

# *Candidatus Dactylopiibacterium carminicum*, a Nitrogen-Fixing Symbiont of *Dactylopius Cochineal* Insects (Hemiptera: Coccoidea: Dactylopiidae)

Arturo Vera-Ponce de León<sup>1</sup>, Ernesto Ormeño-Orrillo<sup>1,2</sup>, Shamayim T. Ramírez-Puebla<sup>1</sup>, Mónica Rosenblueth<sup>1</sup>, Mauro Degli Esposti<sup>1</sup>, Julio Martínez-Romero<sup>1</sup>, and Esperanza Martínez-Romero<sup>1,\*</sup>

<sup>1</sup>Programa de Ecología Genómica, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Mexico

<sup>2</sup>Laboratorio de Ecología Microbiana y Biotecnología, Departamento de Biología, Facultad de Ciencias, Universidad Nacional Agraria La Molina, Lima, Peru

\*Corresponding author: E-mail: emartine@ccg.unam.mx.

Accepted: August 22, 2017

**Data deposition:** This project has been deposited at DDBJ/ENA/GenBank as Bioproject PRJNA355137 under accession numbers MQNN000000000 *Candidatus Dactylopiibacterium carminicum* NFE1, NMRM000000000 *Candidatus Dactylopiibacterium carminicum* NFDO, NMRN000000000 *Candidatus Dactylopiibacterium carminicum* NFDCM.

## Abstract

The domesticated carmine cochineal *Dactylopius coccus* (scale insect) has commercial value and has been used for more than 500 years for natural red pigment production. Besides the domesticated cochineal, other wild *Dactylopius* species such as *Dactylopius opuntiae* are found in the Americas, all feeding on nutrient poor sap from native cacti. To compensate nutritional deficiencies, many insects harbor symbiotic bacteria which provide essential amino acids or vitamins to their hosts. Here, we characterized a symbiont from the carmine cochineal insects, *Candidatus Dactylopiibacterium carminicum* (betaproteobacterium, Rhodocyclaceae family) and found it in *D. coccus* and in *D. opuntiae* ovaries by fluorescent in situ hybridization, suggesting maternal inheritance. Bacterial genomes recovered from metagenomic data derived from whole insects or tissues both from *D. coccus* and from *D. opuntiae* were around 3.6 Mb in size. Phylogenomics showed that dactylopiibacteria constituted a closely related clade neighbor to nitrogen fixing bacteria from soil or from various plants including rice and other grass endophytes. Metabolic capabilities were inferred from genomic analyses, showing a complete operon for nitrogen fixation, biosynthesis of amino acids and vitamins and putative traits of anaerobic or microoxic metabolism as well as genes for plant interaction. *Dactylopiibacterium nif* gene expression and acetylene reduction activity detecting nitrogen fixation were evidenced in *D. coccus* hemolymph and ovaries, in congruence with the endosymbiont fluorescent in situ hybridization location. *Dactylopiibacterium* symbionts may compensate for the nitrogen deficiency in the cochineal diet. In addition, this symbiont may provide essential amino acids, recycle uric acid, and increase the cochineal life span.

**Key words:** nitrogen fixation, scale insect, metagenomics, *Azoarcus*.

## 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 and fungi may provide amino acids, pheromones, lipids, or vitamins to their hosts (Adams et al. 2009; Behmer et al. 2011; Douglas 2012, 2016). In other cases, bacteria may compete with pathogens or confer

protection to insects against parasites (Xie et al. 2014). Some insects have specialized cells called bacteriocytes in which they harbor maternally inherited bacterial endosymbionts. Only rarely can endosymbionts be cultured in the laboratory (Dale and Maudlin 1999; Matthew et al. 2005; Dale et al. 2006). Consequently, the advent of genomics paved the way to detailed studies on these organisms. The first genome from an endosymbiont (that of *Buchnera*, the endosymbiont

of aphids) was published by Shigenobu et al. (2000). Since then, many genomes have become available from endosymbionts belonging to different bacterial phyla (reviewed in McCutcheon and Moran 2011; Rosas-Pérez et al. 2014; Van Leuven et al. 2014; Martínez-Cano et al. 2015; Bennett et al. 2016). Typically, endosymbionts show a reduction in gene content, genome erosion, and high AT content, as well as over-representation of genes encoding essential functions for their insect host (McCutcheon and Moran 2011).

In most cases, insect gut bacteria are less diverse than mammalian gut bacteria (Degli Esposti and Martínez-Romero 2017). Among gut bacteria, some diazotrophs (nitrogen-fixing bacteria) may alleviate insect nitrogen deficiencies and compensate carbon to nitrogen imbalance. Spirochetes fix nitrogen in termite guts (Ohkuma et al. 1999; Desai and Brune 2012). In ant (van Borm et al. 2002), fruit fly (Behar et al. 2005), cockroach (Tai et al. 2016), and some beetle (Morales-Jiménez et al. 2009, 2013) guts, there are *Klebsiella*, *Kluyvera*, and *Roultella* (gammaproteobacteria) as nitrogen-fixing symbionts (van Borm et al. 2002; Behar et al. 2005; Morales-Jiménez et al. 2009, 2013; Pinto-Tomás et al. 2009). However, diazotrophs have not been reported in hemipterans such as *Dactylopius* cochineals. Besides nitrogen fixation, nitrogen recycling occurs in insects (Hirayama et al. 1999; Ulyshen 2015); for example, uric acid (UA) may be used by bacteria or fungi in Malpighian tubules (Sabree et al. 2009). Nitrogen recycling seems to occur in the carmine cochineals *Dactylopius coccus* and *Dactylopius opuntiae* as well (Vera-Ponce de León et al. 2016).

*Dactylopius coccus* is the producer of carmine pigment. This scale insect was domesticated by pre-Hispanic Americans and has considerable economic value for the production of natural pigments that are used for dyeing cosmetics, textiles and food (Perez-Guerra and Kosztarab 1992; Campana et al. 2015; Williams and Ben-Dov 2015). Female carmine cochineals spend their entire life on cactus plants and, like other sap-feeding insects, have nutritional deficiencies that may be compensated by their microbial symbionts. It is not unusual to find a diversity of symbiont species in scale insects (Rosenblueth et al. 2017), and this was the case in *Dactylopius*. At least five species of *Dactylopius* harbor a betaproteobacterium that we named *Candidatus Dactylopiibacterium carminicum* (Ramírez-Puebla et al. 2010), hereafter referred to as *Dactylopiibacterium*. This symbiont was found by a PCR-based approach, also in *D. coccus* eggs (Ramírez-Puebla et al. 2010). Cochineal insects additionally contain *Wolbachia* (Ramírez-Puebla et al. 2015, 2016), *Spiroplasma* (our unpublished results), fungi (Vera-Ponce de León et al. 2016) and proteobacteria commonly associated with plants (James et al. 1997; Kim et al. 1998; López-López et al. 2013), such as *Herbaspirillum*, *Mesorhizobium*, *Massilia*, and *Sphingomonas* (Ramírez-Puebla et al. 2010). However, the latter were absent in *D. coccus* and unevenly found in individual cochineals of a few wild species, thereby suggesting

that they were transient gut symbionts acquired from the cactus sap.

Cactuses are very efficient in photosynthesis, but their sap is nitrogen-poor (Stintzing and Carle 2005). Thus, we wondered whether the carmine cochineal symbiont *Dactylopiibacterium*, which is phylogenetically related to diazotrophs such as *Azoarcus* (Ramírez-Puebla et al. 2010), could fix nitrogen to compensate for the nutritional deficiencies in the cochineals feeding solely on cactus sap. The aim of this study was to analyze the genome sequence and nitrogen-fixing capabilities of *Dactylopiibacterium* and test its location in cochineal reproductive tissues to assess its maternal transmission.

## Materials and Methods

### Fluorescent In Situ Hybridization

To identify *Dactylopiibacterium* location inside the carmine cochineal, fluorescent in situ hybridization (FISH) was performed as described by Koga et al. (2009) with modifications. Adult *D. coccus* and *D. opuntiae* females from different life stages (30 and 80 days old) were detached from the host plant and their ovaries were dissected as described (Vera-Ponce de León et al. 2016). Insect tissues were fixed with Carnoy's solution and autofluorescence was quenched with 6% H<sub>2</sub>O<sub>2</sub> for 7 days. The designed oligonucleotide fluorescent probes were Cy5-END1-390 (5'-Cy5-AAACCATTCTTTCCGACTG-3') that is specific for *Dactylopiibacterium* and Cy5-END1-1081 (5'-Cy5-CTTGCGTAGCAACTAATGATAAGG-3') that matches many betaproteobacterial 16S rRNA genes but not those of *Herbaspirillum* and *Massilia* (betaproteobacteria) that were found inside the carmine cochineals *D. ceylonicus* and *D. confusus* (Ramírez-Puebla et al. 2010). FISH was performed adding the probe (100 nM) diluted with hybridization buffer [Tris-HCl (20 mM); NaCl (180 mM); SDS 10%; formamide 30%] directly to dissected guts and ovaries; slides were incubated overnight at 28°C. After washing with PBS-Tween (0.02%, 40°C for 15 min), the samples were stained with DAPI (24 µg l<sup>-1</sup>, 15 min) and mounted with citifluor antifade solution (Ted Pella, Inc.). Control experiments were performed without a probe or with RNase for 45 min. The samples were analyzed under an Olympus FV100 Multi-photon confocal microscope. Images were processed using FIJI 2.0.0 software (Schindelin et al. 2012).

### Genome Sequencing and Metagenomic Binning

Sequencing of the metagenomes from *D. coccus* was undertaken with different platforms, namely PacBio, pyrosequencing 454 GS-FLX and Illumina HiSeq2000. Illumina and 454 library preparations and sequencing were performed at Macrogen Inc. (Korea) and PacBio at Duke University Genome Sequencing Core Facility (USA). Cells from

hemolymph were used for DNA extraction for the PacBio and 454 metagenomic sequencing from 8 and 20 females, respectively. For the third metagenome, obtained with Illumina, DNA was recovered from macerated organs including ovaries, guts and Malpighian tubules from 40 females. Details of the dissection and extraction procedure were as described in Ramírez-Puebla et al. (2016) and Vera-Ponce de León et al. (2016). The SMRT cell from PacBio resulted in a total of 86,880 reads (of 9,627 bp each on average), 454 GS-FLX sequence yielded 811,305 single reads (of 330 bp each on average) and Illumina 58,146,564 reads (pair-end reads of 100 bp). The *Dactylopiibacterium* whose genome was obtained from the female *D. coccus* metagenomes was designated NFE1. In addition, two more metagenomes of 50 *D. coccus* males and of ten females of the wild species *D. opuntiae* were obtained from whole adult macerates; the dactylopiibacteria whose genomes were obtained from these metagenomes were named NFDCM and NFDO, respectively. Metagenomes were sequenced in MacroGen Inc. (Korea) using Illumina HiSeq 2500 platform with the TruSeq DNA-PCR free Library kit. *D. coccus* male and *D. opuntiae* female metagenomes yielded 251,932,764 and 276,841,126 (pair-end reads of 100 bp), respectively. A metagenomic hybrid assembly from *D. coccus* female metagenomes was performed with Spades v. 3.7 (Bankevich et al. 2012) using Illumina HiSeq 100 bp pair-end reads, FLX-454 single reads and PacBio long reads, previously reported and deposited in GenBank under BioProject PRJNA349489 (Vera-Ponce de León et al. 2016). The resulting assembly, consisting of 332,915,694 contigs, was binned using MaxBin 2.0 (Wu et al. 2016). A set of four bins was recovered. For taxonomic positioning, Metaxa 2.0 (Bengtsson-Palme et al. 2015) was used for locating the small subunit of ribosomal genes in each bin. To refine the bin containing the *Dactylopiibacterium* genome, all contigs were blasted against nonredundant (nr) GenBank database using BLAST-X and those contigs annotated as betaproteobacteria with a higher identity value than 90% and e-value  $<1e^{-5}$  in BLAST results were retained in the final assembly. CheckM v1.0.7 (Parks et al. 2015) was used for further assessing bin quality and to discard possible chimeric contamination.

Scaffolding was performed using the resequencing and hybrid-assembly scaffolding protocol on SMRT Portal with PacBio Reads. A subsequent alignment was performed using pair-end Illumina data with Bowtie2 (Langmead and Salzberg 2012) with--very sensitive parameter. The previous scaffold and mapped reads were assembled with Spades using--trusted contigs option. SSPACE (Boetzer et al. 2011) was then used for final scaffold polish with bwa as aligner and Ciclator 1.4.1 (Hunt et al. 2015) was used to find possible circular scaffolds. All reads from *D. coccus* male and *D. opuntiae* female metagenomes were mapped against the *Dactylopiibacterium* female genome that we previously assembled by Bowtie2 (Langmead and Salzberg 2012) with--

very-sensitive local option. Properly mapped pair reads were recovered using Samtools utilities (Li et al. 2009). Reads from each metagenome were assembled into contigs with IDBA-UD assembler (Peng et al. 2012) with a *k*-mer increment of each iteration of 15 (step arg 15). Scaffolds of the contigs were recovered by SSPACE as above.

Average genome coverage was estimated with coverage from bedtools (Quinlan and Hall 2010) using all reads of each metagenome mapped to the final *Dactylopiibacterium* assembly with Bowtie2 using the--very-sensitive local option.

### *Dactylopiibacterium* Genome Analysis

The RAST server (Meyer et al. 2008) and PROKKA (Seemann 2014) were used for gene prediction and annotation. Manual curation of relevant genes was performed after comparison with sequences deposited in the following databases: nr and Refseq via BLASTP (Camacho et al. 2009), the Conserved Domain Database at GenBank (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>), the Protein families (PFAM) database (Finn et al. 2016) and the Transport Classification Database (Saier et al. 2016). In addition, a pathway profiling of metabolic annotation was obtained from all putative coding genes predicted using GhostKoala tool from KEGG (Kanehisa et al. 2016). Manual annotation was further refined using SmartBLAST of the deduced proteins and their hydrophathy profile analyzed with the online service of Swissprot TMpred ([http://www.ch.embnet.org/cgi-bin/TMPRED\\_form\\_parser](http://www.ch.embnet.org/cgi-bin/TMPRED_form_parser)) as previously described (Degli Esposti et al. 2014). Genome completeness was assessed by the presence of 40 single-copy bacterial widespread orthologue proteins with BUSCO (Simão et al. 2015) and CheckM (Parks et al. 2015). In addition, completeness was assessed by determining that all genes from the Rhodocyclaceae core genome were present in the *Dactylopiibacterium* genome (see below). Prophage sequences were identified, annotated and visualized by PAAST and PHASTER tools (Zhou et al. 2011; Arndt et al. 2016). Genome sequences were deposited at DDBJ/ENA/GenBank under accession numbers MQNN00000000, NMRM00000000, and NMRN00000000.

### Phylogenetics, Phylogenomics, and Comparative Analyses

16S rRNA sequences were identified from genome contigs using the Metaxa2 pipeline (Bengtsson-Palme et al. 2015). Sequences were aligned using ClustalX 2.0 (Larkin et al. 2007) with other 16S rRNA sequences from GenBank. A phylogenetic tree was generated from a maximum likelihood analysis using PhyML 3.1 (Guindon et al. 2010). JModelTest 2.1.10 (Posada 2008) was used to select appropriate models of sequence evolution by the AIC. Model K80+I ( $P_{inv}=0.772$ ) was the chosen model for the 16S rRNA genes.

For comparative genomics, a set of 31 Rhodocyclaceae genomes were retrieved from the NCBI genome database (<https://www.ncbi.nlm.nih.gov/genome/>). Open reading

frames and coding sequences (CDS) of all genomes were predicted using GeneMark (Besemer and Borodovsky 2005). The pangenome and core genome of all strains were obtained by GETHOMOLOGUES version 2.0 software (Contreras-Moreira and Vinuesa 2013), with -A -c -t 0 -M -n 35 parameters. Single copy orthologous genes from the core genome-matrix were aligned with Clustal Omega version 1.2.1 (Sievers et al. 2011) and Protest3 version 3.4.2 (Darriba et al. 2011) was used to select the best amino acid substitution model. A maximum likelihood phylogeny tree was constructed using PhyML 3.1 (Guindon et al. 2010) with the WAG model and the Shimodaira–Hasegawa like procedure for measuring internal branch support (Shimodaira and Hasegawa 1999).

Average amino acid identity (AAI) between *Dactylopiibacterium* genome and other genomes was calculated using all core genes with AAI calculator *enveomics* collection tools (Rodriguez-R and Konstantinidis 2016). In addition, eight selected genomes were aligned with Nucmer and plotted with Mummer plot (Delcher et al. 2002) to determine syntenic blocks of genes between *Dactylopiibacterium* and other Rhodocyclaceae.

To generate a *nifHDK* phylogeny, *Dactylopiibacterium* CDS predicted and annotated as nitrogenase-iron protein, nitrogenase molybdenum–iron protein alpha chain and nitrogenase molybdenum–iron protein subunit beta were aligned to sequences of the corresponding genes from other bacteria deposited in GenBank using Clustal Omega (Sievers et al. 2011). The concatenated alignments of the three *nif* genes were generated in BioEdit version 7.2.5 (Hall 1999) and a maximum likelihood phylogenetic tree was obtained as described above using the LG + I+G + F model ( $\alpha = 0.836$  for gamma distribution;  $P\text{-inv} = 0.225$ ).

In addition, to recover the nitrogenase genes from two available *D. coccus* metagenomes obtained elsewhere from the whole body of cochineals from Oaxaca and Peru (hereafter called DCoax and DCperu metagenomes, respectively, deposited in GenBank under BioProject PRJNA244295, Campana et al. 2015), all reads greater than 100 bp were retrieved from GenBank fastq records and amino acid translations of these sequences were obtained using FragGeneScan v 1.20 (Rho et al. 2010). Nitrogenases genes (*nifH* and *nifD*) from the Campana et al. (2015) filtrated metagenome were searched by BLASTP against *Dactylopiibacterium nif* genes.

#### *Dactylopiibacterium nifH* Gene Expression in *D. coccus*

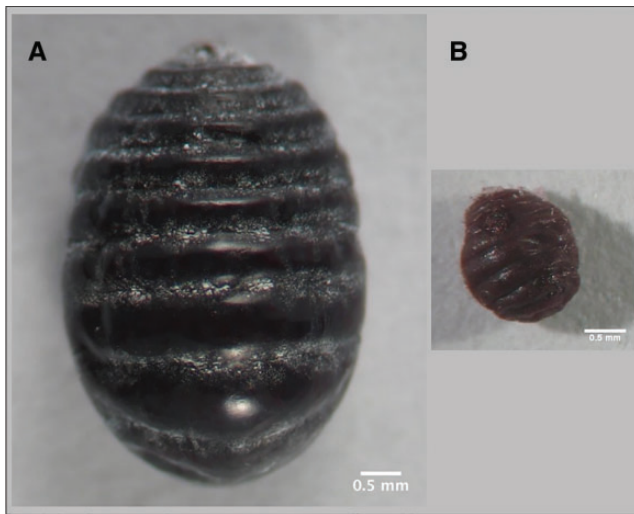
To detect expression of *nif* genes we used reverse transcription of RNA extracted from ten adult cochineal females (~90 days old), previously externally disinfected with 70% ethanol. Hemolymph, guts and ovaries from *D. coccus* were dissected as described (Vera-Ponce de León et al. 2016) and placed separately in 200  $\mu$ l of RNAlater® Solution (Life Technologies,

Carlsbad, CA, USA). Samples were centrifuged 1 min at  $9,300 \times g$ . RNAlater® was removed and RNA was isolated with TRIzol reagent (Invitrogen) following the manufacturer's instructions, digested with DNase and cleaned with the RNeasy column kit (Qiagen, Germantown, MD, USA). RNA purity and integrity was checked by gel electrophoresis and absorbance ratios (260/280 and 260/230). cDNA was synthesized for RT-PCR from purified RNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, USA) following the manufacturer's instructions. cDNA was used for PCR amplification with 0.8 pmol of primers *NifHBeta-9* (5'-AACGTCAATGCGCAATTTACG-3') and *NifHBeta-805* (5'-ATGATGCCGAAGTCCATCAG-3') (this study). The final concentration for 50  $\mu$ l PCR was 0.2 mM dNTPs, 2.5 mM MgCl<sub>2</sub>, 0.5 U Taq polymerase and 1  $\times$  Taq polymerase buffer (Invitrogen Life Technologies, Sao Paulo, Brazil). The cycling conditions were 94°C for 5 min; 35 cycles of 60 s at 94°C, 60 s at 58°C, and 90 s at 72°C and 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). Concentrated RNAs were used as negative controls and genomic DNA of *D. coccus* was used as positive control.

#### Assay for Nitrogen Fixation

To biochemically evaluate nitrogen fixation, we used the conventional acetylene reduction assays with different tissues of *D. coccus*. Five adult females (~60 days old) were detached from the host plant and rinsed in 100% ethanol for wax removal. Insects were then washed twice with sterile deionized water and placed in sterile PBS. 1  $\mu$ l of hemolymph from each individual was obtained by making a fine puncture in the third segment of the insect, the oozing fluid was collected with a micropipette and placed in 10 ml vessels with 5 ml of semi-solid nitrogen-free medium (KH<sub>2</sub>PO<sub>4</sub> 1.1 g l<sup>-1</sup>; K<sub>2</sub>HPO<sub>4</sub> 2.2 g l<sup>-1</sup>; MgSO<sub>4</sub> 0.2 g l<sup>-1</sup>; NaCl 0.1 g l<sup>-1</sup>; CaCl<sub>2</sub> 0.05 g l<sup>-1</sup>; FeSO<sub>4</sub> 0.05 g l<sup>-1</sup>; Na<sub>2</sub>MoO<sub>4</sub> 0.002 g l<sup>-1</sup>; agar 3 g l<sup>-1</sup>; fructose 0.25 g l<sup>-1</sup>; sucrose 0.15 g l<sup>-1</sup>, and trehalose 0.1 g l<sup>-1</sup> and 1.0 ml of a vitamin solution (10 mg of biotin and 20 mg of pyridoxin hydrochloride in 100 ml of distilled water). Ovaries from the same individual were dissected as described earlier (Vera-Ponce de León et al. 2016) and were submerged in the media. Samples were incubated at 28°C for 48 h and acetylene-enriched atmosphere was injected to achieve a final concentration of 10% acetylene. Media with autoclaved organs (121°C for 15 min at 15 psi) were used as negative controls, while positive controls were cultures of *Klebsiella variicola* 6A3 (Rosenblueth et al. 2004). Ethylene formed by nitrogenase was measured as described by Morales-Jiménez et al. (2013) in five replicates for each treatment. The normal distribution of the results was tested using the Shapiro–Wilk test. Differences in ethylene production with respect to sterile controls were analyzed by planned pair comparisons using the





**FIG. 1.**—Dorsal view of adult *Dactylopius* with wax removed with 96% ethanol. (a) *D. coccus*, (b) *D. opuntiae*.

Mann–Whitney test. Statistical analysis was performed with the GraphPad prism v. 7.2 software.

## Results

### *Dactylopiibacterium* in Carmine Cochineal Ovaries

*Dactylopius coccus* and *D. opuntiae* cochineals are shown in figure 1. To verify the presence of *Dactylopiibacterium* in the insects we performed FISH experiments and fluorescent signals were detected in ovaries from the two tested species (fig. 2). Two probes were used, one specific for *Dactylopiibacterium* and the other specific to betaproteobacteria, which gave similar results. RNase treated samples were devoid of signals. Probe signal was frequently observed in ovarioles, in some cases in their distal part (fig. 2*b* and *d*) and magnification showed signals inside nurse cells of the germarium either in 30- or 80-day-old insects (fig. 2*c*, *e*, and *f*).

### Genome Analysis of *Dactylopiibacterium*

The genome of the new endosymbiont *Dactylopiibacterium* was obtained from metagenomic data. One bin obtained from *D. coccus* metagenomes, composed of 193 contigs and 3,694,534 assembled bases, showed the presence of a 16S rRNA gene sequence that was 99–100% identical to our previously reported sequences of the *Dactylopius* betaproteobacterium (supplementary fig. S1, Supplementary Material online) (Ramírez-Puebla et al. 2010). After BLAST annotation of all sequences obtained, we found that 122 contigs of this bin, summing ca. 3.6 Mb (table 1), have betaproteobacterial sequences as their closest matches. Quality check of the 122 contigs by CheckM showed no contaminant sequences. The combined contigs represented a genome with a size of 3,649,266 bases, which was covered 18×, 1×, and

2× times with Illumina, 454-FLX and PacBio metagenomic reads, respectively. Additional *Dactylopiibacterium* genomes (with 21× and 23× coverage) were recovered from *D. coccus* male and *D. opuntiae* female metagenomes in 136 and 150 contigs, respectively. Genome statistics are shown in table 1.

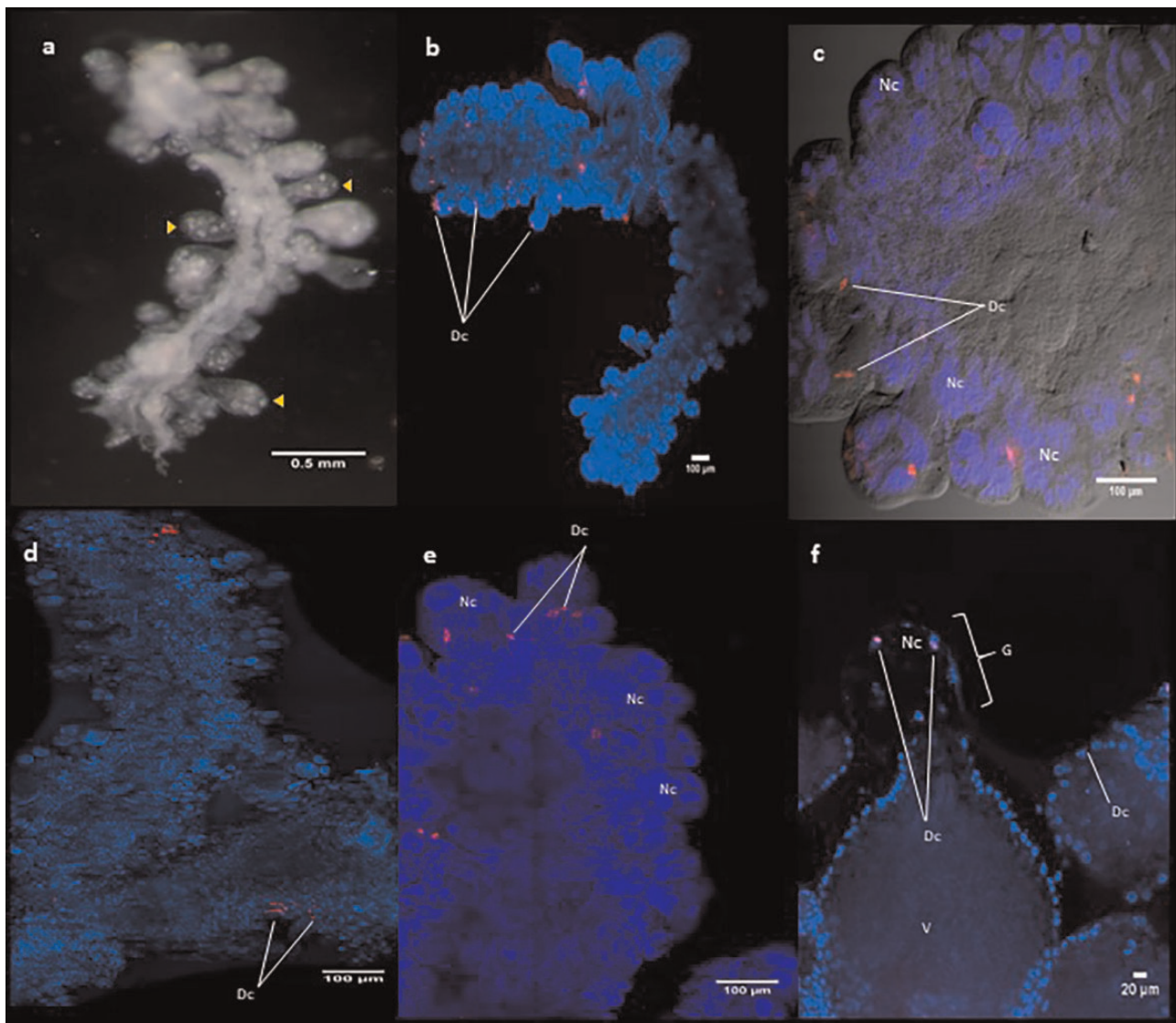
The program BUSCO (Simão et al. 2015) calculated a genome completeness of 97–100%, indicating that the recovered contigs could represent most of the *Dactylopiibacterium* genome. Further analyses indicated that *Dactylopiibacterium* genomes contained the core genome of the Rhodocyclaceae family (29 strains considered) comprising 293 single copy genes (supplementary table S1, Supplementary Material online), further supporting genome completeness. In total, the *Dactylopiibacterium* genome encoded around 3500 CDS genes, 79% of which could be assigned to a known putative function (tables 1 and 2, supplementary table S2, Supplementary Material online). After scaffolding, a set of 107, 120, and 124 scaffolds were retrieved from *Dactylopiibacterium* genomes of *D. coccus* females, *D. coccus* males, and *D. opuntiae* females, respectively.

### Phylogenomics and Comparative Genomics

A phylogenetic analysis of 602 concatenated orthologous genes showed that *Dactylopiibacterium* is a new lineage most closely related to *Uliginosibacterium gangwonense* isolated from a wetland in Korea (Weon et al. 2008). Other *Uliginosibacterium* isolates with no available genomes are plant endophytes (Li et al. 2011, 2013). Both *Uliginosibacterium* and *Dactylopiibacterium* belong to the order Rhodocyclales and have *Azoarcus* (a rice endophyte) and *Thauera* as close genera (fig. 3). The AAI analysis showed that core genome sequences of *Dactylopiibacterium* and *Uliginosibacterium* have 75% identity, but *Dactylopiibacterium* has a smaller genome than that of *U. gangwonense* (Weon et al. 2008), which is ~5 Mb in size (table 2). However, the latter has an unusual large genome compared to those from other Rhodocyclales. Most of *U. gangwonense* unique genetic material comprises transposable elements, prophages and a large plasmid (Weon et al. 2008) which was not found in the *Dactylopiibacterium* genome.

Comparative genomics applied to *Dactylopiibacterium* and close Rhodocyclaceae bacteria indicated that the core genome contains genes for nitrogen fixation and regulation, among others (supplementary table S1, Supplementary Material online). Specifically, 973 *Dactylopiibacterium* genes from NFE1 were not found in the closest relative *U. gangwonense*. Some of these genes encoded for flagella synthesis, pectin and polygalacturonic catabolism and urea metabolism (supplementary table S3, Supplementary Material online).

*Dactylopiibacterium* was not phylogenetically related to endosymbionts from other families of betaproteobacteria



**FIG. 2.**—(a) Ovary dissected from 30-day-old *D. coccus* nymph showing ovarioles in yellow arrowheads. (b–f) FISH using the designed probe Cy5-END1-1081, red and blue signals indicate *Dactylopiibacterium* and insect nuclei, respectively. (b, c) 30-day-old *D. coccus* ovaries, (d, e) 30-day-old *D. opuntiae* ovaries, (f) 80-day-old *D. coccus* ovariole and germarium. Nc, nurse cell; V, vitellarium, G, germarium, Dc, *Dactylopiibacterium*.

such as those found in mealybugs (López-Madrigal et al. 2011), spittlebugs (McCutcheon and Moran 2011), leafhoppers (Bennett et al. 2016), and in the crypts of stinkbugs (Kikuchi et al. 2012).

#### Metabolic Capabilities and Other *Dactylopiibacterium* Characteristics Inferred from Genomic Data

The genome of *Dactylopiibacterium* encoded complete metabolic pathways for the biosynthesis of biotin, thiamine, ubiquinone, cobalamine, heme, coenzyme B12, riboflavin, pyridoxine, NAD(P), folate, lipoic acid and coenzyme A (supplementary table S2, Supplementary Material online).

Moreover, it encoded complete pathways for the biosynthesis of all amino acids except asparagine. However, the genome encoded for an Aspartyl-tRNA(Asn) synthetase, thereby suggesting the ability to produce charged asparaginyl-tRNA from aspartyl-tRNA (supplementary table S2, Supplementary Material online). While the pathways for the biosynthesis of purines and pyrimidines were present, key enzymes for glycolysis and gluconeogenesis appeared to be absent. The pentose phosphate and Entner–Doudoroff pathways were instead fully represented, thus providing alternative routes for carbon metabolism. All genes of the tricarboxylic acid cycle were also present (supplementary table S2, Supplementary Material online).

**Table 1**Genome Features of *Dactylopiibacterium* from *D. coccus* and *D. opuntiae*

| Feature                    | <i>D. coccus</i> Females (NFE1) | <i>D. coccus</i> Males (NFDCM) | <i>D. opuntiae</i> Females (NFDO) |
|----------------------------|---------------------------------|--------------------------------|-----------------------------------|
| Genome accession           | MQNN00000000                    | NMRN00000000                   | NMRM00000000                      |
| Number of contigs          | 122                             | 136                            | 150                               |
| Number of scaffolds        | 107                             | 120                            | 124                               |
| N50 (bp)                   | 50,206                          | 44,208                         | 42,791                            |
| Estimated genome size (bp) | 3, 589, 384                     | 3,578,638                      | 3,550,510                         |
| G+C content (%)            | 62.74                           | 62.74                          | 62.73                             |
| CDS                        | 3,488                           | 3,421                          | 3,396                             |
| rRNA genes                 | 3                               | 3                              | 3                                 |
| tRNAs                      | 46                              | 46                             | 47                                |
| Pseudogenes                | 248                             | 202                            | 216                               |
| Genome completeness        | 100%                            | 97%                            | 97%                               |

**Table 2**Comparison of *Dactylopiibacterium* Genome Features with Related Betaproteobacteria

| Bacterial Genome | <i>Dactylopiibacterium</i> NFE1                     | <i>Dactylopiibacterium</i> NFDO                          | <i>U. gangwonense</i> | <i>Azoarcus</i> sp. BH72                 | <i>Candidatus</i> Tremblaya princeps  | <i>Candidatus</i> Zinderia insecticola      | <i>P. rhizoxinica</i>                      |
|------------------|---|--|-----------------------|--|---------------------------------------|---|--|
| Accession number | MQNN00000000  | NMRM00000000   | GCA_000373965         | GCF_000061505                            | GCF_000219195                         | GCA_000147015                               | GCA_000198775                              |
| Size (Mb)        | 3.6   | 3.5  | 5                     | 3.9                                      | 0.139                                 | 0.209                                       | 3.8  |
| G+C (%)          | 62.74   | 62.73  | 56.57                 | 67.92                                    | 58.83                                 | 13.54                                       | 60.7                                       |
| CDSs             | 3,488   | 3,396  | 4,717                 | 3,992                                    | 140                                   | 202   | 3,878                                      |
| Host             | Domesticated carmine cochineal ( <i>D. coccus</i> ) | Nondomesticated carmine cochineal ( <i>D. opuntiae</i> ) | None                  | Kallar grass ( <i>Leptochloa fusca</i> ) | Mealybug ( <i>Planococcus citri</i> ) | Spittlebug ( <i>Clastoptera arizonana</i> ) | Zygomycete ( <i>Rhizopus microsporus</i> ) |
| Life style       | Insect endosymbiont                                 | Insect endosymbiont                                      | Free-living           | Grass endophyte                          | Insect endosymbiont                   | Insect endosymbiont                         | Fungi endosymbiont                         |
| Diazotroph       | Yes   | Yes  | Yes                   | Yes                                      | No                                    | No  | No   |

Intriguingly, all the three *Dactylopiibacterium* genomes that we report here, had genes for polygalacturonase enzymes, which are absent in the genomes of other members of the Rhodocyclaceae family (supplementary tables S1–S3, Supplementary Material online).

### Anaerobic Metabolism in *Dactylopiibacterium*

The new endosymbiont *Dactylopiibacterium* appears to have distinctive features of anaerobic metabolism. For instance, its genome contains the genes for pyruvate formate-lyase and lactate dehydrogenase, suggesting that the endosymbiont could perform fermentation reactions (supplementary table S1, Supplementary Material online). Other genes associated with anaerobic metabolism include those for nitrite respiration to nitrous oxide (supplementary table S1, Supplementary Material online), as well as two terminal oxidases which are active under very low concentrations of oxygen: the bd-type ubiquinol oxidase and the cbb3-type cytochrome c oxidase (supplementary table S2, Supplementary Material online) (Degli Esposti et al. 2014). Both of these oxidases are over-expressed in proteobacteria that fix N<sub>2</sub> in symbiosis with

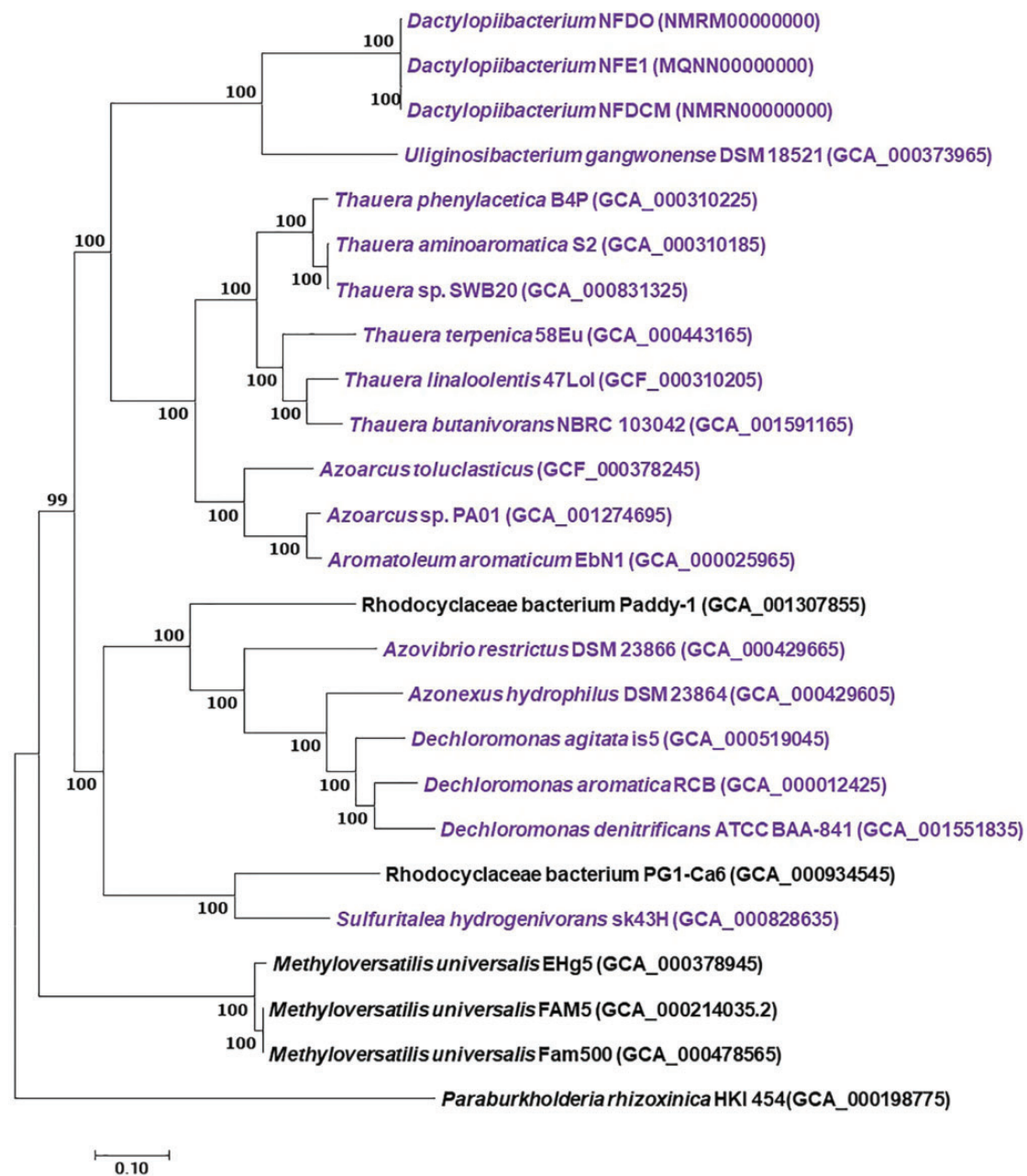
plants (Degli Esposti and Martínez Romero 2016). Conversely, only a few, noncatalytic genes for the subunits of COX operons encoding cytochrome c oxidase were found (supplementary table S2, Supplementary Material online). Nitrite reductase and lactate dehydrogenase are present among the 973 *Dactylopiibacterium* unique genes but absent in the core genome of other Rhodocyclaceae (supplementary tables S2 and S3, Supplementary Material online).

The genome also includes the entire operon of fumarate reductase (supplementary table S2, Supplementary Material online), an enzyme that recycles the reduced membrane quinones that are produced by complex I and other dehydrogenases and is typically present in facultative anaerobic bacteria such as *Escherichia coli*. This anaerobic trait is present in other members of the Rhodocyclaceae family such as *Azoarcus* (Degli Esposti and Martínez Romero 2017).

### Cell Surface and Secretion Systems

Inspection of genes related to membrane and cell envelope biosynthesis indicated that *Dactylopiibacterium* could produce membranes with phosphatidylglycerol, phosphatidylserine,





**FIG. 3.**—Maximum likelihood tree ( $-\ln L = -1403085.372840$ ) of 602 orthologous of *Dactylopiibacterium* and other Rhodocyclaceae genomes. The genome of *Paraburkholderia rhizoxinica* was used as an outgroup. Scale bar indicates 10% estimated sequence divergence. SH-like support values  $\geq 50\%$  are indicated. In purple are genomes with nitrogen-fixing genes.

phosphatidylethanolamine, and cardiolipin but not phosphatidylcholine (supplementary table S2, Supplementary Material online). Its genome encodes genes for lipopolysaccharide biosynthesis, including the O-antigen, a normal peptidoglycan layer and some form of exo- or capsular polysaccharide. *Dactylopiibacterium* was deduced to be potentially motile because its genome has two sets of genes for the flagellar machinery (supplementary tables S2 and S3, Supplementary Material online) and one gene set for chemotaxis. A set of

eight genes that encode FlhA, FlhB, FliH, FliI, FliO, FliP, FliQ, and FliR components of the flagellum-specific protein export apparatus and for type IV pilus biogenesis were also found (supplementary tables S2 and S3, Supplementary Material online).

Secretion systems are present in this genome, including *tra/trb* genes for the mating-pair formation apparatus of a type IV secretion system (T4SS) (supplementary table S2, Supplementary Material online), which are closely related to



those found in the Integrative Conjugative Elements inserted in the chromosomes of beta and gammaproteobacteria (Burrus and Waldor 2004). *Dactylopiibacterium* genome also contains genes for type I and II secretion systems (supplementary table S2, Supplementary Material online).

### Prophages

Analysis by PHASTER identified at least five prophage regions (scaffolds) in the *Dactylopiibacterium* genome from *D. coccus*. Two of these regions encoded intact phage sequences. 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. Moreover, at least three scaffolds in *Dactylopiibacterium* genome from *D. opuntiae* were similar to phages sequences.

### Nitrogen Metabolism in Carmine Cochineals

We found that the *Dactylopiibacterium* genomes contain nitrogenase genes that were phylogenetically related to those of betaproteobacteria (fig. 4), in particular, they were most closely related to *Uliginosibacterium nif* genes either in concatenated *nifHDK* (fig. 4) or in individual *nif* gene phylogenies (supplementary fig. S2, Supplementary Material online). *nif* gene phylogenies showing *Uliginosibacterium* as the closest relative to *Dactylopiibacterium* were congruent with whole core gene phylogenies (figs. 3 and 4). The genes encoding for Mo-Fe nitrogenase (*nifD–nifK*), nitrogenase reductase (*nifH*), nitrogen regulatory proteins and transporters specific for nitrogenase cofactors (*nifQ–nifA* and 4Fe–4S ferredoxin) were found assembled into four contigs of *Dactylopiibacterium*. In addition, we recovered a *nifH* gene sequence from the DCoax metagenome and a *nifD* gene sequence from the DCperu metagenome (Campana et al. 2015) and their phylogenetic analysis indicated that they belonged to *Dactylopiibacterium* (supplementary fig. S3, Supplementary Material online). No other *nif* gene sequences were found in either our samples or in DCoax or DCperu metagenomes.

RT-PCR showed expression of *Dactylopiibacterium nifH* gene in *D. coccus* hemolymph, ovaries, and eggs, but not in guts (fig. 5a). Accordingly, acetylene reduction activity, the standard assay for nitrogen fixation, was detected in hemolymph and ovaries of *D. coccus* ( $38 \pm 5.98$  and  $11.58 \pm 1.38$  nmol ethylene  $\text{h}^{-1}$  tissue  $\text{g}^{-1}$ , respectively). These values were statistically significant when compared with those obtained in autoclaved controls of the same tissues (Mann–Whitney test, fig. 5b).

The endosymbiont genome showed UA recycling genes. UA is the main waste compound of nitrogen catabolism in *D. coccus* (Vera-Ponce de León et al. 2016). The genome of *Dactylopiibacterium* had genes for nitrogen recycling of purine waste products and UA degradation to urea. All genes for

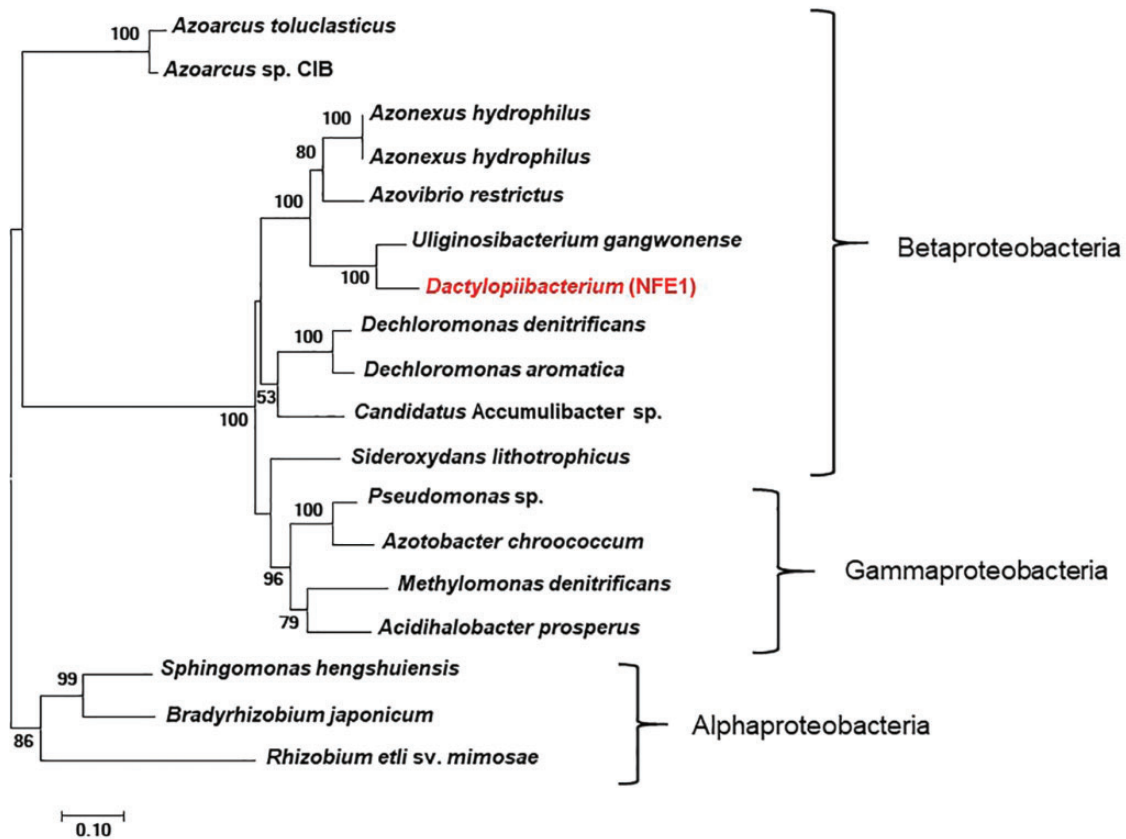
urea metabolism and transport of its metabolites were predicted and annotated (supplementary table S2, Supplementary Material online).

### Discussion

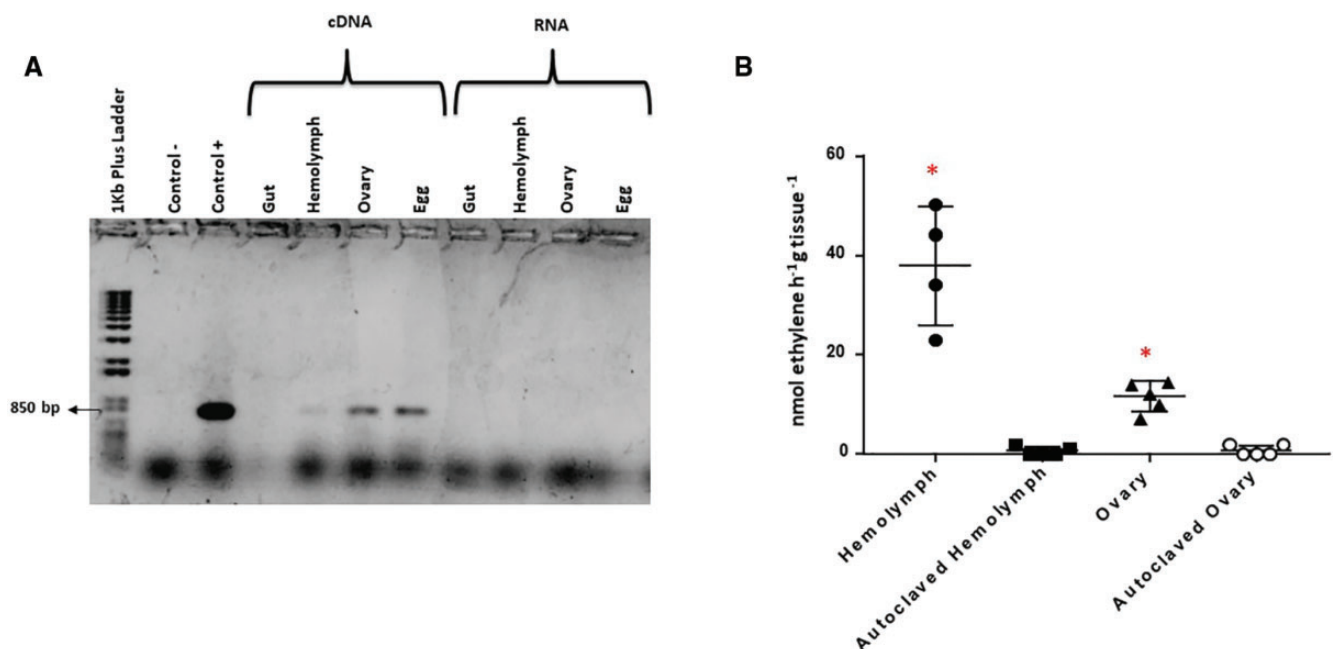
The carmine cochineal was the red gold of the Americas and there is still a large economic interest in carmine even today. Approaches are being considered to produce carmine using novel biotechnological procedures. For a long time, it has been hypothesized that microbes could be involved in carmine production (reviewed in Margulis 2010). To study carmine cochineal microbiota, we undertook a 16S rRNA gene diversity analysis using different Mexican *Dactylopius* species including the domesticated *D. coccus* and the wild species *D. ceylonicus*, *D. confusus*, *D. opuntiae*, and *D. tomentosus* (Ramírez-Puebla et al. 2010). In all the cochineals studied, we uncovered a novel betaproteobacterial 16S rRNA gene sequence and we designated the putative novel symbiont as *Candidatus Dactylopiibacterium carminicum*. Here, we expanded the knowledge on dactylopiibacteria showing their location inside the host, genome sequences, putative metabolic traits, and nitrogen-fixing capabilities.

Until now we have been unable to grow *Dactylopiibacterium* in culture media in the laboratory. Notably, the betaproteobacterium *Azoarcus* which is phylogenetically not distantly related to *Dactylopiibacterium*, attains a nonculturable state in plants (Hurek et al. 2002) and a similar situation may occur in the carmine cochineals with *Dactylopiibacterium*. Consequently, we recovered its genome from different metagenomes obtained from tissues of female *D. coccus*, from whole male cochineals and from female *D. opuntiae*. The *Dactylopiibacterium* genomes (3.6 and 3.5 Mb from *D. coccus* and *D. opuntiae*, respectively) turned out to be larger than those of other endosymbionts (ranging from 0.1 to 2 Mb, table 2). However, the genome was comparable in size to those of gammaproteobacterial symbionts such as *Sodalis* or *Arsenophonus* which are maternally transmitted and considered as secondary (Gil et al. 2008; Oakeson et al. 2014) or primary endosymbionts in some insects (Oakeson et al. 2014). *Sodalis* was considered as an *in statu nascendi* symbiont in the insect *Cicadella viridis* (Michalik et al. 2014) and different *Sodalis* species or strains have variable genome sizes (reviewed in Rosas-Pérez et al. 2017) that may be up to 4.5 Mb in *Sodalis pierantonius* (Oakeson et al. 2014).

In view of the large *Dactylopiibacterium* genomes and similarity, we suggest that dactylopiibacteria may have been recently acquired without much time to evolve reduced or divergent genomes. However, considering that the symbiont was found in all *Dactylopius* species tested, *Dactylopiibacterium* symbiosis was probably established before *Dactylopius* cochineal speciation, unless it is periodically



**FIG. 4.**—Maximum likelihood tree ( $-\ln L = -16717.43208$ ) of concatenated *nifHDK* genes of *Dactylopiibacterium* and other related sequences from GenBank. Scale bar indicates 10% estimated sequence divergence. SH-like support values  $\geq 50\%$  are indicated.



**FIG. 5.**—(a) Amplicons of RT-PCR from *Dactylopiibacterium nifH* transcript. *D. coccus* total DNA as positive control and no-DNA template as negative control. (b) Acetylene reduction activity detected in *D. coccus* tissues. Bar indicates media  $\pm$  SEM. Red asterisks show significant differences between tissues and autoclaved controls and nonsterile tissues (Mann-Whitney test  $P < 0.05$ ).

obtained from plants. Recurrent plant acquisition would be not required if the symbiont is maternally transmitted by cochineals to their progeny.

Maternal transmission of bacteria is commonly found in insects, especially with primary endosymbionts that have largely reduced genomes and complete host dependence (McCutcheon and Moran 2011). Vertical inheritance suggests that there is a benefit from the symbiont to the host as the latter is committed to retain the bacterium for future generations. For maternal inheritance, it seems that only few symbionts would be needed to be transferred to the progeny. Bacteria in germarium and in nurse cells have been found in other scale insects like *Pseudococcus* and *Icerya* (Shinji 1919), as well as in *D. coccus* (Ramírez-Cruz 2012). The presence of *Dactylopiibacterium* in ovaries and eggs in the carmine cochineal (though in low numbers) does suggest maternal transmission, while the expression of *nif* genes and nitrogenase activity in ovary constitute distinctive characteristics of this species. Since *Dactylopiibacterium* looks like a newly acquired symbiont and it is probably maternally inherited, we suggest that maternal inheritance is an early and key step to further evolve specific and highly adapted symbionts. In other endosymbionts, for maternal transmission, flagella and type IV pili are essential for the colonization of host embryos (Dulla et al. 2012; Wang et al. 2013). Indeed, we found these systems encoded in the genome of *Dactylopiibacterium*, suggesting they may participate in the bacterial colonization of ovaries and eggs, as well as in exporting molecules into the cochineal host mediating its symbiotic interactions, as reported for some parasitic or free-living bacteria (Macnab, 2003; Zhou et al. 2015).

A major finding of this study is that bacterial endosymbionts of cochineal insects have the capacity to fix nitrogen, thereby identifying *Dactylopiibacterium* as a novel nitrogen fixing symbiont. The only *nif* gene sequences found in all carmine cochineal metagenomes were those from *Dactylopiibacterium* and included all genes required for nitrogen fixation. Nitrogen-fixing bacteria previously reported to be associated with insects such as ants, bark beetles and fruit flies, are gut symbionts (van Borm et al. 2002; Behar et al. 2005; Pinto-Tomás et al. 2009; Morales-Jiménez et al. 2009, 2013; Desai and Brune 2012). Only in the case of the wood-feeding insect *Anoplophora glabripennis*, *nif* genes, and transcripts from gut bacteria were also found in eggs (Ayayee et al. 2014). In *D. coccus* guts no *Dactylopiibacterium nifH* transcripts were detected by RT-PCR. This result and the very few *Dactylopiibacterium* transcripts found in guts (our unpublished transcriptome results) suggest that dactylopiibacteria are not principally gut symbionts. We further suggest that these bacteria are not directly and continuously acquired from the sap diet, since adult male *Dactylopius* cochineals contain *Dactylopiibacterium* even if they do not exhibit any feeding behavior since pupae.

Genomic analysis indicated that the symbiont is well adapted to a low oxygen life style that is compatible with nitrogen fixation, which is oxygen sensitive. Moreover, the presence of glutamine synthetase genes in the genome of the endosymbiont suggests that nitrogen fixation follows the transformation from  $\text{NH}_3$  to glutamine. Ammonium assimilation constitutes the first step towards the biosynthesis of amino acids. In scale insects, essential amino acids are commonly produced by flavobacteria and gammaproteobacteria (Rosenblueth et al. 2012; Rosas-Pérez et al. 2014; Skidmore and Hansen 2017). Since flavobacteria and enterobacteria are not found in *Dactylopius*, *Dactylopiibacterium*, which has all genes for essential amino acid biosynthesis, might in turn provide essential amino acids to the host. Genes for nonessential amino acids are not found in the aphid endosymbiont *Buchnera*, as nonessential amino acids are provided by the insect host. *Dactylopiibacterium* only lacks one of the biosynthetic genes for the nonessential amino acid, asparagine, which may well be provided by the insect host, or obtained from the plant sap. Alternatively, asparagine may be obtained within the endosymbiont by the conversion of asparagine-charged tRNAs as described in other bacteria (Bailly et al. 2007).

Another notable feature of the *Dactylopiibacterium* genome is that it encodes genes typical of bacterial association with plants, such as those for the catabolism of salicylate (key molecule in plant defense, Klessig et al. 2000), pectinases, polygalacturonases (to degrade plant cell walls), and auxin (plant hormone) transport (supplementary table S3, Supplementary Material online). Genes involved in salicylate and pectin degradation, as well as auxin transport, may be reminiscent of a previous *Dactylopiibacterium* life in plants. Alternatively, their presence may suggest a potential dual life style in symbiosis with insects and plants, as unexpectedly found in other insect endosymbionts (Frago et al. 2012). Bacteria closely related to *Dactylopiibacterium* as *Uliginosibacterium* and *Azoarcus* are grass plant endophytes (Li et al. 2011, 2013; Krause et al. 2006). It is worth noting that nitrogen-fixing bacteria are commonly minor components of bacterial communities (Church et al. 2005; Morris et al. 2012), mainly due to the energy cost of fixing nitrogen. A similar argument may be used to explain the limited numbers of *Dactylopiibacterium* cells that were found inside carmine cochineals.

Finally, we speculate that *Dactylopiibacterium* provides benefits to the host that may extend even to an increase in cochineal longevity for it may produce nitric oxide (NO) via partial denitrification, a process encoded in its genome. NO can act as a signal molecule in plants and animals (Wendehenne et al. 2001) and has been found to extend the life span of worms (Heintz and Mair 2014). In sum, this study has revealed a novel nitrogen fixing endosymbiont of arthropods that may be evolved from an earlier symbiosis with



plants and may fulfill various roles in its association with cochineal insects.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

The authors thank CONACyT Basic Science (grant 253116) and PAPIIT-UNAM (IN207615). They also thank Pilar González-Román for help with dissections; María J. Palma-Martínez for help with RT-PCR; Arturo Ramírez Cruz for his discussions on microscopy from cochineal ovaries; José Luis Marquina for providing *D. coccus*; and Alejandro Sánchez-Flores, Víctor del Moral, Alfredo Hernández and Romualdo Zayas for bioinformatics and computing support. All bioinformatic analyses were performed on CCG-UNAM servers.

## Literature Cited

- Adams AS, Currie CR, Cardoza Y, Klepzig KD, Raffa KF. 2009. Effects of symbiotic bacteria and tree chemistry on the growth and reproduction of bark beetle fungal symbionts. *Can J Forest Res.* 39(6):1133–1147.
- Arndt D, et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44(W1):1–6.
- Ayayee P, et al. 2014. Gut microbes contribute to nitrogen provisioning in a wood-feeding cerambycid. *Environ Entomol.* 43(4):903–912.
- Bailey M, Blaise M, Lorber B, Becker HD, Kern D. 2007. The transamidosome: a dynamic ribonucleoprotein particle dedicated to prokaryotic tRNA-dependent asparagine biosynthesis. *Mol Cell.* 28(2):228–239.
- Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.
- Behar A, Yuval B, Jurkevitch E. 2005. Enterobacteria-mediated nitrogen fixation in natural populations of the fruit fly *Ceratitis capitata*. *Mol Ecol.* 14:2637–2643.
- Behmer ST, Grebenok RJ, Douglas AE. 2011. Plant sterols and host plant suitability for a phloem-feeding insect. *Funct Ecol.* 25(3):484–491.
- Bengtsson-Palme J, et al. 2015. Metaxa2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol Ecol Resour.* 15(6):1403–1414.
- Bennett GM, Abbà S, Kube M, Marzachi C. 2016. Complete genome sequences of the obligate symbionts ‘*Candidatus Sulcia muelleri*’ and ‘*Ca. Nasuia deltocephalinicola*’ from the pestiferous leafhopper *Macrosteles quadripunctulatus* (Hemiptera: Cicadellidae). *Genome Announc.* 4(1):e01604-15.
- Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res.* 33(Web Server):451–454.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4):578–579.
- Burrus V, Waldor MK. 2004. Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol.* 155(5):376–386.
- Camacho C, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Campana MG, Robles García NM, Tuross N. 2015. America’s red gold: multiple lineages of cultivated cochineal in Mexico. *Ecol Evol.* 5(3):607–617.
- Church MJ, Jenkins BD, Karl DM, Zehr JP. 2005. Vertical distributions of nitrogen-fixing phylotypes at Stn ALOHA in the oligotrophic North Pacific Ocean. *Aquatic Microbial Ecol.* 38:3–14.
- Contreras-Moreira B, Vinuesa P. 2013. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol.* 79(24):7696–7701.
- Dale C, Beeton M, Harbison C, Jones T, Pontes M. 2006. Isolation, pure culture, and characterization of “*Candidatus Arsenophonus arthropodicus*” an intracellular secondary endosymbiont from the hippoboscoid louse fly *Pseudolynchia canariensis*. *Appl Environ Microbiol.* 72(4):2997–3004.
- Dale C, Maudlin I. 1999. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. *Culture* 49(1):267–275.
- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27(8):1164–1165.
- Degli Esposti M, et al. 2014. Evolution of mitochondria reconstructed from the energy metabolism of living bacteria. *PLoS One* 9(5):e96566.
- Degli Esposti M, Martínez Romero E. 2016. A survey of the energy metabolism of nodulating symbionts reveals a new form of respiratory complex I. *FEMS Microbiol Ecol.* 92(6):fiw084.
- Degli Esposti M, Martínez Romero E. 2017. The functional microbiome of arthropods. *PLOS One* 12:e0176573.
- Delcher AL, Phillippy A, Carlton J, Salzberg SL. 2002. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* 30(11):2478–2483.
- Desai MS, Brune A. 2012. Bacteroidales ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites. *ISME J.* 6(7):1302–1313.
- Douglas AE. 2012. ‘Can’t live without you.’ essential animal-bacterial relationships. *Microbe Mag.* 7(6):273–277.
- Douglas AE. 2015. Multiorganismal insects: diversity and function of resident microorganisms. *Annu Rev Entomol.* 60:17–34.
- Douglas AE. 2016. How multi-partner endosymbioses function. *Nat Rev Microbiol.* 14(12):731–743.
- Dulla GFJ, Go RA, Stahl DA, Davidson SK. 2012. *Verminephrobacter eiseniae* type IV pili and flagella are required to colonize earthworm nephridia. *ISME J.* 6(6):1166–1175.
- Finn RD, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44(D1):D279–D285.
- Frago E, Dicke M, Godfray HC. 2012. Insect symbionts as hidden players in insect-plant interactions. *Trends Ecol Evol.* 27(12):705–711.
- Gil R, et al. 2008. Massive presence of insertion sequences in the genome of SOPE, the primary endosymbiont of the rice weevil *Sitophilus oryzae*. *Int Microbiol.* 11(1):41–48.
- Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59(3):307–321.
- Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.
- Heintz C, Mair W. 2014. You are what you host: microbiome modulation of the aging process. *Cell* 156(3):408–411.
- Hirayama C, Sugimura M, Shinbo H. 1999. Recycling of urea associated with the host plant urease in the silkworm larvae, *Bombyx mori*. *J Insect Physiol.* 45(1):15–20.
- Hunt M, et al. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* 16:294.
- Hurek T, Handley LL, Reinhold-Hurek B, Piché Y. 2002. *Azoarcus* grass endophytes contribute fixed nitrogen to the plant in an unculturable state. *Mol Plant Microbe Interact.* 15(3):233–242.

- James EK, Olivares FL, Baldani JJ, Döbereiner J. 1997. *Herbaspirillum*, an endophytic diazotroph colonizing vascular tissue *Sorghum bicolor* L. Moench. *J Exp Bot*. 48(3):785–798.
- Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for functional characterization of genome and metagenome Sequences. *J Mol Biol*. 428:726–731.
- Kikuchi Y, et al. 2012. Symbiont-mediated insecticide resistance. *Proc Natl Acad Sci U S A*. 109(22):8618–8622.
- Kim H, et al. 1998. High population of *Sphingomonas* species on plant surface. *J Appl Microbiol*. 85(4):731–736.
- Klessig DF, et al. 2000. Nitric oxide and salicylic acid signaling in plant defense. *Proc Natl Acad Sci U S A*. 97(16):8849–8855.
- Koga R, Tsuchida T, Fukatsu T. 2009. Quenching autofluorescence of insect tissues for *in situ* detection of endosymbionts. *Appl Entomol Zool*. 44(2):281–291.
- Krause A, et al. 2006. Complete genome of the mutualistic, N<sub>2</sub>-fixing grass endophyte *Azoarcus* sp. strain BH72. *Nat Biotechnol*. 24(11):1385–1391.
- Larkin MA, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947–2948.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie. *Nature Methods* 2(9):357–359.
- Li H, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Li YH, Liu QF, Liu Y, Zhu JN, Zhang Q. 2011. Endophytic bacterial diversity in roots of *Typha angustifolia* L. in the constructed Beijing Cuihu Wetland (China). *Res Microbiol*. 162(2):124–131.
- Li YH, et al. 2013. Comparison of the diversity of root-associated bacteria in *Phragmites australis* and *Typha angustifolia* L. in artificial wetlands. *World J Microbiol Biotechnol*. 29(8):1499–1508.
- López-López A, et al. 2013. Native bradyrhizobia from Los Tuxtlas in Mexico are symbionts of *Phaseolus lunatus* (Lima bean). *Syst Appl Microbiol*. 36(1):33–38.
- López-Madrugal S, Latorre A, Porcar M, Moya A, Gil R. 2011. Complete genome sequence of 'Candidatus Tremblaya princeps' strain PCVAL, an intriguing translational machine below the living-cell status. *J Bacteriol*. 193(19):5587–5588.
- Macnab RM. 2003. How bacteria assemble flagella. *Annu Rev Microbiol*. 57:77–100.
- Margulis L. 2010. Symbiogenesis. A new principle of evolution rediscovery of Boris Mikhaylovich Kozo-Polyansky (1890–1957). *Paleontol J. (Paleontologicheskii Zhurnal)* 44:1525–1539.
- Martinez-Cano DJ, Reyes-Prieto M, Martinez-Romero E, Partida-Martinez LP, Latorre A, Moya A. 2015. Evolution of small prokaryotic genomes. *Front Microbiol*. 5:1–23.
- Matthew CZ, Darby AC, Young SA, Hume LH, Welburn SC. 2005. The rapid isolation and growth dynamics of the tsetse symbiont *Sodalis glossinidius*. *FEMS Microbiol Lett*. 248(1):69–74.
- McCutcheon JP, Moran N. 2011. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*. 10(1):13–26.
- Meyer F, et al. 2008. The metagenomics RAST server a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9(1):386.
- Michalik A, Jankowska W, Kot M, Gołas A, Szklarzewicz T. 2014. Symbiosis in the green leafhopper, *Cicadella viridis* (Hemiptera, Cicadellidae). *Association in statu nascendi?* *Arthropod Struct Dev*. 43(6):579–587.
- Morales-Jiménez J, Zúñiga G, Villa-Tanaca L, Hernández-Rodríguez C. 2009. Bacterial community and nitrogen fixation in the red turpentine beetle, *Dendroctonus valens* LeConte (Coleoptera: Curculionidae: Scolytinae). *Microb Ecol*. 58(4):879–891.
- Morales-Jiménez J, et al. 2013. Nitrogen-fixing and uricolytic bacteria associated with the gut of *Dendroctonus rhizophagus* and *Dendroctonus valens* (Curculionidae: Scolytinae). *Microb Ecol*. 66(1):200–210.
- Morris JJ, Lenski RE, Zinser ER. 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *MBio* 3(2):e00036-12.
- Oakeson KF, et al. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biol Evol*. 6(1):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. *Appl Environ Microbiol*. 65(11):4926–4934.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 25:1043–1055.
- Perez-Guerra G, Kosztarab M. 1992. Biosystematics of the family Dactylopiidae (Homoptera: Coccoinea) with emphasis on the life cycle of *Dactylopius coccus* Costa. *Bull Virginia Agric Exp Stn* 92:1–90.
- Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420–1428.
- Pinto-Tomás AA, et al. 2009. Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. *Science* 326(5956):1120–1123.
- Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol*. 25(7):1253–1256.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
- Ramírez-Cruz A. 2012. Ultraestructura del germario de *Dactylopius coccus* Costa, 1829 (Hemiptera: Coccoidea: Dactylopiidae). *Acta Zoológica Mexicana* 28:586–593.
- Ramírez-Puebla ST, et al. 2016. Genomes of *Candidatus Wolbachia bourtzi* wDacA and *Candidatus Wolbachia pipientis* wDacB from the cochineal insect *Dactylopius coccus* (Hemiptera: Dactylopiidae). *G3: Genes Genomes Genet*. 6:3343–3349.
- Ramírez-Puebla ST, et al. 2010. Molecular phylogeny of the genus *Dactylopius* (Hemiptera: Dactylopiidae) and identification of the symbiotic bacteria. *Environ Entomol*. 39(4):1178–1183.
- Ramírez-Puebla ST, et al. 2015. Species in *Wolbachia*? Proposal for the designation of "Candidatus *Wolbachia bourtzi*", "Candidatus *Wolbachia onchocercicola*", "Candidatus *Wolbachia blaxteri*", "Candidatus *Wolbachia brugii*", "Candidatus *Wolbachia taylori*", "Candidatus *Wolbachia collembolicola*" and "Candidatus *Wolbachia multihospitum*" for the different species within *Wolbachia* supergroups. *Syst Appl Microbiol*. 38(6):390–399.
- Rho M, Tang H, Ye Y. 2010. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res*. 38(20):e191.
- Rodríguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints* 4:e1900v1.
- Rosas-Pérez T, et al. 2017. Chapter 5. The Symbiome of *Llaveia cochinealis* (Hemiptera: Coccoidea: Monophlebidae) includes a Gammaproteobacterial cosymbiont *Sodalis* TME1 and the known *Candidatus Walczuchella monophlebidarum*. In: Shields VDC, editor. *Insect physiology and ecology*.
- Rosas-Pérez T, Rosenblueth M, Rincón-Rosales R, Mora J, Martínez-Romero E. 2014. Genome sequence of 'Candidatus Walczuchella monophlebidarum' the flavobacterial endosymbiont of *Llaveia axin axin* (Hemiptera: Coccoidea: Monophlebidae). *Genome Biol Evol*. 6(3):714–726.
- Rosenblueth M, Martínez L, Silva J, Martínez-Romero E. 2004. *Klebsiella variicola*, a novel species with clinical and plant-associated isolates. *Syst Appl Microbiol*. 27(1):27–35.
- Rosenblueth M, et al. 2017. Scale-insect endosymbiotic microorganisms. *TIP Revista Especializada en Ciencias Químico-Biológicas* (in press).

- 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(11):2357–2368.
- Sabree ZL, Kambhampati S, Moran NA. 2009. Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proc Natl Acad Sci U S A.* 106(46):19521–19526.
- Saier MH, Jr, et al. 2016. The transporter classification database (TCDB): recent advances. *Nucleic Acids Res.* 44(D1):D372–D379.
- Schindelin J, et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9(7):676–682.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068–2069.
- Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* 407(6800):81–86.
- Shinji GO. 1919. Embryology of coccids, with especial reference to the formation of the ovary, origin and differentiation of the germ cells, germ layers, rudiments of the midgut, and the intracellular symbiotic organisms. *J Morphol.* 33(1):72–167.
- Shimodaira H, Hasegawa M. 1999. Letter to the editor multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol.* 16(8):1114–1116.
- Sievers F, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 7:539.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
- Skidmore IH, Hansen AK. 2017. The evolutionary development of plant-feeding insects and their nutritional endosymbionts. *Insect Sci.* doi: 10.1111/1744-7917.12463.
- Stintzing FC, Carle R. 2005. Cactus stems (*Opuntia* spp.): a review on their chemistry, technology, and uses. *Mol Nutr Food Res.* 49(2):175–194.
- Tai V, et al. 2016. Genome evolution and nitrogen-fixation in bacterial ectosymbionts of a protist inhabiting wood-feeding cockroaches. *Appl Environ Microbiol.* 82(15):4682–4695.
- Ulyshen MD. 2015. Insect-mediated nitrogen dynamics in decomposing wood. *Ecol Entomol.* 40:97–112.
- 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(1504):2023–2027.
- Van Leuven JT, Meister RC, Simon C, McCutcheon JP. 2014. Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. *Cell* 158(6):1270–1280.
- Vera-Ponce de León A, Sanchez-Flores A, Rosenblueth M, Martínez-Romero E. 2016. Fungal community associated with *Dactylopius* (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.
- Wendehenne D, Pugin A, Klessig DF, Durner J. 2001. Nitric oxide: comparative synthesis and signaling in animal and plant cells. *Trends Plant Sci.* 6(4):177–183.
- Weon HY, et al. 2008. *Uliginosibacterium gangwonense* gen. nov., sp. nov. isolated from a wetland, Yongneup, in Korea. *Int J Syst. Evol Microbiol.* 58(Pt 1):131–135.
- Williams DJ, Ben-Dov Y. 2015. Scale insect species names that have been combined with the genus name *Dactylopius* Costa (Hemiptera: Sternorrhyncha: Coccoomorpha). *Zootaxa* 4006(1):161–170.
- Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32(4):605–607.
- Xie J, Butler S, Sanchez G, Mateos M. 2014. Male killing *Spiroplasma* protects *Drosophila melanogaster* against two parasitoid wasps. *Heredity* 112(4):399–408.
- Zhou M, et al. 2015. More than a locomotive organelle: flagella in *Escherichia coli*. *Appl Microbiol Biotechnol.* 99(21):8883–8890.
- Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res.* 39(suppl):1–6.

Associate editor: Bill F. Martin