


RESEARCH ARTICLE

Open Access



# Functional genomics of a *Spiroplasma* associated with the carmine cochineals *Dactylopius coccus* and *Dactylopius opuntiae*

Arturo Vera-Ponce León<sup>1,2\*</sup> , Marian Dominguez-Mirazo<sup>1,3</sup>, Rafael Bustamante-Brito<sup>1</sup>, Víctor Higareda-Alvear<sup>1</sup>, Mónica Rosenblueth<sup>1</sup> and Esperanza Martínez-Romero<sup>1</sup>

## Abstract

**Background:** *Spiroplasma* is a widely distributed endosymbiont of insects, arthropods, and plants. In insects, *Spiroplasma* colonizes the gut, hemolymph, and reproductive organs of the host. Previous metagenomic surveys of the domesticated carmine cochineal *Dactylopius coccus* and the wild cochineal *D. opuntiae* reported sequences of *Spiroplasma* associated with these insects. However, there is no analysis of the genomic capabilities and the interaction of this *Spiroplasma* with *Dactylopius*.

**Results:** Here we present three *Spiroplasma* genomes independently recovered from metagenomes of adult males and females of *D. coccus*, from two different populations, as well as from adult females of *D. opuntiae*. Single-copy gene analysis showed that these genomes were > 92% complete. Phylogenomic analyses classified these genomes as new members of *Spiroplasma ixodetis*.

Comparative genome analysis indicated that they exhibit fewer genes involved in amino acid and carbon catabolism compared to other spiroplasmas. Moreover, virulence factor-encoding genes (i.e., *glpO*, *spaid* and *rip2*) were found incomplete in these *S. ixodetis* genomes. We also detected an enrichment of genes encoding the type IV secretion system (T4SS) in *S. ixodetis* genomes of *Dactylopius*. A metatranscriptomic analysis of *D. coccus* showed that some of these T4SS genes (i.e., *traG*, *virB4* and *virD4*) in addition to the superoxide dismutase *sodA* of *S. ixodetis* were overexpressed in the ovaries.

**Conclusion:** The symbiont *S. ixodetis* is a new member of the bacterial community of *D. coccus* and *D. opuntiae*. The recovery of incomplete virulence factor-encoding genes in *S. ixodetis* of *Dactylopius* suggests that this bacterium is a non-pathogenic symbiont. A high number of genes encoding the T4SS, in the *S. ixodetis* genomes and the overexpression of these genes in the ovary and hemolymph of the host suggest that *S. ixodetis* use the T4SS to interact with the *Dactylopius* cells. Moreover, the transcriptional differences of *S. ixodetis* among the gut, hemolymph and ovary tissues of *D. coccus* indicate that this bacterium can respond and adapt to the different conditions (e.g., oxidative stress) present within the host. All this evidence proposes that there is a strong interaction and molecular signaling in the symbiosis between *S. ixodetis* and the carmine cochineal *Dactylopius*.

\* Correspondence: [avera@ccg.unam.mx](mailto:avera@ccg.unam.mx)

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

<sup>2</sup>Present Address: Faculty of Biotechnology, Chemistry and Food Science, Norwegian University of Life Sciences, 1433 Ås, Norway

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Insects are associated with a plethora of microorganisms partitioned in different tissues of the host. Particularly, these bacterial symbionts can be distributed in the gut, hemolymph and even in the reproductive organs of the insect host [1]. An example of multi-tissue colonizer bacteria are some members of *Spiroplasma*.

The genus *Spiroplasma* comprises wall-less, motile, and helical bacteria of the class Mollicutes. These bacteria are mainly associated with insects but there are also reports of *Spiroplasma* in arachnids, crustaceans and plants [2, 3]. In insects, *Spiroplasma* can be vertically transmitted from females to offspring [4]. Transmission to plants involves an insect vector, as has been observed with the plant pathogen *S. citri* which is transferred to plants by leafhoppers and psyllids [5]. *Spiroplasma* associated with insects consist of mutualists, commensals, male-killing reproductive parasites and pathogens [6]. There are a few reports of pathogenic spiroplasmas in insects, like *S. apis* and *S. melliferum*, which produce lethal infections in honeybees [7, 8]. Nonetheless, most reported spiroplasmas are beneficial and may be considered facultative symbionts [3]. In addition to lethal pathogenicity, two other phenotypes are induced by spiroplasmas in insects such as protection against parasites (wasps and nematodes) and male-killing. Both phenotypes are produced by *Spiroplasma poulsonii* in *Drosophila* [9–11]. Protection against parasites has been associated with the presence of genes encoding ribosomal inactivating proteins (RIPs) in *S. poulsonii* genome. These RIPs are toxins that damage the ribosomes of *Drosophila* parasitic wasps and nematodes [12, 13]. Likewise, a plasmid-encoded protein (Spaid) seems to be involved in *D. melanogaster* male-killing phenotype produced by *S. poulsonii* [11, 14].

Molecular and phylogenetic classification using either the 16S rRNA or single-copy gene markers have split *Spiroplasma* spp. into four major clades. Three clades are composed of the formally described *Spiroplasma*: Citri-Chrysopicola-Mirum (CPM), Apis, and Ixodetis. The remaining clade (Mycoides-Entomoplasmataceae) contains species from the genera *Mycoplasma*, *Mesoplasma* and *Entomoplasma*, that have lost the helical cell morphology (thus they are not named *Spiroplasma*) [15]. *Spiroplasma* genome size ranges between 0.7 to 2.2 megabase pairs (Mbp). To date, 32 genomes of *Spiroplasma* spp. are available. However, only two belong to the Ixodetis clade, one is from the parasitic wasp *Cephus cinctus* [16] and the other is from the African monarch butterfly *Danaus chrysippus* [17], which makes this clade the least represented at the genomic level, in spite of being present in many insects [18].

Up to now, there are few functional genomic studies of *Spiroplasma*, two of them used quantitative reverse transcriptase PCR (qRT-PCR) to analyze the change of expression in key genes of *S. citri* [19, 20]. Additionally, there are two total transcriptional studies (RNAseq) of spiroplasmas, one analyzed the differences in culture growth of *S. diminutum* and *S. taiwanense* isolated from mosquitos [21], and other described the transcriptional profile of *S. poulsonii* in *Drosophila* (hemolymph) and in culture medium [22]. Differential gene expression was evidenced in all studies under contrasting conditions. Nonetheless, to the best of our knowledge, there is no transcriptome analysis of spiroplasmas colonizing different insect host tissues and the genes involved in the symbiotic interaction. Moreover, there is no transcriptional study of *S. ixodetis*.

*Dactylopius* (Hemiptera: Coccoidea: Dactylopiidae) is a scale insect that feeds on the sap of prickly pear, mainly from the genus *Opuntia* [23]. *Dactylopius* is the main source of carminic acid, a red dye used in cosmetics, drugs, food, and textile industry, achieving economical relevance [24]. The genus *Dactylopius* comprises more than 10 species, but only *D. coccus* is used for the extraction of carminic acid. This species was probably domesticated in Mexico more than one thousand years ago, selected for increased production and quality of the pigment compared to the other 'wild' species (e.g., *D. opuntiae*) [23, 25, 26]. Previous reports using metagenomic, metatranscriptomic and culture approaches showed diverse bacteria and fungi in both the domesticated and wild carmine cochineals [25, 27–30]. Within the microbial community present in *Dactylopius*, we reported a nitrogen-fixing  $\beta$ -proteobacterium *Candidatus* Dactylopiibacterium carminicum [30, 31] as well as two different strains of *Wolbachia* (*wDacA* and *wDacB*) [29], and fungi [32]. As part of the metagenome and metatranscriptome surveys from *D. coccus*, we also reported the presence of *Spiroplasma* sequences [31]. However, there is no study of the phylogenetic classification, genomic capabilities, and the putative roles of *Spiroplasma* bacterium within *Dactylopius*. Here we present the analysis of three *Spiroplasma ixodetis* metagenome-assembly genomes (MAGs) independently recovered from the domesticated *D. coccus* females and males, of two different populations each, as well as from a population of the wild cochineal *D. opuntiae* (females) metagenomes. Additionally, we used the previously reported metatranscriptome data from the gut, hemolymph, and ovary of *D. coccus* [31], to analyze the expression of genes putatively involved in the symbiotic interaction between *S. ixodetis* and *Dactylopius*. Genomic and metatranscriptomic analyses here presented, suggest that *S. ixodetis* may use the type IV secretion system (T4SS) to interact with the

carmine cochineal. Moreover, the incompleteness of genes encoding virulence factors indicates that *S. ixodetis* could be considered as a non-pathogenic symbiont.

## Results and discussion

### New *Spiroplasma ixodetis* symbiont is present in multiple *Dactylopius* spp. metagenomic samples

Mollicute-related MAGs were recovered in metagenomic assembled and binned samples from adult males and females of the domesticated carmine cochineal *D. coccus* from two different populations, as well as in the metagenome from adult females of the wild cochineal species *D. opuntiae*. Analysis of single-copy gene markers showed that these MAGs exhibited high completeness (>92%) and no apparent contamination (Table 1). Mollicute MAGs of *D. opuntiae* females, *D. coccus* females and males were placed into 258, 286 and 353 scaffolds, respectively (Table 1). The estimated genome size of these MAGs ranged from 1.32 to 1.9 Mbp. Around 1215 to 1371 coding sequences (CDS) were identified within these genomes (Table 1). Like in other Mollicutes [33], the Mollicutes-related MAGs from the *Dactylopius* spp. had low G-C % content (~24%, Table 1). The complete sequences of the 16S rRNA gene were obtained for all three MAGs (Table 1). Phylogenetic reconstruction of these 16S rRNA sequences placed the *Dactylopius* Mollicute-related MAGs within the *Spiroplasma* genus in the Ixodetis clade. (Fig. 1a). Additionally, 16S rRNA sequence from *D. coccus* and *D. opuntiae* MAGs showed 99% nucleotide identity with the *S. ixodetis* Y32 16S rRNA sequence from the western black-legged ticks (*Ixodes pacificus*) and to other *S. ixodetis*

(Additional file 1 Fig. S1). We further refer to *Dactylopius* associated *Spiroplasma* as *S. ixodetis* DO (from *D. opuntiae*, female), *S. ixodetis* DCF (from *D. coccus*, female) and *S. ixodetis* DCM (from *D. coccus* male), respectively.

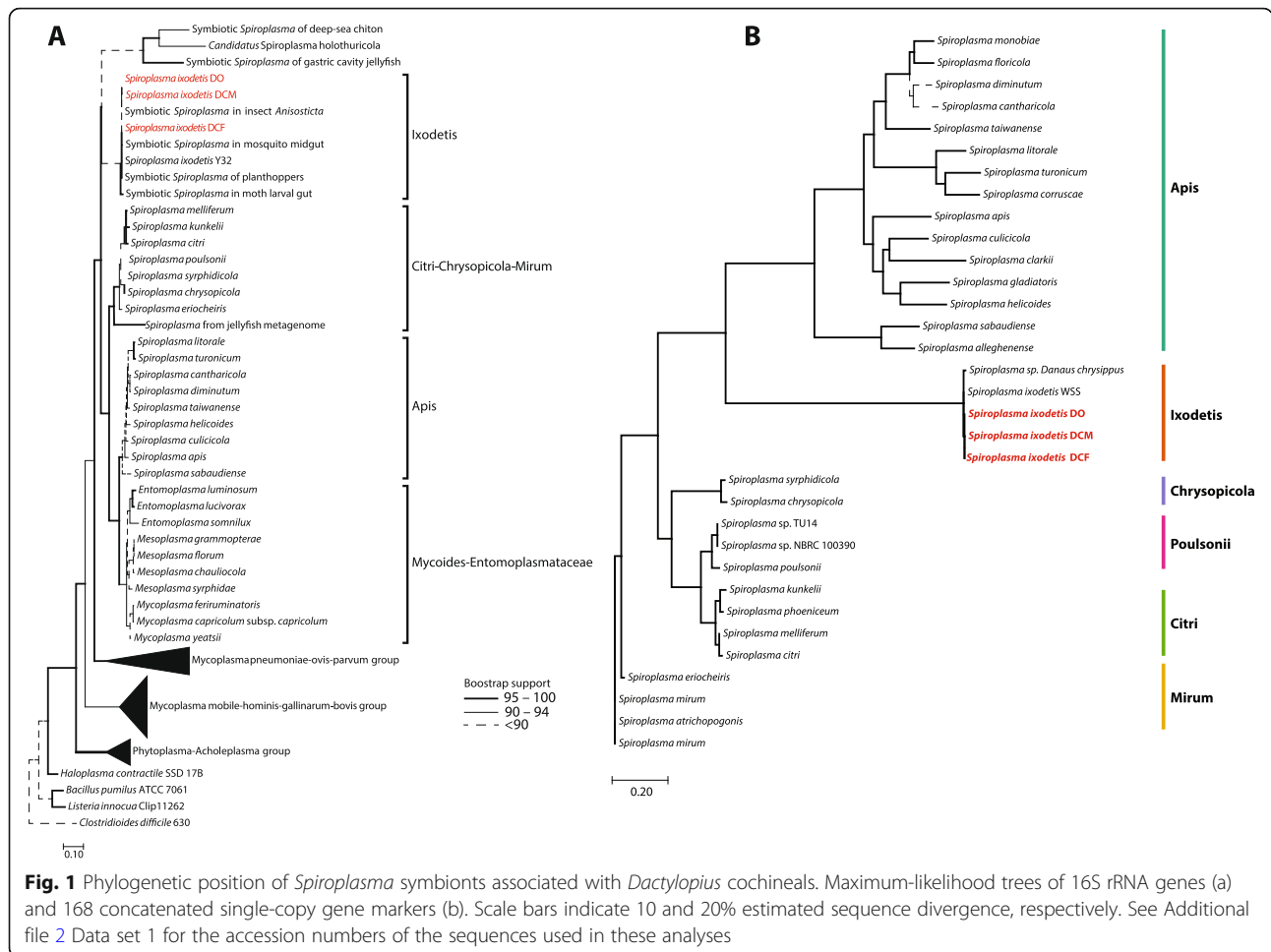
To further classify and compare the *S. ixodetis* from *Dactylopius* cochineals with other spiroplasmas, we performed a pan-genome analysis using 30 public *Spiroplasma* genomes from the Apis, Ixodetis, Chrysopicola, Poulsonii, Citri and Mirum phylogenetic clades (Additional file 2 Data set 1). The pan-genomic analysis showed 11,045 gene clusters and 168 of them corresponded to single copy core-genes within the *Spiroplasma* pan-genome. Robust phylogenomic analysis using the 168 single-copy core genes confirmed the position of spiroplasmas from *Dactylopius* within the Ixodetis clade (Fig. 1b). Average amino acid identity (AAI), using the same 168 core single-copy genes, showed that *S. ixodetis* DO, DCM and DCF shared identities of ~99.5% with the genomes of *Spiroplasma ixodetis* symbiont of the wheat stem sawfly (WSS) *Cephus cinctus* [16] and ~98.5 with the *Spiroplasma* endosymbiont of the African monarch butterfly *Danaus chrysippus* [17] (Additional file 2 Data set 1). AAI analysis also showed that *S. ixodetis* in both female and male *D. coccus* populations were 100% identical (Additional file 2 Data set 1). Moreover, *S. ixodetis* of *D. opuntia* showed a 99.78% AAI value in comparison to *S. ixodetis* of both *D. coccus* populations (Additional file 2 Data set 1). This result suggests that there are slight variations in the bacterial genomes between the two *Dactylopius* species (i.e., *D. coccus* and *D. opuntiae*), although there is no apparent difference between spiroplasmas of the two populations

**Table 1** General genomic features of *Spiroplasma ixodetis* DO, DCF and DCM compared to others insect-associated *Spiroplasma*

Genome ID	<i>S. ixodetis</i> DO	<i>S. ixodetis</i> DCF	<i>S. ixodetis</i> DCM	<i>S. ixodetis</i> WSS	<i>Spiroplasma</i> sp.	<i>S. melliferum</i>	<i>S. sabaudiense</i>
Number of scaffolds/contigs	258	286	353	145	12	6	1
Estimated genome size (Mb)	1.28	1.19	1.32	0.73	1.75	1.29	1.075
Average genomic coverage	1795x	1480x	1112x	0.727x	2000x	11x	600.8x
N50	8196	6014	7774	5160	265,779	741,187	NA
G + C content (%)	24.63	24.21	24.16	24.56	23.7	27.09	30.16
CDS genes	1260	1215	1371	649	1813	1297	933
rRNA (16S, 5S, 23S)	2(1,1,ND)	2(1,1,ND)	3(1,1,1)	NA	4(1,2,1)	3(1,1,1)	6(2,2,2)
tRNA	27	27	27	23	27	31	32
Genome completeness (%) <sup>a</sup>	95.5	92.5	94.7	77.8	95.5	98.1	100
Insect associated	<i>Dactylopius opuntiae</i>	<i>Dactylopius coccus</i>		<i>Cephus cinctus</i>	<i>Danaus chrysippus</i>	<i>Apis mellifera</i>	<i>Aedes</i> sp.

The number of CDS genes, rRNA, tRNA for *Dactylopius* associated *S. ixodetis* was calculated using the Prokka annotation files. Otherwise, information was retrieved from the NCBI GenBank database

<sup>a</sup>Genome completeness was calculated for all organisms with the checkM pipeline  
NA not available, ND not detected



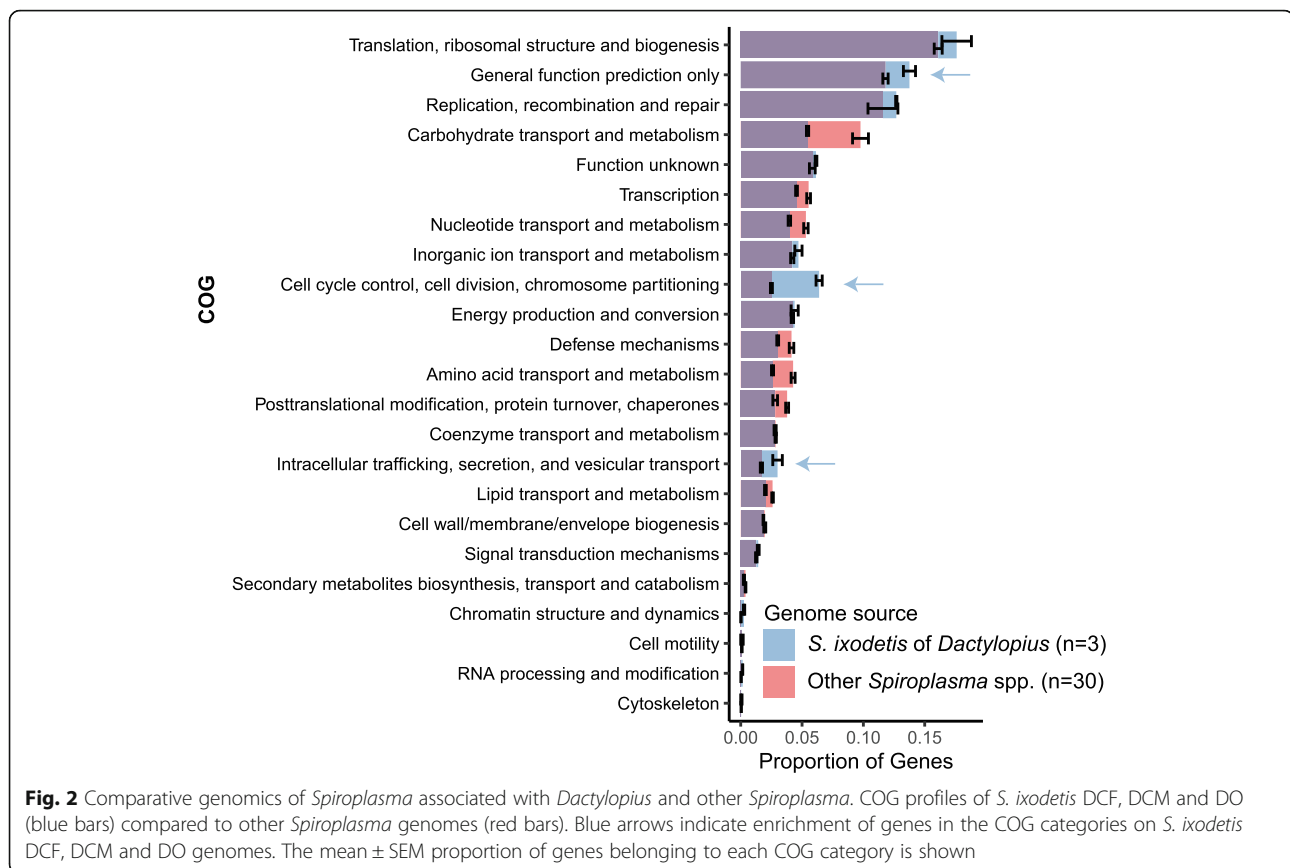
(males and females) of *D. coccus*. Similar variations in core gene identity (ANI ~ 99.5) have been reported between the sister species strains of *S. poulsonii* (sMel and sHy), symbionts of *Drosophila melanogaster* and *D. hydei*, respectively, and it has been linked to a host adaptation process [34]. Likewise, genomic differences between the *S. ixodetis* strains of *Dactylopius* may result from adaptation to different host species (i.e., *D. opuntiae* and *D. coccus*).

To find out if *S. ixodetis* was present in other *Dactylopius*, we further analyzed the metagenomes of *D. coccus* (females), from commercial samples of Mexico and Peru, previously reported by Campana et al., [35]. No complete *S. ixodetis* MAGs were recovered after assembly and binning analyses of these *D. coccus* metagenomes. Nonetheless, we were able to identify long (> 900 base-pairs [bp]) and highly similar (BLASTn identity > 90% and e-value < 0.0005) contigs to those from the *S. ixodetis* DCF genome in both the Mexican (*n* = 56) and Peruvian (*n* = 20) metagenomic assemblies. The above confirms that *S. ixodetis* is distributed not only in different species of *Dactylopius*, but also in different

populations of *D. coccus*. This indicates that this bacterium can form a seemingly symbiotic relationship with the carmine cochineal and may not be a merely sporadic association. As in nature, frequencies of *Spiroplasma* spp. in their insect host are variable [36, 37], further studies are required to elucidate the prevalence of *S. ixodetis* in other *Dactylopius* species/populations.

**Reduced number of genes for amino acid and carbon metabolism were found in *S. ixodetis* DO, DCM and DCF**

To gain information about the general metabolic profiles of *S. ixodetis* and their role in the interaction with *Dactylopius*, a comprehensive comparative genomic analysis was performed between *S. ixodetis* DO, DCM and DCF and other *Spiroplasma* (*n* = 30) genomes recovered from diverse environments (i.e., vertebrates, plants, and other insects). Analysis of the clusters of orthologous groups of proteins (COG) showed that *S. ixodetis* DO, DCM and DCF have fewer genes associated with transport and metabolism of amino acids compared to other *Spiroplasma* spp. genomes (Fig. 2). *Spiroplasma* species are auxotrophs for most of the essential amino acids and



require multiple transporters to obtain them from the host [38]. A smaller number of genes coding for ABC transporters of peptides and oligopeptides as well as fewer genes involved in the biosynthesis of amino acids were detected in *S. ixodetis* genomes from *Dactylopius* in comparison with other *Spiroplasma* spp. (Fig. 2 and Additional file 3 Data set 2). Most of the *Spiroplasma* obtain ATP through arginine metabolism [39, 40]. However, and contrasting with the previous *S. ixodetis* WSS genome [16], no arginine biosynthetic genes were found in *S. ixodetis* of *Dactylopius* (Additional file 3 Data set 2).

A complete set of genes for glycolysis, fructose catabolism and the pentose-phosphate pathway were found in *S. ixodetis* DO, DCM and DCF genomes (Additional file 3 Data set 2). Trehalose, glucose and mannose are abundant sugars in insect hemolymph and most of the insect-associated spiroplasmas encode genes involved in the phosphotransferase system (PTS) to transport these sugars into the bacterial cell [38, 40, 41]. Accordingly, *S. ixodetis* DO, DCM and DCF showed genes encoding glucose (*ptsG*), fructose (*fruA/B*) and N-acetylglucosamine (*nagE*) PTS transporters (Additional file 3 Data set 2). Even though *S. ixodetis* DO, DCM and DCF display genes for maltose and cellobiose PTS systems, no other genes for oligosaccharides catabolism were

found in these genomes (Additional file 3 Data set 2). Particularly, no genes coding for oligosaccharide breakdown and catabolism (e.g., glycoside hydrolases) were found in DO, DCM and DCF genomes (Additional file 3 Data set 2) suggesting that *S. ixodetis* is unable to utilize complex polysaccharides as a carbon source.

In *Spiroplasma* species that are insect pathogens, such as *S. citri*, *S. apis* and *S. mellipherrum*, genes for trehalose utilization, including the PTS, are present [40–42]. However, similar to *S. ixodetis* of *Dactylopius*, genes for trehalose catabolism are absent or found as non-functional pseudogenes in the non-pathogenic species *S. poulsonii*, *S. chrysopicola* and *S. syrphidicola* [40, 43]. The putative inability to metabolize trehalose by *S. ixodetis* DO, DCF and DCM may limit spiroplasma growth in *Dactylopius* tissues as has been previously suggested in the *Drosophila-S. poulsonii* interaction [40].

#### ***S. ixodetis* DO, DCM and DCF genomes have few virulence factors-encoding genes**

Incomplete genes encoding toxin-like proteins and other putative virulence factors were found in the *S. ixodetis* DO, DCM and DCF genomes (Table 2). Ankyrin repeat domains are present in many virulence effector proteins [44]. Particularly, the ankyrin-repeat containing protein

**Table 2** Virulence factor-encoding genes present in *S. poulsonii* and homologous genes in *Dactylopius*-associated *S. ixodetis*

	Annotation in <i>S. poulsonii</i> <sup>a</sup>	<i>S. ixodetis</i>			
		<i>S. poulsonii</i> locus id <sup>a</sup>	DCF locus id	DCM locus id	DO locus id
Toxins	Ankyrin repeat (Spaid)	SMSRO_SFP00290	DC_DC_00469 <sup>b</sup>	KIIIGDCO_00556 <sup>b</sup>	DO_DO_00453 <sup>b</sup>
	ETX-like	SMSRO_SF021610	–	–	–
	RIP1	SMSRO_SF016530	–	–	–
	RIP2	SMSRO_SF018820	DC_DC_00311 <sup>b</sup>	KIIIGDCO_00357 <sup>b</sup>	–
	RIP3	SMSRO_SF023880	–	–	–
	RIP4	SMSRO_SF020720	–	–	–
	RIP5	SMSRO_SF003660	–	–	–
Adhesins	SpARP1 <sup>d</sup>	SMSRO_SF002520	–	–	DO_DO_00287
	SpARP2	SMSRO_SF011850	–	–	–
	SpARP3	SMSRO_SF022680	–	–	DO_DO_00279
	SpARP4	SMSRO_SF024450	–	KIIIGDCO_00941	–
	SpARP5	SMSRO_SFP00390	–	–	–
Spiralins	SpiA	SMSRO_SF013140	–	–	–
	SpiB	SMSRO_SF009660	–	–	–
	SpiC	SMSRO_SF015890	–	–	–
Chitinases	ChiD1	SMSRO_SF008450	–	–	–
	ChiD2	SMSRO_SF013110	–	–	–
Lipid metabolism	CIs	SMSRO_SF001010	DC_DC_00075	KIIIGDCO_00120	DO_DO_00089
	GlpO	SMSRO_SF018440	DC_DC_00364 <sup>c</sup> DC_DC_00365 <sup>c</sup>	KIIIGDCO_00489 <sup>c</sup> KIIIGDCO_00490 <sup>c</sup>	DO_DO_00848 <sup>c</sup> DO_DO_00849 <sup>c</sup>

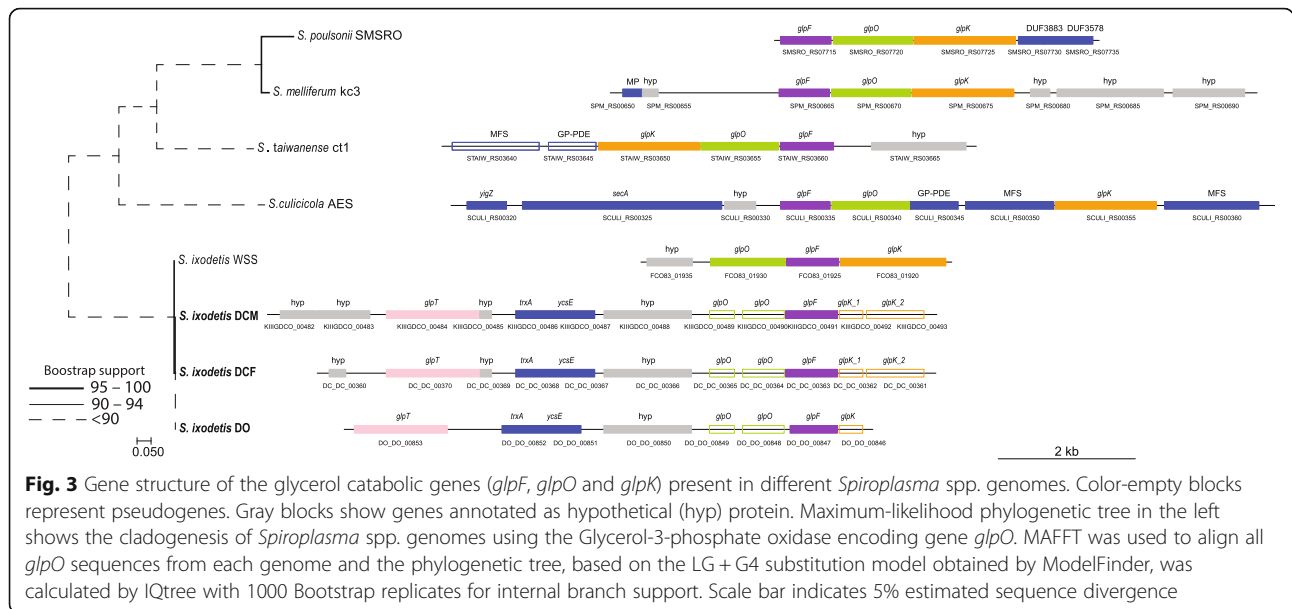
<sup>a</sup>Masson et al., 2018<sup>b</sup>Incomplete or partial protein<sup>c</sup>Pseudogene<sup>d</sup>Annotated as Putative adhesin P89 in *S. citri* (Uniprot ID: Q9EV58\_SPIC1)

Spaid of *S. poulsonii* (locus SMSRO\_SFP00290) contributes to the male-killing phenotype in *D. melanogaster* [11]. Incomplete homologs (> 50% identity, > 75% BLASTp coverage) of *spaid* gene were found in DO, DCM and DCF genomes (Table 2). Similarly, *spaid* homologs were reported in *S. ixodetis* WSS genome [16]. However, *spaid* sequences of *S. ixodetis* WSS lack the N-terminal signal peptide domains present in *S. poulsonii* Spaid protein and thus were classified as not functional proteins [16]. In addition to *spaid*, partial coding sequence, homologs to the ribosome-inactivating protein (RIP2) gene of *S. poulsonii* were found in *S. ixodetis* from the domesticated *D. coccus* but not from the wild species *D. opuntiae* (Table 2). RIP proteins of other spiriplasmas have been implicated in protection against nematodes and parasitic wasp in different species of *Drosophila* [45, 46]. Even though genomic results suggest *S. ixodetis* DO, DCM and DCF encode multiple toxin-like factors, most of these are incomplete or annotated as putative pseudogenes.

In *S. poulsonii* and *S. citri* spiralin protein, encoded by *spiA*, *spiB* and *spiC*, acts as a lectin anchor for binding

to glycoproteins present in the insect cells and is required for efficient colonization of the host [39, 47, 48]. No homologous gene of spiralin *spiA*, *B* or *C* was detected in *S. ixodetis* DO, DCM and DCF genomes (Additional file 3 Data set 2). Nonetheless, homologous genes encoding the adhesins SARP1 and SARP3 from *S. poulsonii* and *S. citri* were found in *S. ixodetis* DO and DCM genomes (Table 2). Adhesins like SARP1 and SARP3 as well as phosphoglycerate kinase (PGK) are involved in cell adherence and invasion of *S. citri* to the leafhoppers *Circulifer tenellus* and *Circulifer haematoceps* [49, 50]. All three *S. ixodetis* DO, DCM and DCF genomes encode PGK (Additional file 3 Data set 2) which in *S. citri* is a key factor for bacterial internalization into the host cells [51, 52].

Glycerol catabolic and transporter genes were found in *S. ixodetis* DO, DCM and DCF genomes (Additional file 3 Data set 2), particularly, the genes *glpO* in addition to *glpF* (membrane channel) and *glpK* (kinase) (Fig. 3). Synteny analysis showed similar genomic architecture of *glpO*, *glpF* and *glpK* genes in *S. ixodetis* DO, DCM and DCF genomes and in *S. poulsonii*, *S. melliferum* and



**Fig. 3** Gene structure of the glycerol catabolic genes (*glpF*, *glpO* and *glpK*) present in different *Spiroplasma* spp. genomes. Color-empty blocks represent pseudogenes. Gray blocks show genes annotated as hypothetical (*hyp*) protein. Maximum-likelihood phylogenetic tree in the left shows the cladogenesis of *Spiroplasma* spp. genomes using the Glycerol-3-phosphate oxidase encoding gene *glpO*. MAFFT was used to align all *glpO* sequences from each genome and the phylogenetic tree, based on the LG + G4 substitution model obtained by ModelFinder, was calculated by IQtree with 1000 Bootstrap replicates for internal branch support. Scale bar indicates 5% estimated sequence divergence

*S. culicicola* (Fig. 3), which agrees with previous reports of *glp* gene organization in other spiroplasmas [53]. Glycerol metabolism may lead to the production of hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) by L-α-glycerophosphate oxidase (GlpO) [3, 54]. In the human pathogen, *Mycoplasma mycoides* subsp. *mycoides*, H<sub>2</sub>O<sub>2</sub> has been linked to tissue inflammation and cellular damage [55, 56]. Thus, in insect associated pathogenic *Spiroplasma* spp. (i.e., *S. taiwanense*, *S. culicicola* and *S. apis*) the presence of *glpO* coding gene has been considered to be a virulence factor [57]. However, unlike other spiroplasmas, *glpO* and *glpK* genes in *S. ixodetis* DO, DCM, and DCF were incomplete and annotated as pseudogenes suggesting that they are unable to produce H<sub>2</sub>O<sub>2</sub> from glycerol (Fig. 3).

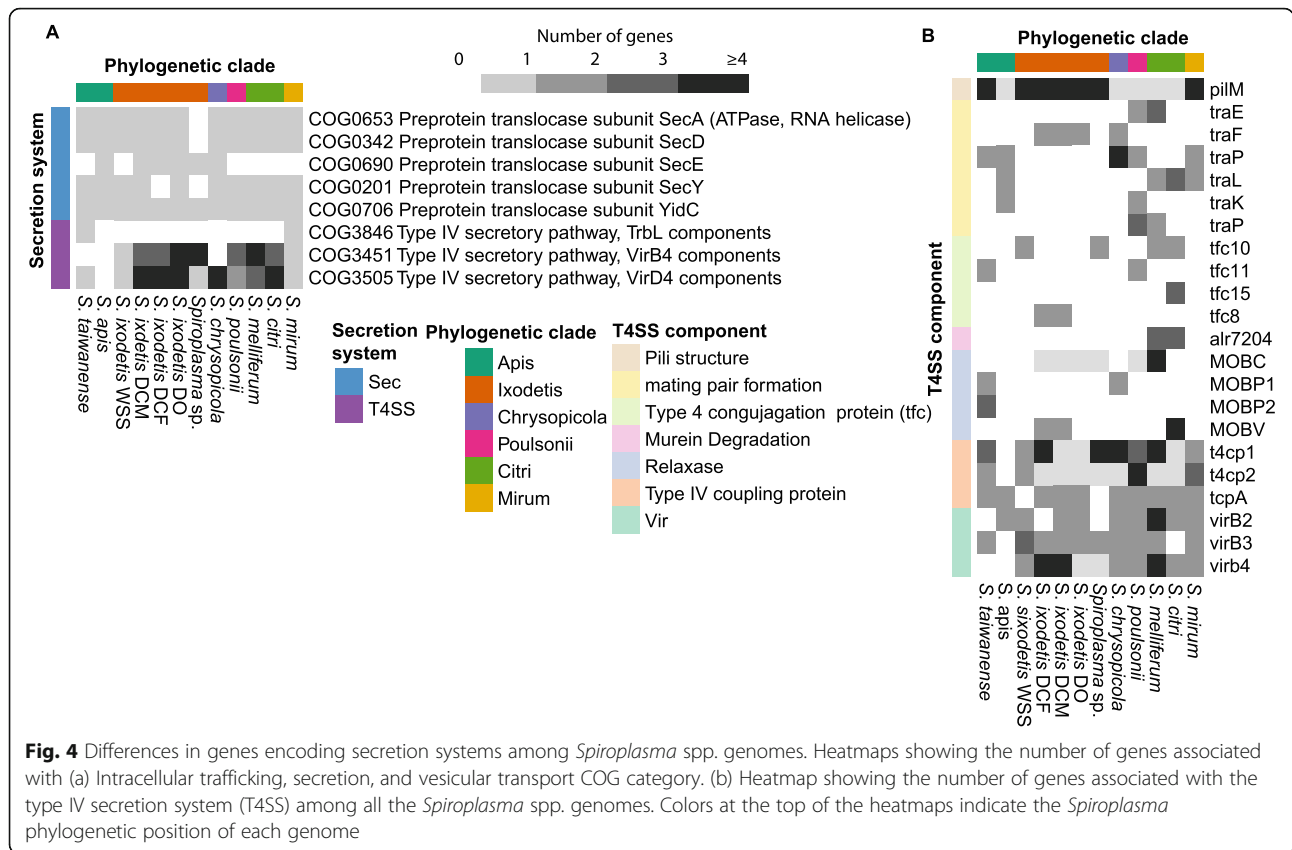
*glpO*, *glpF* and *glpK* genes are commonly recovered from pathogenic *Spiroplasma* like *S. culicicola* and *S. taiwanense*. In those bacteria n-glycerol 3-phosphate (G3P) is taken up by the transporters UgpA, UgpC and UgpE [57]. Even though no *ugpA/C/E* homologous genes were found in *S. ixodetis* genomes (Additional file 3 Data set 2), a *glpT* gene, encoding the G3P-transporter, was found in the DO, DCM and DCF genomes, but not present in other *Spiroplasma* genomes of insects (Fig. 3). Comparative analysis showed that this gene forms an orthologous cluster with the *glpT* of *S. floricola*, *Mycoplasma yeatsii* and *Mycoplasma putrefaciens*. This suggests that in *S. ixodetis* the G3P might be transported using a GlpT transporter instead of the UgpA/C/E system which is associated with pathogenicity.

Altogether, the absence of genes for trehalose catabolism, the pseudogenization of virulence factors (i.e., *spaid* and *glpO*) as well as the presence of genes for insect cell colonization (i.e., the adhesins SARP1, SARP3 and PKG), suggest *S. ixodetis* is adapted to live as a non-pathogenic symbiont inside *Dactylopius*.

**Comparative genomics revealed an enrichment of genes encoding the type IV secretion system (T4SS) in the *Dactylopius* associated *S. ixodetis* in comparison to other *Spiroplasma* spp. genomes**

COG profile comparison between DO, DCM and DCF and other *Spiroplasma* spp. revealed a greater representation of the intracellular trafficking/secretion category in DO, DCM and DCF (Fig. 2). Secretion of macromolecular effectors (i.e., protein and nucleic acids) plays a central role in modulating the interactions between symbiotic (pathogenic and mutualistic) bacteria and their host [58, 59]. Similar to other spiroplasmas associated with insects (e.g., *S. poulsonii*) [40], *S. ixodetis* DO, DCM and DCF showed genes encoding the general secretion (Sec) system. Particularly, the genes *secA*, *secY* (missing in DCF), *secG* and *secE* were found in *S. ixodetis* from *Dactylopius* spp. (Fig. 4a). Other Sec coding genes such as the signal recognition GTPase (*ffh* and *ftsY*) and the translocase (*gidC*) were also present in *S. ixodetis* DO, DCM and DCF genomes (Additional file 3 Data set 2) suggesting DO, DCM and DCF use the Sec system to export proteins.

Additionally, *S. ixodetis* DO, DCM and DCF genomes showed more *virB4-D4* predicted genes, associated with the type IV secretion system (T4SS), than other *Spiroplasma* species associated with insects (Fig. 4a). A phylogeny using the VirB4 ATPase clustered the *virB4* sequences of *S. ixodetis* DO, DCM, and DCF with other *virB4* of *Spiroplasma* spp. (Additional file 1 Fig. S2), discarding a putative chimeric-assembly origin of these genes. Additionally, multiple *virB4-D4* genes were found in scaffolds encoding plasmid-like coding sequences (i.e., *soj* and plasmid replication protein; Additional file 1 Fig. S3 and Additional file 3 Data set S2) in the *S. ixodetis*



DO genome. Differences in coverage of genomic contigs have been used as a proxy to detect plasmids in bacteria [60] and plasmid-like scaffolds with *virB4-D4* sequences of *S. ixodetis* DO showed a higher sequence coverage (> ~ 5000x) than the coverage for other scaffolds in the DO genome (Additional file 1 Fig. S3). Similarly, genes encoding VirB4 T4SS ATPase were detected in plasmids of *S. kunkelii* and *S. citri*, pSKU146 and pSci1-6, respectively [61, 62]. A maximum-likelihood phylogenetic analysis showed that the sequence of *S. ixodetis* DO *virB4* (DO\_DO\_00343) cluster together with the *virB4* genes from plasmids pSKU146 and pSci1-6 (Additional file 1 Fig. S4). All this evidence suggests that like in *S. citri* and *S. kunkelii* the T4SS of *S. ixodetis* DO is plasmid borne. Although the DCM and DCF genomes have copies of genes *soj/parA*, *smc/spoJ* and the replication protein COG5521, involved in plasmid segregation [63–65], no *virB4-D4* genes were found in the plasmid-like scaffolds of these bacteria (Additional file 3 Data set 2).

Annotation using the TXSSdb database (see methods) revealed that other T4SS elements are present in *S. ixodetis* DO, DCM and DCF genomes such as genes coding for the type IV coupling protein (t4cp), the MOB relaxase and *tra* genes (Fig. 4b). These elements are typically used by bacteria in the conjugation process [66]. Previously, the presence of *mob* and *tra* coding genes in *S.*

*citri* plasmid pSKU146 suggested that spiroplasmas are conjugative [62]. Having similar elements in DO, DCM and DCF indicates that these bacteria might be conjugative. Not all the components for the T4SS were found in *S. ixodetis* DO, DCM and DCF genomes (Fig. 4a-b), however, it has been suggested that Gram-positive bacteria, and specifically the cell wall lacking members of the *Mollicutes*, do not require as many T4SS components to secrete macromolecules as the Gram-negative bacteria [61]. Thus, *S. ixodetis* DO, DCM and DCF may have a functional T4SS. In *Wolbachia*, perhaps the most abundant insect intracellular-symbiont, multiple T4SS loci encoding Vir proteins were expressed during all stages of the host development from embryogenesis to adult male and female *Drosophila* flies [59]. Moreover, it has been considered that *Wolbachia* uses the T4SS to secrete as many as 105 proteins with putative effector function to the host (*Drosophila*) cells [53]. Like *Wolbachia* and other symbiotic bacteria (e.g., *Mesorhizobium loti*) [67], T4SS could be used by *S. ixodetis* to interact with the host.

#### ***S. ixodetis* DCF differential gene expression in the gut, hemolymph, and ovary tissues of *D. coccus***

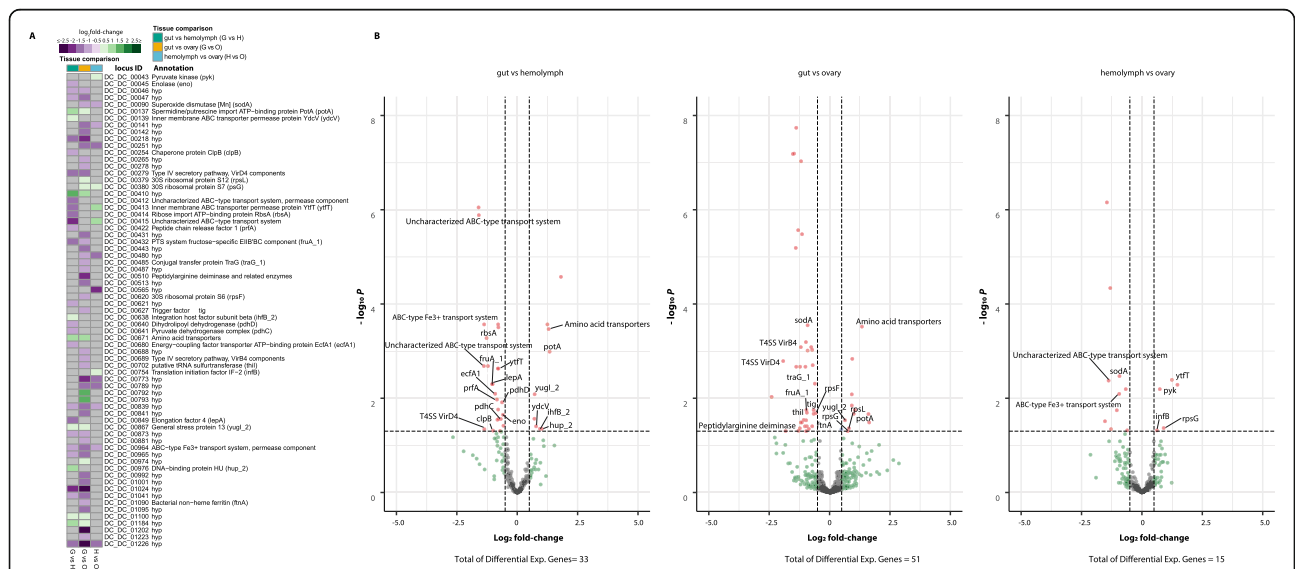
To detect genes seemingly involved in the interaction between *S. ixodetis* and *Dactylopius*, we compared the



transcriptome of *S. ixodetis* DCF recovered from previously reported metatranscriptomic data of *D. coccuus* gut, hemolymph, and ovaries [31]. After ribosomal sequences depletion, ca. 1 million metatranscriptomic reads mapped to *S. ixodetis* DCF genome in each of the *D. coccuus* gut, hemolymph, and ovary metatranscriptomic samples (Additional file 1 Fig. S5a). Principal component analysis (PCA) of *S. ixodetis* DCF transcripts showed that the transcriptomes differed along with the ordination space by intrasample variation (y-axis 21% of variance) and tissue specificity (i.e., gut, hemolymph, and ovary; x-axis 30% of variance; Additional file 1 Fig. S5b). The above suggests that *S. ixodetis* DCF modulate gene expression depending on the allocation in the different host tissues. Pairwise comparisons between each tissue combination (i.e., gut vs. hemolymph, gut vs. ovary and hemolymph vs. ovary) identified 69 unique differentially expressed genes of *S. ixodetis* DCF across all comparisons and 46% matched with a coding product in the bacterial genomic annotation (Fig. 5a and Additional file 3 Data set 2). Among these genes, we detected that the conjugal transfer protein coding gene *traG*, as well as the *virB4* and *virD4* genes of the T4SS of *S. ixodetis* DCF, were up-regulated in the hemolymph and the ovaries in comparison to the gut (Fig. 5b). Additionally, other *S. ixodetis* DCF T4SS related encoding genes such as the type IV pilus inner membrane component *pilM*, the type

four conjugation *tfc*, the relaxase *mob* and the *virB3* were also expressed, but not differentially ( $p$ -value > 0.05), in the *D. coccuus* gut, hemolymph, and ovary (Additional file 3 Data set 2). Expression of *vir* genes, particularly *virB4*, encoding for the T4SS of *Wolbachia* has been previously detected in the ovaries of the wasp *Aso-bara tabida* suggesting that T4SS is active and could participate in the *Wolbachia-A. tabida* interaction. It has been suggested that *Wolbachia* might use T4SS to export effectors for manipulating the wasp cell biology and oogenesis [59, 68]. Similarly, high expression of *virB4* and *virD4* genes in *D. coccuus* ovary suggests that the T4SS of *S. ixodetis* DCF is biologically active. An elevated number of T4SS coding genes in the *S. ixodetis* DCF genome and the up-regulation of T4SS genes in the hemolymph and ovaries of *D. coccuus* (Fig. 5a-b and Additional file 3 Data set 2), indicate that the T4SS might play an important role in the interaction between *S. ixodetis* DCF and the host.

Metatranscriptome analysis also showed that *S. ixodetis* DCF genes encoding the SpaID and GlpO proteins, described as virulence factors, have low average levels of expression ( $\leq 42$  fragments per kilobase of transcript per million mapped reads [FPKM]) and were not differentially expressed in any of the *D. coccuus* tissues analyzed. Moreover, no expression of *S. ixodetis* RIP encoding gene was detected in any of the *D. coccuus*



**Fig. 5** Variation on gene expression of *S. ixodetis* DCF in the gut, ovary, and hemolymph of *D. coccuus*. (a) Heatmap showing all the differentially expressed genes in each tissue comparison (i.e. gut vs hemolymph, gut vs ovary and hemolymph vs ovary). Color range in heatmaps indicates variation in log<sub>2</sub> fold-change. Gray boxes indicate no differential expression observed after DESeq2 analysis (absolute log<sub>2</sub> fold-change  $\geq 0.58$  and  $p$ -adjust value  $\leq 0.05$ ). hyp means genes annotated as hypothetical proteins. (b) Volcano plot of differential gene expression of *S. ixodetis* DCF in gut vs hemolymph, gut vs ovary and hemolymph vs ovary comparisons. Red dots show genes considered as differentially expressed after DESeq2 analysis absolute log<sub>2</sub> fold change > 0.58 (outer broken vertical lines) and  $p$ -adjust value  $\leq 0.05$  ( $-\log_{10} p$ -value  $\geq 1.3$ , broken horizontal line). Green dots show non-differentially expressed genes with an absolute log<sub>2</sub> fold change > 0.58 and  $p$ -adjust value > 0.05. Gray dots show non-differentially expressed genes with an absolute log<sub>2</sub> fold change < 0.58 and  $p$ -adjust value > 0.05 (non-differentially expressed). All annotated genes are represented, red dots without annotation legend represent hypothetical protein coding genes differentially expressed

metatranscriptomes (Additional file 3 Data set 2). In *S. poulsonii*, genes encoding the RIP and GlpO proteins were up-regulated in *D. melanogaster* hemolymph in contrast to in-vitro cultures, indicating an important role of these genes in the fly-*S. poulsonii* interaction [22]. The low expression values of these spiroplasma genes in all the *Dactylopius* tissues, as well as the fact that these genes are incomplete, suggests that genes encoding virulence factors might not have an important role for the *D. coccus*-*S. ixodetis* interaction.

A high expression of *S. ixodetis* DCF genes associated with reactive oxygen species (ROS) catabolism such as the *sodA*, encoding the superoxide dismutase, was observed in the *D. coccus* ovaries relative to the gut and hemolymph. Additionally, genes *pdhC* and *pdhD*, encoding the dihydrolipoyl dehydrogenase, were up-regulated in the hemolymph relative to the gut (Fig. 5a-b and Additional file 3 Data set 2). In the intracellular bacteria *Mycobacterium tuberculosis* the expression of *pdhC* and *pdhD* has been associated with resistance to host-reactive nitrogen intermediates [69]. ROS and reactive nitrogen species (RONS) play a critical role in the establishing, maintenance, and success of the symbiosis [70]. As other eukaryotes, insects control bacterial symbionts load in their tissues by producing RONS. Nonetheless, some bacteria, mainly mutualistic bacteria, can cope with the burst of host RONS by displaying an arsenal of antioxidant enzymes (e.g., superoxide-dismutase and peroxidase) [70]. For example, in the mutualistic association between the bobtail squid *Euprymna scolopes* and the luminous bacteria, *Vibrio fischeri*, bacteria can overcome the squid RONS by over expressing a peculiar NO/oxygen-binding (H-NOX) protein to catabolize the nitric oxide produced by the host. This allows *V. fischeri* to successfully colonize the highly oxidative light-organ tissue of *E. scolopes* [71]. Likewise, *Wolbachia*, possess a single *sod* gene that likely plays a fundamental role in the management of oxidative stress produced by insects' cells [72]. In many insects, *Spiroplasma* reside as extracellular bacteria in the gut and the hemolymph, however, in other tissues such as the fat body and ovaries these bacteria are found as intracellular symbionts [4, 38, 73]. The up-regulation of *S. ixodetis* DCF *sod* gene in the ovaries compared to the gut indicates that like *Wolbachia*, *Spiroplasma* could use the superoxide-dismutase to manage the oxidative stress produced inside the insect ovarian cells. This strategy might allow *S. ixodetis* DCF to colonize the embryonic cells and be vertically transmitted to the *Dactylopius* offspring.

## Conclusions

Microbial associations with insects are considered key to their biological success. In many cases, there are multiple microbes within a single insect, and that is the case of *Dactylopius* spp. which is associated with the

diazotroph bacterium *Candidatus Dactylopiibacterium carminicum* and two *Wolbachia* strains. Here, we described that *Spiroplasma ixodetis* is also a member of the bacterial community of both the wild (*D. opuntiae*) and the domesticated (*D. coccus*) carmine cochineals. The presence of *S. ixodetis* in three independent metagenomic samples of different populations of *Dactylopius* from two species (*D. coccus* and *D. opuntiae*) as well as in the metagenomes of other two different *D. coccus* metagenomes from Peru and Mexico, indicates that *S. ixodetis* is commonly found within *Dactylopius*. However, since these metagenome studies were performed with pools of insects, we cannot conclude that each individual cochineal has *Spiroplasma*. In *Drosophila* not all individuals of same population are infected with *Spiroplasma* and the same phenomenon should be further explored in the carmine cochineals. The genome analysis showed that females and males from different populations of the domesticated cochineals carry the same *S. ixodetis*, with slight differences from that of the wild cochineals, suggesting a common origin of the symbiosis.

The *Dactylopius* associated *S. ixodetis* genomes showed pseudogenization or incompleteness of genes commonly coding for virulence factors in other *Spiroplasma* spp. As no complete virulence genes were detected, *S. ixodetis* in *Dactylopius* could be considered a non-pathogenic symbiont. Transcriptomic analyses of endosymbionts in insects are scarce. Here the transcriptomic analysis of *S. ixodetis* DCF showed that genes involved in the T4SS (e.g., *traG*, *virB4* and *virD4*) were over expressed in hemolymph and ovary of the host *D. coccus*. Moreover, genes involved in oxygen-stress tolerance such as *sodA* were also over expressed in the ovary. These results suggest that a) *S. ixodetis* might use T4SS to interact with *Dactylopius* cells, and b) these bacteria can tolerate oxygen-stress produced inside the host cells. Finally, transcriptional differences of *S. ixodetis* among *Dactylopius* tissues suggest that this bacterium can respond and adapt to the different conditions present within the host tissues indicating that there is a strong interaction and molecular signaling in the symbiosis between *S. ixodetis* and the carmine cochineal *Dactylopius*.

## Methods

### Insect sampling

Females of *Dactylopius coccus* were obtained from Campo Carmin Company, Morelos, Mexico (18°44' 46.7"N, 99°11' 17.8"W) in April 2012. Males from a different population of *D. coccus* were obtained from a commercial farm in Tepoztlán, Morelos, Mexico (18°59' 24.8"N 99°07' 02.9"W) in February 2017. Females of wild cochineal *D. opuntiae* were collected from a cactus farm field in Morelos, Mexico (18°59' 25.3"N 99°07' 02.6"W) in May 2017. Insects were obtained from

*Opuntia* spp. cactus and were transported together with their host plants to the laboratory for further experiments.

#### DNA extraction and metagenomic sequencing

Shotgun metagenomes were previously reported and obtained [29, 32] using different sequencing technologies (i.e., 454-GS-FLx, Illumina HiSeq2000–2500 and PacBio) of total DNA from guts, Malpighian tubules, ovaries and hemolymph of *Dactylopius coccus* females. Details of the insect dissection and DNA extraction procedures from different *Dactylopius* tissues are fully described in [29, 30, 32]. Briefly, for 454 sequencing, 20 *D. coccus* 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 as is described by [74]. Percoll phases were observed under a microscope, and those with cells were selected for DNA extraction. For Illumina HiSeq2000 sequencing guts, ovaries, and Malpighian tubules from 40 females were dissected using sterile forceps under a stereoscopic microscope. We tried to avoid tissue cross-contamination by rinsing the tissues several times with sterile phosphate-buffered saline to remove remnant hemolymph. These organs were pooled, suspended in sterile PBS, and macerated using a sterile plastic pestle. For PacBio sequencing, eight individuals were superficially disinfected with 70% ethanol as above. Guts and exoskeleton were removed with sterile forceps and hemolymph from all individuals was pooled for DNA extraction. In addition, metagenome sequences of wild *Dactylopius opuntiae* females and males of different *D. coccus* population were obtained by Illumina HiSeq2500 platform. For these last metagenomes, two independent pools containing the whole body of 10 *D. opuntiae* females and 30 *D. coccus* males were superficially disinfected as described above. Insects were macerated using sterile pestles and resuspended in sterile PBS. Independently of the sequence technology, DNA from all insects was extracted with the DNeasy Blood & Tissue Kit (QIAGEN) following the manufacturer's instructions. Sequencing was performed at Macrogen Inc. (Korea) for 454 and Illumina HiSeq2000–2500. PacBio sequencing was performed at Duke University Genome Sequencing Core Facility (USA). In all cases, after base calling high-quality metagenomic reads (quality Phred score > Q30) were obtained by adaptor removal and trimming using the TrimmGalore 0.4.4 pipeline ([http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) with the --paired --q 30 options.

#### Binning, genome assembly and annotation

For genomic binning all metagenomic reads from *D. coccus* females (i.e., 454, Illumina and PacBio), *D. coccus* males and *D. opuntiae* females were independently mapped using Bowtie2 2.3.5 [75] versus a concatenated index of *Wolbachia wDacA* (LSYX00000000.1), *Wolbachia wDacB*

(LSYY00000000.1) and *Candidatus Dactylopiibacterium carminicum* (NZ\_MQNN00000000.1) genomes, previously reported bacteria present in *Dactylopius* metagenomes [29, 32], with the --end-to-end and --very-sensitive options. All non-mapped reads from each metagenome (*D. coccus* female, *D. coccus* male and *D. opuntiae*, respectively) were retrieved from the bam files using Samtools 1.7 [76] with the following options: samtools fastq -@ 20 -f 2. To maintain read parity after samtools procedure, all fastq files from each metagenome were analyzed by TrimmGalore 0.4.4 pipeline with the --paired option, this resulted in high-quality paired reads for all libraries. These reads, from each *Dactylopius* species and populations, were then independently assembled using idba-ud 11.1 [77] with the default parameters. The resulting assembled genomic contigs were binned into single metagenomic assembled genomes (MAGs) by MaxBin 2.2.1 [78] with the following parameter -min\_contig\_length 900 for minimum contig length. In each binning the metagenomic contigs were divided into two different MAGs and classified by checkM 1.1.2 pipeline [79] with the lineage\_wf option for a broad taxonomical classification. In the three metagenomes analyzed, checkM detected the presence of a Mollicute related MAG in addition to an insect-related MAG. To improve the assembly, the original high-quality reads were mapped to the MAG classified as Mollicutes lineage in each metagenome using Bowtie2 with the --end-to-end and --very-sensitive options. All paired and mapped reads were once again retrieved using samtools as above and reassembled using idba-ud. Assembled contigs were compared to the previous MAG obtained by MaxBin2 and if any improvement on the total assembly length, N50 and N90 statistics was observed, this new assembly was considered further, if no improvement was observed the original MaxBin2 Mollicute MAG was used instead. Final genome assemblies were then polished by SSPACE [80] with bwa as aligner and Ciclator 1.4.1 [81] was used to find possible circular scaffolds. After the 'best' assembly was obtained, manual curation of the contigs in each MAG was done by performing a BLASTn [82] search against the 'nr' database (downloaded on September 2016) and manually removing those contigs matching to insect or any other no Mollicutes bacteria (>90% identity, <1e-5 e-value, >80% coverage). Protein translated sequences of the three Mollicutes MAGs, one for each metagenome, were retrieved by Prodigal 2.6.3 [83]. The quality (completeness, contamination, and strain heterogeneity) of this manual curated Mollicutes MAG was then assessed by the checkM [79] pipeline using the protein predicted sequences from Prodigal and the following parameters: checkm lineage\_wf --reduced\_tree --genes -t 8 -x faa. Total genomic features prediction and open reading frame (ORFs) annotation on each MAGs were performed using PROKKA v1.11 [84] with the following parameters: --gcode 4 --metagenome --rfam. Metabolic pathways were predicted using the GhostKoala tool from KEGG [85] and manually

curated using the Bioconductor R package KEGGREST 1.26.1 and the KEGG\_parser tool developed for this work ([https://github.com/avera1988/KEGG\\_parser](https://github.com/avera1988/KEGG_parser)). The dbCAN2 pipeline [86] was used to detect genes coding for putative carbohydrate-active enzymes (CAZymes) in the Mollicutes MAGs using the dbCAN-database 7 (downloaded in January 2019). Additionally, all Mollicutes MAGs were annotated using the NCBI Prokaryotic Genome Annotation Pipeline PGAP-4.6 [87] for public availability. Correlations between PROKKA, KEGG, CAZY and PGAP locus identifiers of each MAGs are shown in the Additional file 3 Data Set S2.

### Phylogenetic and comparative genomic analysis

Near-full-length (~ 1400-nucleotide [nt]) 16S rRNA gene sequences were retrieved from the annotations of the Mollicutes MAGs and compared to the NCBI “16S ribosomal RNA sequences” (accessed in February 2020) and Arb-SILVA (accessed in January 2020) databases using BLASTn to identify the nearest neighbors. MAFFT 7.3.10 [88] was used to align all sequences and a maximum-likelihood-based (ML) phylogenetic tree, based on the TVMe+R4 substitution model obtained by ModelFinder [89], was calculated by IQtree 1.6.12 [90] with 1000 Bootstrap replicates for internal branch support.

A total of 30 genomes (Additional file 2 Data set 1) across all the different *Spiroplasma* phylogenetic clades (i.e., Citri-Chrysopicola, Mirum, Apis, and Ixodetis) were downloaded from the NCBI “assembly” database (accessed in January 2020) and used for phylogenomic and comparative analyses. Pan and core genome analyses were conducted using GETHOMOLOGUES 2.0 [91], and OrthoMCL [92] was used for orthologue clustering (parameters: -A -c -t 0 -M -n 12). Single-copy genes from the core-genome of all *Spiroplasma* analyzed were parsed from the GETHOMOLOGUES pan-genome matrix results using custom Bash and Perl scripts (deposited in [https://github.com/avera1988/Comparative\\_genomics](https://github.com/avera1988/Comparative_genomics)).

For phylogenomics, a concatenated protein-based maximum-likelihood (ML) phylogenetic tree was constructed using the protein predicted sequences of 168 single-copy genes of *Spiroplasma* core genome. Protein sequences were concatenated and aligned using the multigenome2blocks pipeline [93]. ML phylogeny, based on the LG + F + R4 amino acid substitution model obtained by ModelFinder [89], was constructed using IQtree 1.6.12 [90] with 1000 Bootstrap replicates for internal branch support. Additionally, these 168 protein-coding genes were used to calculate the average amino acid identity (AAI) between *Spiroplasma* genomes using the AAI calculator from the “enveomics” collection tools [94]. AAI distance matrices were calculated and

visualized by custom Perl and R scripts (deposited in [https://github.com/avera1988/Comparative\\_genomics](https://github.com/avera1988/Comparative_genomics)).

Finally, to compare functional profiles between *Dactylopius* associated and public *Spiroplasma*, all clusters of orthologous groups (COG) of proteins from each genome (Additional file 2 Data set 1) were annotated using the cdd2cog pipeline (<https://github.com/aleimba/bac-genomics-scripts>). Comparisons between proportions of COG genes from *Dactylopius* and non-*Dactylopius* associated *Spiroplasma* were performed using custom R scripts (deposited in [https://github.com/avera1988/COG\\_differential\\_analysis](https://github.com/avera1988/COG_differential_analysis)). The number of genes on each COG category between *Dactylopius* and non-*Dactylopius*-associated *Spiroplasma* genomes were visualized by heatmaps in R 3.6.1.

### Presence of *S. ixodetis* in other *D. coccus* metagenomes from Mexico and Peru

Previously reported metagenomes of *D. coccus* collected from Mexico and Peru by Campana et al., [35] were used to detect the presence of *S. ixodetis* in different *Dactylopius* samples. Metagenomic reads were obtained by the fastq dump tool (<https://github.com/ncbi/sra-tools>) using the SRR1231828 and SRR1231831 accession numbers from the sequence-read archive (SRA) NCBI portal (<https://www.ncbi.nlm.nih.gov/sra>). Paired-reads of each metagenome (i.e., Mexico and Peru) were independently assembled using the idba-ud 11.1 [77] assembler with the default parameters. The assemblies resulted in 8085 (22,546,409 bp, N50 192,783 bp) and 10,617 (25,132,088 bp, N50 99,725 bp) contigs for the Mexican and Peruvian metagenomes, respectively. The resulting assembled metagenomic contigs were binned into MAGs by MaxBin 2.2.1 [78] with the following parameter -min\_contig\_length 900 for minimum contig length. The resulting MAGs were then classified by CheckM 1.1.2 pipeline [79] with the lineage\_wf option for a broad taxonomical classification. In the case that no *Spiroplasma*-like MAGs were recovered, BLASTn searches were performed using as a database the index of *S. ixodetis* DCF genome and as query the total assembled contigs for each Mexican and Peruvian metagenomes with the following parameters: blastn -query contigs.fa -db DCF.fasta -max\_target\_seqs 1 -num\_threads 8 -outfmt 6. Metagenomic contigs from both Mexican and Peruvian samples with an identity value > 90% and e-value < 0.0005 in the BLASTn searches were further classified as *S. ixodetis*-like contigs. To corroborate these contigs were taxonomically close to *Spiroplasma* spp. all *S. ixodetis*-like contigs recovered in each metagenome were annotated by BLASTx searches against the Uniref100 protein database (downloaded November 2020). Those contigs with a *Spiroplasma* match in the database were then scored as *S. ixodetis* contigs.

### Genes encoding components of the type IV secretion system (T4SS) detection

Protein sequence predictions from each *Spiroplasma* genome were compared to the type IV secretion system (T4SS) HMM profiles in the TXSSscan database [95] using the “hmmScan” (parameters: `hmmScan -cpu 40 -domtblout`) tool from HMMER 3.1b2 [96] to identify genes encoding domains that comprised potential components of T4SS. Putative T4SS component domains were manually parsed from `hmmScan` result tables and compared with previous PROKKA-PGAP annotations of each *Spiroplasma* genome.

### RNASeq and metatranscriptomic analysis

Metatranscriptomic data, previously obtained and reported by Bustamante-Brito et al., [31] from hemolymph, ovaries, and gut of female *D. coccus* cochineals, was used to analyze changes in the expression of *Spiroplasma* genes in different tissues of *Dactylopius*. Tissue dissection and RNA extraction procedures are fully described in [31]. Briefly, 30-s instar nymphs of *D. coccus* females were collected and superficially rinsed with ethanol 90% to remove the covering wax. Hemolymph was collected, pooled, and resuspended in 200 µl RNeasy (ThermoFisher) by doing a dorsal puncture in each insect with a 1 ml sterile syringe. After bleeding, all individuals were dissected under sterile conditions and the whole gut (including foregut, midgut, hindgut, and Malpighian tubules) and ovaries were independently collected, pooled, and resuspended in 200 µl of RNeasy (ThermoFisher). As for DNA extraction (see above), we tried to avoid tissue cross-contamination by rinsing the tissues several times with sterile phosphate-buffered saline. Total RNA was obtained from each pool of tissues (i.e., gut, hemolymph, and ovary) using the RNeasy kit (Qiagen) following the modification reported by Guerrero-Castro et al., [97]. High-quality RNA samples were used for cDNA strand-specific library preparation by the TruSeq Stranded Total RNA kit (Illumina), and ribosomal RNA (rRNA) present in the samples was removed with the RiboZero Removal kit for Bacteria (Illumina). Three biological replicates of each tissue were obtained, thus nine metatranscriptomic libraries were generated and sequenced using a single lane on an Illumina HiSeq4000 sequencer at Macrogen Korea, by a 100 bp read length pair-end protocol. Purity-filtered reads were adapter and quality (Phred score  $Q > 30$ ) trimmed by TrimmGalore 0.4.4 pipeline with the `--paired --q 30` options. Reads matching to rRNA sequences in each library were identified by Metaxa2 2.2–1 pipeline [98] using the `-x T` option and manually removed from the fastq files. To maintain read parity, all fastq files obtained after rRNA removal were subjected to a second round of filtering by TrimmGalore 0.4.4 with the `--paired` option and considered further.

### Differential gene expression analysis and quantification

To extract all the reads of *Spiroplasma* from the metatranscriptomes, high-quality filtered reads of each library (gut, hemolymph, and ovary) were independently aligned against the *Spiroplasma ixodetis* DCF genome using Hisat2 2.1.0 [99]. Properly mapped reads were obtained using samtools 1.7 with the following parameters: `samtools fastq -@ 12 -c 6 -F 4 -N`. The number of *Spiroplasma* read counts per gene locus in each library (tissue) was then obtained and summarized by the RSEM 1.3.1 pipeline [100] with the following parameters: `rsem-calculate-expression -p 12 --paired-end --bowtie2 --estimate-rspd --append-names --output-genome-bam` using the *S. ixodetis* DCF genomic annotation. Total matrix count with all abundances from each tissue replicate was obtained with the `abundance_estimates_to_matrix.pl` script from Trinity 2.4.0 [101] and parsed by the `tximport` Bioconductor package [102] in R. For differential expression analysis, treatments were classified depending on the RNAseq library as gut, hemolymph, and ovary. The software DESeq2 [103] was used to detect differentially expressed genes using as contrast each library/tissue comparison (i.e., gut vs hemolymph, gut vs ovary and hemolymph vs ovary). Genes were considered differentially expressed if the adjusted *p*-value (Benjamini-Hochberg [BH] multiple test correction [104]) was less than or equal to 0.05 and an absolute fold-change was above 1.5 (absolute  $\log_2$  fold-change  $\geq 0.58$ ) per each gene in any of the particular contrast comparison analyzed. Evaluation of main differences in expression per library (i.e., tissue) was assessed by a principal component analysis (PCA). PCA was performed and plotted by `plotPCA` function from DESeq2 with the variance stabilizing transformation (`vst`) method. Volcano plots and heatmaps were used to display and visualize all differential expressed genes of *S. ixodetis* DCF in the different libraries using the `phamp` and `EnhancedVolcano` (<https://github.com/kevinblighe/EnhancedVolcano>) packages in R. Fragments per kilobase of transcript per million mapped reads (FPKM) values of *S. ixodetis* DCF genes in each transcriptomic sample were obtained by the “`fpkm`” function of the DESeq2 package in R. All scripts to reproduce the gene expression analysis are deposited in Github ([https://github.com/avera1988/Spiroplasma\\_ixodetis\\_Dactylopius\\_RNAseq](https://github.com/avera1988/Spiroplasma_ixodetis_Dactylopius_RNAseq)).

### Abbreviations

T4SS: Type IV secretion system; ROS: Reactive oxygen species; RIP: Ribosome-inactivating proteins; Mbp: Mega base pairs; qRT-PCR: quantitative reverse transcriptase polymerase chain reaction; RNAseq: Ribonucleic acid sequencing; MAGs: Metagenome assembly genomes; CDS: Coding sequences; AAI: Average amino acid identity; COG: Clusters of orthologous groups of proteins; RNA: Ribonucleic acid; DNA: Deoxyribonucleic acid; KEGG: Kyoto encyclopedia of genes and genomes; PTS: Phosphotransferase system; PacBio: Pacific biosciences; BUSCO: Benchmarking Universal Single-

Copy Orthologs; ML: Maximum-likelihood; FPKM: Fragments Per Kilobase of transcript per Million mapped reads

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07540-2>.

**Additional file 1: Figure S1.** Maximum-likelihood phylogenetic tree of the 16S rRNA from different Mollicutes. In red are the 16S rRNA sequences of *S. ixodetis* DO, DCM and DCF. Scale bar indicates 2% estimated sequence divergence. ModelFinder was used to calculate the TVMe+R4 nucleotide substitution model. Maximum-likelihood tree was constructed by IQTree with 1000 Bootstrap replicates for internal branch support. The 16S rRNA sequences of *Clostridioides difficile*, *Bacillus pumilus* and *Listeria innocua*, were used as outgroup. **Figure S2.** Maximum-likelihood phylogenetic tree of the *virB4* ATPase coding gene from *S. ixodetis* DO, DCF and DCM (in red) and other organisms from the Genbank. Scale bar indicates 50% estimated sequence divergence. Accession numbers of all *virB4* sequences are shown. MAFFT was used to align all sequences and a maximum-likelihood-based (ML) phylogenetic tree, based on the LG + I + G4 substitution model obtained by ModelFinder, was calculated by IQtree with 1000 Bootstrap replicates for internal branch support. **Figure S3.** Plasmid-like scaffolds encoding genes of the type IV secretion system (T4SS) in the *S. ixodetis* DO genome. Arrows represent the structure of the genes. The sequencing coverage per each scaffold is presented. **Figure S4.** Maximum-likelihood phylogenetic tree of plasmid (pink) and chromosomal (black) encoding *virB4* ATPase of *S. ixodetis* DO, DCM, DCF and plasmid encoding *virB4* ATPase of *S. citri* and *S. kunkelii* (green). Scale bar indicates 50% estimated sequence divergence. Accession numbers of all *virB4* sequences are shown. MAFFT was used to align all sequences and a maximum-likelihood (ML) phylogenetic tree, based on the LG + F + G4 substitution model obtained by ModelFinder, was calculated by IQtree with 1000 Bootstrap replicates for internal branch support. **Figure S5.** General transcriptomic features of *S. ixodetis* DCF expressed genes in the gut ovary and hemolymph of *D. coccus*. (a) Number of RNAseq mapped reads to *S. ixodetis* DCF genome in different *Dactylopius* tissues. (b) Principal component analysis (PCA) after DESeq2 normalized transcripts. Colors correspond to different insect tissues: in green from gut (GUT), in orange from hemolymph (HM), and in purple from ovary (OV).

**Additional file 2: Data set 1.** Genbank accession numbers of genomes and sequences used in this study and average amino acid identity (AAI) of *Spiroplasma* genomes.

**Additional file 3: Data set 2.** General genome annotation and differential expression data of *S. ixodetis* from *Dactylopius*. DCF, DCM and DO sheets show all the annotations obtained from different programs and databases (i.e., Prokka, CAZYDB, COG, KEGG, NCBI-PGAP) of each genome. PlasmidLikeScaffolds sheet shows the gene features of putative plasmid-like sequences in DO, DCF and DCM genomes. COGCarbohydrateMetabolism and COGAminoacidMetabolism sheets show the number of genes in each COG category annotated in DCF, DCM, DO genomes, and other 30 *Spiroplasma* spp. from the NCBI Genome database for the Carbohydrate transport and metabolism and Amino acid transport and metabolism categories, respectively. DifferentialExpressedGenes sheet show all the log2FoldChange, *p*-values, *p*-adjust values, and annotations of *S. ixodetis* DCF differential expressed genes in the gut, ovary and hemolymph tissues. Columns show the values of the different pair comparisons (i.e., gut vs hemolymph, gut vs ovary and hemolymph vs ovary). Expre.Values.w.AnnotationDCF shows all the DESeq2 results including the normalized expression as fragments per kilobase of transcript per million mapped reads (FPKM) values for each transcriptomic comparison and the annotation of all *S. ixodetis* DCF genes recovered.

## Acknowledgements

We thank to Benjamin Jahnes for the valuable comments on the manuscript. We also thank to José Luis Marquina for providing *D. coccus*; and Víctor del Moral, Alfredo Hernández and Romualdo Zayas for computing support. All bioinformatic analyses were performed using the CCG-UNAM servers.

## Authors' contributions

AVPL and EMR designed and managed the project. AVPL and MDM performed all bioinformatic analyses. RBB collected insect samples and performed the RNA extractions. MR helped with phylogenetic analyses. AVPL, RBB, VHA, MR and EMR wrote the paper. All authors have read and approved the manuscript.

## Funding

This work was accomplished by the support of PAPIIT-UNAM IN207718 grant. AVPL. was funding by "Programa de becas para estancias postdoctorales en el extranjero del Consejo Nacional de Ciencia y Tecnología CONACyT Mexico" grant 019-000012-01EXTV-00267. RBB. is a graduate student from the Graduate Program in Biochemical Sciences, UNAM with CONACyT fellowship. VHA. is a graduate student from the Graduate Program in Biomedical Sciences, UNAM with CONACyT fellowship.

## Availability of data and materials

All metagenomic data from *D. coccus* females, *D. coccus* males and *D. opuntiae* was deposited with links to BioProject accession numbers PRJNA291435, PRJNA658779 and PRJNA658782 in the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>). Sequences of *S. ixodetis* DCF, DCM and DO genomes were deposited at DDBJ/ENA/GenBank under accession numbers JACSE0000000000, JACSER0000000000 and JACSES0000000000, respectively. The genomic data has been deposited with links to BioProject accession number PRJNA655193 in the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>). *S. ixodetis* DCF raw metatranscriptomic reads from *D. coccus* gut, hemolymph and ovary were deposited in the NCBI Sequence Read Archive (SRA) with links to BioProject accession number PRJNA658344. A description of all software, including scripts and commands, used for the analyses in this paper can be found at <https://github.com/avera1988>.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Programa de Ecología Genómica, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Mexico. <sup>2</sup>Present Address: Faculty of Biotechnology, Chemistry and Food Science, Norwegian University of Life Sciences, 1433 Ås, Norway. <sup>3</sup>Present Address: School of Biology, Georgia Institute of Technology, Atlanta, GA, USA.

Received: 30 August 2020 Accepted: 18 March 2021

Published online: 06 April 2021

## References

- Sudakaran S, Kost C, Kaltenpoth M. Symbiont acquisition and replacement as a source of ecological innovation. *Trends Microbiol.* 2017;25(5):375–90. <https://doi.org/10.1016/j.tim.2017.02.014>.
- Cisak E, Wójcik-Fatla A, Zajac V, Sawczyn A, Sroka J, Dutkiewicz J. *Spiroplasma* – an emerging arthropod-borne pathogen? *Ann Agric Environ Med.* 2015;22(4):589–93. <https://doi.org/10.5604/12321966.1185758>.
- Bolaños LM, Servín-Garcidueñas LE, Martínez-Romero E. Arthropod–*Spiroplasma* relationship in the genomic era. *FEMS Microbiol Ecol.* 2015; 91(2):1–8. <https://doi.org/10.1093/femsec/fiu008>.
- Herren JK, Paredes JC, Schüpfer F, Lemaitre B. Vertical transmission of a drosophila endosymbiont via cooption of the yolk transport and internalization machinery. *MBio.* 2013;4(2):1–8. <https://doi.org/10.1128/mBio.00532-12>.
- Bové JM, Renaudin J, Saillard C, Foissac X, Garnier M. *Spiroplasma citri*, a plant pathogenic mollicute: relationships with its two hosts, the plant and the leafhopper vector. *Annu Rev Phytopathol.* 2003;41(1):483–500. <https://doi.org/10.1146/annurev.phyto.41.052102.104034>.

6. Gasparich GE. Spiroplasmas and phytoplasmas: microbes associated with plant hosts. *Biologicals*. 2010;38(2):193–203. <https://doi.org/10.1016/j.biologicals.2009.11.007>.
7. Mouches C, Bové J, Tully J, Rose D, McCoy R, Carle-Junca P, et al. *Spiroplasma apis*, a new species from the honey-bee *Apis mellifera*. *Ann l'Institut Pasteur / Microbiol*. 1983;134(3):383–97. [https://doi.org/10.1016/0769-2609\(83\)90013-3](https://doi.org/10.1016/0769-2609(83)90013-3).
8. Clark TB, Whitcomb RF, Tully JG, Mouches C, Saillard C, Bove JM, et al. *Spiroplasma melliferum*, a new species from the honeybee (*Apis mellifera*). *Int J Syst Bacteriol*. 1985;35(3):296–308. <https://doi.org/10.1099/00207713-35-3-296>.
9. Mateos M, Winter L, Winter C, Higareda-Alvarez VM, Martínez-Romero E, Xie J. Independent origins of resistance or susceptibility of parasitic wasps to a defensive symbiont. *Ecol Evol*. 2016;6(9):2679–87. <https://doi.org/10.1002/ece3.2085>.
10. Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ. Adaptation via symbiosis: recent spread of a drosophila defensive symbiont. *Science*. 2010;329(5988):212–5. <https://doi.org/10.1126/science.1188235>.
11. Harumoto T, Lemaître B. Male-killing toxin in a bacterial symbiont of *Drosophila*. *Nature*. 2018;557(7704):252–5. <https://doi.org/10.1038/s41586-018-0086-2>.
12. García-Arreaez MG, Masson F, Escobar JCP, Lemaître B. Functional analysis of RIP toxins from the *Drosophila* endosymbiont *Spiroplasma poulsonii*. *BMC Microbiol*. 2019;19(1):46. <https://doi.org/10.1186/s12866-019-1410-1>.
13. Ballinger MJ, Perlman SJ. Generality of toxins in defensive symbiosis: ribosome-inactivating proteins and defense against parasitic wasps in *Drosophila*. *Plos Pathog*. 2017;13:1–19.
14. Masson F, Calderon-Copete S, Schüpfer F, Vigneron A, Rommelaere S, García-Arreaez MG, et al. Blind killing of both male and female *Drosophila* embryos by a natural variant of the endosymbiotic bacterium *Spiroplasma poulsonii*. *Cell Microbiol*. 2020;22:1–12.
15. Gasparich GE, Whitcomb RF, Dodge D, French FE, Glass J, Williamson DL. The genus *Spiroplasma* and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the *Mycoplasmatales* clade. *Int J Syst Evol Microbiol*. 2004;54(3):893–918. <https://doi.org/10.1099/ijs.0.02688-0>.
16. Yeoman CJ, Brutscher LM, Esen ÖC, Ibaoglu F, Fowler C, Eren AM, et al. Genome-resolved insights into a novel *Spiroplasma* symbiont of the wheat stem sawfly (*Cephus cinctus*). *PeerJ*. 2019;7:e7548. <https://doi.org/10.7717/peerj.7548>.
17. Martin SH, Singh KS, Gordon IJ, Omufwoko KS, Collins S, Warren IA, et al. Whole-chromosome hitchhiking driven by a male-killing endosymbiont. *Plos Biol*. 2020;18(2):e3000610. <https://doi.org/10.1371/journal.pbio.3000610>.
18. Binetruy F, Bailly X, Chevillon C, Martin OY, Bernasconi MV, Duron O. Phylogenetics of the *Spiroplasma ixodeti* endosymbiont reveals past transfers between ticks and other arthropods. *Ticks Tick Borne Dis*. 2019;10(3):575–84. <https://doi.org/10.1016/j.ttbdis.2019.02.001>.
19. Machenaud J, Henri R, Dieuaide-Noubhani M, Pracros P, Renaudin J, Eveillard S. Gene expression and enzymatic activity of invertases and sucrose synthase in *Spiroplasma citri* or stolbur phytoplasma infected plants. *Bull Insectology*. 2007;60:219–20.
20. Dubrana M-P, Béven L, Aricaou-Bouvery N, Duret S, Claverol S, Renaudin J, et al. Differential expression of *Spiroplasma citri* surface protein genes in the plant and insect hosts. *BMC Microbiol*. 2016;16(1):53. <https://doi.org/10.1186/s12866-016-0666-y>.
21. Lo W-S, Kuo C-H. Horizontal acquisition and transcriptional integration of novel genes in mosquito-associated *Spiroplasma*. *Genome Biol Evol*. 2017;9(12):3246–59. <https://doi.org/10.1093/gbe/evx244>.
22. Masson F, Calderon Copete S, Schüpfer F, García-Arreaez G, Lemaître B. In vitro culture of the insect endosymbiont *Spiroplasma poulsonii* highlights bacterial genes involved in host-symbiont interaction. *MBio*. 2018;9(2):1–11. <https://doi.org/10.1128/mBio.00024-18>.
23. Chávez-Moreno CK, Tecante A, Casas A. The *Opuntia* (Cactaceae) and *Dactylopius* (Hemiptera: Dactylopiidae) in Mexico: a historical perspective of use, interaction and distribution. *Biodivers Conserv*. 2009;18(13):3337–55. <https://doi.org/10.1007/s10531-009-9647-x>.
24. Deveoglu O, Karadag R, Yurdun T. Qualitative HPLC determination of main anthraquinone and lake pigment contents from *Dactylopius coccus* dye insect. *Chem Nat Compd*. 2011;47(1):103–4. <https://doi.org/10.1007/s10600-011-9842-3>.
25. Ramírez-Puebla ST, Rosenblueth M, Chávez-Moreno CK, Catanho Pereira de Lyra MC, Tecante A, Martínez-Romero E. Molecular phylogeny of the genus *Dactylopius* (Hemiptera: Dactylopiidae) and identification of the symbiotic bacteria. *Environ Entomol*. 2010;39(4):1178–83. <https://doi.org/10.1603/EN10037>.
26. Van Dam AR, Martínez LP, Chavez AJ, May BP. Range wide phylogeography of *Dactylopius coccus* (Hemiptera: Dactylopiidae). *Ann Entomol Soc Am*. 2015;108:299–310.
27. Pankewitz F, Zöllmer A, Hilker M, Gräser Y. Presence of *Wolbachia* in insect eggs containing antimicrobially active anthraquinones. *Microb Ecol*. 2007;54(4):713–21. <https://doi.org/10.1007/s00248-007-9230-5>.
28. Ramírez-Puebla ST, Servín-Garcidueñas LE, Ormeño-Orrillo E, Vera-Ponce de León A, Rosenblueth M, Delaye L, et al. Species in *Wolbachia*? Proposal for the designation of "*Candidatus Wolbachia bourtziisii*", "*Candidatus Wolbachia onchocercicola*", "*Candidatus Wolbachia blaxteri*", "*Candidatus Wolbachia brugii*", "*Candidatus Wolbachia taylori*", "*Candidatus Wolbachia collembol*". *Syst Appl Microbiol*. 2015;38(6):390–9. <http://dx.doi.org/10.1016/j.syapm.2015.05.005>.
29. Ramírez-Puebla ST, Ormeño-Orrillo E, Vera-Ponce de León A, Lozano L, Sanchez-Flores A, Rosenblueth M, et al. Genomes of *Candidatus Wolbachia bourtziisii* wDacA and *Candidatus Wolbachia pipientis* wDacB from the cochineal insect *Dactylopius coccus* (Hemiptera: Dactylopiidae) G3. 2016;6(10):3343–9. <https://doi.org/10.1534/g3.116.031237>.
30. Vera-Ponce de León A, Ormeño-Orrillo E, Ramírez-Puebla ST, Rosenblueth M, Degli Esposti M, Martínez-Romero J, et al. *Candidatus Dactylopiibacterium carminicum*, a nitrogen-fixing symbiont of *Dactylopius* cochineal insects (Hemiptera: Coccoidea: Dactylopiidae). *Genome Biol Evol*. 2017;9(9):2237–50. <https://doi.org/10.1093/gbe/evx156>.
31. Bustamante-Brito R, Vera-Ponce de León A, Rosenblueth M, Martínez-Romero J, Martínez-Romero E. Metatranscriptomic analysis of the bacterial symbiont *Dactylopiibacterium carminicum* from the carmine cochineal *Dactylopius coccus* (Hemiptera: Coccoidea: Dactylopiidae). *Life*. 2019;9:4. doi: <https://doi.org/10.3390/life9010004>.
32. 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. *Front Microbiol*. 2016;7:954. <https://doi.org/10.3389/fmicb.2016.00954>.
33. Lo WS, Gasparich GE, Kuo CH. Convergent evolution among ruminant-pathogenic mycoplasma involved extensive gene content changes. *Genome Biol Evol*. 2018;10(8):2130–9. <https://doi.org/10.1093/gbe/evy172>.
34. Gerth M, Martínez-Montoya H, Ramírez P, Masson F, Griffin J, Aramayo R, et al. Rapid molecular evolution of *Spiroplasma* symbionts of *Drosophila*. 2020. bioRxiv doi:<https://doi.org/10.1101/2020.06.23.165548>
35. Campana MG, Robles García NM, Tuross N. America's red gold: multiple lineages of cultivated cochineal in Mexico. *Ecol Evol*. 2015;5(3):607–17. <https://doi.org/10.1002/ece3.1398>.
36. Kageyama D, Anbutsu H, Watada M, Hosokawa T, Shimada M, Fukatsu T. Prevalence of a non-male-killing *Spiroplasma* in natural populations of *Drosophila hydei*. *Appl Environ Microbiol*. 2006;72(10):6667–73. <https://doi.org/10.1128/AEM.00803-06>.
37. Watts T, Haselkorn TS, Moran NA, Markow TA. Variable incidence of *Spiroplasma* infections in natural populations of *Drosophila* species. *Plos One*. 2009;4(5):e5703. <https://doi.org/10.1371/journal.pone.0005703>.
38. Blow F, Douglas AE. The hemolymph microbiome of insects. *J Insect Physiol*. 2018;2019(115):33–9. <https://doi.org/10.1016/j.jinsphys.2019.04.002>.
39. Duret S, André A, Renaudin J. Specific gene targeting in *Spiroplasma citri*: improved vectors and production of unmarked mutations using site-specific recombination. *Microbiology*. 2005;151(8):2793–803. <https://doi.org/10.1099/mic.0.28123-0>.
40. Paredes JC, Herren JK, Schüpfer F, Marin R, Claverol S, Kuo C-H, et al. Genome sequence of the *Drosophila melanogaster* male-killing *Spiroplasma* strain MSRO endosymbiont. *MBio*. 2015;6(2):1–12. <https://doi.org/10.1128/mBio.02437-14>.
41. Lo W-S, Chen L-L, Chung W-C, Gasparich GE, Kuo C-H. Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. *BMC Genomics*. 2013;14(1):22. <https://doi.org/10.1186/1471-2164-14-22>.
42. Evans JD, Schwarz RS. Bees brought to their knees: microbes affecting honeybee health. *Trends Microbiol*. 2011;19(12):614–20. <https://doi.org/10.1016/j.tim.2011.09.003>.
43. Ku C, Lo W-S, Chen L-L, Kuo C-H. Complete genomes of two dipteran-associated *Spiroplasmas* provided insights into the origin, dynamics, and impacts of viral invasion in *Spiroplasma*. *Genome Biol Evol*. 2013;5(6):1151–64. <https://doi.org/10.1093/gbe/evt084>.

44. Al-Khodor S, Price CT, Kalia A, Abu KY. Functional diversity of ankyrin repeats in microbial proteins. *Trends Microbiol.* 2010;18(3):132–9. <https://doi.org/10.1016/j.tim.2009.11.004>.
45. Hamilton PT, Peng F, Boulanger MJ, Perlman SJ. A ribosome-inactivating protein in a *Drosophila* defensive symbiont. *Proc Natl Acad Sci.* 2016;113(2):350–5. <https://doi.org/10.1073/pnas.1518648113>.
46. Ballinger MJ, Perlman SJ. The defensive *Spiroplasma*. *Curr Opin Insect Sci.* 2019;32:36–41. <https://doi.org/10.1016/j.cois.2018.10.004>.
47. Duret S, Berho N, Danet JL, Garnier M, Renaudin J. Spiralin is not essential for helicity, motility, or pathogenicity but is required for efficient transmission of *Spiroplasma citri* by its leafhopper vector *Circulifer haematoceps*. *Appl Environ Microbiol.* 2003;69(10):6225–34. <https://doi.org/10.1128/AEM.69.10.6225-6234.2003>.
48. Mouches C, Candresse T, Barroso G, Saillard C, Wroblewski H, Bové JM. Gene for spiralin, the major membrane protein of the helical mollicute *Spiroplasma citri*: cloning and expression in *Escherichia coli*. *J Bacteriol.* 1985;164(3):1094–9. <https://doi.org/10.1128/JB.164.3.1094-1099.1985>.
49. Berg M, Melcher U, Fletcher J. Characterization of *Spiroplasma citri* adhesion related protein SARP1, which contains a domain of a novel family designated sarpin. *Gene.* 2001;275(1):57–64. [https://doi.org/10.1016/S0378-1119\(01\)00655-2](https://doi.org/10.1016/S0378-1119(01)00655-2).
50. Duret S, Batailler B, Dubrana MP, Saillard C, Renaudin J, Béven L, et al. Invasion of insect cells by *Spiroplasma citri* involves spiralin relocalization and lectin/glycoconjugate-type interactions. *Cell Microbiol.* 2014;16(7):1119–32. <https://doi.org/10.1111/cmi.12265>.
51. Labroussaa F, Arricau-Bouvery N, Dubrana M-P, Saillard C. Entry of *Spiroplasma citri* into *Circulifer haematoceps* cells involves interaction between *Spiroplasma* Phosphoglycerate kinase and leafhopper actin. *Appl Environ Microbiol.* 2010;76(6):1879–86. <https://doi.org/10.1128/AEM.02384-09>.
52. Labroussaa F, Dubrana M-P, Arricau-Bouvery N, Béven L, Saillard C. Involvement of a minimal actin-binding region of *Spiroplasma citri* phosphoglycerate kinase in spiroplasma transmission by its leafhopper vector. *Plos One.* 2011;6(2):e17357. <https://doi.org/10.1371/journal.pone.0017357>.
53. Rice DW, Sheehan KB, Newton ILG. Large-scale identification of *Wolbachia pipientis* effectors. *Genome Biol Evol.* 2017;9(7):1925–37. <https://doi.org/10.1093/gbe/evx139>.
54. Blötz C, Stülke J. Glycerol metabolism and its implication in virulence in *Mycoplasma*. *FEMS Microbiol Rev.* 2017;41(5):640–52. <https://doi.org/10.1093/femsre/fux033>.
55. Bischof DF, Janis C, Vilei EM, Bertoni G, Frey J. Cytotoxicity of *Mycoplasma mycoides* subsp. *mycoides* small colony type to bovine epithelial cells. *Infect Immun.* 2008;76(1):263–9. <https://doi.org/10.1128/IAI.00938-07>.
56. Vilei EM, Frey J. Genetic and biochemical characterization of glycerol uptake in *Mycoplasma mycoides* subsp. *mycoides* SC: its impact on H<sub>2</sub>O<sub>2</sub> production and virulence. *Clin Diagnostic Lab Immunol.* 2001;8(1):85–92. <https://doi.org/10.1128/CDLI.8.1.85-92.2001>.
57. Chang T-H, Lo W-S, Ku C, Chen L-L, Kuo C-H. Molecular evolution of the substrate utilization strategies and putative virulence factors in mosquito-associated *Spiroplasma* species. *Genome Biol Evol.* 2014;6(3):500–9. <https://doi.org/10.1093/gbe/evu033>.
58. Tseng TT, Tyler BM, Setubal JC. Protein secretion systems in bacterial-host associations, and their description in the gene ontology. *BMC Microbiol.* 2009;9(SUPPL. 1):1–9.
59. Bhattacharya T, Newton ILG. Mi casa es su casa: how an intracellular symbiont manipulates host biology. *Environ Microbiol.* 2019;21(9):3188–96. <https://doi.org/10.1111/1462-2920.13964>.
60. Arredondo-Alonso S, Willems RJ, van Schaik W, Schürch AC. On the (im) possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb Genomics.* 2017;3(10):e000128. <https://doi.org/10.1099/mgen.0.000128>.
61. Davis RE, Dally EL, Jomantiene R, Zhao Y, Roe B, Lin S, et al. Cryptic plasmid pSKU146 from the wall-less plant pathogen *Spiroplasma kunkelii* encodes an adhesin and components of a type IV translocation-related conjugation system. *Plasmid.* 2005;53(2):179–90. <https://doi.org/10.1016/j.plasmid.2004.09.002>.
62. Saillard C, Carle P, Duret-Nurbel S, Henri R, Killiny N, Carrère S, et al. The abundant extrachromosomal DNA content of the *Spiroplasma citri* GIL3-3X genome. *BMC Genomics.* 2008;9(1):195. <https://doi.org/10.1186/1471-2164-9-195>.
63. Hester CM, Lutkenhaus J. Soj (ParA) DNA binding is mediated by conserved arginines and is essential for plasmid segregation. *Proc Natl Acad Sci.* 2007;104(51):20326–31. <https://doi.org/10.1073/pnas.0705196105>.
64. Dunham TD, Xu W, Funnell BE, Schumacher MA. Structural basis for ADP-mediated transcriptional regulation by P1 and P7 ParA. *EMBO J.* 2009;28(12):1792–802. <https://doi.org/10.1038/emboj.2009.120>.
65. Britton RA, Lin DC-H, Grossman AD. Characterization of a prokaryotic SMC protein involved in chromosome partitioning. *Genes Dev.* 1998;12(9):1254–9. <https://doi.org/10.1101/gad.12.9.1254>.
66. Guglielmini J, Quintais L, Garcillán-Barcia MP, de la Cruz F, Rocha EPC. The repertoire of ice in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *Plos Genet.* 2011;7(8):e1002222. <https://doi.org/10.1371/journal.pgen.1002222>.
67. Nelson MS, Sadowsky MJ. Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes. *Front Plant Sci.* 2015:1–11. doi:<https://doi.org/10.3389/fpls.2015.00491>.
68. Rancès E, Voronin D, Tran-Van V, Mavingui P. Genetic and functional characterization of the type IV secretion system in *Wolbachia*. *J Bacteriol.* 2008;190(14):5020–30. <https://doi.org/10.1128/JB.00377-08>.
69. Venugopal A, Bryk R, Shi S, Rhee K, Rath P, Schnappinger D, et al. Virulence of *Mycobacterium tuberculosis* depends on lipamide dehydrogenase, a member of three multienzyme complexes. *Cell Host Microbe.* 2011;9(1):21–31. <https://doi.org/10.1016/j.chom.2010.12.004>.
70. Moné Y, Monnin D, Kremer N. The oxidative environment: a mediator of interspecies communication that drives symbiosis evolution. *Proc R Soc B Biol Sci.* 2014;281(1785):20133112. <https://doi.org/10.1098/rspb.2013.3112>.
71. Wang Y, Dufour YS, Carlson HK, Donohue TJ, Marletta MA, Ruby EG. H-NOX-mediated nitric oxide sensing modulates symbiotic colonization by *Vibrio fischeri*. *Proc Natl Acad Sci.* 2010;107(18):8375–80. <https://doi.org/10.1073/pnas.1003571107>.
72. Brennan LJ, Keddle BA, Braig HR, Harris HL. The endosymbiont *Wolbachia pipientis* induces the expression of host antioxidant proteins in an *Aedes albopictus* cell line. *Plos One.* 2008;3(5):e2083. <https://doi.org/10.1371/journal.pone.0002083>.
73. Anbutsu H, Fukatsu T. Tissue-specific infection dynamics of female-killing and nonmale-killing spiroplasmas in *Drosophila melanogaster*. *FEMS Microbiol Ecol.* 2006;57(1):40–6. <https://doi.org/10.1111/j.1574-6941.2006.00087.x>.
74. Charles H, Ishikawa H. Physical and genetic map of the genome of *Buchnera*, the primary endosymbiont of the pea aphid *Acyrtosiphon pisum*. *J Mol Evol.* 1999;48(2):142–50. <https://doi.org/10.1007/PL00006452>.
75. Langmead B, Slazberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2013;9:357–9.
76. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
77. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics.* 2012;28(11):1420–8. <https://doi.org/10.1093/bioinformatics/bts174>.
78. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. MaxBin: an automated binning method to recover individual genomes from metagenomes using. *Microbiome.* 2014;2(13):4904–9. <https://doi.org/10.1073/pnas.1402564111>.
79. Imelfort M, Skennerton CT, Parks DH, Tyson GW, Hugenholtz P. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25:1043–55.
80. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics.* 2011;27(4):578–9. <https://doi.org/10.1093/bioinformatics/btq683>.
81. Hunt M, De SN, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol.* 2015;16(1):294. <https://doi.org/10.1186/s13059-015-0849-0>.
82. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10(1):421. <https://doi.org/10.1186/1471-2105-10-421>.
83. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11(1). <https://doi.org/10.1186/1471-2105-11-119>.
84. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30(14):2068–9. <https://doi.org/10.1093/bioinformatics/btu153>.
85. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol.* 2016;428(4):726–31. <https://doi.org/10.1016/j.jmb.2015.11.006>.
86. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 2018;46:W95–101. <https://doi.org/10.1093/nar/gky418>.



87. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016; 44(14):6614–24. <https://doi.org/10.1093/nar/gkw569>.
88. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013; 30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
89. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14(6):587–9. <https://doi.org/10.1038/nmeth.4285>.
90. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37(5):1530–4. <https://doi.org/10.1093/molbev/msaa015>.
91. Contreras-Moreira B, Vinuesa P. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol.* 2013;79(24):7696–701. <https://doi.org/10.1128/AEM.02411-13>.
92. Li L. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003;13(9):2178–89. <https://doi.org/10.1101/gr.1224503>.
93. Vera-Ponce de León A, Jahnes BC, Duan J, Camuy-Vélez LA, Sabree ZL. Cultivable, Host-specific bacteroidetes symbionts exhibit diverse polysaccharolytic strategies. *Appl Environ Microbiol.* 2020;86. <https://doi.org/10.1128/AEM.00091-20>.
94. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr.* 2016;4:e1900v1. <https://doi.org/10.7287/peerj.preprints.1900v1>.
95. Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EPC. Identification of protein secretion systems in bacterial genomes. *Sci Rep.* 2016;6(1):23080. <https://doi.org/10.1038/srep23080>.
96. Eddy SR. Accelerated profile HMM searches. *Plos Comput Biol.* 2011;7(10): e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
97. Guerrero-Castro J, Lozano L, Sohlenkamp C. Dissecting the acid stress response of *Rhizobium tropici* CIAT 899. *Front Microbiol.* 2018;1–14. doi: <https://doi.org/10.3389/fmicb.2018.00846>.
98. Bengtsson-Palme J, Hartmann M, Eriksson KM, Pal C, Thorell K, Larsson DGJ, et al. metaxa2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol Ecol Resour.* 2015;15: 1403–14. <https://doi.org/10.1111/1755-0998.12399>.
99. Kim D, Langmead B, Salzberg SL. HISAT : a fast spliced aligner with low memory requirements; 2015. p. 12.
100. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12(1):323. <https://doi.org/10.1186/1471-2105-12-323>.
101. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat Protoc.* 2013;8(8):1494–512. <https://doi.org/10.1038/nprot.2013.084>.
102. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* 2015;4: 1521. <https://doi.org/10.12688/f1000research.7563.2>.
103. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>.
104. Yoav Benjamini, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Statist Soc B.* 1995; 57:289–300.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

