for some effector coding genes. Mutants in polyamine biosynthesis are affected in intracellular colonization and survival and may be complemented by adding polyamines to the medium [72]. Furthermore, the modulation of a TTSS by a spermidine transporter has been reported in *Pseudomonas aeruginosa*. Exogenous addition of spermidine to the wild *P. aeruginosa* strain increased the expression of genes that produce effector proteins [73]. TME1 has all genes for spermidine and putrescine biosynthesis as well as for the excretion of spermidine. Polyamines may regulate host defense responses as do some effectors secreted by TTSS. This remains to be tested.

Uric acid and uricase activity were detected in *L. axin axin* females. Uric acid is the final product of purine metabolism. Only few insects are capable of degrading uric acid into other products. In plant-feeding insects, bacterial and fungal symbionts are capable of recycling uric acid into other nitrogen sources [74–76]. *Sodalis* TME1 has uricase and allantoinase-codifying genes, and the latter was highly expressed in bacteriomes suggesting that *Sodalis* TME1 could participate in providing nitrogen to the host by uric acid recycling.

By reverse transcriptase-polymerase chain reaction (RT-PCR) using primers directed to *Sodalis*, we found sequences from *Sodalis* in the bacteriome (our own unpublished results), thus we may suppose that *Sodalis* are localized in bacteriomes such as *Walczuchella*. In *Llaveia*, in addition to *Walczuchella* and *Sodalis* we found sequences of alphaproteobacteria that are related to *Rickettsia* and *Wolbachia* and several fungi that are reported elsewhere (Vera Ponce de León, submitted). Coincidently, the seal lice with a *Sodalis* endosymbiont also harbor a *Rickettsia* that is very abundant. The role of the very little abundant *Rickettsia*-like bacterium

in *Llaveia* is unknown. *Wolbachia* is found in members of the Coelostomidiidae family that is closely related to Monophlebidae insect family that contains the Mexican wax cochineals.

Here, we used the term symbiome [27] to refer to the group of primary and secondary (cosymbionts) endosymbionts (and/or their genomes), residing in a host. We consider that the term symbiome is more adequate than the terms endosymbiotic community or consortium that are sometimes used instead.

The cosymbionts of different Flavobacteria in scale insects are diverse lineages of related Gammaproteobacteria [39]. Similarly, the cosymbionts of *Sulcia* (a flavobacterium as *Walczuchella*) are varied and may be different even in related hosts [7, 36]. *Sulcia* cosymbionts may belong to alpha, beta, or gammaproteobacteria, with alpha and betaproteobacteria looking like the oldest symbionts. It was reported that *Candidatus* Zinderia insecticola, the Betaproteobacteria of spittelbugs was probably substituted by a *Sodalis*-like symbiont in members of the Philaenini tribe of the spittelbugs [33, 45]. The displacement of betaproteobacterial cosymbionts by the gammaproteobacterium *Sodalis* seems recent and was described as an event "*in statu nascendi*" (in the stage of being born) in *Cycadella viridis* [77]. There are other examples where that one endosymbiont may substitute another one or is on the way toward displacement of a highly reduced-genome endosymbiont [33, 77–79]. Distinct (apparently replaceable) cosymbionts may fulfill the different needs of insects that may change overtime and conditions specially if the insect changes habit [22], otherwise there may be cosymbiont redundancy, with different bacteria performing the same or very similar role (e.g., the synthesis of essential amino acids). The *Sodalis* cosymbiont in the wax cochineals seems to be recently acquired as in *C. viridis*. The insect symbiome seems plastic or dynamic with cosymbionts playing a key role in this plasticity. Here, we enlarged the list of putative functions of *Sodalis* that may include uric acid recycling, polyamine biosynthesis, or detoxification.

Acknowledgement

We thank CONACyT for funding.

Author details

Tania Rosas-Pérez[1,2], Arturo Vera-Ponce de León[2], Mónica Rosenblueth[2,*], Shamayim T. Ramírez-Puebla[2], Reiner Rincón-Rosales[3], Julio Martínez-Romero[2], Michael F. Dunn[2] and Esperanza Martínez-Romero[2]

*Address all correspondence to: mrosen@ccg.unam.mx

1 Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Spain

2 Center for Genomics Sciences, National Autonomous University of Mexico, Cuernavaca, Morelos, Mexico

3 Technological Institute of Tuxtla Gutiérrez, Tuxtla Gutiérrez, Chiapas, Mexico

01 **References**

02 [1] McCutcheon JP. The bacterial essence of tiny symbiont genomes. Current Opinion in
03 Microbiology. 2010;13:73–78. DOI: 10.1016/j.mib.2009.12.002.

04 [2] Martínez-Cano DJ, Reyes-Prieto M, Martínez-Romero E, Partida-Martínez LP, Latorre A,
05 Moya A, Delaye L. Evolution of small prokaryotic genomes. Frontiers in Microbiology.
06 2015; **5**:742. DOI: 10.3389/fmicb.2014.00742.

07 [3] Husník F, Chrudimský T, Hypša V. Multiple origins of endosymbiosis within the
08 Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches.
09 BMC Biology. 2011;**9**:87. DOI: 10.1186/1741-7007-9-87.

10 [4] Ramírez-Puebla ST, Servín-Garcidueñas LE, Jiménez-Marín B, Bolaños LM, Rosenblueth
11 M, Martínez J, Rogel MA, Ormeño-Orrillo E, Martínez-Romero E. Gut and root micro-
12 biota commonalities. Applied and Environmental Microbiology. 2013; **79**:2–9. DOI:
13 10.1128/AEM.02553-12.

14 [5] Trujillo ME, Alonso-Vega P, Rodríguez R, Carro L, Cerda E, Alonso P, Martínez-
15 Molina E. The genus *Micromonospora* is widespread in legume root nodules: the
16 example of *Lupinus angustifolius*. ISME Journal. 2010; **4**:1265–1281. DOI: 10.1038/
17 ismej.2010.55.

18 [6] Trujillo ME, Riesco R, Benito P, Carro L. Endophytic Actinobacteria and the interaction
19 of *Micromonospora* and nitrogen fixing plants. Frontiers in Microbiology. 2015; **6**:1341.
20 DOI: 10.3389/fmicb.2015.01341.

21 [7] Noda H, Watanabe K, Kawai S, Yukuhiro F, Miyoshi T, Tomizawa M, Koizumi Y,
22 Nikoh N, Fukatsu T. Bacteriome-associated endosymbionts of the green rice leafhop-
23 per *Nephotettix cincticeps* (Hemiptera: Cicadellidae). Applied Entomology and Zoology.
24 2012; **47**:217–225. DOI:10.1007/s13355-012-0110-1

25 [8] Bennett GM, Moran NA. Small, smaller, smallest: the origins and evolution of ancient
26 dual symbioses in a Phloem-feeding insect. Genome Biology and Evolution. 2013;
27 **5**:1675–1688. DOI: 10.1093/gbe/evt118.

28 [9] Tsuchida T, Koga R, Meng XY, Matsumoto T, Fukatsu T. Characterization of a facultative
29 endosymbiotic bacterium of the pea aphid *Acyrthosiphon pisum*. Microbial Ecology. 2005;
30 **49**:126–133. DOI: 10.1007/s00248-004-0216-2.

31 [10] Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, Kimura N, Ito M, Fukatsu T.
32 Primary gut symbiont and secondary, *Sodalis*-allied symbiont of the Scutellerid stinkbug
33 *Cantao ocellatus*. Applied and Environmental Microbiology. 2010; **76**:3486–3494. DOI:
34 10.1128/AEM.00421-10.

35 [11] Chrudimský T, Husník F, Nováková E, Hypša V. *Candidatu*s Sodalis melophagi *sp. nov.*:
36 phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glos-*
37 *sinidius*. PLoS One. 2012; **7**:e40354. DOI: 10.1371/journal.pone.0040354.

01 [12] Happ GM, Happ CM, Barras SJ. Fine structure of the prothoracic mycangium, a cham-
02 ber for the culture of symbiotic fungi, in the southern pine beetle, *Dendroctonus frontalis*.
03 Tissue and Cell. 1971; **3**:295–308. DOI: 10.1016/S0040-8166(71)80024-1

04 [13] Kuechler SM, Gibbs G, Burckhardt D, Dettner K, Hartung V. Diversity of bacterial endo-
05 symbionts and bacteria-host co-evolution in Gondwanan relict moss bugs (Hemiptera:
06 Coleorrhyncha: Peloridiidae). Environmental Microbiology. 2013; **15**:2031–2042. DOI:
07 10.1111/1462-2920.12101.

08 [14] Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky
09 JM, Dunbar HE, Tran PL, Moran NA, Eisen JA. Metabolic complementarity and genom-
10 ics of the dual bacterial symbiosis of sharpshooters. PLoS Biology. 2006; **4**:e188. DOI:
11 10.1371/journal.pbio.0040188.

12 [15] Cottret L, Milreu PV, Acuña V, Marchetti-Spaccamela A, Stougie L, Charles H, Sagot
13 MF. Graph-based analysis of the metabolic exchanges between two co-resident intra-
14 cellular symbionts, *Baumannia cicadellinicola* and *Sulcia muelleri*, with their insect host,
15 *Homalodisca coagulata*. PLoS Computational Biology. 2010; **6**:e1000904.

16 [16] López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R. Mealybugs nested endosymbio-
17 sis: going into the 'matryoshka' system in *Planococcus citri* in depth. BMC Microbiology
18 2013; **13**:74. DOI: 10.1186/1471-2180-13-74

19 [17] Rao Q, Rollat-Farnier PA, Zhu DT, Santos-Garcia D, Silva FJ, Moya A, Latorre A, Klein
20 CC, Vavre F, Sagot MF, Liu SS, Mouton L, Wang XW. Genome reduction and potential
21 metabolic complementation of the dual endosymbionts in the whitefly *Bemisia tabaci*.
22 BMC Genomics. 2015; **16**:226. DOI: 10.1186/s12864-015-1379-6.

23 [18] Scarborough CL, Ferrari J, Godfray HC. Aphid protected from pathogen by endosymbi-
24 ont. Science. 2005; **310**:1781–1781. DOI: 10.1126/science.1120180.

25 [19] Moran NA, Tran P, Gerardo NM. Symbiosis and insect diversification: an ancient symbiont
26 of sap-feeding insects from the bacterial phylum Bacteroidetes. Applied Environmental
27 Microbiology. 2005; **71**:8802–8810. DOI:10.1128/AEM.71.12.8802-8810.2005.

28 [20] Nakabachi A, Ueoka R, Oshima K, Teta R, Mangoni A, Gurgui M, Oldham NJ, van Echten-
29 Deckert G, Okamura K, Yamamoto K, Inoue H, Ohkuma M, Hongoh Y, Miyagishima SY,
30 Hattori M, Piel J, Fukatsu T. Defensive bacteriome symbiont with a drastically reduced
31 genome. Current Biology. 2013; **23**:1478–1484. DOI: 10.1016/j.cub.2013.06.027.

32 [21] Leclair M, Pons I, Mahéo F, Morlière S, Simon J-C, Outreman Y. Diversity in symbiont con-
33 sortia in the pea aphid complex is associated with large phenotypic variation in the insect
34 host. Evolutionary Ecology. 2016; DOI:10.1007/s10682-016-9856-1 (published on line).

35 [22] Oliver KM, Degnan PH, Burke GR, Moran NA. Facultative symbionts in aphids and the
36 horizontal transfer of ecologically important traits. Annual Review of Entomology. 2010;
37 **55**:247–266. DOI 10.1146/annurev-ento-112408-085305.

38 [23] Chen DQ, Montllor CB, Purcell A H. Fitness effects of two facultative endosymbiotic bacte-
39 ria on the pea aphid, *Acyrthosiphon pisum*, and the blue alfalfa aphid *A. kondoi*. Entomologia
40 Experimentalis et Applicata. 2000; **95**:315–323. DOI: 10.1046/j.1570-7458.2000.00670.x

01 [24] Russell JA, Moran NA. Costs and benefits of symbiont infection in aphids: varia-
02 tion among symbionts and across temperatures. Proceedings of the Royal Society B:
03 Biological Sciences. 2006; **273**:603–610. DOI:10.1098/rspb.2005.3348.

04 [25] Montllor C B, Maxmen A, Purcell AH. Facultative bacterial endosymbionts benefit pea
05 aphids *Acyrthosiphon pisum* under heat stress. Ecological Entomology. 2002; **27**:189–195.
06 DOI:10.1046/j.1365-2311.2002.00393.x

07 [26] Teixeira L, Ferreira A, Ashburner M. The bacterial symbiont *Wolbachia* induces resis-
08 tance to RNA viral infections in *Drosophila melanogaster*. PLoS Biology. 2008; **6**:e2. DOI:
09 10.1371/journal.pbio.1000002

10 [27] Bouclas DG, Kariithi HM, Bourtzis K, Schneider DI, Kelley K, Miller WJ, Parker AG,
11 Abd-Alla AM. Transgenerational transmission of the *Glossina pallidipes* hytrosavirus
12 depends on the presence of a functional symbiome. PLoS One. 2013; **8**:e61150. DOI:
13 10.1371/journal.pone.0061150.

14 [28] Chrostek E, Marialva MS, Esteves SS, Weinert LA, Martinez J, Jiggins FM, Teixeira L.
15 *Wolbachia* variants induce differential protection to viruses in *Drosophila melanogaster*: a
16 phenotypic and phylogenomic analysis. PLoS Genetics. 2013; **9**:e1003896. DOI: 10.1371/
17 journal.pgen.1003896

18 [29] Werren JH, Baldo L, Clark ME. *Wolbachia*: master manipulators of invertebrate biology.
19 Nature Reviews Microbiology. 2008; **6**:741–751. DOI: 10.1038/nrmicro1969.

20 [30] Koga R, Tsuchida T, Fukatsu T. Changing partners in an obligate symbiosis: a faculta-
21 tive endosymbiont can compensate for loss of the essential endosymbiont Buchnera in
22 an aphid. Proceedings of the Royal Society B: Biological Sciences. 2003; **270**:2543–2550.
23 DOI: 10.1098/rspb.2003.2537

24 [31] Hosokawa T, Kaiwa N, Matsuura Y, Kikuchi Y, Fukatsu T. Infection prevalence
25 of *Sodalis* symbionts among stinkbugs. Zoological Letters. 2015; **1**:5. DOI: 10.1186/
26 s40851-014-0009-5.

27 [32] McCutcheon JP, Moran NA. Functional convergence in reduced genomes of bacterial
28 symbionts spanning 200 My of evolution. Genome Biology and Evolution. 2010: **2:**708–
29 718. DOI:10.1093/gbe/evq055

30 [33] Koga R, Bennett GM, Cryan JR, Moran NA. Evolutionary replacement of obligate sym-
31 bionts in an ancient and diverse insect lineage. Environmental Microbiology. 2013;
32 **15**:2073–2081. DOI: 10.1111/1462-2920.12121.

33 [34] Van Leuven JT, Meister RC, Simon C, McCutcheon JP. Sympatric speciation in a bac-
34 terial endosymbiont results in two genomes with the functionality of one. Cell. 2014;
35 **158**:1270–1280. DOI: 10.1016/j.cell.2014.07.047.

36 [35] Bennett GM, Abbà S, Kube M, Marzachì C. Complete genome sequences of the obli-
37 gate symbionts "*Candidatus* Sulcia muelleri" and "*Ca*. Nasuia deltocephalinicola" from
38 the Pestiferous Leafhopper *Macrosteles quadripunctulatus* (Hemiptera: Cicadellidae).
39 Genome Announcements. 2016; **4** pii:e01604–15. DOI: 10.1128/genomeA.01604-15.

01 [36] Kobiałka M, Michalik A, Walczak M, Junkiert Ł, Szklarzewicz T. Sulcia symbiont of
02 the leafhopper Macrosteles laevis (Ribaut, 1927) (Insecta, Hemiptera, Cicadellidae:
03 Deltocephalinae) harbors Arsenophonus bacteria. Protoplasma. 2016; **253**:903–912. DOI:
04 10.1007/s00709-015-0854-x.

05 [37] Dhami MK, Turner AP, Deines P, Beggs JR, Taylor MW. Ultrastructural and molecu-
06 lar characterization of a bacterial symbiosis in the ecologically important scale insect
07 family Coelostomidiidae. FEMS Microbiology Ecology. 2012; **81**:537–546. DOI:
08 10.1111/j.1574-6941.2012.01378.x

09 [38] Gruwell ME, Morse GE, Normark BB. Phylogenetic congruence of armored scale
10 insects (Hemiptera: Diaspididae) and their primary endosymbionts from the phy-
11 lum Bacteroidetes. Molecular Phylogenetics and Evolution. 2007; **44**:267–280. DOI:
12 org/10.1016/j.ympev.2007.01.014

13 [39] Rosenblueth M, Sayavedra L, Sámano-Sánchez H, Roth A., Martínez-Romero E.
14 Evolutionary relationships of flavobacterial and enterobacterial endosymbionts with
15 their scale insect hosts (Hemiptera: Coccoidea). Journal of Evolutionary Biology. 2012;
16 **25**:2357–2368. DOI: 10.1111/j.1420-9101.2012.02611.x

17 [40] Koteja J. Essay on the prehistory of the scale insects (Homoptera, Coccinea). In Annales
18 Zoologici. Państwowe Wydawnictwo Naukowe: Warzawa, Wroclaw; 1985; **38**. p.
19 461–503.

20 [41] Dhami MK, Buckley TR, Beggs JR, Taylor MW. Primary symbiont of the ancient scale
21 insect family Coelostomidiidae exhibits strict cophylogenetic patterns. Symbiosis. 2013;
22 **61**:77–91. DOI: 10.1007/s13199-013-0257-8

23 [42] Boyd BM, Allen JM, Koga R, Fukatsu T, Sweet AD, Johnson KP, Reed DL. Two bacte-
24 rial genera, *Sodalis* and *Rickettsia*, associated with the seal louse *Proechinophthirus fluctus*
25 (Phthiraptera: Anoplura). Applied and Environmental Microbiology. 2016; **82**:3185–
26 3197. DOI: 10.1128/AEM.00282-16.

27 [43] Toju H, Koga R, Nikoh N, Meng XY, Kimura N, Fukatsu T. "*Candidatus* Curculioniphilus
28 buchneri" a novel clade of bacterial endocellular symbionts from weevels of the genus
29 *Curculio*. Applied and Environmental Microbiology. 2010; **76**:275–282. DOI: 10.1128/
30 AEM.02154-09

31 [44] Oakeson KF, Gil R, Clayton AL, Dunn DM, von Niederhausern AC, Hamil C, Aoyagi A,
32 Duval B, Baca A, Silva FJ, Vallier A, Jackson DG, Latorre A, Weiss RB, Heddi A, Moya A,
33 Dale C. Genome degeneration and adaptation in a nascent stage of symbiosis. Genome
34 Biology and Evolution. 2014; **6:**76–93. DOI: 10.1093/gbe/evt210.

35 [45] Koga R, Moran NA. Swapping symbionts in spittlebugs: evolutionary replacement
36 of a reduced genome symbiont. ISME Journal. 2014; **8**:1237–1246. DOI: 10.1038/
37 ismej.2013.235.

01 [46] Dale C, Maudlin I. *Sodalis gen. nov*. and *Sodalis glossinidius sp. nov*., a microaerophilic sec-
02 ondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. International Journal
03 of Systematic Bacteriology. 1999; **49**:267–275. DOI: 10.1099/00207713-49-1-267.

04 [47] Smith WA, Oakeson KF, Johnson KP, Reed DL, Carter T, Smith KL, Koga R, Fukatsu T,
05 Clayton DH, Dale C. Phylogenetic analysis of symbionts in feather-feeding lice of the
06 genus *Columbicola*: evidence for repeated symbiont replacements. BMC Evolutionary
07 Biology. 2013; **13**:109. DOI: 10.1186/1471-2148-13-109.

08 [48] Gil García R, Belda E, Gosalbes MJ, Delaye L, Vallier A, Vincent Monégat C, Heddi A,
09 Silva GJ, Moya A, Latorre A. Massive presence of insertion sequences in the genome
10 of SOPE, the primary endosymbiont of the rice weevil "*Sitophilus oryzae*". International
11 Microbiology. 2008; **11**:4–48. DOI: 10.2436/20.1501.01.43.

12 [49] Lo WS, Huang YY, Kuo CH. Winding paths to simplicity: genome evolution in faculta-
13 tive insect symbionts. FEMS Microbiology Reviews. 2016; pii:fuw028. DOI: org/10.1093/
14 femsre/fuw028

15 [50] Rosas-Pérez T, Rosenblueth M, Rincón-Rosales R, Mora J, Martínez-Romero E. Genome
16 sequence of "*Candidatus* Walczuchella monophlebidarum" the flavobacterial endosym-
17 biont of *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae). Genome Biology and
18 Evolution. 2014; **6**:714–726. DOI: 10.1093/gbe/evu049

19 [51] Williams M, MacVean CM. Ethnococcidology: use of the giant margarodids, *Llaveia* spp.
20 (Homoptera: Coccoidea: Margarodidae), by indigenous peoples of Mesoamerica in their
21 culture, medicine and arts. Israel Journal of Entomology. 1995; **29**:147–148.

22 [52] Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for
23 functional characterization of genome and metagenome sequences. Journal of Molecular
24 Biology. 2016; **428**:726–731. DOI: 10.1016/j.jmb.2015.11.006

25 [53] Besemer J, Borodovsky M. GeneMark: Web software for gene finding in prokaryotes,
26 eukaryotes and viruses. Nucleic Acids Research. 2005; **33**:451–454. DOI:10.1093/nar/
27 gki487

28 [54] Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package
29 for scalable and robust microbial pangenome analysis. Applied and Environmental
30 Microbiology. 2013; **79**:7696–7701. DOI:10.1128/AEM.02411-13

31 [55] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Higgins DG. Fast, scalable
32 generation of high-quality protein multiple sequence alignments using Clustal Omega.
33 Molecular Systems Biology. 2011; **7**:539. DOI:10.1038/msb.2011.75

34 [56] Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models
35 of protein evolution. Bioinformatics. 2011; **27**:1164–1165. DOI:10.1093/bioinformatics/
36 btr088

[57] Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology. 2010; **59**:307–321. DOI:10.1093/sysbio/syq010

[58] Shimodaira H, Hasegawa M. Letter to the editor multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution. 1999; **16**:1114–1116. DOI:10.1177/0148607109348061

[59] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009; **10**:421. DOI:10.1186/1471-2105-10-421

[60] Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A. Microbial species delineation using whole genome sequences. Nucleic Acids Research. 2015; **43**:6761–6771. DOI: 10.1093/nar/gkv657.

[61] Kal AJ, van Zonneveld AJ, Benes V, van den Berg M, Koerkamp MG, Albermann K, et al. Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. Molecular Biology of the Cell. 1999; **10**:1859–1872. DOI:10.1091/mbc.10.6.1859

[62] Vera-Ponce de León A, Sanchez-Flores A, Rosenblueth M, Martínez-Romero E. Fungal community associated with *Dactylopius* (Hemiptera: Coccoidea: Dactylopiidae) and its role in uric acid metabolism. Frontiers in Microbiology. 2016; **7**:954. DOI: 10.3389/fmicb.2016.00954

[63] Chari A, Oakeson KF, Enomoto S, Jackson DG, Fisher MA, Dale C. Phenotypic characterization of *Sodalis praecaptivus sp. nov*., a close non-insect-associated member of the *Sodalis*-allied lineage of insect endosymbionts. International Journal of Systematic and Evolutionary Microbiology. 2015; **65**:1400–1405. DOI: 10.1099/ijs.0.000091.

[64] Thornalley PJ. The glyoxalase system: new developments towards functional characterization of a metabolic pathway fundamental to biological life. The Biochemical Journal. 1990; **269**:1–11. DOI: 10.1042/bj2690001

[65] Hentschel U, Hand SC, Felbeck H. The contribution of nitrate respiration to the energy budget of the symbiont-containing clam *Lucinoma aequizonata*: A calorimetric study. The Journal of Experimental Biology. 1996; **199**:427–433.

[66] Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M, Aksoy S. Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Research. 2006; **16**:149–156. DOI:10.1101/gr.4106106

[67] Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, von Niederhausern AC, Weiss RB, Fisher M, Dale C. A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect–bacterial symbioses. PLoS Genetics 2012; **8**:e1002990. DOI: 10.1371/journal.pgen.1002990

[68] Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proceedings of the National Academy of Sciences of the United States of America. 2005; **102**:2567–2572. DOI:10.1073/pnas.0409727102.

[69] Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proceedings of the National Academy of Sciences of the United States of America. 2009; **106:**19126–19131. DOI: 10.1073/pnas.0906412106.

[70] Sabree ZL, Huang CY, Okusu A, Moran NA, Normark BB. The nutrient supplying capabilities of *Uzinura*, an endosymbiont of armoured scale insects. Environmental Microbiology. 2013; **15**:1988–1999. DOI: 10.1111/1462-2920.12058.

[71] Dale C, Young SA, Haydon DT, Welburn SC. The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proceedings of the National Academy of Sciences of the United States of America. 2001; **98**:1883–1888. DOI:10.1073/pnas.98.4.1883

[72] Jelsbak L, Thomsen LE, Wallrodt I, Jensen PR, Olsen JE. Polyamines are required for virulence in *Salmonella enterica* serovar Typhimurium. PLoS One. 2012; **7**:e36149. DOI: 10.1371/journal.pone.0036149.

[73] Zhou L, Wang J, Zhang LH. Modulation of bacterial Type III secretion system by a spermidine transporter dependent signaling pathway. PLoS One. 2007;**2**:e1291. DOI: 10.1371/journal.pone.0001291.

[74] Potrikus CJ, Breznak JA. Anaerobic degradation of uric Acid by gut bacteria of termites. Applied and Environmental Microbiology. 1980; **40(1)**:125–132.

[75] Morales-Jiménez J, Vera-Ponce de León A, García-Domínguez A, Martínez-Romero E, Zúñiga G, Hernández-Rodríguez C. Nitrogen-fixing and uricolytic bacteria associated with the gut of *Dendroctonus rhizophagus* and *Dendroctonus valens* (Curculionidae: Scolytinae). Microbial Ecology. 2013; **66**:200–210. DOI:10.1007/s00248-013-0206-3

[76] Patiño-Navarrete R, Piulachs M-D, Belles X, Moya A, Latorre A, Peretó J. The cockroach Blattella germanica obtains nitrogen from uric acid through a metabolic pathway shared with its bacterial endosymbiont. Biology Letters. 2014; **10**:7–10. DOI:10.1098/rsbl.2014.0407.

[77] Michalik A, Jankowska W, Kot M, Gołas A, Szklarzewicz T. Symbiosis in the green leafhopper, Cicadella viridis (Hemiptera, Cicadellidae). Association in statu nascendi? Arthropod Structure and Development. 2014; **43**:579–587. DOI: 10.1016/j.asd.2014.07.005.

[78] Lamelas A, Gosalbes MJ, Manzano-Marín A, Peretó J, Moya A, Latorre A. *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. PLoS Genetics. 2011; **7**:e1002357. DOI: 10.1371/journal.pgen.1002357.

[79] Lefèvre C, Charles H, Vallier A, Delobel B, Farrell B, Heddi A. Endosymbiont phylogenesis in the dryophthoridae weevils: evidence for bacterial replacement. Molecular Biology and Evolution. 2004; **21**:965–973. DOI:10.1093/molbev/msh063

Systematic and Applied Microbiology 38 (2015) 390–399

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Species in *Wolbachia*? Proposal for the designation of '*Candidatus* Wolbachia bourtzisii', '*Candidatus* Wolbachia onchocercicola', '*Candidatus* Wolbachia blaxteri', '*Candidatus* Wolbachia brugii', '*Candidatus* Wolbachia taylori', '*Candidatus* Wolbachia collembolicola' and '*Candidatus* Wolbachia multihospitum' for the different species within *Wolbachia* supergroups

Shamayim T. Ramírez-Puebla[a], Luis E. Servín-Garcidueñas[a], Ernesto Ormeño-Orrillo[a], Arturo Vera-Ponce de León[a], Mónica Rosenblueth[a], Luis Delaye[b], Julio Martínez[a], Esperanza Martínez-Romero[a,*]

[a] *Centro de Ciencias Genómicas, UNAM, Cuernavaca, Morelos, Mexico*
[b] *Departamento de Ingeniería Genética, CINVESTAV-Irapuato, Irapuato, Guanajuato, Mexico*

ARTICLE INFO

Article history:
Received 22 October 2014
Received in revised form 21 May 2015
Accepted 27 May 2015

Keywords:
Wolbachia supergroups
Wolbachia Candidatus
Wolbachia taxonomy
ANI
DDH

ABSTRACT

Wolbachia are highly extended bacterial endosymbionts that infect arthropods and filarial nematodes and produce contrasting phenotypes on their hosts. *Wolbachia* taxonomy has been understudied. Currently, *Wolbachia* strains are classified into phylogenetic supergroups. Here we applied phylogenomic analyses to study *Wolbachia* evolutionary relationships and examined metrics derived from their genome sequences such as average nucleotide identity (ANI), *in silico* DNA–DNA hybridization (DDH), G + C content, and synteny to shed light on the taxonomy of these bacteria. Draft genome sequences of strains *w*DacA and *w*DacB obtained from the carmine cochineal insect *Dactylopius coccus* were included. Although all analyses indicated that each *Wolbachia* supergroup represents a distinct evolutionary lineage, we found that some of the analyzed supergroups showed enough internal heterogeneity to be considered as assemblages of more than one species. Thus, supergroups would represent supraspecific groupings. Consequently, *Wolbachia pipientis nomen* species would apply only to strains of supergroup B and we propose the designation of '*Candidatus* Wolbachia bourtzisii', '*Candidatus* Wolbachia onchocercicola', '*Candidatus* Wolbachia blaxterii', '*Candidatus* Wolbachia brugii', '*Candidatus* Wolbachia taylorii', '*Candidatus* Wolbachia collembolicola' and '*Candidatus* Wolbachia multihospitis' for other supergroups.
© 2015 Elsevier GmbH. All rights reserved.

Introduction

Wolbachia is a genus of endosymbiotic bacteria that are wide spread in nature. *Wolbachia* endosymbionts do not have a free-living phase and are under confinement to particular hosts. It is estimated that *Wolbachia* may be found in 40% of arthropod species [106], while a previous report calculated 60% [44]. *Wolbachia* endosymbionts have been found associated with nematodes from the Onchocercidae family [22,54]. Interactions with their hosts range from parasitism to mutualism. In arthropods are mostly considered as parasites since they may manipulate host reproduction by mechanisms like parthenogenesis, feminization, male killing, and cytoplasmic incompatibility [12,80,100]. However, *Wolbachia* symbiosis has been implicated in host fitness [15,94], or as being necessary for oogenesis [25]; in nematodes they are regarded as mutualistic and essential for survival [58]. *Wolbachia* symbiosis is outstanding as it may cause host speciation events [11].

Some insects and their endosymbionts have a parallel evolutionary history, and cospeciation events have been described for both host and bacteria, especially primary endosymbionts [1,10,21,79,87]. For endosymbionts that have cospeciated with

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

http://dx.doi.org/10.1016/j.syapm.2015.05.005
0723-2020/© 2015 Elsevier GmbH. All rights reserved.

their hosts, endosymbionts in different hosts would be distinct species. It seems that cospeciation is rare in *Wolbachia* and insects as their phylogenies are usually not congruent [1,49,86,88]. Thus, adaptations to different hosts would not necessarily mean bacterial speciation. *Wolbachia* infections in insects may be recent in some cases [41,48], implying a short host–symbiont interaction that would not lead to speciation. Recent *Wolbachia* acquisitions may come from horizontal transfers from close or even distant insects [41]. In filarial nematode–*Wolbachia* associations, congruence between *Wolbachia* phylogenies and those of their host has been documented [9,17,32]. In this case, cospeciation between bacteria and their worm hosts seems to have occurred and a single origin of this symbiosis for supergroups C and D has even been proposed [32]. *Wolbachia* have become essential for nematode development and play an important role in host embryogenesis [58]. Nematodes treated with antibiotics cannot reach adulthood [13,93].

Wolbachia pipientis Hertig 1936 [43], was first observed in cells of the *Culex pipiens* mosquito [42]. Heterogeneity within *Wolbachia* has been revealed by sequence analysis of 16S rRNA genes and protein-coding genes, resulting in its distribution into sixteen phylogenetic supergroups, ten of which are found in arthropods (A, B, E, H, I, K, M, N, P, Q), five in nematodes (C, D, J, L, O) and one comprising both arthropod and nematode endosymbionts (F) [4]. The strains of *Wolbachia* detected in Australian spiders [81], were designed as Supergroup G but it was later revealed that it has a *wsp* gene that is a recombinant between those of A and B supergroups rather than being a distinct new supergroup [8]. A phylogenetic tree based on a multilocus analysis has been recently published giving insight about the relationships between *Wolbachia* supergroups [37]. A consensus of whether supergroups represent lineages of *W. pipientis* or distinct species has not been reached. Sequence divergence between supergroups seemed to indicate that each represented a species [78], however, other studies have indicated that they do not represent isolated genetic entities [7,99], as would be expected from *bona fide* species [61]. *Wolbachia* have been described as highly recombinogenic bacteria [6,7,99]. Multiple infections with different *Wolbachia* are frequent in the same insect individual [98,104], affording the opportunity for recombination between different strains, including not closely related ones [6,104]. Nevertheless, a recent study found that recombination is far higher within supergroups than between them [30]. Recombination events between supergroups are limited to small DNA fragments.

Endosymbiont confinement in a host leads to an inevitable dependence on the host. This is evident upon inspection of endosymbiont genomes, which generally lack many functions required for independent living. Classic taxonomy relying on phenotypic characterization of pure cultures as well as establishing genomic relatedness by DNA–DNA hybridization (DDH) experiments could not be applied to non-cultivable endosymbionts like *Wolbachia*. In the genomic era, however, metrics based on genome sequences like ANI (average nucleotide identity) and *in silico* DDH can be used as replacements for wet lab DDH [3,38], thus allowing the use of similar taxonomic criteria for both cultivable and non-cultivable prokaryotes. Furthermore, it is increasingly acknowledged that phenotypes should not be given as much importance for species delineation as they currently are [20,70,96].

Here, we evaluated the diversity of *Wolbachia* by performing phylogenomic analyses and by analyzing genome-derived metrics like ANI, *in silico* DDH, G + C content and synteny in order to shed light on the taxonomy of these endosymbionts. Additionally, we increased the genomic database of *Wolbachia* by reporting sequences from two strains recently obtained from the carmine cochineal insect *Dactylopius coccus*.

Materials and methods

Genome sequences

Sequences of all reported *Wolbachia* genomes were retrieved from GenBank database, except those of strains *w*Di and *w*Ls, which were available at http://nematodes.org/genomes/index_filaria.html [22]. Genomes of *Wolbachia* strains *w*DacA (Bioproject PRJNA274701) and *w*DacB (Bioproject PRJNA274698) were sequenced by a metagenomic approach from dissected cochineal insects of *Dactylopius coccus*. Detailed functional analyses of those genomic sequences will be reported elsewhere (Ramírez-Puebla et al., in preparation). For G + C content determination, contigs of each genome were concatenated, the number of G plus C nucleotides counted and the sum divided by the genome length. Genome of strain *w*Men was obtained from the *Strepsiptera* Genome Project [68], and genomes of strains *w*Fol, *w*Oc and *w*Cte were only deposited like Sequence Read Archive (SRA) so they were not included in G + C determination because they were not completely sequenced [36].

Phylogenomic analyses

Predicted proteomes were obtained from annotated genomes deposited at GenBank if available. The RAST server was also used for annotating and comparing whole genome sequences [5]. The AMPHORA2 pipeline [103], was used to identify a set of 31 conserved bacterial proteins from complete or draft genomes. Sequencing reads were obtained from the Sequence Read Archive (SRA) database to obtain phylogenetic markers for strains *w*Men, *w*Fol, *w*Oc, and *w*Cte. Reads were mapped against individual marker genes obtained from fully sequenced *Wolbachia* genomes using the runMapping option from Newbler (Roche). The obtained mapped reads were processed to obtain the markers for these strains by performing tblastn searches against reference protein sequences corresponding to the markers from other sequenced strains. Protein sequences were concatenated using the EMBOSS union web tool (http://emboss.bioinformatics.nl/cgi-bin/emboss/union). The concatenated sequences were then aligned using MUSCLE v.3.8.31 [29], and the resulting alignment was processed with Gblocks [18], to obtain conserved protein blocks and eliminate poorly aligned positions and divergent regions. The edited alignment contained 9151 amino acid positions. A maximum-likelihood analysis was then performed using the JTT substitution model under PhyML 3.0 [40]. Branch support values are based on 100 bootstrap replicates. The genomes from *Ehrlichia canis* (GenBank CP000107) and *Anaplasma marginale* (GenBank CP001079) were used as outgroups.

In silico *DDH and ANI calculations*

DDH estimates were computed using the Genome-to-Genome Distance Calculator version 2.0 [65], as recommended by Auch et al. [2,3], and Meier-Kolthoff et al. [65]. BLAST+ was used for alignment and formula 2 for genome distance estimation. ANI values were calculated as previously proposed [38], using the ANI calculator from the Kostas lab (http://enve-omics.ce.gatech.edu/ani/) with default parameters.

Synteny

Syntenic blocks between ten finished *Wolbachia* genomes were identified by BLASTN. Only blocks at least 3000 bp in length and with 80% identity or higher were used to construct the graphs using the Artemis comparison tool [16].

Results and discussion

Genome-based relationships

Predicted evolutionary relationships between all 34 complete or almost complete *Wolbachia* genomes currently available and our two *Wolbachia* strains from *D. coccus*, *w*DacA and *w*DacB (Table 1) based on a set of conserved proteins is shown in Fig. 1. The distinctiveness of *Wolbachia* supergroups was evidenced by each forming a different and well-supported clade. The A and B supergroups clustered together in a branch separated from the C, D and F supergroups as previously observed by Nikoh et al. [69], for a set of six genomes using 52 ribosomal proteins. The phylogenetic reconstructions were also in agreement with a previous analysis obtained with a set of 90 orthologous genes from only eleven sequenced *Wolbachia* strains [22]. All the *Wolbachia* strains associated with *Culex* mosquitoes (*W. pipientis* representatives) tightly clustered in a single clade within supergroup B. The *Wolbachia* strain *w*Bol1-b from *Hypolimnas bolina* was also phylogenetically close to the *W. pipientis* strains. *Wolbachia* strains from *D. coccus* clustered into distinct supergroups. *w*DacA was resolved as the most phylogenetically distant strain within supergroup A, whereas *w*DacB was a member of supergroup B having *w*VitB from *Nasonia vitripennis* as its closest sequenced relative. The strain *w*Fol associated with the springtail *Folsomia candida* was found placed as the most distant of all the analyzed *Wolbachia* supergroups as reported previously [36].

G + C content

The G + C content of *Wolbachia* genomes ranged from 32.1% to 38.4% (Table 1) evidencing an enrichment of AT nucleotides as is common in endosymbionts [67]. Mean G + C contents of supergroups A, B, C and D were 35.5%, 34.0%, 32.4% and 33.4%, respectively. The sequenced representative of supergroup F had a G + C content of 36.3%. Although the analyzed genomes may not completely comprise the natural variation present in *Wolbachia* supergroups, it is worth noting that each supergroup seems to have a characteristic G + C content (Table 1).

In silico *DDH and ANI estimates*

DDH is the "gold standard" for species delineation in prokaryotes but it is not applicable for non-cultivable bacteria like *Wolbachia*. ANI represents a suitable surrogate for wet lab DDH as correlation analyses indicate that strains showing ANI higher or equal than 95–96% shared DDH values higher or equal than 70% and are thus considered to be of the same species [38]. Genome sequences also allow the estimation of *in silico* DDH values, and

Table 1
Characteristics of the sequenced *Wolbachia* genomes used in this work.

Wolbachia strain	Host species	GenBank accession number	Genome status	Number of contigs	Genome size (bp)	G + C%	Super group	Reference
*w*Mel	*Drosophila melanogaster*	AE017196	Complete	1	1,267,782	35.2	A	[102]
*w*MelPop	*Drosophila melanogaster*	AQQE00000000	WGS	80	1,239,155	35.1	A	[101]
*w*Ri	*Drosophila simulans* Riverside	CP001391	Complete	1	1,445,873	35.2	A	[52]
*w*Ha	*Drosophila simulans*	CP003884	Complete	1	1,295,804	35.1	A	[30]
*w*Sim	*Drosophila simulans*	AAGC00000000	WGS	629	1,063,100	35.4	A	[84]
*w*Au	*Drosophila simulans*	LK055284	Complete	1	1,268,461	35.2	A	[91]
*w*Rec	*Drosophila recens*	JQAM00000000	WGS	43	1,126,656	35.2	A	[66]
*w*Suzi	*Drosophila suzukii*	CAOU02000000	WGS	110	1,415,350	35.2	A	[89]
*w*Dwi	*Drosophila willistoni*	AAQP00000000	WGS	260	1,145,915	38.4	A	Remington et al. (unpublished)
*w*Ana	*Drosophila ananassae*	AAGB00000000	WGS	464	1,440,750	35.7	A	[84]
*w*Gmm	*Glossina morsitans morsitans*	AWUH00000000	WGS	241	1,019,510	35.2	A	[14]
*w*Uni	*Muscidifurax uniraptor*	ACFP00000000	WGS	256	867,873	35.2	A	[52]
*w*DacA	*Dactylopius coccus*	PRJNA274701	WGS	456	933,576	35.0	A	This study
*w*No	*Drosophila simulans*	CP003883	Complete	1	1,301,823	34.0	B	[30]
*w*Pip_Pel	*Culex quinquefasciatus*	AM999887	Complete	2	1,482,455	34.2	B	[51]
*w*Pip_JHB	*Culex quinquefasciatus*	ABZA00000000	WGS	21	1,543,661	34.2	B	[85]
*w*Pip_Mol	*Culex molestus*	HG428761	Complete	1	1,340,443	33.9	B	[74]
*w*Pip	*Culex pipiens molestus*	CACK00000000	WGS	888	1,479,531	34.3	B	Sinkins et al. (unpublished)
*w*Di	*Diaphorina citri*	AMZJ00000000	WGS	124	1,240,904	34.0	B	[83]
*w*Bol1-b	*Hypolimnas bolina*	CAOH00000000	WGS	144	1,377,933	33.9	B	[28]
*w*AlbB	*Aedes albopictus*	CAGB00000000	WGS	156	1,162,431	33.7	B	[63]
*w*DacB	*Dactylopius coccus*	PRJNA274698	WGS	321	1,282,277	34.2	B	This study
*w*VitB	*Nasonia vitripennis*	AERW00000000	WGS	523	1,107,643	33.9	B	[50]
*w*Cte	*Ctenocephalides felis*	SRR1222150	Raw data	–	–	–	B	[36]
*w*Oo	*Onchocerca ochengi*	HE660029	Complete	1	957,990	32.1	C	[24]
*w*Ov	*Onchocerca volvulus* strain Cameroon	HG810405	Complete	1	960,618	32.1	C	Cotton et al. (unpublished)
*w*Di	*Dirofilaria immitis*	PRJEB4154[a]	WGS	2	921,012	32.7	C	[22]
*w*Bm strain TRS	*Brugia malayi*	AE017321	Complete	1	1,080,084	34.2	D	[35]
*w*Bn	*Wuchereria bancrofti*	ADHD00000000	WGS	763	1,052,327	34.0	D	[26]
*w*Ls	*Litomosoides sigmodontis*	PRJEB4155[a]	WGS	10	1,048,936	32.1	D	[22]
*w*Fol	*Folsomia candida*	SRR1222159	Raw data	–	–	–	E	[36]
*w*Cle	*Cimex lectularius*	AP013028	Complete	1	1,250,060	36.3	F	[69]
*w*Oc	*Osmia caerulescens*	SRR1221705	Raw data	–	–	–	F	[36]
*w*Men	*Mengenilla moldrzyki*	SRX095325	WGS	–	–	–	F	[68]

[a] Accessions numbers correspond to the European Nucleotide Archive database as submitted by the original authors.

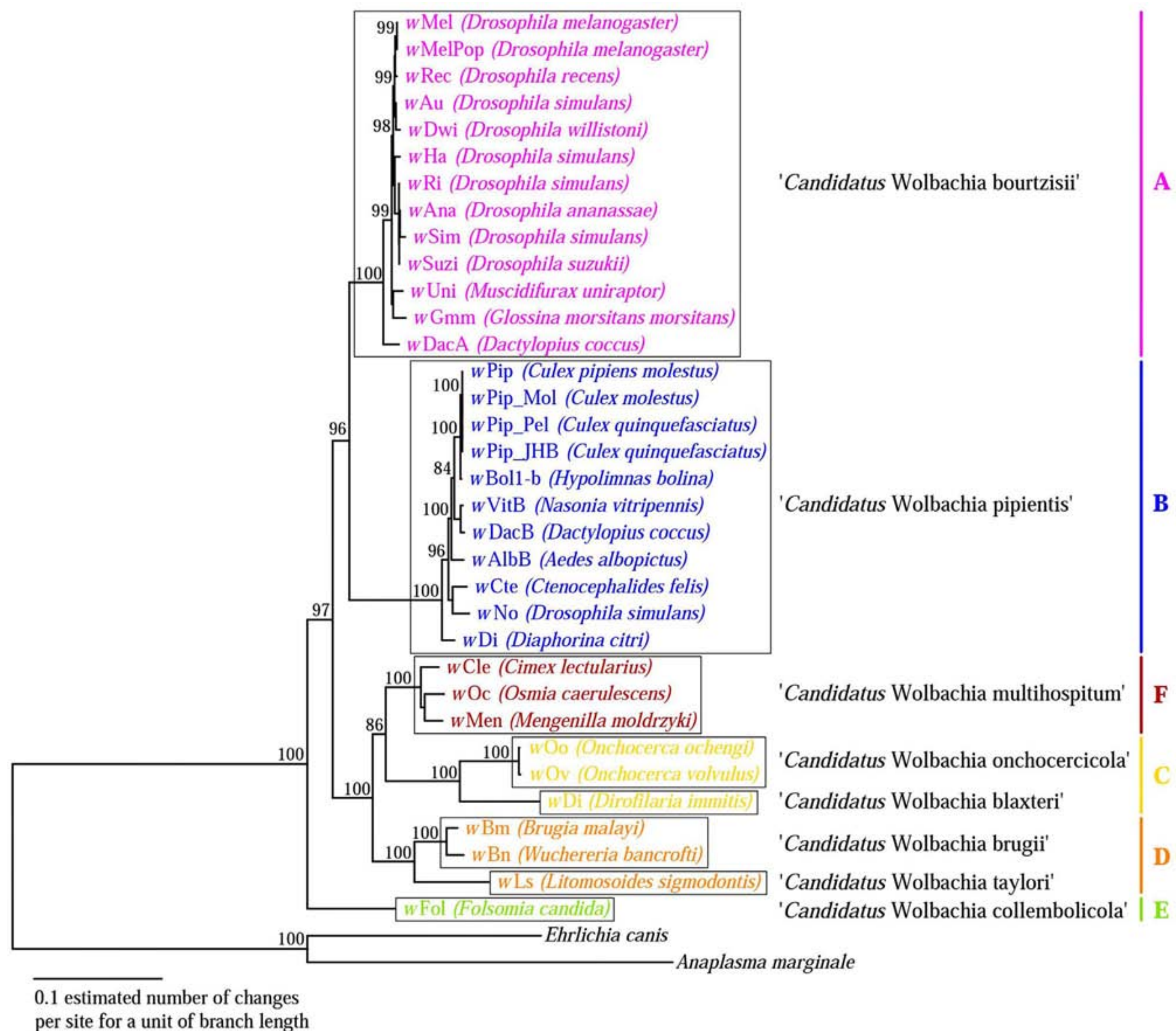


Fig. 1. Phylogenomic tree showing evolutionary relationships between *Wolbachia* strains inferred with PhyML based on a concatenated alignment of 31 marker proteins detected with AMPHORA2 and analyzed with the JTT substitution model. Hosts for each strain are indicated in forceps. Strains included in the new designations for each species name are boxed. Proposed species names are shown next to the boxes. Supergroups are shown to the right. Numbers at the branch points represent bootstrap support values based on 100 replicates. The scale bar represents the estimated number of amino acid changes per site for a unit of branch length.

these correlate very well with wet lab DDH [3]. Based in these criteria other bacteria as *Ensifer* and *Rhizobium* have been reclassified [71].

In silico DDH and ANI values were calculated for all pairs of analyzed *Wolbachia* genomes (Tables 2 and 3, respectively). Strains from different supergroups showed maximum ANI and *in silico* DDH values of 86.8% and 34.6%, respectively indicating that *Wolbachia* is comprised of several different species as previously suggested based on fewer ANI comparisons [75]. Within each supergroup, most members showed ANI values over 96% (Table 3) and *in silico* DDH values over 70% (Table 2), relatedness levels that are consistent with single species. However, in some supergroups there were strains with enough differences to put them below or near the borderline for species delineation. In supergroup A *Dactylopius* strain *w*DacA, and in supergroup B *Drosophila simulans* strain *w*No and *Diaphorina citri* strain *w*Di, showed *in silico* DDH values <62%, below the species circumscription level with all members of their own supergroups and ANI values just above of the species cut-off level (Table 2). In the phylogenomic analyses, strains *w*DacA, *w*No, *w*Di occupied peripheral positions within their supergroups (Fig. 1). *w*AlbB from *Aedes albopictus* and *w*DacB from *D. coccus* in supergroup B, also showed *in silico* DDH values below or very close to 70% with their supergroup neighbors, although these differences were not reflected by low ANI values (Table 2).

Within supergroup B, *in silico* DDH and ANI values were high among representative sequenced strains of *Wolbachia pipientis* (*w*Pip strains). Values were also high between *w*Bol1-b and *W. pipientis*. In contrast, comparisons between *w*Pip strains and other members of supergroup B produced values that were just over or slightly under the cutoff limits recommended for species delineation. The genome from strain *w*VitB had ANI value of around 97.4% when compared with the *w*Pip strain (Table 3) but most of their DDH estimates were below 70% (Table 2).

Clear examples of the existence of different species within supergroups were observed for nematode strains *w*Di of

Table 2
In silico DNA–DNA hybridization (DDH).

in silico DNA-DNA[a] hybridization

Strain	Supergroup	*w*Mel	*w*MelPop	*w*Rec	*w*Au	*w*Dwi	*w*Uni	*w*Gmm	*w*Ana	*w*Ri	*w*Suzi	*w*Sim	*w*Ha	*w*DacA
*w*Mel	A	100.0	77.7	89.3	93.9	70.5	75.8	71.2	72.8	71.9	72.1	75.1	73.3	57.0
*w*MelPop	A		100.0	95.3	93.6	92.6	76.2	71.7	72.8	72.1	72.3	74.9	73.2	57.6
*w*Rec	A			100.0	90.6	94.0	75.4	72.0	73.0	70.9	71.2	74.3	75.2	57.6
*w*Au	A				100.0	64.7	75.6	70.8	72.4	70.8	71.2	74.3	73.4	57.0
*w*Dwi	A					100.0	75.4	70.5	73.4	72.0	71.9	74.4	72.6	57.0
*w*Uni	A						100.0	74.5	73.5	73.7	73.8	73.3	73.6	58.6
*w*Gmm	A							100.0	67.1	65.1	66.1	67.9	69.5	56.4
*w*Ana	A								100.0	98.0	97.9	95.6	72.0	57.1
*w*Ri	A									100.0	98.6	85.2	69.3	55.9
*w*Suzi	A										100.0	90.7	69.5	55.9
*w*Sim	A											100.0	72.1	56.9
*w*Ha	A												100.0	56.2
*w*DacA	A													100.0

Strain	Supergroup	*w*Pip	wPip_Mol	*w*Pip_Pel	*w*Pip_JHB	*w*Bol1-b	*w*VitB	*w*DacB	*w*Di	*w*AlbB	*w*No	*w*Oo	*w*Ov	*w*Di	*w*Bm	*w*Bn	*w*Ls	*w*Cle
*w*Pip	B	100.0	100.0	96.7	96.4	83.1	71.5	69.5	58.8	63.7	57.3							
*w*Pip_Mol	B		100.0	84.4	67.6	81.5	70.8	68.5	58.2	63.0	56.6							
*w*Pip_Pel	B			100.0	86.3	78.0	68.6	67.5	56.7	61.9	55.2							
*w*Pip_JHB	B				100.0	77.2	68.8	67.5	56.3	61.3	55.0							
*w*Bol1-b	B					100.0	69.8	68.0	58.1	61.9	57.6							
*w*VitB	B						100.0	83.4	61.5	66.0	61.1							
*w*DacB	B							100.0	59.6	62.6	58.6							
*w*Di	B								100.0	59.0	54.5							
*w*AlbB	B									100.0	58.9							
*w*No	B										100.0							
*w*Oo	C											100.0	96.1	29.2				
*w*Ov	C												100.0	29.1				
*w*Di	C													100.0				
*w*Bm	D														100.0	69.8	36.5	
*w*Bn	D															100.0	38.7	
*w*Ls	D																100.0	
*w*Cle	F																	100.0

[a] Values <70% are shaded.

Table 3
Average nucleotide identity (ANI) percentage values between Wolbachia strains.

ANI[a] percentages for strain

Strain	Supergroup	*w*Mel	*w*MelPop	*w*Rec	*w*Au	*w*Dwi	*w*Uni	*w*Gmm	*w*Ana	*w*Ri	*w*Suzi	*w*Sim	*w*Ha	*w*DacA
*w*Mel	A	100.0	100.0	99.6	99.6	99.4	97.9	97.3	97.4	97.8	97.7	97.9	97.7	95.0
*w*MelPop	A		100.0	99.6	99.6	99.4	97.9	97.4	97.5	97.8	97.7	97.8	97.6	94.9
*w*Rec	A			100.0	99.6	99.5	97.9	97.4	97.5	97.6	97.6	97.6	97.8	94.9
*w*Au	A				100.0	99.9	98.0	97.4	97.4	97.6	97.6	97.7	97.6	94.9
*w*Dwi	A					100.0	97.8	97.2	97.2	97.5	97.4	97.7	97.5	94.5
*w*Uni	A						100.0	97.6	97.7	97.6	97.6	97.6	97.5	95.0
*w*Gmm	A							100.0	96.9	96.9	96.9	96.9	97.0	94.6
*w*Ana	A								100.0	99.9	99.9	99.8	97.7	94.7
*w*Ri	A									100.0	100.0	99.9	97.7	94.7
*w*Suzi	A										100.0	99.9	97.7	94.7
*w*Sim	A											100.0	97.5	94.8
*w*Ha	A												100.0	94.8
*w*DacA	A													100.0

Strain	Supergroup	*w*Pip	*w*Pip_Mol	*w*Pip_Pel	*w*Pip_JHB	*w*Bol1-b	*w*VitB	*w*DacB	*w*Di	*w*AlbB	*w*No	*w*Oo	*w*Ov	*w*Di	*w*Bm	*w*Bn	*w*Ls	*w*Cle
*w*Pip	B	100.0	100.0	99.9	99.9	99.1	97.4	97.3	95.8	96.3	95.3							
*w*Pip_Mol	B		100.0	99.8	99.8	99.0	97.2	97.2	95.7	96.1	95.2							
*w*Pip_Pel	B			100.0	99.9	98.9	97.2	97.1	95.6	96.1	95.1							
*w*Pip_JHB	B				100.0	98.9	97.1	97.2	95.6	96.1	95.2							
*w*Bol1-b	B					100.0	97.2	97.1	95.7	96.2	95.3							
*w*VitB	B						100.0	98.8	95.7	96.2	95.6							
*w*DacB	B							100.0	95.5	96.1	95.4							
*w*Di	B								100.0	95.4	94.8							
*w*AlbB	B									100.0	95.3							
*w*No	B										100.0							
*w*Oo	C											100.0	99.4	86.1				
*w*Ov	C												100.0	86.1				
*w*Di	C													100.0				
*w*Bm	D														100.0	96.3	88.8	
*w*Bn	D															100.0	88.8	
*w*Ls	D																100.0	
*w*Cle	F																	100.0

[a] Values <96% are shaded.

Dirofilaria immitis and *w*Ls in supergroups C and D, respectively (Tables 2 and 3). In these cases, both strains occupied peripheral positions within their supergroups (Fig. 1). Also, *in silico* DDH and ANI values supported the differentiation of these strains from other members of their respective supergroups supporting that they belong to different species.

Cospeciation between nematodes and *Wolbachia* is reflected in the fact that more distantly related nematodes [32], also have more distantly related *Wolbachia*. In supergroup D, *Wolbachia* from related nematode genera *Brugia* and *Wuchereria* are more closely related to each other than to *Wolbachia* from the more distant genus *Litomosoides* (Fig. 1, Tables 2 and 3). It is worth noting that *Wolbachia* infecting *Brugia* and *Wuchereria* are at the borderline for species definition by *in silico* DDH. Similarly, in supergroup C, *Wolbachia* from two species of *Onchocerca* are related but showed very low *in silico* DDH and ANI values to the endosymbiont of *D. immitis* (Fig. 1, Tables 2 and 3). Taken together, these data seem to indicate that each nematode genus harbors a different *Wolbachia* species. Nevertheless, the presence of supergroup F *Wolbachia* in both nematodes and insects, and its loss in some filarial species in

Strain: wPipMol, wPipPel, wNo, wMel, wRi, wHa, wOo, wOv, wBm, wCl
Supergroup: B, A, C, D, F

Fig. 2. Synteny between complete genomes of *Wolbachia* strains. Sequence blocks oriented in the same or inverted directions are shown in dark or light green, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

the onchocercids, contravene the idea of cospeciation in general [33].

Genome synteny

Structural genome comparisons in other bacteria like rhizobia have shown that chromosome synteny is very well conserved within a species and less maintained between different species [39,77]. Synteny was used as a further criterion to distinguish *Wolbachia* species. It has been observed that levels of genome synteny are higher within than between *Wolbachia* supergroups [30], as it is evidenced in Fig. 2 for ten finished genomes. *Wolbachia* genomes have high levels of repetitive DNA and mobile genetic elements that lead to DNA rearrangements that diminish synteny even between related strains [30,35,52,102]. Genome rearrangements in other organisms represent recombination barriers and could lead to genetic isolation [76]. Strain *w*No showing significant divergence by *in silico* DDH and ANI values is less syntenic with its supergroup siblings. It would be worth investigating if a speciation process could start within a supergroup by *Wolbachia* strains developing novel genomic rearrangements, as discussed previously [30].

Conclusions

We showed here that *Wolbachia* supergroups represent distinct evolutionary lineages based on phylogenomics, G + C content, ANI, *in silico* DDH and synteny. Our results support the previous proposal that *Wolbachia* from different supergroups should be considered as genetically distinct clades not only from implications related to host confinement and their biology [72], but on the basis of molecular evidence [30,53,78]. Furthermore, we found heterogeneity within supergroups. The more divergent strains within each supergroup were recovered as outliers in the phylogenomic analyses. Not all strains, however, seem to have accumulated enough nucleotide sequence differences to show ANI values lower than 95–96%, used to delineate different species, with other distant strains in their supergroups (Table 3). Nevertheless, within a supergroup significant genome content differences were evidenced by low *in silico* DDH (less than 70%) and genome synteny among supergroup members was not always high. Thus, our novel analyses indicate that different species may occur inside a supergroup. Consequently, supergroups would have a supraspecific status.

Given the evidence reviewed and presented here, the name *W. pipientis* (Hertig 1936) [43], should be applied only to supergroup B strains. As *Wolbachia* are still uncultivable, the proper designation for supergroup B strains should be '*Candidatus* Wolbachia pipientis'. In order to distinguish the different species within *Wolbachia*, we propose the designation '*Candidatus* Wolbachia bourtzisii' for strains in supergroup A, '*Candidatus* Wolbachia onchocercicola' for endosymbionts of genus *Onchocerca* in supergroup C, '*Candidatus* Wolbachia blaxterii' for endosymbionts of *D. immitis* in supergroup C, '*Candidatus* Wolbachia brugii' for endosymbionts of nematodes from *Brugia* and *Wuchereria* species in supergroup D, '*Candidatus* Wolbachia taylorii' for endosymbionts of nematodes from *Litomosoides* species in supergroup D, '*Candidatus* Wolbachia collembola' for endosymbionts of Collembola arthropods in supergroup E and '*Candidatus* Wolbachia multihospitis' for *Wolbachia* strains hosted by nematodes and arthropods in supergroup F.

Description of '*Candidatus* Wolbachia bourtzisii'

'*Candidatus* Wolbachia bourtzisii' (bourt.zi'si.i. N.L. gen. n. *bourtzisii*, of Bourtzis, in honor of Kostas Bourtzis, as a recognition for his studies on *Wolbachia* and other bacteria associated with arthropods).

The description of the species '*Candidatus* Wolbachia bourtzisii' is based on the studies reported by Louis and Nigro [62], Sacchi et al. [82], Texeira et al. [94], and Zhukova and Kiseleva [105]. Cell size is 0.5 μm in *D. simulans*, and 0.5–1.0 μm in *D. melanogaster*. Cells are roundish and less frequently rod shaped and are surrounded by three enveloping membranes. The first is the plasmatic membrane and the second represents the outer part of the cell wall. The third one, closely related to the cytoplasm of the host cell, forms a vacuole for each single microorganism. Ribosomes and nucleic acid fibrils are observed in the cytoplasm. In *D. melanogaster* individual bacterial cells are distributed throughout the host cell cytoplasm, occasionally occurring as small groups. Bacteria occur in the ovarioles in high numbers and in germline cells like cytocytes, oogonia, oocytes and nurse cells.

The percentage of apoptotic cells in germaria are increased in *D. melanogaster* infected with *w*MelPop. Tetracycline treatments accelerated the time to death in *D. melanogaster* infected with *Drosophila* C virus (DCV) as the bacteria confer resistance to DCV by interfering with virus proliferation. The DNA G + C content is between 35.0 and 38.4 mol% as calculated from genomic sequences. Most strains exhibit a DNA G + C content of 35.2 mol%.

Description of '*Candidatus* Wolbachia onchocercicola'

'*Candidatus* Wolbachia onchocercicola' [on.cho.cer.ci'co.la. N.L. fem. n. Onchocerca a filarial nematode genus; L. suffix – *cola* (from L. masc. or fem. n. *incola*), a dweller, an inhabitant; N.L. fem. n. *onchocercicola*, a dweller of *Onchocerca*].

The description of the species '*Candidatus* Wolbachia onchocercicola' is based on the studies reported by Determan et al. [27], Egyed et al. [31], Horeauf et al. [46], Kozek and Marroquin [55], and Langworthy et al. [59]. Cell size is 0.3 up to 0.8 μm in diameter and 1.5 up to 1.8 μm in length. Cells are generally round or spherical shaped. Bacteria are located in the cytoplasm surrounded by a membrane-bound vacuole. In *Onchocerca lupi* each vacuole contains only one bacterium surrounded by a double membrane. In contrast, in *Onchocerca volvulus* they often form clusters and in *Onchocerca ochengi* some of them contain up to seven bacteria. Wolbachia live in the subcutaneous and connective tissues of their hosts, usually enclosed in fibrous cysts or nodules. In adults and larvae bacterial cells occur in the lateral cords, and in germinal tissues in females. Depletion of the endosymbiont by oxytetracycline in *O. ochengi* results in the death of adults and microfilaria. Also, there is a decline in the quantity of embryos and an increase in the proportion of embryos showing abnormal morphology. In *O. volvulus* doxycycline treatment blocks embryogenesis. The DNA G + C content is 32.1 mol% as calculated from genomic sequences.

Description of '*Candidatus* Wolbachia blaxteri'

'*Candidatus* Wolbachia blaxteri' (blax'ter.i. N.L. gen. n. *blaxteri*, of Blaxter, in honor of Mark Blaxter, in recognition of his molecular studies on nematodes and their associated *Wolbachia* symbionts).

The description of the species '*Candidatus* Wolbachia baxteri' is based on the studies reported by Kozek [55,56], McLaren et al., [64], and Sironi et al. [90]. Cell size is 0.3–1.0 μm in diameter and 4.5 μm in length. Cells are spherical or ovoid shaped. Bacteria are contained in an individual membrane-bounded host vacuole. Some bacterial cells have condensations of dense material within their cytoplasm. In *D. immitis* bacteria occur in the reproductive tract mainly in the ovary and the proximal region of the uterus, and are also found in oocytes and in all embryonic stages of microfilariae developing in the uterus. In lateral cords of adults, they occur as clusters that can

fill most of the hypodermal tissue. Often they appear to surround the hypodermal nuclei. In embryos, five to ten bacteria per host cell are found. Also, bacteria are abundant in oogonia, eggs and early dividing embryos. Treatment with tetracycline blocks embryo development. The DNA G + C content is 32.7 mol% as calculated from genomic sequences.

Description of '*Candidatus* Wolbachia brugii'

'*Candidatus* Wolbachia brugii' (bru'gi.i. N.L. gen. n. *brugii*, of Brug, named after S. L. Brug, a Dutch parasitologist who first described the filarial nematode *Brugia malayi*, a model for the study of *Wolbachia*–nematode relationships).

The description of the species '*Candidatus* Wolbachia brugii' is based on the studies reported by Fischer et al. [34], and Landmann et al. [57,58], and Taylor et al. [92]. Cell size is 0.5 μm up to 1 μm. Cells are spherical or have an elongated shape and are surrounded with a double membrane. Bacteria are contained within membrane-bound vacuoles. In *Brugia malayi* clusters of bacteria are detected in microfilaria. In larvae L2, bacterial cells are detected in the hypodermis and in L3 and L4 larvae in the cells of lateral chord, in high numbers. In adult female worms, bacteria are commonly found in the lateral hypodermal cords, in hypodermis, and close to or inside the ovaries. Bacteria are also seen in the cells surrounding the basal lamina of the oviduct. In adult male worms, microfilariae, and third-stage larvae bacteria are detected in the lateral cord, but in lower numbers compared with females and dispersed in focal groups or as individual bacteria. They are also detected in testis and the border of vas deferens. In *Wuchereria bancrofti* bacteria show a similar distribution as in *B. malayi*, in small clusters or as a single bacterium.

Tetracycline treatments dramatically reduce the endosymbiont population in female adults of *B. malayi*. Pyknotic nuclei are observed throughout the ovaries and uteri in the female germline. Microfilaria resulting from a completed embryogenesis after antibiotic treatments, showed defects as abnormal muscle quadrants. Apoptotic nuclei are detected in the ovaries of treated females and become more numerous as the uteri is filled with embryos. The DNA G + C content is between 34.0 and 34.2 mol% as calculated from genomic sequences.

Description of '*Candidatus* Wolbachia taylori'

'*Candidatus* Wolbachia taylori' (tay'lo.ri. N.L. gen. n. *taylori*, of Taylor, in honor of Mark J. Taylor, in recognition of his studies on the role of *Wolbachia*–nematode symbiosis in human diseases and his search for treatments).

The description of the species '*Candidatus* Wolbachia taylori' is based on the studies reported by Chagas-Moutinho et al. [19], and Horeauf et al. [45]. Cell size is approximately1 μm and round shaped. Cells present a reduced cell wall and not a typical septum during cell division. Cells are surrounded by a host-derived vacuolar membrane. In *Litomosoides chagasfilhoi*, bacterial cells occur in regions of the hypoderm, in the oocytes, early-stage embryos and complete developed intrauterine microfilariae close to the cell host nucleus. In other filarial tissues, bacteria are found in intracellular vacuoles associated to the nuclear envelope. They are also observed in proximity to the endoplasmic reticulum. TEM suggested a single bacterium per vacuole.

Depletion by tetracycline results in infertility by blocking female worm development and early embryogenesis in *Litomosoides sigmodontis*. The DNA G + C content is 32.1 mol% as calculated from genomic sequences.

Description of '*Candidatus* Wolbachia collembolicola'

'*Candidatus* Wolbachia collembolicola' [col.lem.bo.li'co.la. N.L. n. pl. *Collembola* a lineage of hexapods; L. suffix – *cola* (from L. masc. or fem. n. *incola*), a dweller, an inhabitant; N.L. fem. n. *collembolicola*, a dweller of *Collembola*].

The description of the species '*Candidatus* Wolbachia collembolicola' is based on the studies reported by Czarnetzki and Tebbe [23], Pike and Kingcombe [73], Timmermans and Ellers [95], and Vandekerckhove et al. [97]. Cells detected in hexapod species of the order Collembola. Cell size is 0.2 μm up to 1.4 μm. Cells are pleomorphic from curved to almost hairpin-shaped. Cell wall lacks detectable peptidoglycan layer. Periplasmic space is of around 5–15 nm. Cells are surrounded by a host-derived vacuolar membrane. DNA filaments are visible in a rather diffuse network dispersed throughout the cell and interspersed with ribosomes. Cells occur in aggregations and are found mostly in close association with the rough endoplasmic reticulum in the ovaries. Fat bodies and interstitial cells as detected by TEM techniques or restricted to the ovary and brain as detected by FISH techniques.

Infection is obligatory for host offspring survival. The endosymbiont is sensitive to high-dose of rifampicin and heat treatments. High-dose tetracycline treatment is inefficient for removing cell infections. Bacteria obligate role early in the parthenogenetic developmental process includes egg hatching.

Description of '*Candidatus* Wolbachia multihospitum'

'*Candidatus* Wolbachia multihospitis' (mul.ti.hos'pi.tum. L. adj. *multus* many, numerous; L. n. *hospes -itis*, he who entertains a stranger, a host; N.L. gen. pl. n. multihospitum of numerous hosts, referring to the occurrence of the bacterium on various species of arthropods and nematodes).

The description of '*Candidatus* Wolbachia multihospitis' is based on the studies reported by Ferri et al. [33], Hosokawa et al. [47], and Lefoulon et al. [60]. In *Cimex lectularius* cell size is 0.5 up to 1.2 μm. Cells are rod-shaped. In males, bacterial cells are located in the testis-associated bacteriome, whereas in females they are located in bacteriomes and ovaries. Cells are also detected in the nutritive cord and developing oocytes. In the nematode *Madathamugadia hiepei*, they are detected in young and late embryos. In adult females they are observed in the ovaries and the intestinal wall. In contrast with other nematodes they are absent in the hypodermal lateral chord. In *Cercopithifilaria japonica* and *Mansonella perforata* bacteria are located in the epithelial somatic gonad and in the intestinal wall.

Elimination of the endosymbiont by rifampicin treatments in *C. lectularius* resulted in deformed developing eggs, reduction in the adult emergence rate and prolonged nymphal period. The DNA G + C content is 36.3 mol% as calculated from genomic sequences.

Acknowledgements

This work was supported by CONACyT basic science grant 154453. Results presented in this paper are part of the PhD thesis by S. T. Ramírez-Puebla, A. Vera-Ponce de León and L. E. Servín-Garcidueñas. They are students from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and each was recipient of a scholarship from Consejo Nacional de Ciencia y Tecnología (CONACYT), México. We thank Mariana Mateos for discussion and Michael Dunn for critical reading of this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.syapm.2015.05.005

References

[1] Ahmed, M.Z., De Barro, P.J., Ren, S.X., Greeff, J.M., Qiu, B.L. (2013) Evidence for horizontal transmission of secondary endosymbionts in the *Bemisia tabaci* cryptic species complex. PLoS ONE 8, e53084.

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

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

[4] Augustinos, A.A., Santos-García, D., Dionyssopoulou, E., Moreira, M., Papapanagiotou, A., Scarvelakis, M., Doudoumis, V., Ramos, S., Aguiar, A.F., Borges, P.A.V., Khadem, M., Latorre, A., Tsiamis, G., Bourtzis, K. (2011) Detection and characterization of *Wolbachia* infections in natural populations of aphids: is the hidden diversity fully unraveled? PLoS ONE 6, e28695.

[5] Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O. (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9, 75.

[6] Baldo, L., Lo, N., Werren, J.H. (2005) Mosaic nature of the *Wolbachia* surface protein. J. Bacteriol. 187, 5406–5418.

[7] Baldo, L., Bordenstein, S., Wernegreen, J.J., Werren, J.H. (2006) Widespread recombination throughout *Wolbachia* genomes. Mol. Biol. Evol. 23, 437–449.

[8] Baldo, L., Werren, J.L. (2007) Revisiting *Wolbachia* supergroup typing based on WSP: spurious lineages and discordance with MLST. Curr. Microbiol. 55, 81–87.

[9] Bandi, C., Anderson, T.J., Genchi, C., Blaxter, M.L. (1998) Phylogeny of *Wolbachia* in filarial nematodes. Proc. Biol. Sci. 265, 2407–2413.

[10] Baumann, P., Moran, N.A., Baumann, L. (2000) Bacteriocyte associated endosymbionts of insects. In: Dworkin, M. (Ed.), The prokaryotes, Springer, New York, pp. 155–189.

[11] Bordenstein, S.R. (2003) Symbiosis and the origin of species. In: Bourtzis, K., Miller, T.A. (Eds.), Insect Symbiosis, CRC Press, Boca Raton, pp. 283–304.

[12] Bordenstein, S.R., Paraskevopoulos, C., Dunning-Hotopp, J.C., Sapountzis, P., Lo, N., Bandi, C., Tettelin, H., Werren, J.H., Bourtzis, K. (2009) Parasitism and mutualism in *Wolbachia*: what the phylogenomic trees can and cannot say. Mol. Bio. Evol. 26, 231–241.

[13] Bosshardt, S.C., McCall, J.W., Coleman, S.U., Jones, K.L., Petit, T.A., Klei, T.R. (1993) Prophylactic activity of tetracycline against *Brugia pahangi* infection in jirds (*Meriones unguiculatus*). J. Parasitol. 79, 775–777.

[14] Brelsfoard., C., Tsiamis, G., Falchetto, M., Gomulski, L.M., Telleria, E., Alam, U., Doudoumis, V., Scolari, F., Benoit, J.B., Swain, M., Takac, P., Malacrida, A.R., Bourtzis, K., Aksoy, S. (2014) Presence of extensive *Wolbachia* symbiont insertions discovered in the genome of its host *Glossina morsitans morsitans*. PLoS Negl. Trop. Dis. 8, e2728.

[15] Brownlie, J.C., Cass, B.N., Riegler, M., Witsenburg, J.J., Iturbe-Ormaetxe, I., McGraw, E.A., O'Neill, S.L. (2009) Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipientis*, during periods of nutritional stress. PLoS Pathog. 5, e1000368.

[16] Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., Parkhill, J. (2005) ACT: the Artemis Comparison Tool. Bioinformatics 21, 3422–3423.

[17] Casiraghi, M., Bain, O., Guerrero, R., Martin, C., Pocacqua, V., Gardner, S.L., Franceschi, A., Bandi, C. (2004) Mapping the presence of *Wolbachia pipientis* on the phylogeny of filarial nematodes: evidence for symbiont loss during evolution. Int. J. Parasitol. 34, 191–203.

[18] Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

[19] Chagas-Moutinho, V.A., Silva, R., de Souza, W., Motta, M.C. (2015) Identification and ultrastructural characterization of the *Wolbachia* symbiont in *Litomosoides chagasfilhoi*. Parasit. Vector 8, 74.

[20] Chan, J.Z.-M., Halachev, M., Loman, N., Constantinidou, C., Pallen, M. (2012) Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*. BMC Microbiol. 12, 302.

[21] Chen, X., Li, S., Aksoy, S. (1999) Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. J. Mol. Evol. 48, 49–58.

[22] Comandatore, F., Sassera, D., Montagna, M., Kumar, S., Koutsovoulos, G., Thomas, G., Repton, C., Babayan, S.A., Gray, N., Cordaux, R., Darby, A., Makepeace, B., Blaxter, M. (2013) Phylogenomics and analysis of shared genes suggest a single transition to mutualism in *Wolbachia* of nematodes. Genome Biol. Evol. 5, 1668–1674.

[23] Czarnetzki, A.B., Tebbe, C.C. (2004) Detection and phylogenetic analysis of *Wolbachia* in *Collembola*. Environ. Microbiol. 6, 35–44.

[24] Darby, A.C., Armstrong, S.D., Bah, G.S., Kaur, G., Hughes, M.A., Kay, S.M. (2012) Analysis of gene expression from the *Wolbachia* genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. Genome Res. 22, 2467–2477.

[25] Dedeine, F., Vavre, F., Fleury, F., Loppin, B., Hochberg, M.E., Boulétreau, M. (2001) Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. Proc. Natl. Acad. Sci. 98, 6247–6252.

[26] Desjardins, C.A., Cerqueira, G.C., Goldberg, J.M., Dunning-Hotopp, J.C., Haas, B.J., Zucker, J., Ribeiro, J.M.C., Saif, S., Levin, J.Z., Fan, L., Zeng, Q., Russ, C., Wortman, J.R., Fink, D.L., Birren, B.W., Nutman, T.B. (2013) Genomics of *Loa loa*, a *Wolbachia*-free filarial parasite of humans. Nat. Genet. 45, 495–500.

[27] Determann, A., Mehlhorn, H., Ghaffar, F. (1997) Electron microscope observations on *Onchocerca ochengi* and *O. fasciata* (Nematoda: Filarioidea). Parasitol. Res. 83, 591–603.

[28] Duplouy, A., Iturbe-Ormaetxe, I., Beatson, S.A., Szubert, J.M., Brownlie, J.C., McMeniman, C.J., McGraw, E.A., Hurst, G.D.D., Charlat, S., O'Neill, S.L., Woolfit, M. (2013) Draft genome sequence of the male-killing *Wolbachia* strain *w*Bol1 reveals recent horizontal gene transfers from diverse sources. BMC Genomic 14, 20.

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

[30] Ellegaard, K.M., Klasson, L., Näslund, K., Bourtzis, K., Andersson, S.G. (2013) Comparative genomics of *Wolbachia* and the bacterial species concept. PLoS Genet. 9, e1003381.

[31] Egyed, Z., Sréter, T., Széll, Z., Nyirő, G., Dobos-Kovács, M., Márialigeti, K., Varga, I. (2002) Electron microscopic and molecular identification of *Wolbachia* endosymbionts from *Onchocerca lupi*: implications for therapy. Vet. Parasitol. 106, 75–82.

[32] Fenn, K., Blaxter, M. (2004) Are filarial nematode *Wolbachia* obligate mutualist symbionts? Trends Ecol. Evol. 19, 163–166.

[33] Ferri, E., Bain, O., Barbuto, M., Martin, C., Lo, N., Uni, S., Landmann, F., Baccei, S.G., Guerrero, R., de Souza Lima, S., Bandi, C., Wanji, S., Diagne, M., Casiraghi, M. (2011) New insights into the evolution of *Wolbachia* infections in filarial nematodes inferred from a large range of screened species. PLoS ONE 6, e20843.

[34] Fischer, K., Beatty, W.L., Jiang, D., Weil, G.J., Fischer, P.U. (2011) Tissue and stage-specific distribution of *Wolbachia* in *Brugia malayi*. PLoS Negl. Trop. Dis. 5, e1174.

[35] Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., Bhattacharyya, A., Kapatral, V., Kumar, S., Posfai, J., Vincze, T., Ingram, J., Moran, L., Lapidus, A., Omelchenko, M., Kyrpides, N., Ghedin, E., Wang, S., Goltsman, E., Joukov, V., Ostrovskaya, O., Tsukerman, K., Mazur, M., Comb, D., Koonin, E., Slatko, B. (2005) The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. PLoS Biol. 3, e121.

[36] Gerth, M., Gansauge, M-T., Weigert, A., Bleidron, C. (2014) Phylogenomic analyses uncover origin and spread of the *Wolbachia* pandemic. Nat. Commun. 5, 5117.

[37] Glowska, E., Dragun-Damian, A., Dabert, M., Gerth, M. (2015) New *Wolbachia* supergroups detected in quill mites (Acari: Syringophilidae). Infect. Genet. Evol. 30, 140–146.

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

[39] Guerrero, G., Peralta, H., Aguilar, A., Díaz, R., Villalobos, M.A., Medrano-Soto, A., Mora, J. (2005) Evolutionary, structural and functional relationships revealed by comparative analysis of syntenic genes in Rhizobiales. BMC Evol. Biol. 55, 55.

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

[41] Heath, B.D., Butcher, R.D.J., Whitfield, W.G.F., Hubbard, S.F. (1999) Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. Curr. Biol. 9, 313–316.

[42] Hertig, M., Wolbach, S.B. (1924) Studies in *Rickettsia*-like micro-organisms in insects. J. Med. Res. 44, 329–374.

[43] Hertig, M. (1936) The rickettsia, *Wolbachia pipientis* (gen. et sp. n.) and associated inclusions of the mosquito, Culex pipiens. Parasitology 28, 453–486.

[44] Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A., Werren, J.H. (2008) How many species are infected with *Wolbachia*? A statistical analysis of current data. FEMS Microbiol. Lett. 281, 215–220.

[45] Hoerauf, A., Volkmann, L., Nissen-Paehle, K., Schmetz, C., Autenrieth, I., Büttner, D.W., Fleischer, B. (2000) Targeting of *Wolbachia* endobacteria in *Litomosoides sigmodontis*: comparison of tetracyclines with chloramphenicol, macrolides and ciprofloxacin. Trop. Med. Int. Health 5, 275–279.

[46] Hoerauf, A., Mand, S., Adjei, O., Fleischer, B., Büttner, D.W. (2001) Depletion of *Wolbachia* endobacteria in *Onchocerca volvulus* by doxycycline and microfilariadermia after ivermectin treatment. The Lancet 357, 1415–1416.

[47] Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X-Y., Fukatsu, T. (2010) *Wolbachia* as a bacteriocyte-associated nutritional mutualist. PNAS 107, 769–774.

[48] Huigens, M.E., de Almeida, R.P., Boons, P.A.H., Luck, R.F., Stouthamer, R. (2004) Natural interspecific and intraspecific horizontal transfer of parthenogenesis-inducing *Wolbachia* in *Trichogramma* wasps. Proc. Biol. Sci. 271, 509–515.

[49] Jäckel, R., Mora, D., Dobler, S. (2013) Evidence for selective sweeps by *Wolbachia* infections: phylogeny of *Altica* leaf beetles and their reproductive parasites. Mol. Ecol. 22, 4241–4255.

[50] Kent, B.N., Salichos, L., Gibbons, J.G., Rokas, A., Newton, I.L., Clark, M.E., Bordenstein, S.R. (2011) Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. Genome Biol. Evol. 3, 209–218.

[51] Klasson, L., Walker, T., Sebaihia, M., Sanders, M.J., Quail, M.A., Lord, A., Sanders, S., Earl, J., O'Neill, S.L., Thomson, N., Sinkins, S.P., Parkhill, J. (2008) Genome evolution of *Wolbachia* strain *w*Pip from the *Culex pipiens* group. Mol. Biol. Evol. 25, 1877–1887.

[52] Klasson, L., Westberg, J., Sapountzis, P., Näslund, K., Lutnaes, Y., Darby, A.C., Veneti, Z., Chen, L., Braig, H.R., Garrett, R., Bourtzis, K., Andersson, S.G.E. (2009) The mosaic genome structure of the *Wolbachia w*Ri strain infecting *Drosophila simulans*. Proc. Natl. Acad. Sci. U. S. A. 106, 5725–5730.

[53] Konstantinidis, K.T., Roselló-Móra, R. (2015) Classifying the uncultivated microbial majority: a place for metagenomics data in the *Candidatus* proposal. Syst. Appl. Microbiol. 38, 223–230.

[54] Koutsovoulos, G., Makepeace, B., Tanya, V.N., Blaxter, M. (2014) Palaeosymbiosis revealed by genomic fossils of *Wolbachia* in a Strongyloidean nematode. PLoS Genet. 10, e1004397.

[55] Kozek, W.J., Marroquin, H.F. (1977) Intracytoplasmic bacteria in *Onchocerca volvulus*. Am J. Trop. Med. Hyg. 26, 663–678.

[56] Kozek, W.J. (2005) What is new in the *Wolbachia*/*Dirofilaria* interaction? Vet. Parasitol. 133, 127–132.

[57] Landmann, F., Voronin, D., Sullivan, W., Taylor, M.J. (2011) Anti-filarial activity of antibiotic therapy is due to extensive apoptosis after *Wolbachia* depletion from filarial nematodes. PLoS Pathog. 7, e1002351.

[58] Landmann, F., Foster, J.M., Michalski, M.L., Slatko, B.E., Sullivan, W. (2014) Co-evolution between an endosymbiont and its nematode host: *Wolbachia* asymmetric posterior localization and AP polarity establishment. PLoS Negl. Trop. Dis. 8, e3096.

[59] Langworthy, N.G., Renz, A., Mackenstedt, U., Henkle-du, K., Bronsvoort, M.B.C., Tanya, V.N., Donnelly, M.J., Trees, A.J. (2000) Macrofilaricidal activity of tetracycline against the filarial nematode *Onchocerca ochengi*: elimination of *Wolbachia* precedes worm death and suggests a dependent relationship. Proc. R. Soc. Lond. 267, 1063–1069.

[60] Lefoulon, E., Gavotte, L., Junker, K., Barbuto, M., Uni, S., Landmann, F., Laaksonen, S., Saari, S., Nikander, S., Souza, S.D., Casiraghi, M., Bain, O., Martin, C. (2012) A new type F *Wolbachia* from Splendidofilariinae (Onchocercidae) supports the recent emergence of this supergroup. Int. J. Parasitol. 42, 1025–1036.

[61] Lo, N., Paraskevopoulos, C., Bourtzis, K., O'Neill, S.L., Werren, J.H., Bordenstein, S.R., Bandi, C. (2007) Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. Int. J. Syst. Evol. Microbiol. 57, 654–657.

[62] Louis, C., Nigro, L. (1989) Ultrastructural evidence of *Wolbachia* rickettsiales in *Drosophila simulans* and their relationships with unidirectional cross-incompatibility. J. Invertebr. Pathol. 24, 39–44.

[63] Mavingui, P., Valiente Moro, C., Tran-Van, V., Wisniewski-Dyé, F., Raquin, V., Minard, G., Tran, F.H., Voronin, D., Rouy, Z., Bustos, P., Lozano, L., Barbe, V., González, V. (2012) Whole-genome sequence of *Wolbachia* strain *w*AlbB, an endosymbiont of tiger mosquito vector *Aedes albopictus*. J. Bacteriol. 194, 1840.

[64] McLaren, D.J., Worms, M.J., Laurence, B.R., Simpson, M.G. (1975) Micro-organisms in filarial larvae (Nematoda). Trans. R. Soc. Trop. Med. Hyg. 69, 509–514.

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

[66] Metcalf, J.A., Jo, M., Bordenstein, S.R., Jaenike, J., Bordenstein, S.R. (2014) Recent genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism. PeerJ 2, e529.

[67] Moran, N.A. (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U. S. A. 93, 2873–2878.

[68] Niehuis, O., Hartig, G., Grath, S., Pohl, H., Lehmann, J., Tafer, H., Donath, A., Krauss, V., Eisenhardt, C., Hertel, J., Petersen, M., Mayer, C., Meusemann, K., Peters, R.S., Stadler, P.F., Beutel, R.G., Bornberg-Bauer, E., McKenna, D.D., Misof, B. (2012) Genomic and morphological evidence converge to resolve the enigma of *Strepsiptera*. Curr. Biol. 22, 1309–1313.

[69] Nikoh, N., Hosokawa, T., Moriyama, M., Oshima, K., Hattori, M., Fukatsu, T. (2014) Evolutionary origin of insect-*Wolbachia* nutritional mutualism. Proc. Natl. Acad. Sci. U. S. A. 111, 10257–10262.

[70] Ormeño-Orrillo, E., Martínez-Romero, E. (2013) Phenotypic tests in *Rhizobium* species description: an opinion and (a sympatric speciation) hypothesis. Syst. Appl. Microbiol. 36, 145–147.

[71] Ormeño-Orrillo, E., Servín-Garcidueñas, L.E., Rogel, M.A., González, V., Peralta, H., Mora, J., Martínez-Romero, J., Martínez-Romero, E. (2014) Taxonomy of rhizobia and agrobacteria from the Rhizobiaceae family in light of genomics. Syst. Appl. Microbiol. 38, 287–291.

[72] Pfarr, K., Foster, J., Slatko, B., Hoerauf, A., Eisen, J.A. (2007) On the taxonomic status of the intracellular bacterium *Wolbachia pipientis*: should this species name include the intracellular bacteria of filarial nematodes? Int. J. Syst. Evol. Microbiol. 57, 1677–1678.

[73] Pike, N., Kingcombe, R. (2009) Antibiotic treatment leads to the elimination of *Wolbachia* endosymbionts and sterility in the diplodiploid collembolan *Folsomia candida*. BMC Biol. 7, 54.

[74] Pinto, S.B., Stainton, K., Harris, S., Kambris, Z., Sutton, E.R., Bonsall, M.B., Parkhill, J., Sinkins, S.P. (2013) Transcriptional regulation of *Culex pipiens* mosquitoes by *Wolbachia* influences cytoplasmic incompatibility. PLoS Pathog. 9, e1003647.

[75] Richter, M., Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. 106, 19126–19131.

[76] Rieseberg, L.H. (2001) Chromosomal rearrangements and speciation. Trends Ecol. Evol. 16, 351–358.

[77] Rogel, M.A., Bustos, P., Santamaría, R.I., González, V., Romero, D., Cevallos, M.A., Lozano, L., Castro-Mondragón, J., Martínez-Romero, J., Ormeño-Orrillo, E., Martínez-Romero, E. (2014) Genomic basis of symbiovar mimosae in *Rhizobium etli*. BMC Genomic 15, 575.

[78] Ros, V.I., Fleming, V.M., Feil, E.J., Breeuwer, J.A. (2009) How diverse is the genus *Wolbachia*? Multiple-gene sequencing reveals a putatively new *Wolbachia* supergroup recovered from spider mites (Acari: Tetranychidae). App. Environ. Microbiol. 75, 1036–1043.

[79] Rosenblueth, M., Sayavedra, L., Sámano-Sánchez, H., Roth, A., Martínez-Romero, E. (2012) Evolutionary relationships of flavobacterial and enterobacterial endosymbionts with their scale insect hosts (Hemiptera: Coccoidea). J. Evol. Biol. 25, 2357–2368.

[80] Rousset, F., Bouchon, D., Pintureau, B., Juchault, P., Solignac, M. (1992) *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. Proc. R. Soc. Lond. B 250, 91–98.

[81] Rowley, S.M., Raven, R.J., McGraw, E.A. (2004) *Wolbachia pipientis* in Australian spiders. Curr. Microbiol. 49, 208–214.

[82] Sacchi, L., Genchi, M., Clementi, E., Negri, I., Alma, A., Ohler, S., Sassera, D., Bourtzis, K., Bandi, C. (2010) Bacteriocyte-like cells harbour *Wolbachia* in the ovary of *Drosophila melanogaster* (Insecta, Diptera) and *Zyginidia pullula* (Insecta, Hemiptera). Tissue Cell 42, 328–333.

[83] Saha, S., Hunter, W.B., Reese, J., Morgan, J.K., Marutani-Hert, M., Huang, H., Lindeberg, M. (2012) Survey of endosymbionts in the *Diaphorina citri* metagenome and assembly of a *Wolbachia w*Di draft genome. PLoS ONE 7, e50067.

[84] Salzberg, S.L., Dunning Hotopp, J.C., Delcher, A.L., Pop, M., Smith, D.R., Eisen, M.B., Nelson, W.C. (2005) Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. Genome Biol. 6, R23.

[85] Salzberg, S.L., Puiu, D., Sommer, D.D., Nene, V., Lee, N.H. (2009) Genome sequence of the *Wolbachia* endosymbiont of *Culex quinquefasciatus* JHB. J. Bacteriol. 191, 1725.

[86] Schilthuizen, M., Stouthamer, R. (1997) Horizontal transmission of parthenogenesis-inducing microbes in *Trichogramma* wasps. Proc. R. Soc. Lond. B 264, 361–366.

[87] Schröder, D., Deppisch, H., Obermayer, M., Krohne, G., Stackebrandt, E., Hôlldobler, B., Goebel, W., Gross, R. (1996) Intracellular endosymbiotic bacteria of *Camponotus* species (carpenter ants): systematics, evolution and ultrastructural characterization. Mol. Microbiol. 21, 479–489.

[88] Shoemaker, D.D., Machado, C.A., Molbo, D., Werren, J.H., Windsor, D.M., Herre, E.A. (2002) The distribution of *Wolbachia* in fig wasps: correlations with host phylogeny, ecology and population structure. Proc. R. Soc. Lond. B 269, 2257–2267.

[89] Siozios, S., Cestaro, A., Kaur, R., Pertot, I., Rota-Stabelli, O., Anfora, G. (2013) Draft genome sequence of the *Wolbachia* endosymbiont of *Drosophila suzukii*. Genome Announc. 1, e00032–13.

[90] Sironi, M., Bandi, C., Sacchi, L., Di Sacco, B., Damiani, G., Genchi, C. (1995) Molecular evidence for a close relative of the arthropod endosymbiont *Wolbachia* in a filarial worm. Mol. Biochem. Parasitol. 74, 223–227.

[91] Sutton, E.R., Harris, S.R., Parkhill, J., Sinkins, S.P. (2014) Comparative genome analysis of *Wolbachia* strain *w*Au. BMC Genomic 15, 928.

[92] Taylor, M.J., Bilo, K., Cross, H.F., Archer, J.P., Underwood, A.P. (1999) 16S rDNA phylogeny and ultrastructural characterization of *Wolbachia* intracellular bacteria of the filarial nematodes *Brugia malayi*, *B. pahangi*, and *Wuchereria bancrofti*. Exp. Parasitol. 91, 356–361.

[93] Taylor, M.J., Bandi, C., Hoerauf, A.M., Lazdins, J. (2000) *Wolbachia* bacteria of filarial nematodes: a target for control? Parasitol. Today 16, 179–180.

[94] Teixeira, L., Ferreira, A., Ashburner, M. (2008) The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. PLoS Biol. 6, e1000002.

[95] Timmermans, M.J.T.N., Ellers, J. (2009) *Wolbachia* endosymbiont is essential for egg hatching in a parthenogenetic arthropod. Evol. Ecol. 23, 931–942.

[96] Vandamme, P., Peeters, C. (2014) Time to revisit polyphasic taxonomy. Antonie van Leeuwenhoek 106, 57–65.

[97] Vandekerckhove, T.T., Watteyne, S., Willems, A., Swings, J.G., Mertens, J., Gillis, M. (1999) Phylogenetic analysis of the 16S rDNA of the cytoplasmic bacterium *Wolbachia* from the novel host *Folsomia candida* (Hexapoda, Collembola) and its implications for wolbachial taxonomy. FEMS Microbiol. Lett. 180, 279–286.

[98] Werren, J.H., Windsor, D., Guo, L.R. (1995) Distribution of *Wolbachia* among neotropical arthropods. Proc. R. Soc. Lond. B 262, 197–204.

[99] Werren, J.H., Bartos, J.D. (2001) Recombination in *Wolbachia*. Curr. Biol. 11, 431–435.

[100] Werren, J.H., Baldo, L., Clark, M.E. (2008) *Wolbachia*: master manipulators of invertebrate biology. Nat. Rev. Microbiol. 6, 741–751.

[101] Woolfit, M., Iturbe-Ormaetxe, I., Brownlie, J.C., Walker, T., Riegler, M., Seleznev, A., Popovici, J., Rancès, E., Wee, B.A., Pavlides, J., Sullivan, M.J., Beatson, S.A., Lane, A., Sidhu, M., McMeniman, C.J., McGraw, E.A., O'Neill, S.L. (2013)

Genomic evolution of the pathogenic *Wolbachia* strain, *w*MelPop. Genome Biol. Evol. 5, 2189–2204.

[102] Wu, M., Sun, L.V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J.C., McGraw, E.A., Martin, W., Esser, C., Ahmadinejad, N., Wiegand, C., Madupu, R., Beanan, M.J., Brinkac, L.M., Daugherty, S.C., Durkin, A.S., Kolonay, J.F., Nelson, W.C., Mohamoud, Y., Lee, P., Berry, K., Young, M.B., Utterback, T., Weidman, J., Nierman, W.C., Paulsen, I.T., Nelson, K.E., Tettelin, H., O'Neill, S.L., Eisen, J.A. (2004) Phylogenomics of the reprodutive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. 2, 327–341.

[103] Wu, M., Scott, A.J. (2012) Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics 28, 1033–10334.

[104] Yang, X.H., Zhu, D.H., Liu, Z., Zhao, L., Su, C.Y. (2013) High levels of multiple infections, recombination and horizontal transmission of *Wolbachia* in the *Andricus mukaigawae* (Hymenoptera; Cynipidae) communities. PLoS ONE 8, e78970.

[105] Zhukova, M., Kiseleva, E. (2012) The virulent *Wolbachia* strain *w*MelPop increases the frequency of apoptosis in the female germline cells of *Drosophila melanogaster*. BMC Microbiol. 12, S15.

[106] Zug, R., Hammerstein, P. (2012) Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. PLoS ONE 7, e38544.

frontiers in Genetics

REVIEW
published: 17 December 2015
doi: 10.3389/fgene.2015.00348

The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics

*Alejandra Escobar-Zepeda[1], Arturo Vera-Ponce de León[2] and Alejandro Sanchez-Flores[1]**

[1] Unidad de Secuenciación Masiva y Bioinformática, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, México, [2] Programa de Ecología Genómica, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, México

The study of microorganisms that pervade each and every part of this planet has encountered many challenges through time such as the discovery of unknown organisms and the understanding of how they interact with their environment. The aim of this review is to take the reader along the timeline and major milestones that led us to modern metagenomics. This new and thriving area is likely to be an important contributor to solve different problems. The transition from classical microbiology to modern metagenomics studies has required the development of new branches of knowledge and specialization. Here, we will review how the availability of high-throughput sequencing technologies has transformed microbiology and bioinformatics and how to tackle the inherent computational challenges that arise from the DNA sequencing revolution. New computational methods are constantly developed to collect, process, and extract useful biological information from a variety of samples and complex datasets, but metagenomics needs the integration of several of these computational methods. Despite the level of specialization needed in bioinformatics, it is important that life-scientists have a good understanding of it for a correct experimental design, which allows them to reveal the information in a metagenome.

Keywords: metagenomics, bioinformatics, high-throughput sequencing, taxonomy, functional genomics, microbiology

OPEN ACCESS

Edited by:
Yasset Perez-Riverol, European Bioinformatics Institute, UK

Reviewed by:
Philippe Rocca-Serra, Oxford e-Research Centre, UK
Christian M. Zmasek, Sanford-Burnham Medical Research Institute, USA

***Correspondence:**
Alejandro Sanchez-Flores
alexsf@ibt.unam.mx

Specialty section:
This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Genetics

Received: *05 June 2015*
Accepted: *27 November 2015*
Published: *17 December 2015*

Citation:
Escobar-Zepeda A, Vera-Ponce de León A and Sanchez-Flores A (2015) The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. Front. Genet. 6:348. doi: 10.3389/fgene.2015.00348

BRIEF HISTORY OF MICROBIAL COMMUNITIES STUDY

From various definitions of microbial communities, the one proposed by Begon et al. (1986) defines it as the set of organisms (in this case, microorganisms) coexisting in the same space and time. The study of microbial communities has changed from the first report of microbes made by Leeuwenhoek and their oral organisms in 1676 (Schierbeek, 1959), to the characterization using the current molecular techniques. Pioneer scientists tried to isolate these "invisible" organisms, and like Robert Koch, they started by using nutrients in a solid phase like potato slices or gelatine to cultivate and isolate microorganisms in order to count and visualize them. Ultimately, these isolation techniques helped scientists to understand the microorganisms' physiologies (Blevins and Bronze, 2010).

Soon, the microscope became the principal tool to study microorganisms and their interactions. Development of practical staining techniques such as Gram, Ziehl–Neelsen, and Schaeffer and Fulton (Beveridge, 2001; Blevins and Bronze, 2010) significantly improved the resolution of microscopy techniques. Something evident to microbiologist was that the number of observed

microorganisms in a microscope did not correspond with number of microorganism obtained in culture plates (Staley and Konopka, 1985). Although the explanation to this observation was not evident at that time, the conclusion was that the microorganisms need special conditions to grow, and based on this, Winogradsky emulated environments for culture media production that resembled native growing conditions (McFall-Ngai, 2008). Winogradsky's ideas and his contribution to ecology revolutionized microbiology and gave birth to a new concept named "microbial ecology," which refers to the study of microorganisms and their environmental roles (Ackert, 2012).

For almost 300 years (**Figure 1**), the study of microorganisms was based on morphology features, growth, and selection of some biochemical profiles (Roszak et al., 1984; Oliver et al., 1991; Colwell et al., 1996). These techniques provided an insight into the microbial world, but nowadays, they provide only a limited resolution for other applications.

In the late 1970s, Carl Woese proposed the use of ribosomal RNA genes as molecular markers for life classification (Woese and Fox, 1977). This idea in conjunction with the Sanger automated sequencing (Sanger et al., 1977) method revolutionized the study and classification of microorganisms. Some decades later, advances in molecular techniques were applied to microbial diversity description and granted access to a "new uncultured world" of microbial communities. Some of these techniques, which had a remarkable impact, were the polymerase chain reaction (PCR), rRNA genes cloning and sequencing, fluorescent *in situ* hybridization (FISH), denaturing gradient gel electrophoresis (DGGE and TGGE), restriction-fragment length polymorphism, and terminal restriction-fragment length polymorphism (T-RFLP). However, in spite all these improvements, there were many other observations in microbiology that remained unanswered like those related to the microorganisms' metabolic and ecological function. Characterization of certain functions in a particular environment was possible only after gene cloning from total DNA of a certain habitat and when its heterologous expressed product was associated with a given metabolic function (i.e., nitrogenases, cellulases, oxidoreductases, laccases, etc.). This implied the development of gene expression techniques using other microorganisms as systems to test gene function and roles in the microbial community. In addition, a window of opportunity was open to discover new genes, functions, and metabolic products with technological application, thereby giving birth to biotechnology. Products such as "terragines" from *Streptomyces lividians* (Wang et al., 2000) or genes related to broad-spectrum antibiotics were cloned from soil-DNA libraries (Gillespie et al., 2002) were achievements that set the foundation to a new area named "metagenomics analysis," which was later defined as the theoretical collection of all genomes from members in a microbial community from a specific environment (Handelsman et al., 1998). Even if these approaches led to the discovery of new molecules and identification of new microbial communities members (Giovannoni et al., 1990), more recently, some problems have been spotted. Cloning biases (Morgan et al., 2010), sampling biases, misidentification of "decorating enzymes" and incorrect promoter sites in genomes, and dispersion of genes involved in secondary metabolite production (Keller and Zengler, 2004) are some of the problems found in metagenomics. Therefore, it is important to evaluate

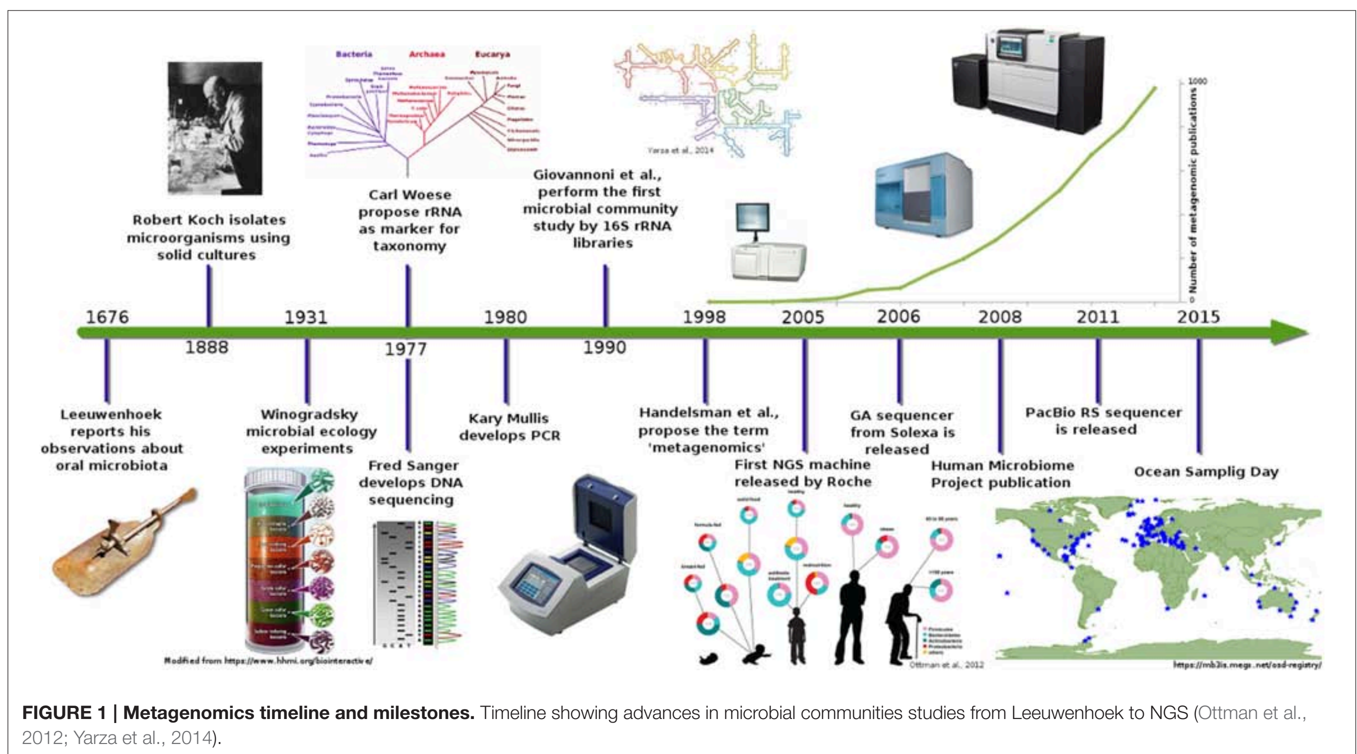


FIGURE 1 | Metagenomics timeline and milestones. Timeline showing advances in microbial communities studies from Leeuwenhoek to NGS (Ottman et al., 2012; Yarza et al., 2014).

and correct these biases with statistical methods to have a better understanding of the species richness and know the difference between the expected and the observed microbial diversity.

CONCEPTS OF MICROBIAL DIVERSITY AND SPECIES RICHNESS

"Species diversity" is an attribute of any biological community (Krebs, 2014), but how we quantify it, is not trivial. The simplest idea to describe and quantify a microbial community (e.g., a metagenome) is the species richness concept, which refers to the number of species in a specified region. Another idea that can be applied to metagenomics is the evenness concept or differential abundance proposed by Simpson (1949). The evenness measurement attempts to quantify the unequal representation in communities where there are few dominant species and many species that are relatively uncommon. This could be tested against a hypothetical community in which all species are equally common. Therefore, when comparing two communities, if both have the same number of species (equal species richness) but different abundances, then the consortia with the shortest difference between the observed and hypothetical distribution (even abundance) will be the more diverse. Hence, it should be considered that species richness should not be the only parameter to define diversity.

In order to describe and compare communities in a better way, there are other metrics that have been adapted to metagenomics and that can complement the aforementioned. Alpha (α) is a metric for local diversity of a community; opposite to it, we have Gamma (γ), which measures the total regional diversity that includes many communities, and finally Beta (β) metric tells us how different community samples are in an area, linking Alpha and Gamma metrics (Krebs, 2014).

In the Alpha diversity assessment, the accumulation of species or Operational Taxonomic Units (OTUs) plots have been used to evaluate the sample efficiency and to correct sampling problems. Although a species accumulation curve could present an asymptotic trend after using a bigger sample size, the maximum species number could not be reached. This is why a statistical approach has to be performed, i.e., rarefaction curves, which are useful to estimate the real maximum species or OTUs number observed in the sample and to compare samples with different sizes (Sanders, 1968; Heck et al., 1975; Colwell and Coddington, 1994).

Another alternative to calculate species diversity quantitatively is the use of statistical estimators. Particularly, non-parametric estimators have been used for microbial communities' studies. These estimators do not depend on the statistical behavior of the sample and can consider low abundance species. On one hand, the simplest non-parametric diversity estimator is the Simpson's index (D), which is based on the probability of assigning two independent individuals taken randomly from the community into the same species (Simpson, 1949). On the other hand, Shannon–Wiener function or Shannon–Weaver index H' (Shannon, 1948) is an entropy measurement that increases with the number of species in the sample. Simpson and Shannon–Wiener indices are used as heterogeneity measurements and differ mainly in calculation of the taxa abundance for the final richness estimation. Simpson index gives a higher weight to species with more frequency in a sample, whereas Shannon–Wiener gives more weight to rare species (Krebs, 2014).

The development of molecular biology provided a new vision of microbial ecology and allowed the study of highly complex communities in a short period of time. However, the application of diversity estimators in metagenomics projects has been evaluated by some authors with divided ideas about their results.

Some authors concluded that microbial diversity estimation based on molecular markers is possible and can be used for comparison with some precautions (Gihring et al., 2012). They recommended the use of Simpson or Shannon–Wiener estimators as the best descriptors for species richness at high-level taxa in metagenomes (Haegeman et al., 2013; Chernov et al., 2015). However, in nature, the microbial communities have a large number of rare species that can be detected only if an exhaustive sampling is performed (Colwell and Coddington, 1994; Kemp and Aller, 2004; Bonilla-Rosso et al., 2012). Therefore, the use of such estimators is unsuccessful for very complex microbial communities. This problem has generated the creation of new diversity indexes for species that analyse statistically the behavior of the sample. For example, the tail statistic (τ) estimates the number of undiscovered species from a rank abundance curve, giving a higher weight to the low abundant taxa and increasing the sensitivity of the analysis of complex samples (Li et al., 2012).

The use of diversity indexes is a better approach to quantify and compare microbial diversity among samples. Such comparison should be done cautiously because it could be uninformative unless biases related to sampling and criteria for species or OTU definition are minimized (Bonilla-Rosso et al., 2012).

NEXT GENERATION SEQUENCING TECHNOLOGIES TO EXPLORE MICROBIAL COMMUNITIES

As previously mentioned, Sanger sequencing technology had a great impact on the early stage of microbial community studies. Nowadays, the sequencing yield and sequence length have changed a lot since Sanger sequencing (**Table 1**). Currently, Sanger sequencing can retrieve up to 96 sequences per run with an average length of 650 bp, which might be enough for phylogenetic marker analysis. However, low-cost platforms known as Next Generation Sequencing technologies (NGS) are capable of parallel sequencing millions of DNA molecules with different yields and sequence lengths (**Table 1**; Logares et al., 2012; Fichot and Norman, 2013; Glenn, 2014; Sanchez-Flores and Abreu-Goodger, 2014) having a positive impact in different areas.

The first of these technologies that revolutionized the genomics and metagenomics areas was the 454 sequencing platform or "pyrosequencing." The principle of this technology is

TABLE 1 | Direct comparison among sequencing technologies suitable for metagenomics.

	Roche 454	IonTorrent PGM	Illumina	PacBio RSII[a]
Maximum read length (bp)	1200	400	300[b]	50,000
Output per run (Gb)	1	2	1000[c]	1
Amplification for library construction	Yes	Yes	Yes	No
Cost/Gb (USA Dollar)	\$9538.46	\$460.00	\$29.30	\$600
Error kind	Indel	Indel	Substitution	Indel
Error rate (%)	1	∼1	∼0.1	∼13
Run time	20 h	7.3 h	6 days	2 h

Adapted from Glenn, T. 2014 NGS Field Guide—Table 2a—Run time, Reads, Yield|The Molecular Ecologist. Available online at: http://www.molecularecologist.com/next-gen-fieldguide-2014/ (Accessed Aug 17, 2015).
[a]P6-C4 chemistry.
[b]MiSeq read length.
[c]Illumina HiSeq 2500 Dual flowcell yield.

a one-by-one nucleotide addition cycle, where the pyrophosphate (PPi) released from the DNA polymerization reaction is transformed in a luminous signal. The light emission from a plate with millions of microwells containing a given DNA fragment is detected by the machine and is translated to nucleotide sequences with an associated base quality value (Margulies et al., 2005). This technology offered a higher yield than Sanger sequencing at a lower cost but with shorter read lengths (**Table 1**). The main bias of this technology is artificial insertions and deletions due to long homopolymeric regions. In spite of the advantages that this technology provided to metagenomics, it is now obsolete. Recent announcements by Roche (current owner of the technology) reported the shutdown of 454 division, ceasing the platform support by mid-2016 (Karow, 2013). Nevertheless, all the software that has been developed so far to analyse 454 data could be adapted to analyse data obtained by another platforms.

The Ion Torrent platform is an analogous technology to 454 that produces a similar yield and a read length to those obtained at its middle stage of development. The Ion Torrent PGM is considered as the smallest potentiometer that exists and can detect the change in hydrogen potential generated each time a proton is released after a nucleotide is added in the sequencing reaction occurring in millions of microwells (Rothberg et al., 2011). The maximum Ion Torrent yield is ∼500 million reads with a mode length of 400 bp (**Table 1**) (Glenn, 2014). In this case, there is a clear benefit in terms of cost reduction, since Ion Torrent sequencing is just a tenth of the pyrosequencing cost (Whiteley et al., 2012).

However, read length reduction in return for higher yields and error-rates is another trade-off observed in some platforms in order to reduce the sequencing costs, i.e., the case of the Illumina technology, which has become one of the most popular technologies due to its low cost and high yield. The basis of Illumina chemistry is the reversible-termination sequencing by synthesis with fluorescently labeled nucleotides. In a nutshell, DNA fragments are attached and distributed in a flow cell, where the sequencing reaction occurs by adding a labeled nucleotide. When the labeled nucleotide is incorporated and its fluorescent molecule is excited by a laser, the signal is registered by the machine. Afterwards, the fluorophore molecule is removed and the next nucleotide can be incorporated. DNA fragments can be sequenced from one or both sides giving single end or pair-end sequencing, respectively, with a maximum read length of 300 base pairs per read (Bennett, 2004). The output of this technology is currently the highest among the second generation sequencing technologies and makes it suitable for multiplexing hundreds of samples (**Table 1;** Glenn, 2014).

Currently, the technologies already mentioned are the most used for metagenome projects, but the development of sequencing was kept going for the last 5 years in order to solve the known biases of these technologies and to offer a better trade-off between yield, cost, and read length. At present, the so called third generation sequencing technologies such as PacBio RS from Pacific Bioscience (Fichot and Norman, 2013) or the Oxford Nanopore (Kasianowicz et al., 1996), which are single-molecule, real-time technologies, reduced the amplification bias and also the short read length problem. The time and cost reduction offered by these technologies is also a valuable asset. However, the error rate is higher compared to other technologies but correctable if the sequencing depth is high enough. In terms of computational tools, there is virtually no software that can be used for metagenomics analysis.

One of the great improvements of second and third generation sequencing technologies is that the library preparation does not require DNA cloning vectors or bacterial hosts, simplifying the library preparation and reducing DNA contamination from other organisms that are not part of the metagenome.

Although new generation sequencing technologies are powerful and have allowed us to discover novel microbial worlds and explore new environments, they present particular limitations and biases that have to be circumvented (**Table 1**). It is important to consider that data obtained from second or third generation sequencing technologies have certain computational requirements for their analysis. The bigger the dataset generated, the higher computational resources and more complex bioinformatics analyses are necessary. In addition, large data storage is needed to archive and process the data (Logares et al., 2012). In terms of bioinformatic analysis, not only high-end servers are required but also UNIX operative system skills are needed. Programming and scripting knowledge are desirable to run and install the available metagenomics software for parsing and interpreting the results. Thus, it is suggested that biologists or biological scientists should develop basic computational skills in order to take an advantage of metagenomic data.

Quality Control (QC) Procedures for Metagenomics

Assessing the output quality from any of the previously mentioned sequencing technologies will be always a crucial step before starting any analysis. Each sequencing platform presents a particular bias product of the intrinsic mechanism to detect each nucleotide, which conforms the DNA polymer that is being

analyzed (**Table 1**). The error rate from each technology varies, affecting the characterization of a microbial community (Luo et al., 2012). Filtering low quality reads considerably improves metagenome analyses such as taxonomical classification and α and β diversity calculation (Bokulich et al., 2013). There are several programs that can be used for sequencing read QC analysis as described in **Table 1**. In general, they provide information about the sequencing output (number or reads, length, GC content, overrepresented sequences, etc.) and some of them include tools to modify the reads (adapter removal, quality filtering or trimming). These QC operations need an interpretation depending on the analysis. For example, a GC content analysis can be used to anticipate the presence of organisms with different GC content, but a single GC distribution does not imply that our sample has very low diversity, just a bias toward the GC content of the most abundant organisms. Removal of low quality bases or entire reads can be beneficial in terms of mapping, but for metagenome assembly (or any other genome assembly), none of the current assembly programs use or interpret base quality within the assembly process. For Illumina sequencing, removal of optical or PCR duplicates can increase the quality of abundance analysis from whole metagenome shotgun DNA sequencing. However, this QC control has no sense at all in amplicon sequence analysis. Therefore, there are some compulsory QC processes that need to be performed before analysing our data, but depending on the approach, we have to design specific QC steps to improve our results.

RECONSTRUCTING THE GENOMIC CONTENT OF THE MICROBIAL COMMUNITY FROM NGS DATA

The main questions to answer in microbial ecology are "Who is out there?" and "What are they doing?" In fact, metagenomics can answer both questions. Particularly, microbial diversity can be determined using two different approaches: (1) Amplicon sequencing or (2) Shotgun metagenomics. In the first approach, specific regions of DNA from communities are amplified using taxonomical informative primer targets such as 16S rRNA gene for prokaryotes and intergenic transcribed spacers (ITS) or the large ribosomal subunit (LSU) gene for eukaryotes (Sharpton, 2014; Tonge et al., 2014). In the second approach, shotgun metagenomics can help to reconstruct large fragments or even complete genomes from organisms in a community without previous isolation, allowing the characterization of a large number of coding and non-coding sequences that can be used as phylogenetic markers.

Amplicon Sequencing Analysis

First of all, the term "metagenomics" should not be used to refer amplicon sequence analysis, as this analysis is based on just one gene instead of the collection of all the genes in the available genomes from all the organisms in a sample. A better term proposed is "metaprofiling," and it should be interpreted in the rest of this text as the study of all members in a microbial community based on one gene or marker (i.e., 16S rRNA gene) for taxonomy or phylogenetic purposes.

Metaprofiling has been widely used due to its convenience to perform taxonomic and phylogenetic classification in large and complex samples within organisms from different life domains. In addition, it could be performed using almost all mentioned sequencing technologies (**Table 1**).

Moneywise, metaprofiling is currently the best option for 16S rRNA amplicon library preparation and sequencing by platforms such as the Illumina MiSeq or the Ion Torrent PGM. These benchtop sequencers allow microbial ecologists to perform diversity studies at their labs, using multiple replicates and samples from longitudinal time studies. Previous comparisons between HiSeq 2000 and MiSeq technologies have shown that despite the yield difference between them (>50 Gb per day against 5 Gb), the number of OTUs obtained are not significantly different on using both the technologies (Caporaso et al., 2012; Luo et al., 2012).

The advantages of amplicon sequencing are contrasted by the bias generated from using only one phylogenetic marker such as the 16S ribosomal gene or a variable region from it. Some of the pitfalls are low resolution at the species level (Petrosino et al., 2009; Nalbantoglu et al., 2014), a range in gene copy number in many species (Acinas et al., 2004), horizontal transfer of 16S rRNA genes (Schouls et al., 2003; Bodilis et al., 2012), and the fact that <0.1% of the total genome are ribosomal genes, hindering the amplification of this marker from very low abundant genomes in a sample.

The ribosomal genes as phylogenetic markers have been used for the last 40 years or so, resulting in a wide representation of this marker in many databases, allowing the taxonomic annotation of almost any microorganisms present in a metagenomic sample. Some database examples are Greengenes (DeSantis et al., 2006), the Ribosomal Database Project (Wang et al., 2007), and Silva (Quast et al., 2013). The latter includes a great catalog of eukaryotic LSU sequences and is convenient to analyse fungi or other metazoan microorganisms. However, amplicon-dependent techniques are prone to sequencing errors, such as result discrepancy from using different ribosomal variable regions, primers bias, and OTU assignment errors (Fox et al., 1992; Logares et al., 2012; Poretsky et al., 2014).

Most of the earlier amplicon analysis programs were designed for Sanger or 454 ribosomal pyrotag sequences. For example, Mothur (Schloss et al., 2009), QIIME (Caporaso et al., 2010), MEGAN (Huson and Weber, 2013), and CARMA (Krause et al., 2008) are some of the legacy software still available. Nowadays, the software development for metagenomics considers short sequences like Illumina reads or very long sequences such as PacBio reads (**Table 2**).

Once the species level taxonomic annotation objective is covered, metagenome projects can focus on the functional information mining. This could be achieved from the taxonomical information by extrapolating the functional annotation of related reference genomes (De Filippo et al., 2012). To our knowledge, PICRUSt (Langille et al., 2013) is the only available software that connects the taxonomic classification from metaprofiling results with metabolic information (**Table 2**).

TABLE 2 | Examples of software used in metagenomic and metaprofiling analysis.

Software	Application	References	Link (website)
FastQC	Quality control tool for high-throughput sequence data using modular options and giving graphic results of quality per base sequence, GC content, N numbers, duplication, and over represent	Andrews, 2015	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Fastx-Toolkit	Command line tools for Short-reads quality control. These allow processing, cutting, format conversion, and collapsing by sequence length and identity	NP	http://hannonlab.cshl.edu/fastx_toolkit/index.html
PRINTSEQ	Quality control tool for sequence trimming based in dinucleotide occurrence and sequence duplication (mainly 5′/3′)	Schmieder and Edwards, 2011	http://prinseq.sourceforge.net/
NGS QC Toolkit	Tool for quality control analysis performed in parallel environment	Patel and Jain, 2012	http://www.nipgr.res.in/ngsqctoolkit.html
Meta-QC-Chain	Parallel environment tool for quality control. This performs a mapping against 18S rRNA databases for removing eukaryotic contaminant sequences	Zhou et al., 2014	http://www.computationalbioenergy.org/qc-chain.html
Mothur	From reads quality analysis to taxonomic classification, calculus of diversity estimators and ribosomal gene metaprofiling comparison	Schloss et al., 2009	http://www.mothur.org/
QIIME	Quality pre-treatment of raw reads, taxonomic annotation, calculus of diversity estimators, and comparison of metaprofiling or metagenomic data	Caporaso et al., 2010	http://qiime.org/
MEGAN	Taxonomy and functional analysis of metagenomic reads. It based on BLAST output of short reads and performs comparative metagenomics. Graphical interface	Huson and Weber, 2013	http://ab.inf.uni-tuebingen.de/software/megan5/
CARMA	Phylogenetic classification of reads based on Pfam conserved domains	Krause et al., 2008	http://omictools.com/carma-s1021.html
PICRUSt	Predictor of metabolic potential from taxonomic information obtained of 16S rRNA metaprofiling projects	Langille et al., 2013	http://picrust.github.io/picrust/
Parallel-meta	Taxonomic annotation of ribosomal gene markers sequences obtained by metaprofiling or metagenomic reads. Functional annotation based on BLAST best hits results. Comparative metagenomics	Su et al., 2014	http://www.computationalbioenergy.org/parallel-meta.html
MOCAT	Pipeline that includes quality treatment of metagenomic reads, taxonomic annotation based on single copy marker genes classification, and gene-coding prediction	Kultima et al., 2012	http://vm-lux.embl.de/~kultima/MOCAT2/index.html
TETRA	Taxonomic classification by comparison of tetranucleotide patterns. Web service available	Teeling et al., 2004	http://omictools.com/tetra-s1030.html
PhylophytiaS	Composition-based classifier of sequences based on reference genomes signatures	McHardy et al., 2007	http://omictools.com/phylopythia-s1455.html
MetaclusterTA	Taxonomic annotation based on binning of reads and contigs. Dependent of reference genomes	Wang et al., 2014	http://i.cs.hku.hk/~alse/MetaCluster/
MaxBin	Unsupervised binning of metagenomic short reads and contigs	Wu et al., 2014	http://sourceforge.net/projects/maxbin/
Amphora and Amphora2	Metagenomic phylotyping by single copy phylogenetic marker genes classification	Wu and Eisen, 2008; Wu and Scott, 2012	http://pitgroup.org/amphoranet/
BWA	Algorithm for mapping short-low-divergent sequences to large references. Based on Burrows–Wheeler transform	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/
Bowtie	Fast short read aligner to long reference sequences based on Burrows–Wheeler transform	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/index.shtml
Genometa	Taxonomic and functional annotation of short-reads metagenomic data. Graphical interface	Davenport and Tümmler, 2013	http://genomics1.mh-hannover.de/genometa/
SORT-Items	Taxonomic annotation by alignment-based orthology of metagenomic reads	Monzoorul Haque et al., 2009	http://metagenomics.atc.tcs.com/binning/SOrt-ITEMS

(Continued)

TABLE 2 | Continued

Software	Application	References	Link (website)
DiScRIBinATE	Taxonomic assignment by BLASTx best hits classification of reads	Ghosh et al., 2010	http://metagenomics.atc.tcs.com/binning/DiScRIBinATE/
IDBA-UD	Assembler *de novo* of metagenomic sequences with uneven depth	Peng et al., 2012	http://i.cs.hku.hk/~alse/hkubrg/projects/idba_ud/
MetaVelvet	*De novo* assembler of metagenomic short reads	Namiki et al., 2012	http://metavelvet.dna.bio.keio.ac.jp/
Ray Meta	Assembler of *de novo* of metagenomic reads and taxonomy profiler by Ray Communities	Boisvert et al., 2012	http://denovoassembler.sourceforge.net/
MetaGeneMark	Gene coding sequences predictor from metagenomic sequences by heuristic model	Zhu et al., 2010	http://exon.gatech.edu/index.html
GlimmerMG	Gene coding sequences predictor from metagenomic sequences by unsupervised clustering	Kelley et al., 2012	http://www.cbcb.umd.edu/software/glimmer-mg/
FragGeneScan	Gene coding sequences predictor from short reads	Rho et al., 2010	http://sourceforge.net/projects/fraggenescan/
CD-HIT	Clustering and comparing sequences of nucleotides or protein	Li and Godzik, 2006	http://weizhongli-lab.org/cd-hit/
HMMER3	Hidden Markov models applied in sequences alignments	Eddy, 2011	http://hmmer.janelia.org/
BLASTX	Basic local alignment of translated sequences	Altschul et al., 1997	http://blast.ncbi.nlm.nih.gov/blast/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome
MetaORFA	Assembly of peptides obtained from predicted ORFs	Ye and Tang, 2008	NA
MinPath	Reconstruction of pathways from protein family predictions	Ye and Doak, 2009	http://omics.informatics.indiana.edu/MinPath/
MetaPath	Identification of metabolic pathways differentially abundant among metagenomic samples	Liu and Pop, 2011	http://metapath.cbcb.umd.edu/
GhostKOALA	KEGG's internal annotator of metagenomes by k-number assignment by GHOSTX searches against a non-redundant database of KEGG genes	NP	http://www.kegg.jp/ghostkoala/
RAMMCAP	Metagenomic functional annotation and data clustering	Li, 2009	http://weizhong-lab.ucsd.edu/rammcap/cgi-bin/rammcap.cgi
ProViDE	Analysis of viral diversity in metagenomic samples	Ghosh et al., 2011	http://metagenomics.atc.tcs.com/binning/ProViDE/
Phyloseq	Tool-kit to row reads pre-processing, diversity analysis and graphics production. R, Bioconductor package	McMurdie and Holmes, 2014	https://joey711.github.io/phyloseq/
MetagenomeSeq	Analysis of differentially abundance of 16S rRNA gene in metaprofiling data. R, Bioconductor package	Paulson et al., 2013	http://bioconductor.org/packages/release/bioc/html/metagenomeSeq.html
ShotgunFunctionalizeR	Metagenomic functional comparison at level of individual genes (COG and EC numbers) and complete pathways. R, Bioconductor package	Kristiansson et al., 2009	http://shotgun.math.chalmers.se/
Galaxy portal	Web repository of computational tools that can be run without informatic expertise. Graphical interface and free service	Goecks et al., 2010	https://usegalaxy.org/
MG-RAST	Taxonomic and functional annotation, comparative metagenomics. Graphical interface, web portal, and free service	Meyer et al., 2008	http://metagenomics.anl.gov/
IMG/M	Functional annotation, phylogenetic distribution of genes and comparative metagenomics. Graphical interface, web portal, and free service	Markowitz et al., 2012	https://img.jgi.doe.gov/cgi-bin/m/main.cgi

NP, Not published in an indexed Journal; NA, Not web site available.

PICRUSt uses an evolutionary modeling to generate functional predictions from ribosomal (16S rRNA) genes databases, which allows to obtain a general vision of microbial functions in a microbiome. However, it only works adequately for those environments where the results have large numbers of organisms with annotated reference genomes available. Finally, PICRUSt is only designed to analyse prokaryotes, ignoring a large amount of metabolic features performed by eukaryotes.

Shotgun Metagenomics

As mentioned, after deciphering the microbial diversity of a metagenome, it would be very convenient to understand its metabolic potential. This can be achieved by using a whole metagenome approach where total DNA is obtained to prepare whole shotgun libraries. As discussed, the sequencing platform choice will be somehow influenced by the computational resources and available software to handle and process the sequencing output (**Table 2**). It should be noted that the impact and potential of shotgun metagenomics would be also reflected in taxonomy species level classification. The many microorganisms obtained from whole metagenome shotgun sequencing will probably deliver new genes with novel functions.

Assessment of Taxonomy Based on Markers

Theoretically, when a whole metagenome shotgun sequencing approach is performed, we can obtain a representation of all the genomes in the sample. This permit us not only to choose from a wide range of phylogenetic markers in order to perform taxonomic annotation but also we can obtain the ribosomal markers or any other used in the amplicon sequencing approach.

A multithreading software option to extract ribosomal marker genes from metagenomic sequences to conduct the taxonomic annotation is Parallel-meta (Su et al., 2014). The program collects ribosomal sequences from short reads by using a Hidden Markov Models (HMM)-based reconstruction algorithm (De Fonzo et al., 2007). Then it maps the reconstructed sequences to different 16S gene databases using Megablast (http://www.ncbi.nlm.nih.gov/blast/html/megablast.html). As discussed in the metaprofiling analysis section, taxonomical annotation could be improved by using more than one phylogenetic marker. Therefore, in whole metagenome shotgun sequencing, we can use software to search single copy marker genes in other databases. Two examples of programs using these approaches are MOCAT (Kultima et al., 2012), which uses the RefMG database (Ciccarelli et al., 2006) constituted by a collection of 40 single copy marker genes, and AMPHORA (Wu and Eisen, 2008), which includes a database containing around 31 single copy universal markers (**Table 2**). After the single copy marker identification, such pipelines perform an OTU multiple sequence alignment, distance calculation, and clustering. Finally, the taxonomical annotation is performed using reference genomes giving a species resolution in many cases.

The Binning Strategy

Binning classification is a quick and handy method to predict taxonomical composition using the information contained in the reads. These could be performed using either reads or assembled sequences. Binning algorithms use different strategies to get the taxonomic assignment: (a) sequence composition classification or (b) sequence alignment against references.

The first one is based on k-mer frequencies methods, which uses short words (k-mers) to represent a vector-like sequence and then to obtain the similarity among all words in the query. This representation can be considered as a "genomic signature" and was widely used by Karlin and Burge (1995) to explore evolutionary conservation among species. Examples of software that perform sequence classification by composition are TETRA (Teeling et al., 2004), PhylophytiaS (McHardy et al., 2007), and MetaclusterTA (Wang et al., 2014) (**Table 2**).

Other methods have more than one strategy to support the correct binning of sequences as in the case of MaxBin (Wu et al., 2014) and Amphora2 (Wu and Scott, 2012), which rely on finding single copy marker genes, k-mer signatures, GC content, and coverage information to perform contig and read binning.

In spite of the binning approach facilitating taxonomic classification, it should be considered that this strategy have some problems with horizontally transferred sequences, where genes from an organism appear in another. This could lead to an aggravated misclassification if it occurs between non-described organisms (Sharpton, 2014).

However, other methods based on reference read alignment are based on Burrows–Wheeler Transform indexes like BWA (Li and Durbin, 2009) or Bowtie (Langmead and Salzberg, 2012). These fast and accurate alignment methods can assess species richness and abundance in metagenomes by mapping reads directly to individual reference genomes or many concatenated genomes (pangenomes) or sequences. This last approach is used in the Genometa software (Davenport and Tümmler, 2013) and allows us to obtain OTUs for metagenome samples by grouping genomic islands, operons, or orthologous genes present in reference pangenomes. Furthermore, if long reads are available, then it is possible to do a taxonomic assignment by translating them and use all potential coding sequences to perform searches in annotated protein databases using local alignment tools, i.e., BLAST. In addition, some programs like SORT-Items (Monzoorul Haque et al., 2009), Megan, or Discribinate (Ghosh et al., 2010) (**Table 2**) can recover the lowest common ancestor (LCA) of a certain sequence from BLAST results.

Finally, we should consider that the more information we have for supporting taxonomic or functional results, the more reliable will be our conclusions. This is why it is always advisable to use more than one approach to assess taxonomic or functional annotation, if possible.

Functional Metagenomics Analysis

Reconstruction of metabolic pathways from enzyme-coding genes is a relevant matter in the metagenome analysis. Generally, there are two options to perform functional annotation from shotgun sequences, one is using sequencing reads directly and another is by read assembly.

Read assembly

Assembly is more efficient for genome reconstruction in low complex samples and when closely related species reference

genomes are present (Teeling and Glöckner, 2012; Luo et al., 2013). However, the task is hampered when the read coverage is low and when there is high frequency of polymorphisms and repetitive regions (De Filippo et al., 2012). Nowadays, there are *ad hoc* assemblers for metagenome reads (**Table 2**) such as IDBA-UD (Peng et al., 2012) and MetaVelvet (Namiki et al., 2012). Both are based on de Bruijn graph construction methods and consider different coverage peaks, which are expected in a community composed by several different organisms (Thomas et al., 2012).

An extension of this algorithm is the use of the so called "colored" de Brujin graphs. This computational implementation can perform a genome assembly and variant calling at the same time (Iqbal et al., 2012). An assembler that incorporates this technique is Ray genome assembler that presents a different implementation such as RayMeta for *de novo* assembly of metagenomes and RayCommunities that calculates microbe abundance and taxonomical profiling (**Table 2**) (Boisvert et al., 2012).

Some advantages of assembling metagenomes are: (1) The possibility of analysing the genome context (i.e., operons); (2) Increasing the probability of complete genes and genomes reconstruction, arising the confidence of sequence annotation; (3) Analysis simplification by mapping long contigs instead of short reads (Thomas et al., 2012; Luo et al., 2013; Segata et al., 2013).

Prediction of gene coding sequences

After metagenome assembly, gene prediction and annotation are similar to the framework followed in whole genome characterization (Yandell and Ence, 2012; Richardson and Watson, 2013). For metagenomics, it is recommended to predict genes using algorithms that consider di-codons frequency, preferential bias in codon usage, patterns in the use of start and stop codons and, if possible, incorporates the information of species-specific ribosome-binding sites patterns, Open Reading Frame (ORF) length, and GC content of coding-sequences (Liu et al., 2013).

To assess such tasks, some gene predictors have been designed particularly for metagenomic contig ORFs calling (**Table 2**). For example, MetaGeneMark (Zhu et al., 2010) or GlimmerMG (Kelley et al., 2012) uses *ab initio* gene identification by "heuristical model" methods and second-order Markov chains for coding-sequence prediction training.

However, it is not always possible to get a good assembly, especially for complex metagenomes with a great number of low abundance species. A workaround would be the use of FragGeneScan tool, which predicts partial ORFs from short reads of at least 60 bp length (Rho et al., 2010).

With predicted genes, we can continue to analyse the translations of such predictions and obtain a product and functional annotation.

Function assignment and databases

Function assignment of predicted ORFs could be performed on either nucleotide or translated sequences. In both cases, homology detection is probably the easiest and most frequent annotation method, despite being computationally demanding and time consuming. Using algorithms like BLAST against databases such as Swiss-Prot or NCBI-nr retrieve a list of related hits with a certain annotation that can be used to mine taxonomical information as well. However, a limitation of this approach is the size and phylogenetic coverage of the database (Carr and Borenstein, 2014).

Searches in customized databases such as CAZY, dbCAN, or MetaBioMe are alternative to avoid time consuming and the use of excessive computational resources in the annotation of genes related to a metabolic pathway (Teeling and Glöckner, 2012; Yang et al., 2014). In any case, reducing computational workload is useful to remove redundant sequences using algorithms such as CD-HIT (Li and Godzik, 2006) to make the ORF or read annotation process more efficient.

Usually, when protein function assignment by homology is not possible due to low sequence identity values ($<20\%$ of identity), HMM searches (Eddy, 2011) can be used for interrogating protein functional domain profiles using databases like the Conserved Domain Database of NCBI, PFAM, or SEED. Apart from solving the remote homology problem, this approach has helped us to find the regional or functional domains in proteins, in addition to the product annotation that sometimes could be cryptic.

Homology-based or HMM strategies can deliver a great number of false negatives especially when using short reads (Scholz et al., 2012; Yang et al., 2014). It is noteworthy that for functional annotation, the longer the sequence, the more information is provided, which makes the sequence search easier (Carr and Borenstein, 2014). The use of short reads to perform direct searches has low sensitivity and specificity for homologous identification (Wommack et al., 2008); therefore, *E*-value threshold should be adjusted in order to obtain correct results (Carr and Borenstein, 2014).

Another option is sequence clustering using BLASTX (Altschul et al., 1997). This strategy allows us to search directly from reads or contigs, since the program will perform all the possible translations. This has been implemented by Ye and Tang (2008) in the MetaORFA pipeline, where the translations (ORFome) are used to search homologs in the databases (**Table 2**). However, this could be very inefficient if a large set of reads is being analyzed.

A workflow summary for functional annotation could be as follows: get the best possible metagenome assembly (highest N50, N90, and contig/scaffold ave. length) to perform the ORF prediction and then assign function to a set of translated sequences by homology against well-curated databases of both protein and conserved domains. Finally, mine the functional and taxonomical information obtained from the search results based on the target sequences.

An alternative to avoid dealing with local software and computational resources is web portals such as Galaxy (Goecks et al., 2010), MG-RAST (Meyer et al., 2008), and IMG-M portal (Markowitz et al., 2012). These web servers are dedicated to perform taxonomical and functional analysis of metagenomes via a graphical user-friendly interface (**Table 2**). Unfortunately, these portals sometimes are saturated and the analysis parameters are

not customizable. Finally, the internet bandwidth to transfer very large datasets could be a bottleneck for some users.

Metabolic pathway reconstruction

Pathway reconstruction of the metagenome data is one of the annotation goals. The concept of metabolic pathway in microbial ecology should be understood as the flow of information through different species. Therefore, the term "inter-organismic meta-routes" or "meta-pathways" has been proposed for this kind of analysis (De Filippo et al., 2012).

In order to perform a reliable metabolic reconstruction, a good functional annotation should be achieved in the first place. This has to be used to find each gene in an appropriate metabolic context, filling missing enzymes in pathways and find optimal metabolic states to perform the best pathway reconstructions. Examples of programs available are MinPath (Ye and Doak, 2009) and MetaPath (Liu and Pop, 2010). Both use information deposited in KEGG (Ogata et al., 1999) and MetaCyc (Caspi et al., 2014) repositories (**Table 2**).

However, most of the metabolic information comes from model organisms, but not all the enzymes or pathways are conserved among all species or environments. That is why most of the current platforms fail in metabolic reconstruction of variant pathways (de Crécy-Lagard, 2014) and most are designed to analyse single genomes.

A web service implementation by KEGG for metagenome analysis is GhostKOALA (Kanehisa Laboratories; http://www.kegg.jp/ghostkoala/). It relates taxonomic origin with their respective functional annotation, and the user is able to visualize metabolic pathways from different taxa in the same map.

Metabolic pathway reconstruction could be completed with information provided by the data context such as gene function interactions, synteny, and copy number of annotated genes to integrate the metabolic potential of consortium.

Bottlenecks in functional annotation: The ORFans problem

There are some relevant issues to consider in the whole metagenome shotgun sequencing annotation. Protocols based on sequence similarity searching assume that each read will be mapped to a homologous gene of some closely related species. However, depending on the database quality and size, different results could be obtained. For example, if direct DNA searches are performed, then it is probable to get matches against intergenic regions or non-coding genes (as a tRNA). In addition, alignments could retrieve best hits from a sequence in a potentially distant genome (Carr and Borenstein, 2014), affecting the taxonomic annotation if the search results are used for this endeavor (i.e., MEGAN).

In spite of the annotation method, it is known that metagenomes will have around 50% of protein sequences with no annotation or unknown function (referred as ORFans). This percentage increases when the species richness is high in the community. ORFans can be classified into three categories: (1) spurious genes produced by errors in the gene prediction; (2) genes with homology at secondary or tertiary structure level but not at nucleotide sequence level, or (3) real new genes with no homology to other genes, hence with unknown functions.

An option to deal the ORF prediction errors is to use the rate of possible non-synonymous and synonymous substitutions (ka/ks) as a criterion to select probable genes. If *ka*/*ks* value is close to 1, then it indicates that such sequence is not under selective pressure, suggesting a low probability to code for a real protein (Yooseph et al., 2008). To confirm a candidate for a novel gene, the appropriate strategy should include a *de novo* secondary and tertiary structure predictions using tools like I-TASSER (Yang et al., 2015), QUARK (Xu and Zhang, 2012), or RaptorX (Källberg et al., 2012) and perform a protein structure comparison using tools like STRAP (Gille et al., 2014). Nevertheless, this will reveal the protein tertiary structure but not necessarily its function. In fact, from more than six millions of putative enzymes identified by 454-sequencing in metagenome projects, only less than a few hundred proteins have a reliable functional annotation (Guazzaroni et al., 2010). Finally, the best way to confirm novel genes or discover new functions is through experimental procedures such as heterologous expression, biochemical characterization, and proteomics.

Pseudogenes are also a problem in metagenome functional annotation, and they could represent up to 35% in prokaryotic genomes (Liu et al., 2004). To address this annotation challenge, there are databases like BactPepDB (Rey et al., 2014) and Pseudogene.org for short sequences and pseudogenes of prokaryotic and eukaryotic organisms (Karro et al., 2007). A search in such databases before further analysis could be useful to discard non-coding sequences.

COMPARATIVE METAGENOMICS

In either of the metaprofiling or shotgun sequencing, the species richness or OTUs profiling could be contrasted among samples based on species diversity comparison (beta-diversity).

Two types of beta-diversity indices, such as incidence type and abundance type, could be used. The former, such as Jaccard and Sørensen indices, treats the common and rare species equally and just compares the number of shared and unique taxa between the samples. The abundance-type index contemplates abundance similarity, thereby treating individuals not species equally; some examples are the Morisita-type and Bray–Curtis dissimilarity indices (Chao et al., 2006). Such indexes are affected by sampling size. An excellent review of beta-diversity fundamentals were done by Tuomisto (2010). Alternatively, UniFrac is a method for comparing microbial communities through phylogenetic distance information contained in marker genes as the 16S ribosomal rRNA (Lozupone and Knight, 2005). This method has been well accepted in metagenomics pipelines and implemented in some R-Bioconductor packages such as phyloseq (McMurdie and Holmes, 2013) and metagenomeSeq (Paulson et al., 2013). The latter implemented a novel algorithm for normalization as alternative to rarefaction.

In metaprofiling analysis, some modular pipelines such as Mothur and QIIME are capable of analysing raw reads and performing taxonomical annotation. In addition, they can compute sample comparisons and the calculation of some indexes mentioned in the Section Concepts of Microbial Diversity and Species Richness. In order to improve diversity

estimation, a lot of specialized software have been developed (**Table 2**) like ProViDE, which is designed for viral diversity estimation (Ghosh et al., 2011).

For whole metagenome shotgun projects, where gene protein coding information is available, functional comparative metagenomics is possible. It is based on identifying differential feature abundance (pathways, subsystems, or functional roles) between two or more conditions following a statistical procedure with some normalization step (Rodriguez-Brito et al., 2006; Pookhao et al., 2015). Some useful tools to perform robust comparative functional metagenomics are Parallel-meta and MEGAN. Other more specialized software are capable of returning graphical representations of metabolic abundances and taxonomic correlations as heatmaps or PCA plots of communities cluster genes. Two examples that compares metabolic pathways are ShotgunFunctionalizeR, which use a binomial and hypergeometric test to perform comparisons (Kristiansson et al., 2009), and MetaPath, a tool implemented in Perl that identifies and compares differentially abundant pathways in metagenomes (Liu and Pop, 2011).

THE NEGLECTED WORLD OF EUKARYOTES IN METAGENOMICS

Eukaryotes play important roles in almost all ecological niches in the earth; however, the study of eukaryotic domain is mostly biased toward animals, plants, and fungi, thereby resulting in a narrow view of the great eukaryotic diversity. Microscopic eukaryotes (regularly named protists) are the real bulk of most of the eukaryotic lineages (Burki, 2014). Microeukarya are poorly studied, but it is estimated that around 10% percent of protrist species are already described and were found in the ocean (Mann and Droop, 1996; Norton et al., 1996). Meanwhile, a 1.2–10 million species have been predicted as host-associated protista from which only 6000 have been reported (Burki, 2014).

Studying these organisms by NGS techniques has been a challenge because they are not well represented in the sequence databases. The lack of reference eukaryotic genomes is due in part the difficulty of their genome assembly and annotation (Gilbert and Dupont, 2011). In spite of the lack of information, it is important to remark the importance of microeukaryotes in the environment. They are responsible for CO_2 fixing in the oceans, and they are the principal organic matter degraders in soils, and some of them are symbionts of other eukaryotes (Burki, 2014).

Diversity studies of the "eukaryotome" have been done using 18S rRNA gene amplicons (Andersen et al., 2013), and some programs include tools to analyse them such as Parallel-meta and QIIME, which have an option for mapping reads against eukaryotic Silva small ribosomal subunit (SSU) database. The SSU is commonly used for diversity analysis as universal phylogenetic marker for eukaryotic genes, but there are issues to reach a species classification level due to their little variation that limits the taxonomical position, especially for some fungi and protists (Schoch et al., 2012).

Nowadays, new strategies have been developed based on other phylogenetic markers to evaluate the eukaryotic fraction in a sample. The LSU or ITS regions are good alternatives to classify organisms at the species level with high accuracy. Ecologists interested in analysing the eukaryotic fraction are using NGS platforms like the Ion Torren PGM or the Illumina MiSeq sequencers, which generate 400 bp single reads or 300 bp paired end reads, respectively (**Table 1**). Both platforms deliver enough yield to perform the analysis of LSU or ITS amplicons at a very high depth (Lindahl and Kuske, 2013; Hugerth et al., 2014; Tonge et al., 2014).

Regarding the metabolic association of eukaryotic genes in a certain pathway, it can be a greater challenge than bacterial annotation. Eukaryotic genomes are typically 6–10 times larger than the average bacterial genome (about 3–5 Mb) size, plus they can have different genome ploidy states. It is worthy to mention that the eukaryotic genes contain introns, which may have differential splicing patterns under particular environmental conditions, thereby increasing the amount of products (isoforms) with different functions to annotate. Moreover, high percentage of intergenic non-coding sequences that are represented differently in a shotgun sequenced metagenome can represent a problem if they were not assembled correctly leaving them out of their gene context. A strategy to further characterize coding regions in a eukaryotic metagenome is to isolate some mRNA to perform a metatranscriptomics analysis. Enriched mRNA from eukaryotic organisms (Qi et al., 2011; Keeling et al., 2014) can be *de novo* assembled or mapped to related reference genomes in order to elucidate the functions from these transcripts.

CONCLUDING REMARKS

Here, we have reviewed the evolution of Microbiology into Metagenomics to describe exhaustively a microbial community in terms of taxonomic diversity and metabolic potential. Metagenomics allows us to discover new genes and proteins or even the complete genomes of non-cultivable organisms in less time and with better accuracy than classical microbiology or molecular methods. However, there are no standard methods or universal tools that can answer all of our questions in metagenomics. In fact, the lack of standards reduces the reproducibility and comparison between similar projects, making metagenomics a case by case study. It is noteworthy that each metagenome project has specific requirements depending on its experimental design, and hence, the sequencing technology and computational tools should be chosen carefully. In spite of the serendipity that is present in science, we have to bear in mind that the experimental design is the most important part and should fit each project objectives in order to reach them and answer the biological question behind the project.

A metagenome usually represents a snapshot of a community at a certain time when its DNA is obtained. As mentioned, a good experimental design is necessary to explore the complete population dynamics by combining different approaches like culture methods, DNA and RNA analysis, protein studies, and if possible, the metabolic profile. Consequently, integration of several tools to microbiology (such as molecular biology, genetics, bioinformatics, and statistics) is necessary to answer the

questions related to microbial diversity and ecology in a greater extent.

In our opinion, the development of more bioinformatics tools for metagenomics analysis is necessary, but the experience of scientists to manipulate such tools and interpret their results is the key to a sensible biological conclusion. The bioinformatics expertise is a necessity, as the sequencing platforms are delivering a massive yield at a very low cost, increasing the amount of information to analyse. Finally, the near future challenge will reside in the manipulation and analysis of the data deluge and how we can interpret them in a more integrative way that could reflect the biodiversity present in our world.

ACKNOWLEDGMENTS

AE and AV are Ph.D. students from Programa de Doctorado en Ciencias Bioquímicas and Programa de Doctorado en Ciencias Biomédicas Universidad Nacional Autónoma de México (UNAM) with scholarship from Consejo Nacional de Ciencia y Tecnología (México).

REFERENCES

Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V., and Polz, M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J. Bacteriol.* 186, 2629–2635. doi: 10.1128/JB.186.9.2629-2635.2004

Ackert, L. (2012). *Sergei Vinogradskii and the Cycle of Life: From the Thermodynamics of Life to Ecological Microbiology, 1850-1950.* Netherlands: Springer Science & Business Media.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Andersen, L. O., Vedel Nielsen, H., and Stensvold, C. R. (2013). Waiting for the human intestinal Eukaryotome. *ISME J.* 7, 1253–1255. doi: 10.1038/ismej.2013.21

Andrews, S. (2015). *Babraham Bioinformatics—FastQC A Quality Control Tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed July 17, 2015).

Begon, M., Harper, J. L., and Townsend, C. R. (1986). *Ecology: Individuals, Populations and Communities*. Oxford: Blackwell Scientific Publications.

Bennett, S. (2004). Solexa Ltd. *Pharmacogenomics* 5, 433–438. doi: 10.1517/14622416.5.4.433

Beveridge, T. J. (2001). Use of Gram stain in microbiology. *Biotech. Histochem.* 76, 111–118. doi: 10.1080/bih.76.3.111.118

Blevins, S. M., and Bronze, M. S. (2010). Robert Koch and the "golden age" of bacteriology. *Inter. J. Infect. Diseases* 14, e744–e751. doi: 10.1016/j.ijid.2009.12.003

Bodilis, J., Nsigue-Meilo, S., Besaury, L., and Quillet, L. (2012). Variable copy number, intra-genomic heterogeneities and lateral transfers of the 16S rRNA gene in pseudomonas. *PLoS ONE* 7:e35647. doi: 10.1371/journal.pone.0035647

Boisvert, S., Raymond, F., Godzaridis, É., Laviolette, F., and Corbeil, J. (2012). Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biol.* 13:R122. doi: 10.1186/gb-2012-13-12-r122

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Meth.* 10, 57–59. doi: 10.1038/nmeth.2276

Bonilla-Rosso, G., Eguiarte, L. E., Romero, D., Travisano, M., and Souza, V. (2012). Understanding microbial community diversity metrics derived from metagenomes: performance evaluation using simulated data sets. *FEMS Microbiol. Ecol.* 82, 37–49. doi: 10.1111/j.1574-6941.2012.01405.x

Burki, F. (2014). The Eukaryotic Tree of Life from a Global Phylogenomic Perspective. *Cold Spring Harb. Perspect. Biol.* 6:a016147. doi: 10.1101/cshperspect.a016147

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8

Carr, R., and Borenstein, E. (2014). Comparative analysis of functional metagenomic annotation and the mappability of short reads. *PLoS ONE* 9:e105776. doi: 10.1371/journal.pone.0105776

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucl. Acids Res.* 42, D459–D471. doi: 10.1093/nar/gkt1103

Chao, A., Chazdon, R. L., Colwell, R. K., and Shen, T. J. (2006). Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* 62, 361–371. doi: 10.1111/j.1541-0420.2005.00489.x

Chernov, T. I., Tkhakakhova, A. K., and Kutovaya, O. V. (2015). Assessment of diversity indices for the characterization of the soil prokaryotic community by metagenomic analysis. *Eurasian Soil Sc.* 48, 410–415. doi: 10.1134/S1064229315040031

Ciccarelli, F. D., Doerks, T., Von Mering, C., Creevey, C. J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287. doi: 10.1126/science.1123061

Colwell, R. K., and Coddington, J. A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philos. Trans. R. Soc. Lond. B* 345, 101–118. doi: 10.1098/rstb.1994.0091

Colwell, R. R., Brayton, P., Herrington, D., Tall, B., Huq, A., and Levine, M. M. (1996). Viable but non-culturable Vibrio cholerae O1 revert to a cultivable state in the human intestine. *World J. Microbiol. Biotechnol.* 12, 28–31. doi: 10.1007/BF00327795

Davenport, C. F., and Tümmler, B. (2013). Advances in computational analysis of metagenome sequences: *In silico* metagenome analysis. *Environ. Microbiol.* 15, 1–5. doi: 10.1111/j.1462-2920.2012.02843.x

de Crécy-Lagard, V. (2014). Variations in metabolic pathways create challenges for automated metabolic reconstructions: Examples from the tetrahydrofolate synthesis pathway. *Comput. Struct. Biotechnol. J.* 10, 41–50. doi: 10.1016/j.csbj.2014.05.008

De Filippo, C., Ramazzotti, M., Fontana, P., and Cavalieri, D. (2012). Bioinformatic approaches for functional annotation and pathway inference in metagenomics data. *Brief. Bioinformatics* 13, 696–710. doi: 10.1093/bib/bbs070

De Fonzo, V., Aluffi-Pentini, F., and Parisi, V. (2007). Hidden markov models in bioinformatics. *Curr. Bioinform.* 2, 49–61. doi: 10.2174/157489307779314348

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

Fichot, E. B., and Norman, R. S. (2013). Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform. *Microbiome* 1:10. doi: 10.1186/2049-2618-1-10

Fox, G. E., Wisotzkey, J. D., and Jurtshuk, P. (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int. J. Syst. Bacteriol.* 42, 166–170. doi: 10.1099/00207713-42-1-166

Ghosh, T. S., Mohammed, M. H., Komanduri, D., and Mande, S. S. (2011). ProViDE: A software tool for accurate estimation of viral diversity in metagenomic samples. *Bioinformation* 6, 91–94. doi: 10.6026/97320630006091

Ghosh, T. S., Monzoorul Haque, M., and Mande, S. S. (2010). DiScRIBinATE: a rapid method for accurate taxonomic classification of metagenomic sequences. *BMC Bioinformatics* 11 (Suppl. 7):S14. doi: 10.1186/1471-2105-11-S7-S14

Gihring, T. M., Green, S. J., and Schadt, C. W. (2012). Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity

comparisons due to variable library sizes. *Environ. Microbiol.* 14, 285–290. doi: 10.1111/j.1462-2920.2011.02550.x

Gilbert, J. A., and Dupont, C. L. (2011). Microbial metagenomics: beyond the genome. *Ann. Rev. Mar. Sci.* 3, 347–371. doi: 10.1146/annurev-marine-120709-142811

Gille, C., Fähling, M., Weyand, B., Wieland, T., and Gille, A. (2014). Alignment-Annotator web server: rendering and annotating sequence alignments. *Nucl. Acids Res.* 42, W3–W6. doi: 10.1093/nar/gku400

Gillespie, D. E., Brady, S. F., Bettermann, A. D., Cianciotto, N. P., Liles, M. R., Rondon, M. R., et al. (2002). Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl. Environ. Microbiol.* 68, 4301–4306. doi: 10.1128/AEM.68.9.4301-4306.2002

Giovannoni, S. J., DeLong, E. F., Schmidt, T. M., and Pace, N. R. (1990). Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. *Appl. Environ. Microbiol.* 56, 2572–2575.

Glenn, T. (2014). *2014 NGS Field Guide: Overview|The Molecular Ecologist*. Available online at: http://www.molecularecologist.com/next-gen-fieldguide-2014/ (Accessed August 24, 2015).

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11:R86. doi: 10.1186/gb-2010-11-8-r86

Guazzaroni, M. E., Beloqui, A., Vieites, J. M., Al-ramahi, Y., Cortés, N. L., Ghazi, A., et al. (2010). "Metagenomic mining of enzyme diversity," in *Handbook of Hydrocarbon and Lipid Microbiology*, ed K. N. Timmis (Berlin; Heidelberg: Springer), 2911–2927. Available online at: http://link.springer.com/referenceworkentry/10.1007/978-3-540-77587-4_216 (Accessed May 17, 2015).

Haegeman, B., Hamelin, J., Moriarty, J., Neal, P., Dushoff, J., and Weitz, J. S. (2013). Robust estimation of microbial diversity in theory and in practice. *ISME J.* 7, 1092–1101. doi: 10.1038/ismej.2013.10

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* 5, R245–R249. doi: 10.1016/s1074-5521(98)90108-9

Heck, K. L. Jr., Belle, G., and van Simberloff, D. (1975). Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology* 56, 1459–1461. doi: 10.2307/1934716

Hugerth, L. W., Muller, E. E. L., Hu, Y. O. O., Lebrun, L. A. M., Roume, H., Lundin, D., et al. (2014). Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. *PLoS ONE* 9:e95567. doi: 10.1371/journal.pone.0095567

Huson, D. H., and Weber, N. (2013). Microbial community analysis using MEGAN. *Meth. Enzymol.* 531, 465–485. doi: 10.1016/B978-0-12-407863-5.00021-6

Iqbal, Z., Caccamo, M., Turner, I., Flicek, P., and McVean, G. (2012). De novo assembly and genotyping of variants using colored de Bruijn graphs. *Nat. Genet.* 44, 226–232. doi: 10.1038/ng.1028

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., et al. (2012). Template-based protein structure modeling using the RaptorX web server. *Nat. Protocols* 7, 1511–1522. doi: 10.1038/nprot.2012.085

Karlin, S., and Burge, C. (1995). Dinucleotide relative abundance extremes: a genomic signature. *Trends Genet.* 11, 283–290. doi: 10.1016/S0168-9525(00)89076-9

Karow, J. (2013). *Following Roche's Decision to Shut Down 454, Customers Make Plans to Move to Other Platforms*. GenomeWeb. Available online at: https://www.genomeweb.com/sequencing/following-roches-decision-shut-down-454-customers-make-plans-move-other-platform (Accessed October 29, 2015).

Karro, J. E., Yan, Y., Zheng, D., Zhang, Z., Carriero, N., Cayting, P., et al. (2007). Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic Acids Res.* 35, D55–D60. doi: 10.1093/nar/gkl851

Kasianowicz, J. J., Brandin, E., Branton, D., and Deamer, D. W. (1996). Characterization of individual polynucleotide molecules using a membrane?channel. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13770–13773. doi: 10.1073/pnas.93.24.13770

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., et al. (2014). The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* 12:e1001889. doi: 10.1371/journal.pbio.1001889

Keller, M., and Zengler, K. (2004). Tapping into microbial diversity. *Nat. Rev. Micro.* 2, 141–150. doi: 10.1038/nrmicro819

Kelley, D. R., Liu, B., Delcher, A. L., Pop, M., and Salzberg, S. L. (2012). Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. *Nucleic Acids Res.* 40, e9. doi: 10.1093/nar/gkr1067

Kemp, P. F., and Aller, J. Y. (2004). Estimating prokaryotic diversity: When are 16S rDNA libraries large enough? *Limnol. Oceanogr. Methods* 2, 114–125. doi: 10.4319/lom.2004.2.114

Krause, L., Diaz, N. N., Goesmann, A., Kelley, S., Nattkemper, T. W., Rohwer, F., et al. (2008). Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Res.* 36, 2230–2239. doi: 10.1093/nar/gkn038

Krebs, C. (2014). "CHAPTER 12—species diversity measures," in *Ecological Methodology* (Boston: Addison-Wesley Educational Publishers, Inc.). Available online at: http://www.zoology.ubc.ca/~krebs/downloads/krebs_chapter_13_2014.pdf (Accessed August 3, 2015).

Kristiansson, E., Hugenholtz, P., and Dalevi, D. (2009). ShotgunFunctionalizeR: an R-package for functional comparison of metagenomes. *Bioinformatics* 25, 2737–2738. doi: 10.1093/bioinformatics/btp508

Kultima, J. R., Sunagawa, S., Li, J., Chen, W., Chen, H., Mende, D. R., et al. (2012). MOCAT: a metagenomics assembly and gene prediction toolkit. *PLoS ONE* 7:e47656. doi: 10.1371/journal.pone.0047656

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821. doi: 10.1038/nbt.2676

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* 9, 357–359. doi: 10.1038/nmeth.1923

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, K., Bihan, M., Yooseph, S., and Methé, B. A. (2012). Analyses of the microbial diversity across the human microbiome. *PLoS ONE* 7:e32118. doi: 10.1371/journal.pone.0032118

Li, W. (2009). Analysis and comparison of very large metagenomes with fast clustering and functional annotation. *BMC Bioinformatics* 10:359. doi: 10.1186/1471-2105-10-359

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Lindahl, B. D., and Kuske, C. R. (2013). "Metagenomics for study of fungal ecology," in *The Ecological Genomics of Fungi*, ed Francisrtin (John Wiley & Sons, Inc), 279–303. Available online at: http://onlinelibrary.wiley.com/doi/10.1002/9781118735893.ch13/summary (Accessed May 28, 2015).

Liu, B., and Pop, M. (2010). "Identifying differentially abundant metabolic pathways in metagenomic datasets," in *Bioinformatics Research and Applications Lecture Notes in Computer Science*, eds M. Borodovsky, J. P. Gogarten, T. M. Przytycka, and S. Rajasekaran (Berlin; Heidelberg: Springer), 101–112. Available online at: http://link.springer.com/chapter/10.1007/978-3-642-13078-6_12 (Accessed May 17, 2015).

Liu, B., and Pop, M. (2011). MetaPath: identifying differentially abundant metabolic pathways in metagenomic datasets. *BMC Proc.* 5:S9. doi: 10.1186/1753-6561-5-S2-S9

Liu, Y., Guo, J., Hu, G., and Zhu, H. (2013). Gene prediction in metagenomic fragments based on the SVM algorithm. *BMC Bioinformatics* 14:S12. doi: 10.1186/1471-2105-14-S5-S12

Liu, Y., Harrison, P. M., Kunin, V., and Gerstein, M. (2004). Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol.* 5:R64. doi: 10.1186/gb-2004-5-9-r64

Logares, R., Haverkamp, T. H. A., Kumar, S., Lanzén, A., Nederbragt, A. J., Quince, C., et al. (2012). Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms

and bioinformatics approaches. *J. Microbiol. Methods* 91, 106–113. doi: 10.1016/j.mimet.2012.07.017

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005

Luo, C., Rodriguez,-R. L. M., and Konstantinidis, K. T. (2013). "Chapter twenty-three—a user's guide to quantitative and comparative analysis of metagenomic datasets," in *Methods in Enzymology Microbial Metagenomics, Metatranscriptomics, and Metaproteomics*, ed E. F. DeLong (Academic Press), 525–547. Available online at: http://www.sciencedirect.com/science/article/pii/B978012407863500023X (Accessed February 17, 2015).

Luo, C., Tsementzi, D., Kyrpides, N., Read, T., and Konstantinidis, K. T. (2012). Direct comparisons of illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS ONE* 7:e30087. doi: 10.1371/journal.pone.0030087

Mann, D. G., and Droop, S. J. M. (1996). 3. Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia* 336, 19–32. doi: 10.1007/BF00010816

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in open microfabricated high density picoliter reactors. *Nature* 437, 376–380. doi: 10.1038/nature03959

Markowitz, V. M., Chen, I. M. A., Chu, K., Szeto, E., Palaniappan, K., Grechkin, Y., et al. (2012). IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res.* 40, D123–D129. doi: 10.1093/nar/gkr975

McFall-Ngai, M. (2008). Are biologists in "future shock"? Symbiosis integrates biology across domains. *Nat. Rev. Micro.* 6, 789–792. doi: 10.1038/nrmicro1982

McHardy, A. C., Martín, H. G., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. (2007). Accurate phylogenetic classification of variable-length DNA fragments. *Nat. Methods* 4, 63–72. doi: 10.1038/nmeth976

McMurdie, P. J., and Holmes, S. (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 8:e61217. doi: 10.1371/journal.pone.0061217

McMurdie, P. J., and Holmes, S. (2014) Shiny-phyloseq: web application for interactive microbiome analysis with provenance tracking. *Bioinformatics* 31, 282–283. doi: 10.1093/bioinformatics/btu616

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. doi: 10.1186/1471-2105-9-386

Monzoorul Haque, M., Ghosh, T. S., Komanduri, D., and Mande, S. S. (2009). SOrt-ITEMS: Sequence orthology based approach for improved taxonomic estimation of metagenomic sequences. *Bioinformatics* 25, 1722–1730. doi: 10.1093/bioinformatics/btp317

Morgan, J. L., Darling, A. E., and Eisen, J. A. (2010). Metagenomic sequencing of an *in vitro*-simulated microbial community. *PLoS ONE* 5:e10209. doi: 10.1371/journal.pone.0010209

Nalbantoglu, U., Cakar, A., Dogan, H., Abaci, N., Ustek, D., Sayood, K., et al. (2014). Metagenomic analysis of the microbial community in kefir grains. *Food Microbiol.* 41, 42–51. doi: 10.1016/j.fm.2014.01.014

Namiki, T., Hachiya, T., Tanaka, H., and Sakakibara, Y. (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res.* 40, e155. doi: 10.1093/nar/gks678

Norton, T. A., Melkonian, M., and Andersen, R. (1996). Algal biodiversity. *Phycologia* 35, 308–326. doi: 10.2216/i0031-8884-35-4-308.1

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34. doi: 10.1093/nar/27.1.29

Oliver, J. D., Nilsson, L., and Kjelleberg, S. (1991). Formation of nonculturable Vibrio vulnificus cells and its relationship to the starvation state. *Appl. Environ. Microbiol.* 57, 2640–2644.

Ottman, N., Smidt, H., de Vos, W. M., and Belzer, C. (2012). The function of our microbiota: who is out there and what do they do? *Front. Cell. Infect. Microbiol.* 2:104. doi: 10.3389/fcimb.2012.00104

Patel, R. K., and Jain, M. (2012). NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* 7:e30619. doi: 10.1371/journal.pone.0030619

Paulson, J. N., Stine, O. C., Bravo, H. C., and Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nat. Meth.* 10, 1200–1202. doi: 10.1038/nmeth.2658

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. doi: 10.1093/bioinformatics/bts174

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A., and Versalovic, J. (2009). Metagenomic Pyrosequencing and Microbial Identification. *Clin. Chem.* 55, 856–866. doi: 10.1373/clinchem.2008.107565

Pookhao, N., Sohn, M. B., Li, Q., Jenkins, I., Du, R., Jiang, H., et al. (2015). A two-stage statistical procedure for feature selection and comparison in functional analysis of metagenomes. *Bioinformatics* 31, 158–165. doi: 10.1093/bioinformatics/btu635

Poretsky, R., Rodriguez,-R. L. M., Luo, C., Tsementzi, D., and Konstantinidis, K. T. (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS ONE* 9:e93827. doi: 10.1371/journal.pone.0093827

Qi, M., Wang, P., O'Toole, N., Barboza, P. S., Ungerfeld, E., Leigh, M. B., et al. (2011). Snapshot of the eukaryotic gene expression in muskoxen rumen—a metatranscriptomic approach. *PLoS ONE* 6:e20521. doi: 10.1371/journal.pone.0020521

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Rey, J., Deschavanne, P., and Tuffery, P. (2014). BactPepDB: a database of predicted peptides from a exhaustive survey of complete prokaryote genomes. *Database (Oxford).* 2014:bau106. doi: 10.1093/database/bau106

Rho, M., Tang, H., and Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* 38, e191. doi: 10.1093/nar/gkq747

Richardson, E. J., and Watson, M. (2013). The automatic annotation of bacterial genomes. *Brief. Bioinformatics* 14, 1–12. doi: 10.1093/bib/bbs007

Rodriguez-Brito, B., Rohwer, F., and Edwards, R. A. (2006). An application of statistics to comparative metagenomics. *BMC Bioinformatics* 7:162. doi: 10.1186/1471-2105-7-162

Roszak, D. B., Grimes, D. J., and Colwell, R. R. (1984). Viable but nonrecoverable stage of Salmonella enteritidis in aquatic systems. *Can. J. Microbiol.* 30, 334–338. doi: 10.1139/m84-049

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348–352. doi: 10.1038/nature10242

Sanchez-Flores, A., and Abreu-Goodger, C. (2014). A practical guide to sequencing genomes and transcriptomes. *Curr. Top. Med. Chem.* 14, 398–406. doi: 10.2174/1568026613666131204142353

Sanders, H. L. (1968). Marine benthic diversity: a comparative study. *Am. Nat.* 102, 243–282. doi: 10.1086/282541

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463–5467. doi: 10.1073/pnas.74.12.5463

Schierbeek, A. (1959). *Measuring the Invisible World: The Life and Works of Antoni van Leeuwenhoek*. London: Abelard-Schuman. Available online at: https://www.questia.com/library/73684/measuring-the-invisible-world-the-life-and-works (Accessed August 9, 2015).

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., et al. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6241–6246. doi: 10.1073/pnas.1117018109

Scholz, M. B., Lo, C. C., and Chain, P. S. (2012). Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr. Opin. Biotechnol.* 23, 9–15. doi: 10.1016/j.copbio.2011.11.013

Schouls, L. M., Schot, C. S., and Jacobs, J. A. (2003). Horizontal transfer of segments of the 16S rRNA genes between species of the Streptococcus anginosus group. *J. Bacteriol.* 185, 7241–7246. doi: 10.1128/JB.185.24.7241-7246.2003

Segata, N., Boernigen, D., Tickle, T. L., Morgan, X. C., Garrett, W. S., and Huttenhower, C. (2013). Computational meta'omics for microbial community studies. *Mol. Syst. Biol.* 9, 666. doi: 10.1038/msb.2013.22

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Front. Plant Sci.* 5:209. doi: 10.3389/fpls.2014.00209

Simpson, E. (1949). Measurement of diversity. *Nature* 163:688. doi: 10.1038/163688a0

Staley, J. T., and Konopka, A. (1985). Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.* 39, 321–346. doi: 10.1146/annurev.mi.39.100185.001541

Su, X., Pan, W., Song, B., Xu, J., and Ning, K. (2014). Parallel-META 2.0: enhanced metagenomic data analysis with functional annotation, high performance computing and advanced visualization. *PLoS ONE* 9:e89323. doi: 10.1371/journal.pone.0089323

Teeling, H., and Glöckner, F. O. (2012). Current opportunities and challenges in microbial metagenome analysis–a bioinformatic perspective. *Brief. Bioinformatics* 13, 728–742. doi: 10.1093/bib/bbs039

Teeling, H., Waldmann, J., Lombardot, T., Bauer, M., and Glöckner, F. O. (2004). TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics* 5:163. doi: 10.1186/1471-2105-5-163

Thomas, T., Gilbert, J., and Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microb. Inform. Exp.* 2:3. doi: 10.1186/2042-5783-2-3

Tonge, D. P., Pashley, C. H., and Gant, T. W. (2014). Amplicon-based metagenomic analysis of mixed fungal samples using proton release amplicon sequencing. *PLoS ONE* 9:e93849. doi: 10.1371/journal.pone.0093849

Tuomisto, H. (2010). A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography* 33, 2–22. doi: 10.1111/j.1600-0587.2009.05880.x

Wang, G. Y., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., et al. (2000). Novel Natural Products from Soil DNA Libraries in a Streptomycete Host. *Org. Lett.* 2, 2401–2404. doi: 10.1021/ol005860z

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07

Wang, Y., Leung, H., Yiu, S., and Chin, F. (2014). MetaCluster-TA: taxonomic annotation for metagenomic data based on assembly-assisted binning. *BMC Genomics* 15(Suppl. 1):S12. doi: 10.1186/1471-2164-15-S1-S12

Whiteley, A. S., Jenkins, S., Waite, I., Kresoje, N., Payne, H., Mullan, B., et al. (2012). Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *J. Microbiol. Methods* 91, 80–88. doi: 10.1016/j.mimet.2012.07.008

Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5088–5090. doi: 10.1073/pnas.74.11.5088

Wommack, K. E., Bhavsar, J., and Ravel, J. (2008). Metagenomics: read length matters. *Appl. Environ. Microbiol.* 74, 1453–1463. doi: 10.1128/AEM.02181-07

Wu, M., and Eisen, J. A. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* 9:R151. doi: 10.1186/gb-2008-9-10-r151

Wu, M., and Scott, A. J. (2012). Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* 28, 1033–1034. doi: 10.1093/bioinformatics/bts079

Wu, Y. W., Tang, Y. H., Tringe, S. G., Simmons, B. A., and Singer, S. W. (2014). MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2:26. doi: 10.1186/2049-2618-2-26

Xu, D., and Zhang, Y. (2012). Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* 80, 1715–1735. doi: 10.1002/prot.24065

Yandell, M., and Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* 13, 329–342. doi: 10.1038/nrg3174

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat. Meth.* 12, 7–8. doi: 10.1038/nmeth.3213

Yang, Y., Jiang, X. T., and Zhang, T. (2014). Evaluation of a hybrid approach using UBLAST and BLASTX for metagenomic sequences annotation of specific functional genes. *PLoS ONE* 9:e110947. doi: 10.1371/journal.pone.0110947

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K. H., et al. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Micro.* 12, 635–645. doi: 10.1038/nrmicro3330

Ye, Y., and Doak, T. G. (2009). A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput. Biol.* 5:e1000465. doi: 10.1371/journal.pcbi.1000465

Ye, Y., and Tang, H. (2008). An ORFome assembly approach to metagenomics sequences analysis. *Comput. Syst. Bioinformatics Conf.* 7, 3–13. doi: 10.1142/9781848162648_0001

Yooseph, S., Li, W., and Sutton, G. (2008). Gene identification and protein classification in microbial metagenomic sequence data via incremental clustering. *BMC Bioinformatics* 9:182. doi: 10.1186/1471-2105-9-182

Zhou, Q., Su, X., Jing, G., and Ning, K. (2014). Meta-QC-chain: comprehensive and fast quality control method for metagenomic data. *Genomics Proteomics Bioinformatics* 12, 52–56. doi: 10.1016/j.gpb.2014.01.002

Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucl. Acids Res.* 38, e132–e132. doi: 10.1093/nar/gkq275

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Escobar-Zepeda, Vera-Ponce de León and Sanchez-Flores. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

INVESTIGATION

Genomes of *Candidatus* Wolbachia bourtzisii *w*DacA and *Candidatus* Wolbachia pipientis *w*DacB from the Cochineal Insect *Dactylopius coccus* (Hemiptera: Dactylopiidae)

Shamayim T. Ramírez-Puebla,*[,1] Ernesto Ormeño-Orrillo,†[,1] Arturo Vera-Ponce de León,* Luis Lozano,* Alejandro Sanchez-Flores,‡ Mónica Rosenblueth,* and Esperanza Martínez-Romero*[,2]
*Centro de Ciencias Genómicas and ‡Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, CP 62210, Mexico and †Laboratorio de Ecología Microbiana y Biotecnología "Marino Tabusso", Universidad Nacional Agraria La Molina, Lima 12, Peru

ABSTRACT *Dactylopius* species, known as cochineal insects, are the source of the carminic acid dye used worldwide. The presence of two *Wolbachia* strains in *Dactylopius coccus* from Mexico was revealed by PCR amplification of *wsp* and sequencing of 16S rRNA genes. A metagenome analysis recovered the genome sequences of *Candidatus* Wolbachia bourtzisii *w*DacA (supergroup A) and *Candidatus* Wolbachia pipientis *w*DacB (supergroup B). Genome read coverage, as well as 16S rRNA clone sequencing, revealed that *w*DacB was more abundant than *w*DacA. The strains shared similar predicted metabolic capabilities that are common to *Wolbachia*, including riboflavin, ubiquinone, and heme biosynthesis, but lacked other vitamin and cofactor biosynthesis as well as glycolysis, the oxidative pentose phosphate pathway, and sugar uptake systems. A complete tricarboxylic acid cycle and gluconeogenesis were predicted as well as limited amino acid biosynthesis. Uptake and catabolism of proline were evidenced in *Dactylopius Wolbachia* strains. Both strains possessed WO-like phage regions and type I and type IV secretion systems. Several efflux systems found suggested the existence of metal toxicity within their host. Besides already described putative virulence factors like ankyrin domain proteins, VlrC homologs, and patatin-like proteins, putative novel virulence factors related to those found in intracellular pathogens like *Legionella* and *Mycobacterium* are highlighted for the first time in *Wolbachia*. Candidate genes identified in other *Wolbachia* that are likely involved in cytoplasmic incompatibility were found in *w*DacB but not in *w*DacA.

KEYWORDS
endosymbiont
scale insect

Many insects possess vertically-transmitted bacterial symbionts that provide them with amino acids and vitamins (Moran 2006). While most insect endosymbionts belong to the Gammaproteobacteria there are others in many other phyla (Moran *et al.* 2008). A remarkable case is the *Wolbachia* endosymbiont that infects between 40% (Zug and Hammerstein 2012) to 66% (Hilgenboecker *et al.* 2008) of arthropod species. *Wolbachia* are phylogenetically affiliated to the Alphaproteobacteria, not distantly related to *Rickettsia*, *Ehrlichia*, and *Anaplasma* (Williams *et al.* 2007). There are 16 phylogenetic supergroups of *Wolbachia* identified, and 10 of them are associated with arthropods (Augustinos *et al.* 2011). Based on phylogenomic analysis, six *Wolbachia* supergroups have been separated in eight species (Ramírez-Puebla *et al.* 2015).

Wolbachia are nematode as well as arthropod symbionts (Hilgenboecker *et al.* 2008; Sommer and Streit 2011), and have different effects in their hosts ranging from parasitism to mutualism with spatial and temporal spread of infections in some insect populations (Vavre and Charlat 2012). In nematodes, *Wolbachia* provide vitamins, energy, help in embryo development, and are capable of evading the host immune response (Darby *et al.* 2012; Landmann

Copyright © 2016 Ramírez-Puebla *et al.*
doi: 10.1534/g3.116.031237
Manuscript received May 16, 2016; accepted for publication August 18, 2016; published Early Online August 19, 2016.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Supplemental material is available online at www.g3journal.org/lookup/suppl/doi:10.1534/g3.116.031237/-/DC1
[1]These authors contributed equally to this work.
[2]Corresponding author: Centro de Ciencias Genómicas, UNAM, Av. Universidad SN, Col. Chamilpa, CP 62210, Cuernavaca, Morelos, Mexico. E-mail: emartine@ccg.unam.mx

et al. 2014). In arthropods, *Wolbachia* have been found infecting many tissues inside the insect body including reproductive tracts and somatic cells as bacteriocytes (Dobson *et al.* 1999; Clark *et al.* 2005; Hosokawa *et al.* 2010; Sacchi *et al.* 2010; Saha *et al.* 2012). They alter the host reproduction by induction of parthenogenesis (Stouthamer *et al.* 1999), male-killing (Duplouy *et al.* 2013), feminization (Stouthamer *et al.* 1999), and strain incompatibility (Rousset *et al.* 1992). However, it is also known that *Wolbachia* may confer benefits to insects by playing an important role in insect development and survival (Dedeine *et al.* 2001). For example, removal of *Wolbachia* with antibiotics in *Asobara tabida* wasps inhibits maturation of oocytes (Dedeine *et al.* 2001). In *Drosophila*, *Wolbachia* may confer protection against virus infections (Teixeira *et al.* 2008; Chrostek *et al.* 2013) and provide a fecundity benefit to females when subjected to low or high iron diets (Brownlie *et al.* 2009). Thus, *Wolbachia* inside insects may not be only facultative symbionts, but can also be obligate endosymbionts necessary for survival (Dedeine *et al.* 2001).

There are 12 *Dactylopius* species (Ben-Dov 2006; Van Dam and May 2012). Six of them are present in Mexico, including the smallest and most distantly related *Dactylopius tomentosus* (Portillo and Vigueras 2006; Chávez-Moreno *et al.* 2009). *Dactylopius* insects feed exclusively on the sap of cactus plants of the genera *Opuntia* and *Nopalea* (Pérez-Guerra and Kosztarab 1992). Females of these scale insects spend all their lives on the host plant surface, whereas males are winged and short lived. These insects feed on a poor nutritional and low-calorie diet since cactus sap consists mainly of water (88–95%) and is low in nitrogen (0–0.5%) (Stintzing and Carle 2005). The red pigment carmine is obtained from cochineal insects of the genus *Dactylopius*, especially from *D. coccus*, which is a domesticated species. Carmine has been used as a natural dye to color food, medicines, cosmetics, textiles, and artworks, is considered safe for human consumption (Dapson 2005), and has antimicrobial and insecticidal properties (Eisner *et al.* 1980; Pankewitz *et al.* 2007).

Previously, we described a betaproteobacterium, *Candidatus* Dactylopiibacterium carminicum, and other diverse bacterial species associated with *Dactylopius* species present in Mexico (Ramírez-Puebla *et al.* 2010). Here, we extend the knowledge of *Dactylopius* endosymbionts by reporting the presence and genome sequences of two strains of *Wolbachia*, *Candidatus* Wolbachia bourtzisii *w*DacA (supergroup A) and *Candidatus* Wolbachia pipientis *w*DacB (supergroup B) obtained from Mexican *D. coccus*.

MATERIALS AND METHODS

Sample collection

D. coccus insects were provided by Campo Carmín Greenhouse (Morelos, Mexico) and were maintained on cactus plants (*Opuntia ficus indica* var. Campo Carmín) in a growth room with controlled photoperiod (12L:12D), temperature (25°), and humidity (40–60%). Other *Dactylopius* species were collected from different states in Mexico: *D. confusus* from Tlaxcala, *D. ceylonicus* from Estado de México, *D. opuntiae* from Querétaro and Mexico City, and *D. tomentosus* from Hidalgo.

DNA extraction for detection of Wolbachia in Dactylopius individuals

DNA from the whole bodies of adult females of *Dactylopius* species collected in Mexico were extracted and purified with DNeasy Blood & Tissue Kit (QIAGEN) following the manufacturer's instructions. PCRs were performed using primer pairs wsp81F/wsp691R (Braig *et al.* 1998) and 27F/1492R (Lane 1991) directed to the *wsp* and 16S rRNA genes, respectively.

Recovery of Wolbachia genomes

Sequence and assembly of metagenomic DNA from samples of pooled *D. coccus* individuals, as well as the recovery of the *Wolbachia* genomes from the metagenome, were previously reported (Ramírez-Puebla *et al.* 2016). For 454 sequencing, 2 g (20 individuals) of adult females were superficially disinfected with 70% ethanol, rinsed with sterile distilled water, and dissected with sterile forceps to remove the exoskeleton and guts. Cells in the hemolymph and debris were separated by centrifugation in a Percoll gradient (adapted from Charles and Ishikawa 1999), phases were observed under a microscope, and those with cells were selected for DNA extraction. For PacBio sequencing, eight individuals were superficially disinfected as previously described. Guts and exoskeleton were removed with sterile forceps. Hemolymph from all individuals was pooled for DNA extraction. For Illumina sequencing, guts, ovaries, and Malpighian tubules from 40 females were dissected using sterile forceps under a stereoscopic microscope. These organs were pooled, suspended in PBS, and macerated using a sterile plastic pestle. In all cases, DNA was extracted with DNeasy Blood & Tissue Kit (QIAGEN) following the manufacturer´s instructions. Sequencing was performed at Macrogen Inc. (Korea) for Illumina and 454 and at Duke University Genome Sequencing Core Facility (USA) for PacBio.

Genome analysis

The RAST server was used for gene prediction and annotation (Aziz *et al.* 2008). Manual curation of relevant genes was performed after comparisons with sequences deposited in the following databases: nr and Refseq via BLASTX (Benson *et al.* 2013), the Conserved Domain Database at GenBank (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), the Protein families (PFAM) database (Finn *et al.* 2014), and the Transport Classification Database (Saier *et al.* 2014). Genome completeness was assessed by the presence of single-copy widespread orthologs with BUSCO (Simão *et al.* 2015). Four out of the forty genes evaluated by BUSCO are absent in all the sequenced *Wolbachia* genomes and this was taken into consideration for the calculations.

Data availability

The genome sequences of *Wolbachia* strains *w*DacA and *w*DacB have been deposited in the GenBank database under accession numbers LSYX00000000 and LSYY00000000, respectively.

RESULTS

Wolbachia in Mexican Dactylopius spp.

Previously, no *Wolbachia* sequences were found by PCR amplification of the 16S rRNA gene with primer pair fD1/rD1 (Weisburg *et al.* 1991) from several *Dactylopius* samples collected in Mexico and Brazil (Ramírez-Puebla *et al.* 2010). We reassessed the presence of these endosymbionts by PCR amplification of the *Wolbachia*-specific *wsp* gene (Braig *et al.* 1998). Amplicons of the expected size were obtained from *D. ceylonicus*, *D. coccus*, *D. confusus*, *D. opuntiae*, and *D. tomentosus*, although not all individuals of each species gave a positive reaction. To identify the *Wolbachia* inhabiting *D. coccus* in Mexico, 16S rRNA gene amplification with primer pair 27F/1492R and sequencing was performed. Two divergent sequences were found; one affiliated with *Wolbachia* supergroup A and the other with supergroup B. Sequences of the latter supergroup were more abundant in all surveyed *D. coccus* individuals (Table 1).

Table 1 Abundance of sequences matching *Candidatus* Wolbachia bourtzisii (*w*DacA) and *Candidatus* Wolbachia pipientis (*w*DacB) in *D. coccus* individuals

	Number of 16S rRNA Gene Sequences Assigned to	
Individual	*w*DacA	*w*DacB
Female 1	0	15
Female 2	1	9
Female 3	1	6
Embryo 1	7	8
Embryo 2	2	11

Divergence between Wolbachia from Dactylopius from different countries

We have recently reported the recovery of two contig bins matching *Wolbachia* from a metagenome of *D. coccus* (Ramírez-Puebla *et al.* 2015, 2016). A phylogenomic analysis of those bins (Ramírez-Puebla *et al.* 2015), confirmed that they corresponded to the genomes of two different *Wolbachia* strains belonging to *Candidatus* Wolbachia bourtzisii (supergroup A) and *Candidatus* Wolbachia pipientis (supergroup B), which will be referred to here as *w*DacA and *w*DacB, respectively.

Wolbachia from supergroups A and B were previously reported in *Dactylopius* sp. collected in Lanzarote, Canary Islands, Spain (Pankewitz *et al.* 2007). The Canarian and Mexican *Wolbachia* from supergroups A and B showed, respectively, 99.8% and 98.3% identity at the *ftsZ* gene, and 100% and 98.3% identity at the *wsp* gene. Thus, *Wolbachia* infecting *Dactylopius* sp. populations in the Canary Islands are closely related but distinct to the Mexican *Wolbachia*, the divergence being more pronounced among supergroup B representatives. Recently, a *Wolbachia* genome was recovered during a genome sequencing of *D. coccus* (Campana *et al.* 2015). The reported *w*Coc1 genome was found to belong to supergroup B by *ftsZ* gene sequence analysis, but no analysis of the genome was provided. *w*Coc1 showed 92.4% and 98.2% ANI values with *w*DacA and *w*DacB, respectively, indicating that *w*Coc1 and *w*DacB belong to the same species. No further comparison against our strains was performed because the *w*Coc1 genome assembly was highly fragmented (1064 contigs, N50 size = 1387 bp), and also because that genome may represent a chimera as it is the product of sequences originating from two different and geographically distant *D. coccus* populations, one from Oaxaca in Mexico and the other from Peru.

Genomes sequences of Wolbachia strains wDacA and wDacB

The number of contigs and N50 sizes of genome assemblies of *w*DacA and *w*DacB were 157 and 13.7 kb and 198 and 14.5 kb, respectively (Table 2), values that were average in comparison to released WGS genomes of *Wolbachia*. Genome completeness assessed with BUSCO (Simão *et al.* 2015) indicated that the recovered genomes of *w*DacA and *w*DacB represented 92% and 94%, respectively, of their whole genomes. It should be pointed out that closed *Wolbachia* genomes are reported as 92–94% complete by BUSCO because from one to three of the evaluated genes are either missing or fragmented in any given genome. Read coverage was widely different between both genomes, 2700 × in *w*DacB *vs.* 174 × in *w*DacA, indicating that the first *Wolbachia* strain is predominant in the tissues of *D. coccus* used in this study. Detection of each *Wolbachia* strain by 16S rRNA gene PCR amplification and sequencing in isolated individuals of *D. coccus* seemed to corroborate that *w*DacB is more abundant than *w*DacA (Table 1). As in other *Wolbachia* strains, *w*DacA and *w*DacB strains showed reduced genomes and low G + C contents (Table 2). Hypothetical genes represented 35% and 23% of the CDS genes in *w*DacA and *w*DacB, respectively. *Wolbachia* strains show high genome plasticity compared with other insect endosymbionts. The presence of a high proportion of mobile DNA and insertion sequences (Bordenstein and Reznikoff 2005; Cordaux *et al.* 2008) may promote this plasticity. The two *Wolbachia* strains of *D. coccus* were not exceptions, although it is worth mentioning that the genome of *w*DacB has a higher number of genes annotated as coding for mobile genetic elements and transposases (404, 24% of the CDS genes) in comparison to *w*DacA (120, 9% of the CDS genes).

Vitamin, coenzymes, cofactors, and nucleotide synthesis

Both *Wolbachia* strains from *D. coccus* seemed able to synthesize riboflavin and ubiquinone (coenzyme Q). They also had genes required for purine and pyrimidine nucleotide biosynthesis. They lacked complete biosynthesis genes for biotin, thiamine, coenzyme A, NAD, and folic acid. Nevertheless, an uptake system for biotin and a gene for folate salvage were found encoded in each genome. Both strains also possessed a bacterioferritin gene and heme biosynthesis genes.

Metabolism

The set of genes for the tricarboxylic cycle was complete in both genomes. There were genes for the pentose phosphate pathway but not the oxidative reactions. The phosphofructokinase gene is absent, suggesting that there may be gluconeogenesis but not glycolysis. Cytochrome c oxidase, as well as components of the respiratory complex, were found in both strains.

As has been observed in other *Wolbachia* and other Rickettsiales, most amino acid biosynthesis pathways were incomplete. However, catabolic genes for proline, aspartate, glutamate, and possibly cysteine were identified in both strains. Genes for glutamate dehydrogenase, glutamine synthetase (GS), and glutamate synthase (GOGAT) required for ammonia assimilation were also present. NifU was identified but no other nitrogen fixation genes. Nif proteins involved in the formation of FeS clusters or other metallo clusters can be found in organisms that do not fix nitrogen.

A complete set of genes for fatty acid biosynthesis were present in both genomes as well as for the synthesis of the phospholipids phosphatidylethanolamine, phosphatidylglycerol, and phosphatidylserine. No genes for lipopolysaccharide biosynthesis were found in either genome. *w*DacA and *w*DacB had genes for peptidoglycan synthesis but no transpeptidase genes for chain cross-linking were found.

Transport

Both genomes encoded genes for ATP-binding cassette (ABC) transporters for uptake of phosphate (*pstABCS* genes), ferric iron, zinc, and possibly lipids; one for export of heme; and one gene for a Mg^{+2} (or Co^{+2}) transporter-E (MgtE) family importer. Several genes for putative amino acid symporters were shared by both genomes including five of the major facilitator superfamily (MFS), three of the alanine/glycine: cation symporter (AGCS) family, and one of the dicarboxylate/amino acid:cation symporter (DAACS) family. Strain *w*DacA but not *w*DacB had genes coding for an ABC uptake transporter for glutamine/glutamate. On the other hand, strain *w*DacB possessed three uptake systems of the drug/metabolite transporter (DMT) superfamily, and two genes for organophosphate:phosphate MFS antiporters that were not present in *w*DacA. The former DMT transporters were > 75% similar to the S-adenosylmethionine (SAM) uptake transporter of *Rickettsia prowazekii*

Table 2 Genome features of *Candidatus* Wolbachia bourtzisii *w*DacA and *Candidatus* Wolbachia pipientis *w*DacB from *D. coccus*

Feature	*w*DacA	*w*DacB	Other *Wolbachia* Genomes [Median (Range)]
Number of contigs	157	198	195 (9 – 1064)[a]
N50 (kbp)	13.7	14.5	13.2 (1.4 – 466)[a]
Estimated genome size (Mbp)	1.170	1.498	1.224 (0.444 – 1.542)
G + C content (%)	35.1	34.1	34.3 (32.1 – 36.3)
CDS genes	1040	1246	1216 (660 – 1587)[b]
With function	720	810	
Hypothetical	320	436	
RNA genes			
rRNA	2	2	3 (0 – 4)[b]
tRNA	31	39	34 (14 – 39)[b]

*w*DacA, *Candidatus* Wolbachia bourtzisii (supergroup A); *w*DacB, *Candidatus* Wolbachia pipientis (supergroup B); CDS, coding sequence; rRNA, ribosomal RNA; tRNA, transfer RNA.
[a]Calculated only for unfinished genomes.
[b]Number of CDS, rRNA, and tRNA genes *de novo* predicted with prodigal, RNAmmer 1.2, and tRNAscan-SE v.1.3.1, respectively.

(Tucker *et al.* 2003), and the highest similarities (~49%) of the latter MFS antiporters were to proteins of *R. prowazekii* which have been implicated in triose phosphate uptake used for phospholipid biosynthesis (Frohlich and Audia 2013). No hexose transporter genes were found, supporting the theory that there is no glycolysis in both strains.

Few export transporters were found in both genomes. Besides the heme exporter, both genomes encoded an ABC transporter putatively involved in organic solvent resistance, a CorC-family transporter for magnesium or cobalt efflux, and a cation diffusion-facilitator (CDF) family exporter for zinc or cadmium. In addition, the *w*DacA genome encoded two ABC superfamily transporters, one of the heavy metal transporter (HMT) family related to exporters for phytochelatins-Cd complexes and the other of the multidrug resistance (MDR) family.

Secretion systems

Of the two systems for protein export into the periplasm, only the general secretion *sec* system was found encoded in *w*DacA and *w*DacB genomes. Protein secretion into the extracellular environment is accomplished by several types of secretion systems, of which only two were found in the *Wolbachia* strains of *Dactylopius*. Both genomes coded for the inner membrane component and the membrane fusion protein of a type I secretion system (T1SS) whose products were 95% and 83% identical, respectively, between the strains. The outer membrane TolC, a channel that acts in conjunction with the other T1SS components, was coded elsewhere in the genomes.

Both strains possessed one type IV secretion system (T4SS). The gene organization was similar to that found in other *Wolbachia* with two separated clusters, one including *virB3*, *virB4*, and four copies of *virB6*, and another cluster with *virB8*, *virB9*, *virB10*, *virB11*, and *virD4*. As it is also observed in other *Wolbachia*, there was one paralogue of each of *virB4*, *virB8*, and *virB9* coded elsewhere in the genomes. Genes *virB1*, *virB2*, *virB5*, *and virB7* have been reported as being absent in *Wolbachia* and in Rickettsiales in general (Pichon *et al.* 2009). However, we found four and three homologs of the pilin *virB2* gene in *w*DacA and *w*DacB, respectively. The *virB2* homologs were not clustered with each other or with other *vir* genes. BLAST searches recovered *virB2* homologs in many *Wolbachia* genomes (data not shown) that are annotated mostly as hypothetical or membrane proteins.

In the symbiotic *w*Bm strain of the nematode *Brugia malayi*, the transcriptional regulators *w*BmxR1 and *w*BmxR2 bind to the promoter regions of some *vir* genes (Li and Carlow 2012). *w*BmxR1 seems to regulate the *virB8* operon (which includes the upstream riboflavin biosynthesis gene *ribA*) and the second copy of *virB9*, while *w*BmxR2 controls the expression of the second copy of *virB4* (Li and Carlow 2012). Homologs coding for proteins > 74% identical to the *w*BmxR1 product were found in both our *Wolbachia* strains, while a homolog to *w*BmxR2 was found only in *w*DacA (78% identity).

Stress response

Although living in a relatively protected environment inside their host cells, endosymbionts still retain genes required to cope with stressful conditions. Potassium homeostasis is important to react to changes in osmotic pressure and pH changes. One TrkG potassium uptake protein was found encoded in *w*DacA, while there were two in *w*DacB. Both strains had a glutathione-regulated potassium-efflux system KefKL. An HtrA protease/chaperone for degradation of misfolded or mislocalized cell-envelope proteins was encoded in each genome. Genes to contend with oxidative stress, like those for a Fe superoxide dismutase, an alkyl hydroperoxide reductase, three glutaredoxins, and glutathione biosynthesis, were also found. A single gene for a bacterial flavohemoglobin in each genome may be used to contend with nitrosative stress. Common proteins used for temperature stress, such as DnaK-DnaJ-GrpE composing the DnaK chaperone system and GroEL-GroES composing the GroE chaperonin machinery, were found encoded in both genomes as well as a CspA-family cold shock protein. A single sigma factor RpoH protein was encoded in each genome and may be used for stress response.

Virulence factors

Ankyrin domains are involved in protein–protein interactions and, by interacting with specific regions of the host chromatin, can modulate host gene transcription in other bacteria (Iturbe-Ormaetxe *et al.* 2005; Siozios *et al.* 2013). Genes coding proteins with ankyrin domains were found in both strain genomes, although *w*DacB had double the number of genes in comparison to *w*DacA (34 *vs.* 17 genes). Neither strain possessed genes for chemotaxis or motility via flagella or type IV pilus.

Two of the five MFS transporters of *w*DacA and *w*DacB belonged to the phagosomal nutrient transporter (Pht) family (Chen *et al.* 2008). Amino acid and pyrimidine transporters of this family are required for intracellular survival of *Legionella pneumophila* in macrophages (Sauer *et al.* 2005; Fonseca *et al.* 2014). Homologs to Pht transporters were found encoded in many *Wolbachia* genomes and they were distributed in two major phylogenetic clusters which suggests functional divergence (Supplemental Material, Figure S1).

One gene in each *Wolbachia* strain coded for a protein bearing the mammalian cell entry (MCE) domain. Proteins in this family have been identified as necessary for intracellular colonization and survival by *Mycobacterium tuberculosis* and *M. bovis* (Arruda *et al.* 1993; Flesselles

et al. 1999). Several genomes of *Wolbachia* (Table S1) as well as other Rickettsiales encoded homologs for MCE proteins. Although not restricted to intracellular bacteria, MCE homologs are present in other facultative or obligate endosymbiotic and parasitic bacteria (Table S2).

Finally, genes coding for proteins that have been highlighted as candidates to induce cytoplasmic incompatibility were also found. Both genomes possessed two copies of the DNA-binding protein HU beta (Beckmann *et al.* 2013). *w*DacB but not *w*DacA had genes coding for proteins that are homologs to WPIP0282, which seems to be present only in *Wolbachia* strains inducing cytoplasmic incompatibility (Beckmann and Fallon 2013). *w*DacB possessed two homologs of the transcriptional regulator *wtrM* gene whose product is able to upregulate the expression of a host gene implicated in cytoplasmic incompatibility (Pinto *et al.* 2013). The *wtrM* gene of *w*DacA was split into two halves by a frameshift mutation.

Phages

Several genes of gene clusters encoding phage proteins were found in both *Wolbachia* genomes. A complete cluster including phage head-baseplate or head-baseplate-tail genes was not recovered; clusters included either head, baseplate, or tail genes. Paralogous genes located in different head or baseplate clusters in each genome suggested that each strain possesses more than one phage, although it was not possible to determine if any of these phages is complete. The tail clusters, one in each genome, were associated with putative virulence genes: two homologs of the VlrC protein in *w*DacA, and an ankyrin domain protein and a patatin-like protein in *w*DacB.

DISCUSSION

The presence of *Wolbachia* in *D. coccus* (Campana *et al.* 2015) and *Dactylopius* sp. (Pankewitz *et al.* 2007) has been reported. In this study, *Wolbachia* PCR products were obtained from DNA extracted from Mexican samples of *D. ceylonicus*, *D. coccus*, *D. confusus*, *D. opuntiae*, and *D. tomentosus*. These data show that *Wolbachia* might have started its endosymbiotic state with cochineal insects before the genus *Dactylopius* had diverged.

We report the presence of two *Wolbachia* strains, *w*DacA and *w*DacB, in Mexican individuals of *D. coccus*. The large difference in read coverage between the genomes of *w*DacA and *w*DacB indicates that the latter strain is prevalent in *D. coccus*, at least in the tissue samples used here. Interestingly, *w*DacB but not *w*DacA possessed homologs coding for proteins that are likely involved in causing cytoplasmic incompatibility, a mechanism promoting persistence and dissemination of *Wolbachia* in their hosts. In wasps, double infections of supergroup A and group B strains have been found to influence reproductive and ecological isolation among sibling *Nasonia* species; therefore, *Wolbachia* has been implicated in wasp speciation (Bordenstein *et al.* 2001).

Dactylopius insects feed on low-nutrient cactus sap and therefore have to develop strategies to acquire nutrients lacking in their diet from their symbiotic relationships with bacteria. In other insects, riboflavin is produced by their endosymbionts such as the gammaproteobacterium *Wigglesworthia* for Tsetse flies (Akman *et al.* 2002) and *Buchnera* for aphids (Nakabachi and Ishikawa 1999). Recently, it was demonstrated that *Wolbachia* contributes to the growth, survival, and reproduction of bedbugs by riboflavin provisioning (Moriyama *et al.* 2015). Furthermore, it has been postulated that *Wolbachia* strains can also act as heme providers and/or helpers in maintaining iron homeostasis in the host (Brownlie *et al.* 2009; Kremer *et al.* 2009). Both *Wolbachia* strains from *D. coccus* possessed genes for the biosynthesis of riboflavin, heme, and the iron-storage protein bacterioferritin. All these genes are common in *Wolbachia* from insects, even in the sex-manipulator strains that negatively affect the host fitness. For riboflavin, Moriyama *et al.* (2015) have found evidence that its provisioning can counteract the negative effects caused by *Wolbachia* in their hosts.

As has been observed in other *Wolbachia*, *w*DacA and *w*DacB do not have genes to produce most amino acids that their insect hosts require, which may be provided by other bacteria present in *D. coccus* (Ramírez-Puebla *et al.* 2010) that could act as symbionts. The lack of most amino acid biosynthetic capabilities suggests the dependence of *Wolbachia* on its host or on other endosymbionts. Retention of amino acid biosynthesis defines primary insect symbionts and its absence seems to be a characteristic of secondary symbionts (Darby *et al.* 2012). Lack of a functional glycolysis pathway and the presence of several amino acid uptake systems indicate that *Wolbachia* utilizes amino acids instead of sugars as nutrients. Many of the MFS transporters may be proline uptake systems that, together with the presence of PutA for proline catabolism, suggest that this amino acid could be a major nutrient for *Wolbachia*. In fact, high-level expression of PutA has been demonstrated by proteomic analysis of *Wolbachia* (Baldridge *et al.* 2014). Interestingly, proline is an excellent precursor for riboflavin production in the legume endosymbiont *Sinorhizobium* (Phillips *et al.* 1999).

Symbiotic and pathogenic bacteria can use effector proteins delivered to their host via the T4SS to promote intracellular colonization and persistence (Juhas *et al.* 2008). T4SS is widely found in *Wolbachia* strains (Pichon *et al.* 2009) and was also found in *w*DacA and *w*DacB. It was surprising to note that *virB2*, coding for the major T-pilus component, was reported as being absent in *Wolbachia* and other Rickettsiales (Rancès *et al.* 2008; Pichon *et al.* 2009). We found several *virB2* homologs in *w*DacA and *w*DacB, as well as in many *Wolbachia* genomes. This is in agreement with recent data obtained in other Rickettsiales, like *Anaplasma phagocytophilum* (Dugat *et al.* 2014) and *Neorickettsia risticii* (Lin *et al.* 2009) which do possess several *virB2* paralogues. In *N. risticii*, VirB2 is located on the cell surface in agreement with its function as the major T4SS pilus protein (Lin *et al.* 2009). In *A. phagocytophilum*, the AnkA protein is exported via a T4SS (Lin *et al.* 2007). Although T4SS are known to transport proteins and/or DNA, an intriguing possibility is that they can act as nutrient transporters in *Wolbachia* given the scarcity of nutrient export systems in the genomes of these bacteria.

Several efflux systems for heavy metals were found in both genomes suggesting that *Wolbachia* from *D. coccus* have to cope with metal toxicity, perhaps contributed by their host diet. In relation to this, the mucilage of *Opuntia* cacti acts as a good water-soluble chelating polymer (polyelectrolyte) able to remove heavy metals from water (Barka *et al.* 2013), and metal-bound phytochelatin can be found in *Opuntia* shoots (Landero Figueroa *et al.* 2007). Besides heavy metals, other harmful conditions are likely acting on *w*DacA and *w*DacB as both their genomes carried several genes to contend with abiotic stresses. Proteomic profiling of a mosquito *Wolbachia* strain has revealed a profile dominated by chaperones and stress proteins (Baldridge *et al.* 2014).

Another secretion system used by bacteria to interact with eukaryotes is the T1SS. In pathogenic bacteria, virulence factors such as hemolysins are secreted via this system. Secretion of some ankyrin domain proteins by T1SS has been reported in *Rickettsia* (Kaur *et al.* 2012) and *Ehrlichia* (Wakeel *et al.* 2011). Proteins bearing typical T1SS-secretion motifs could not be found in either of our *Wolbachia* genomes, but it is worth noting that several ankyrin domain proteins were coded near the gene for the T1SS inner membrane component in *w*DacB.

In both *Wolbachia* strains, we found transporters of the Pht family, which have been described as virulence factors in *L. pneumophila* (Sauer *et al.* 2005; Fonseca *et al.* 2014). These transporters are present in other Legionellales, in Chlamydiales, as well as in other Rickettsiales besides *Wolbachia*, all having intracellular lifestyles. A protein encoded by *w*DacA and *w*DacB showed homology to the virulence factor Mce of *Mycobacterium* which, when expressed in nonpathogenic *Escherichia coli*, confers the ability to invade and survive within macrophages (Haile *et al.* 2002). The presence of all these putative virulence factors has not been previously pointed out in *Wolbachia*.

ACKNOWLEDGMENTS

The authors wish to thank Michael Dunn and Julio Martínez for technical support and for reading the manuscript, and Campo Carmín Greenhouse for providing *D. coccus* insects. Consejo Nacional de Ciencia y Tecnología (CONACyT) grant 154453 provided financial support. S.T.R.-P. and A.V.-P.L. were in the Programa de Doctorado en Ciencias Biomedicas, Universidad Nacional Autónoma de México and received a scholarship from CONACyT. All the authors declare no conflict of interest, financial or otherwise, that might potentially bias this work.

LITERATURE CITED

Akman, L., A. Yamashita, H. Watanabe, K. Oshima, T. Shiba *et al.*, 2002 Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat. Genet. 32: 402–407.

Arruda, S., G. Bomfim, R. Knights, T. Huima-Byron, and L. W. Riley, 1993 Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. Science 261: 1454–1457.

Augustinos, A. A., D. Santos-Garcia, E. Dionyssopoulou, M. Moreira, A. Papapanagiotou *et al.*, 2011 Detection and characterization of *Wolbachia* infections in natural populations of aphids: is the hidden diversity fully unraveled? PLoS One 6: e28695.

Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz *et al.*, 2008 The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9: 75.

Baldridge, G. D., A. S. Baldridge, B. A. Witthuhn, L. Higgins, T. W. Markowski *et al.*, 2014 Proteomic profiling of a robust *Wolbachia* infection in an *Aedes albopictus* mosquito cell line. Mol. Microbiol. 94: 537–556.

Barka, N., M. Abdennouri, M. El Makhfouk, and S. Qourzal, 2013 Biosorption characteristics of cadmium and lead onto eco-friendly dried cactus (*Opuntia ficus indica*) cladodes. J. Environ. Chem. Eng. 1: 144–149.

Beckmann, J. F., and A. M. Fallon, 2013 Detection of the *Wolbachia* protein WPIP0282 in mosquito spermathecae: Implications for cytoplasmic incompatibility. Insect Biochem. Mol. Biol. 43: 867–878.

Beckmann, J. F., T. W. Markowski, B. A. Witthuhn, and A. M. Fallon, 2013 Detection of the *Wolbachia*-encoded DNA binding protein, HU beta, in mosquito gonads. Insect Biochem. Mol. Biol. 43: 272–279.

Ben-Dov, Y., 2006 *A systematic catalogue of eight scale insect families (Hemiptera: Coccoidea) of the world. Aclerdidae, Asterolecaniidae, Beesoniidae, Carayonemidae, Conchaspididae, Dactylopiidae, Kerriidae and Lecanodiaspididae*, Elsevier, Oxford, UK.

Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman *et al.*, 2013 GenBank. Nucleic Acids Res. 41: D36–D42.

Bordenstein, S. R., and W. S. Reznikoff, 2005 Mobile DNA in obligate intracellular bacteria. Nat. Rev. Microbiol. 3: 688–699.

Bordenstein, S. R., F. P. O'Hara, and J. H. Werren, 2001 *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. Nature 409: 707–710.

Braig, H. R., W. Zhou, S. L. Dobson, and S. L. O'Neill, 1998 Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. J. Bacteriol. 180: 2373–2378.

Brownlie, J. C., B. N. Cass, M. Riegler, J. J. Witsenburg, I. Iturbe-Ormaetxe *et al.*, 2009 Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipientis*, during periods of nutritional stress. PLoS Pathog. 5: e1000368.

Campana, M. G., N. M. Robles García, and N. Tuross, 2015 America's red gold: multiple lineages of cultivated cochineal in Mexico. Ecol. Evol. 5: 607–617.

Charles, H., and H. Ishikawa, 1999 Physical and genetic map of the genome of *Buchnera*, the primary endosymbiont of the pea aphid *Acyrthosiphon pisum*. J. Mol. Evol. 48: 142–150.

Chávez-Moreno, C. K., A. Tecante, and A. Casas, 2009 The *Opuntia* (Cactaceae) and *Dactylopius* (Hemiptera: Dactylopiidae) in Mexico: a historical perspective of use, interaction and distribution. Biodivers. Conserv. 18: 3337–3355.

Chen, D. E., S. Podell, J.-D. Sauer, M. S. Swanson, and M. H. Saier, 2008 The phagosomal nutrient transporter (Pht) family. Microbiology 154: 42–53.

Chrostek, E., M. S. Marialva, S. S. Esteves, L. A. Weinert, J. Martinez *et al.*, 2013 *Wolbachia* variants induce differential protection to viruses in *Drosophila melanogaster*: a phenotypic and phylogenomic analysis. PLoS Genet. 9: e1003896.

Clark, M. E., C. L. Anderson, J. Cande, and T. L. Karr, 2005 Widespread prevalence of *Wolbachia* in laboratory stocks and the implications for *Drosophila* research. Genetics 170: 1667–1675.

Cordaux, R., S. Pichon, A. Ling, P. Pérez, C. Delaunay *et al.*, 2008 Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. Mol. Biol. Evol. 25: 1889–1896.

Dapson, R., 2005 A method for determining identity and relative purity of carmine, carminic acid and aminocarminic acid. Biotech. Histochem. 80: 201–205.

Darby, A. C., S. D. Armstrong, G. S. Bah, G. Kaur, M. A. Hughes *et al.*, 2012 Analysis of gene expression from the *Wolbachia* genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. Genome Res. 22: 2467–2477.

Dedeine, F., F. Vavre, F. Fleury, B. Loppin, M. E. Hochberg *et al.*, 2001 Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp. Proc. Natl. Acad. Sci. USA 98: 6247–6252.

Dobson, S. L., K. Bourtzis, H. R. Braig, B. F. Jones, W. Zhou *et al.*, 1999 *Wolbachia* infections are distributed throughout insect somatic and germ line tissues. Insect Biochem. Mol. Biol. 29: 153–160.

Dugat, T., V. Loux, S. Marthey, M. Moroldo, A.-C. Lagree *et al.*, 2014 Comparative genomics of first available bovine *Anaplasma phagocytophilum* genome obtained with targeted sequence capture. BMC Genomics 15: 973.

Duplouy, A., I. Iturbe-Ormaetxe, S. A. Beatson, J. M. Szubert, J. C. Brownlie *et al.*, 2013 Draft genome sequence of the male-killing *Wolbachia* strain *w*Bol1 reveals recent horizontal gene transfers from diverse sources. BMC Genomics 14: 20.

Eisner, T., S. Nowicki, M. Goetz, and J. Meinwald, 1980 Red cochineal dye (carminic acid): its role in nature. Science 208: 1039–1042.

Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt *et al.*, 2014 Pfam: the protein families database. Nucleic Acids Res. 42: D222–D230.

Flesselles, B., N. N. Anand, J. Remani, S. M. Loosmore, and M. H. Klein, 1999 Disruption of the mycobacterial cell entry gene of *Mycobacterium bovis* BCG results in a mutant that exhibits a reduced invasiveness for epithelial cells. FEMS Microbiol. Lett. 177: 237–242.

Fonseca, M. V., J.-D. Sauer, S. Crepin, B. Byrne, and M. S. Swanson, 2014 The *phtC-phtD* locus equips *Legionella pneumophila* for thymidine salvage and replication in macrophages. Infect. Immun. 82: 720–730.

Frohlich, K. M., and J. P. Audia, 2013 Dual mechanisms of metabolite acquisition by the obligate intracytosolic pathogen *Rickettsia prowazekii* reveal novel aspects of triose phosphate transport. J. Bacteriol. 195: 3752–3760.

Haile, Y., D. A. Caugant, G. Bjune, and H. G. Wiker, 2002 *Mycobacterium tuberculosis* mammalian cell entry operon (*mce*) homologs in *Mycobacterium* other than tuberculosis (MOTT). FEMS Immunol. Med. Microbiol. 33: 125–132.

Hilgenboecker, K., P. Hammerstein, P. Schlattmann, A. Telschow, and J. H. Werren, 2008 How many species are infected with *Wolbachia*? – a statistical analysis of current data. FEMS Microbiol. Lett. 281: 215–220.

Hosokawa, T., R. Koga, Y. Kikuchi, X.-Y. Meng, and T. Fukatsu, 2010 *Wolbachia* as a bacteriocyte-associated nutritional mutualist. Proc. Natl. Acad. Sci. USA 107: 769–774.

Iturbe-Ormaetxe, I., G. R. Burke, M. Riegler, and S. L. O'Neill, 2005 Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. J. Bacteriol. 187: 5136–5145.

Juhas, M., D. W. Crook, and D. W. Hood, 2008 Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. Cell. Microbiol. 10: 2377–2386.

Kaur, S. J., M. S. Rahman, N. C. Ammerman, M. Beier-Sexton, S. M. Ceraul *et al.*, 2012 TolC-dependent secretion of an ankyrin repeat-containing protein of *Rickettsia typhi*. J. Bacteriol. 194: 4920–4932.

Kremer, N., D. Voronin, D. Charif, P. Mavingui, B. Mollereau *et al.*, 2009 *Wolbachia* interferes with ferritin expression and iron metabolism in insects. PLoS Pathog. 5: e1000630.

Landero Figueroa, J. A., S. Afton, K. Wrobelac, K. Wrobelac, and J. A. Caruso, 2007 Analysis of phytochelatins in nopal (*Opuntia ficus*): a metallomics approach in the soil–plant system. J. Anal. At. Spectrom. 22: 897–904.

Landmann, F., J. M. Foster, M. L. Michalski, B. E. Slatko, and W. Sullivan, 2014 Co-evolution between an endosymbiont and its nematode host: Wolbachia asymmetric posterior localization and AP polarity establishment. PLoS Negl. Trop. Dis. 8: e3096.

Lane, D. J., 1991 16S/23S rRNA sequencing, pp. 115–147 in *Nucleic Acid Techniques in Bacterial Systematics*, edited by Stackebrandt, E., and M. Goodfellow. John Wiley and Sons, Chichester, UK.

Li, Z., and C. K. S. Carlow, 2012 Characterization of transcription factors that regulate the type IV secretion system and riboflavin biosynthesis in *Wolbachia* of *Brugia malayi*. PLoS One 7: e51597.

Lin, M., A. den Dulk-Ras, P. J. Hooykaas, and Y. Rikihisa, 2007 *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. Cell. Microbiol. 9: 2644–2657.

Lin, M., C. Zhang, K. Gibson, and Y. Rikihisa, 2009 Analysis of complete genome sequence of *Neorickettsia risticii*: causative agent of Potomac horse fever. Nucleic Acids Res. 37: 6076–6091.

Moran, N. A., 2006 Symbiosis. Curr. Biol. 16: R866–R871.

Moran, N. A., J. P. McCutcheon, and A. Nakabachi, 2008 Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. 42: 165–190.

Moriyama, M., N. Nikoh, T. Hosokawa, and T. Fukatsu, 2015 Riboflavin provisioning underlies *Wolbachia*'s fitness contribution to its insect host. MBio 6: e01732-15.

Nakabachi, A., and H. Ishikawa, 1999 Provision of riboflavin to the host aphid, *Acyrthosiphon pisum*, by endosymbiotic bacteria, *Buchnera*. J. Insect Physiol. 45: 1–6.

Pankewitz, F., A. Zollmer, M. Hilker, and Y. Graser, 2007 Presence of *Wolbachia* in insect eggs containing antimicrobially active anthraquinones. Microb. Ecol. 54: 713–721.

Pérez-Guerra, G., and M. Kosztarab, 1992 *Biosystematics of the family Dactylopiidae (Homoptera: Coccinea) with emphasis on the life cycle of Dactylopius coccus Costa*, Virginia Polytechnic Institute and State University, Blacksburg.

Phillips, D. A., C. M. Joseph, G. P. Yang, E. Martinez-Romero, J. R. Sanborn *et al.*, 1999 Identification of lumichrome as a *Sinorhizobium* enhancer of alfalfa root respiration and shoot growth. Proc. Natl. Acad. Sci. USA 96: 12275–12280.

Pichon, S., D. Bouchon, R. Cordaux, L. Chen, R. A. Garrett *et al.*, 2009 Conservation of the type IV secretion system throughout *Wolbachia* evolution. Biochem. Biophys. Res. Commun. 385: 557–562.

Pinto, S. B., K. Stainton, S. Harris, Z. Kambris, E. R. Sutton *et al.*, 2013 Transcriptional regulation of *Culex pipiens* mosquitoes by *Wolbachia* influences cytoplasmic incompatibility. PLoS Pathog. 9: e1003647.

Portillo, M. I., and A. L. Vigueras, 2006 A review on the cochineal species in Mexico, hosts and natural enemies. Acta Hortic. (728): 249–256.

Ramírez-Puebla, S. T., M. Rosenblueth, C. K. Chávez-Moreno, M. C. de Lyra, A. Tecante *et al.*, 2010 Molecular phylogeny of the genus *Dactylopius* (Hemiptera: Dactylopiidae) and identification of the symbiotic bacteria. Environ. Entomol. 39: 1178–1183.

Ramírez-Puebla, S. T., L. E. Servín-Garcidueñas, E. Ormeño-Orrillo, A. V.-P. d. León, M. Rosenblueth *et al.*, 2015 Species in *Wolbachia*? Proposal for the designation of '*Candidatus* Wolbachia bourtzisii', '*Candidatus* Wolbachia onchocercicola', '*Candidatus* Wolbachia blaxterii', '*Candidatus* Wolbachia brugii', '*Candidatus* Wolbachia taylorii', '*Candidatus* Wolbachia collembolicola' and '*Candidatus* Wolbachia multihospitis' for the different species within *Wolbachia* supergroups. Syst. Appl. Microbiol. 38: 390–399.

Ramírez-Puebla, S. T., L. E. Servín-Garcidueñas, E. Ormeño-Orrillo, A. Vera-Ponce de León, M. Rosenblueth *et al.*, 2016 A response to Lindsey *et al.* "*Wolbachia pipientis* should not be split into multiple species: A response to Ramírez-Puebla *et al.* Syst. Appl. Microbiol. 39: 223–225.

Rancès, E., D. Voronin, V. Tran-Van, and P. Mavingui, 2008 Genetic and functional characterization of the type IV secretion system in *Wolbachia*. J. Bacteriol. 190: 5020–5030.

Rousset, F., D. Bouchon, B. Pintureau, P. Juchault, and M. Solignac, 1992 *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. P. Roy. Soc. B-Bol. Sci. 250: 91–98.

Sacchi, L., M. Genchi, E. Clementi, I. Negri, A. Alma *et al.*, 2010 Bacteriocyte-like cells harbour *Wolbachia* in the ovary of *Drosophila melanogaster* (Insecta, Diptera) and *Zyginidia pullula* (Insecta, Hemiptera). Tissue Cell 42: 328–333.

Saha, S., W. B. Hunter, J. Reese, J. K. Morgan, M. Marutani-Hert *et al.*, 2012 Survey of endosymbionts in the *Diaphorina citri* metagenome and assembly of a *Wolbachia* wDi draft genome. PLoS One 7: e50067.

Saier, M. H., V. S. Reddy, D. G. Tamang, and Å. Västermark, 2014 The transporter classification database. Nucleic Acids Res. 42: D251–D258.

Sauer, J.-D., M. A. Bachman, and M. S. Swanson, 2005 The phagosomal transporter A couples threonine acquisition to differentiation and replication of *Legionella pneumophila* in macrophages. Proc. Natl. Acad. Sci. USA 102: 9924–9929.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212.

Siozios, S., P. Ioannidis, L. Klasson, S. G. E. Andersson, H. R. Braig *et al.*, 2013 The diversity and evolution of *Wolbachia* ankyrin repeat domain genes. PLoS One 8: e55390.

Sommer, R. J., and A. Streit, 2011 Comparative genetics and genomics of nematodes: genome structure, development, and lifestyle. Annu. Rev. Genet. 45: 1–20.

Stintzing, F. C., and R. Carle, 2005 Cactus stems (*Opuntia* spp.): A review on their chemistry, technology, and uses. Mol. Nutr. Food Res. 49: 175–194.

Stouthamer, R., J. A. Breeuwer, and G. D. Hurst, 1999 *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. Annu. Rev. Microbiol. 53: 71–102.

Teixeira, L., A. Ferreira, and M. Ashburner, 2008 The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. PLoS Biol. 6: e2.

Tucker, A. M., H. H. Winkler, L. O. Driskell, and D. O. Wood, 2003 S-adenosylmethionine transport in *Rickettsia prowazekii*. J. Bacteriol. 185: 3031–3035.

Van Dam, A. R., and B. May, 2012 A new species of Dactylopius Costa (*Dactylopius gracilipilus* sp. nov.) (Hemiptera: Coccoidea: Dactylopiidae) from the Chihuahuan Desert, Texas, U.S.A. Zootaxa 3573: 33–39.

Vavre, F., and S. Charlat, 2012 Making (good) use of *Wolbachia*: what the models say. Curr. Opin. Microbiol. 15: 263–268.

Wakeel, A., A. den Dulk-Ras, P. J. Hooykaas, and J. W. McBride, 2011 *Ehrlichia chaffeensis* tandem repeat proteins and Ank200 are type 1 secretion system substrates related to the repeats-in-toxin exoprotein family. Front. Cell. Infect. Microbiol. 1: 22.

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

Williams, K. P., B. W. Sobral, and A. W. Dickerman, 2007 A robust species tree for the Alphaproteobacteria. J. Bacteriol. 189: 4578–4586.

Zug, R., and P. Hammerstein, 2012 Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. PLoS One 7: e38544.

Communicating editor: B. Oliver