SPIROPLASMA AND ITS INTERACTION WITH *DROSOPHILA*: GENOME SEQUENCING AND ANALYSIS OF POTENTIAL FITNESS EFFECTS OF NATURALLY INFECTED POPULATIONS

A Dissertation

by

HUMBERTO MARTINEZ MONTOYA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Mariana Mateos
Committee Members,	Rodolfo Aramayo
	Charles Criscione
	Cecilia Tamborindeguy
Head of Department,	Michael P. Masser

May 2017

Major Subject: Wildlife and Fisheries Sciences

ABSTRACT

Intimate associations of maternally-inherited endosymbiotic bacteria and arthropods are ancient and taxonomically diverse. The nature of such relationships includes reproductive parasitism, as well as nutritional and defensive mutualisms. *Spiroplasma* are a wall-less bacterium (Class Mollicutes) associated with plants and arthropods. The genus *Spiroplasma* includes several strains that are heritable endosymbionts of several *Drosophila* species and other insects. At least four clades of *Spiroplasma* independently invaded *Drosophila*: Poulsonii, Citri Tenebrosa and Ixodetis. *Spiroplasma* frequencies are relatively high in certain populations of several *Drosophila* species as observed in the defensive strains associated to *D. hydei*, *D. neotestacea D. melanogaster*. Several Poulsonii strains act as reproductive manipulators and/or as defensive partners against parasitic wasps and/or nematodes. Herein, we present a draft genome of the defensive symbiont *Spiroplasma poulsonii* hyd1, a strain that enhances survival of *Drosophila hydei* flies against parasitism by the wasp *Leptopilina heterotoma* (Lh14). Additionally, we present a draft genome of the citri clade strain *Spiroplasma citri* moj, associated to *D. hydei* and *D. mojavensis*. Comparison with other available *Spiroplasma* genomes either associated to other *Drosophila* species or more distant arthropods allowed us to identify shared and unique genes and reveal interesting aspects about their metabolic capacity and potential factors involved in their phenotypes. Additionally, we performed phylogenomic analyses, which supported some of the

previously hypothesized relationships but revealed incongruences regarding the validity of certain clades.

On the other hand, members of the Citri clade of *Spiroplasma* which are restricted to the repleta species group of *Drosophila*, can also reach relatively high infection prevalence. No evidence of fitness benefits or reproductive manipulation by members of this clade has been reported to date, although our preliminary work suggests that *Spiroplasma* from this clade do not confer protection. In the present study, we examined whether two citri-clade strains (moj and ald2) confer protection to their respective natural host species against two parasitic wasps: the cosmopolitan generalist *Leptopilina heterotoma* (Lh14; Figitidae); and *Asobara* sp. (Aw35; Braconidae; from Texas). Additionally, we tested whether the citri-clade strains (moj and hyd2; the latter harbored by *D. hydei*) induce reproductive phenotypes. We assessed oviposition rate in both strains and cytoplasmic incompatibility for moj by examining the outcome of reciprocal crosses of *Spiroplasma*-infected and *Spiroplasma*-free individuals. Cytoplasmic incompatibility (CI) was also tested; CI is the embryonic failure of the progeny from crosses between an uninfected female and an infected male, or between a female and male carrying incompatible symbiont strains. Although this reproductive parasitism has been observed in *D. simulans* and *D. melanogaster* infected with *Wolbachia,* we do not find evidence of CI induced by *Spiroplasma* in *Drosophila* hosts.

DEDICATION

To Xochitl F., Humberto F. and Xalli D.

AKNOWLEDGEMENTS

I would like to thank all people involved in this project. Special thanks to my committee chair, Dr. Mariana Mateos. I would like to thank to all committee members, Dr. Rodolfo Aramayo, Dr. Charles Criscione and Dr. Cecilia Tamborindeguy for their valuable feedback and advise.

Thanks to the institutions that supported our research with fellowships and research grants; the Wildlife and Fisheries Sciences Department, *El Consejo Tamaulipeco de Ciencia y Tecnologia* (COTACyT) and the Mexican *Consejo Nacional de Ciencia y Tecnologia* (CONACyT). I am so grateful with the institutions that support my PhD studies, including the Tom Slick Fellowship, the joint research grant provided by TAMU-CONACyT and the TAMU Genomics Seed Grant, the Ecology and Evolutionary Biology travel grant. Thanks to the Food and Agriculture Organization (FAO) and International Atomic Energy Agency (IAEA) for the funds given for research and travel through the Coordinated Research Project "Use of Symbiotic Bacteria to Reduce Mass-Rearing Costs and Increase Mating Success in Selected Fruit Pests in Support of SIT Application" given to Mariana Mateos.

We would like to thank Texas A&M Institute for Genome Sciences and Society for the valuable computing resources used in our genome assemblies. We also want to express our gratitude to the University of Texas at Austin, Texas State University and the University of Southern California for allowing us to collect samples in their facilities. Thanks to El Colegio de la Frontera Sur (ECOSUR), Dr. Pablo Liedo, Dr.

Silvia Cepeda (MOSCAFRUT) and Dr. Jorge Toledo for their support in our *Wolbachia* project.

Finally, I would like to thank all the undergrad, grad students and amazing people who share with me their experience and friendship during my time as a PhD student; Dr. Jialei Xie, Dr. Nadisha Silva, Paul Ramirez, Caitlyn Winter, Lauryn Winter, Mildred Wong, Young Zheng, Vojtech Tichavsky, Laura Carlton, Michael Liston, Brittany Witt, Lorena Interiano, Katharina Moeller, Alejandra Ibarra, Jack Harling, Brandy Chellette, Brenda Moran, Angeles Palomeque and Santiago Izaguirre. Many thanks to my friends for all their support in the good and bad times, Israel Estrada, Iván Rodríguez and Saúl Rodríguez (*Gracias por las Tepoli, keep brewing that good stuff!*).

NOMENCLATURE

CI	Cytoplasmic Incompatibility
HPC	High Performance Research Computing
MK	Male Killer
NGS	Next Generation Sequencing
MSRO	Melanogaster Sex Ratio Organism
NSRO	Nebulosa Sex Ratio Organism
PacBio	Pacific Biosciences
PCR	Polymerase Chain Reaction
SMRT	Single Molecule, Real-Time Sequencing
WSRO	Willistoni Sex Ratio Organism

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Natural associations of maternally-inherited endosymbiotic bacteria and arthropods are pervasive and taxonomically diverse (Moran et al. 2008), and involve a variety of interactions, including reproductive parasitism, nutritional mutualism, and defensive mutualism (Clark et al. 2010). Due to the heritable nature of these associations, the fitness of the endosymbiont is tightly coupled to that of its female host. Numerous heritable associations are obligate, in that both partners are completely dependent of each other for survival and reproduction. Consequently, prevalence of such endosymbionts in natural host populations is 100%. Obligate symbionts, also known as primary endosymbionts or P-symbionts (Moran et al. 2008), are often associated with specialized host cells (e.g. bacteriocytes), and typically occur in hosts that have nutritionally poor diets, such as those that feed on plant sap or cellulose (Ferrari and Vavre 2011). It is estimated that ~10% of insect species harbor obligate mutualistic endosymbionts that provide essential nutrients to their host (Douglas 1989, 2009). Many of these obligate associations are regarded as ancient, based on a long phylogenetic history of co-divergence. For example, the association of *Buchnera aphidicola* with aphids is estimated at ~160–180 Mya (Clark et al. 2000), whereas the association between *Wigglesworthia glossinidia* and tsetse flies is estimated at ~50–80 Mya (Chen et al. 1999).

In contrast to obligate endosymbionts, facultative heritable symbionts (also referred to as secondary or S-symbionts) are not absolutely essential for host survival and reproduction (Pontes and Dale 2006), and may induce different effects on their hosts ranging from nutritional or defensive mutualisms (Clark et al. 2010) to reproductive parasitism, e.g. cytoplasmic incompatibility in *Ceratitis capitata* by *Wolbachia* (Zabalou et al. 2004). Facultative endosymbionts are not uniformly distributed in their hosts and can be found in the gut or body cavities, depending of their nature. In some cases, however, they may invade the bacteriocytes where they can coexist with, or even exclude, obligate endosymbionts (Moran et al. 2008). Additionally, facultative endosymbionts typically exhibit imperfect maternal transmission (Moran et al. 2008). They may infect new hosts and, in some cases, establish novel vertically transmitted infections, as observed in *Wolbachia*, which shows lack of phylogenetic congruence with its hosts, as a result of frequent horizontal transmission and establishment in new (sometimes distantly related) host lineages (Werren et al. 2008).

Despite well-known examples, most effects of facultative endosymbionts on their hosts are not well understood. Their persistence in natural populations with imperfect maternal transmission requires that the production of infected females in a population be greater than the production of uninfected females (Bull 1983). To achieve this goal, many facultative endosymbionts manipulate their host reproduction to enhance their transmission and consequently their own survival. Several forms of reproductive parasitism exist: parthenogenesis; male feminization; male killing; and cytoplasmic incompatibility (CI).

Not all facultative endosymbionts cause reproductive parasitism, and a growing number of studies indicate that many facultative endosymbionts confer a fitness benefit to their host through nutritional or defensive mechanisms. For example, the facultative symbiont *Nardonella* enhances the growth and development of its weevil host *Euscepes postfasciatus* (Hosokawa et al. 2010), and the bacteriocyte-associated *Wolbachia* confers nutritional benefits to its bedbug host *Cimex lectularius* (Hosokawa et al. 2010; Kuriwada et al. 2010). Similarly the *w*Mel strain of *Wolbachia* is reported to improve fecundity *Drosophila melanogaster* during high or low diet iron conditions (Brownlie et al. 2009).

In contrast to nutritional mutualisms, defensive symbioses are interactions that require at least three species: host; symbiont; and enemy (Clay 2014). To persist under this scheme, bacterial symbionts must improve substantially their host fitness compared to that of non-infected hosts, in the presence of natural enemies (Hussain et al. 2013). Defensive mutualistic associations reported to date involve taxonomically diverse symbiotic bacteria, hosts, and natural enemies; the latter range from predators to parasitic/pathogenic wasps, nematodes, viruses, and fungi. For example, the Gammaproteobacterium *Hamiltonella defensa* protects the pea aphid against the parasitic wasp *Aphidius ervi* (Braconidae) (Oliver et al. 2003). Similarly, certain strains of the widespread bacterial symbiont *Wolbachia* (Alphaproteobacteria) confer protection against RNA viruses in *Drosophila* flies (Hedges et al. 2008), and reduce replication and transmission of Dengue (Bull and Turelli 2013) and West Nile viruses by mosquitoes (Hussain et al. 2013). In contrast, recent evidence in natural *Wolbachia* infected *Cx.*

pipiens lines suggest that long-term associations of host and *Wolbachia* may improve the role of *Cx. pipiens* as vector for malaria compared to non-infected lines (Zele et al. 2014). Three strains of the only other heritable bacterial genus associated with *Drosophila* (i.e., *Spiroplasma*; class Mollicutes) are reported to confer protection against parasitoid wasps or nematodes in *D. melanogaster*, *D. neotestacea* and *D. hydei* (Jaenike et al. 2010; Xie et al. 2010; Xie et al. 2014; Haselkorn and Jaenike 2015), but whether or not other strains of *Spiroplasma* that associate with *Drosophila* are defensive is not known.

The genus *Spiroplasma* belongs to the class Mollicutes, which are the simplest self-replicating organisms known, and are characterized by small genomes with low GC contents (Carle et al. 1995). Although they lack a cell wall, the class Mollicutes falls within the Gram-positive clade (Woese 1987; Razin et al. 1998). Their small genomes lack genes coding for basic metabolic pathways (Ammar and Hogenhout 2006; Harris et al. 2010). *Spiroplasma* strains are associated intra- and extra-cellularly with a variety of plants and arthropods, including several species of *Drosophila* (Anbutsu and Fukatsu 2011). Nineteen species of *Drosophila* are known to harbor members of *Spiroplasma* (Figure 1). *Spiroplasma* strains that associate with *Drosophila* fall within four separate clades (i.e., ixodetis; tenebrosa; poulsonii; and citri, (Haselkorn 2010a) that likely represent independent invasions. The phylogenetic patterns of *Drosophila* and *Spiroplasma* suggest several instances of horizontal transmission among distantly related hosts (e.g. members of the poulsonii clade are found in divergent members of *Drosophila*). Of note, although the citri clade of *Spiroplasma* (as defined in Figure 1)

has members that do not associate with *Drosophila* (e.g. horizontally transmitted plant and bee pathogens), the four citri-clade strains harbored by *Drosophila* (i.e., moj, hyd2, ald1, ald2) are exclusively found in members of the repleta species group, a monophyletic group comprised of mostly cactophilic *Drosophila* (Markow and O'Grady 2006).

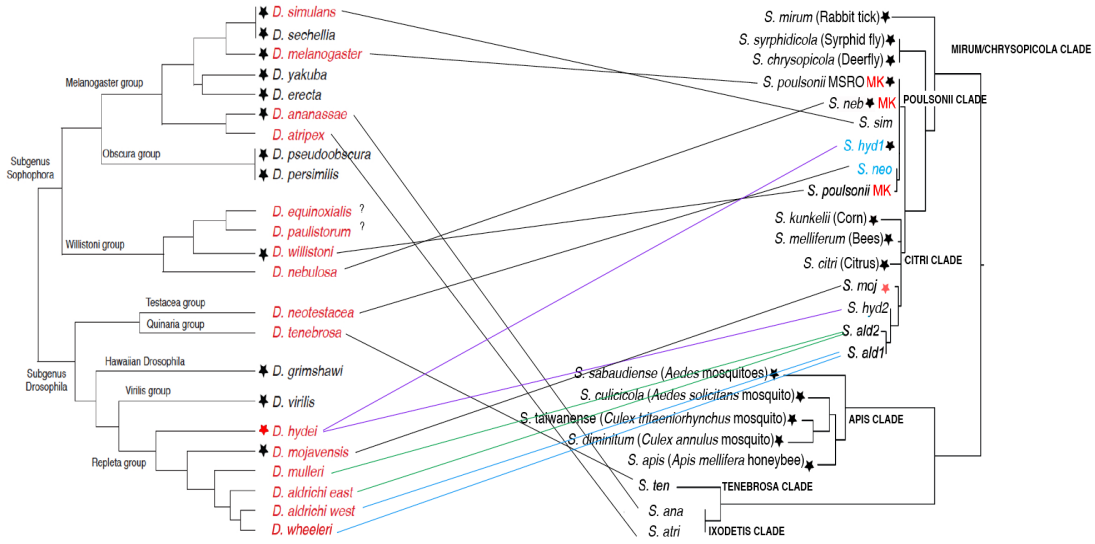


Figure 1. Phylogenetic tree of representative species of *Drosophila* and 16S phylogeny of *Spiroplasma.* Left: Phylogenetic tree of representative species of *Drosophila,* reviewed in Mateos et al. (2006). In red font, *Drosophila* species naturally infected with *Spiroplasma*. Right: 16S Phylogenetic tree of *Spiroplasma*. In blue font, strains known to give protection to their respective *Drosophila* host; MK indicates male killing strains. In both trees, black star indicates complete genome sequence available, red star indicates genome sequencing in process. Lines connect host species and symbiont species/strains. Colored lines indicate one host species associated with two divergent symbionts (purple) or one Spiroplasma strain associated with two host species (blue and green).

Although the effects of the majority of *Spiroplasma* strains that associate with *Drosophila* remain unknown, the phenotypes reported to date are diverse and may

explain persistence of *Spiroplasma* in natural populations of *Drosophila*. The known phenotypic effects include reproductive manipulation in the form of male killing, defensive mutualism against parasitic wasps and nematodes, and other more subtle effects on lifespan and timing of reproductive functions (Malogolowkin-Cohen and Rodrigues-Pereira 1975; Haselkorn 2010a; Jaenike et al. 2010; Xie et al. 2010; Anbutsu and Fukatsu 2011; Haselkorn and Jaenike 2015). The phylogenetic distribution of *Spiroplasma* phenotypes reveals two preliminary patterns. Firstly, all of the male-killing strains that have been genetically characterized to date (e.g. from *D. nebulosa*, *D. melanogaster* and *D. willistoni*) fall within the poulsonii clade, but not all strains in this clade are male killers. Lack of genetic data precludes determination of the phylogenetic positions of the male-killing strains found in *D. paulistorum*, *D. equinoxialis*, *D. ornatifrons*, *D. neocardini*, and *D. paraguayensis*. Secondly, all of the defensive strains known to date also belong to the poulsonii clade. Chapter 3 of this dissertation is aimed at elucidating aspects of the phenotypes exerted by members of the citri clade, including a newly discovered strain.

Few aspects of the mechanisms by which *Spiroplasma* exerts its phenotypic effects have been elucidated. The male killing mechanism occurs during embryonic development and appears to exploit the host machinery utilized for dosage compensation (Harumoto et al. 2014; Cheng et al. 2016). Nonetheless, the *Spiroplasma* encoded elements involved in this process have not been identified. Recent studies have revealed that the defensive mechanism against nematodes (and possibly wasps) might involve a *Spiroplasma*-encoded toxin (Hamilton et al. 2016). Understanding of these mechanisms

and of the evolutionary history of *Spiroplasma* would be greatly aided by availability of sequenced genomes. Chapters 2 and the Supplementary Chapter were aimed at sequencing, assembling, and annotating the genomes of two representative strains (one from the poulsonii clade and one from the citri clade), and comparing their features to those of other *Spiroplasma* genomes.

CHAPTER II

GENOME SEQUENCING OF *SPIROPLASMA POUSLONII* HYD1, A DEFENSIVE SYMBIOTIC BACTERIUM OF *DROSOPHILA HYDEI*

2.1 Introduction

2.1.1 History of Spiroplasma *genomics*

Large plasmids and associated viruses of *S. kunkelii* and *S. citri* were the first targets of the sequence elucidation within the *Spiroplasma* genus (Renaudin et al. 1987b, a; Ye et al. 1994; Joshi et al. 2005). However, *S. citri* GII-3X had the first chromosome sequence draft reported by Carle et al. (2010). *S. citri* genome size 1.8 Mbp, representing 92% of the total chromosome size. It showed a low G+C content (26.1%), a high abundance of plasmids (10 to 14 copies per cell) and phage sequences inserted in the chromosome (Saillard et al. 2008; Carle et al. 2010); 1908 coding sequences (CDS) were found representing an overall coding density of 74%. Nevertheless, the majority of CDS in *S. citri* are hypothetical proteins with unknown function (Carle et al. 2010).

The first arthropod-associated (non-plant-pathogenic) *Spiroplasma* genome draft released was that of *S. melliferum* KC3. This assembly was based on the SOLiD 4 sequencing system (Alexeev et al. 2012). The final assembly was 1.26 Mbp; approximately 5000 CDS were predicted but the annotated CDS were 1,023, representing 81% of total nucleotide sequence. GC content in this species is 27.3%. Additionally, *S. melliferum* IPMB4A was also sequenced (Lo et al. 2013a) and shares

some properties with KC3 and GII-3X, but its genome size is smaller than its close relative KC3 and there is no evidence of extrachromosomal plasmids (Lo et al. 2013a).

In the last three years, genomic studies of the *Spiroplasma* genus have increased exponentially thanks to the development and availability of next generation sequencing (NGS) techniques such as Illumina and PacBio. Multiple draft and complete genomes of cultivable *Spiroplasma* strains belonging to three main clades (apis, citri and mirum-chrysopicola) are now available (Table 1).

Despite the large number of *Spiroplasma* genomes available, to date only four *Spiroplasma* strains associated to *Drosophila* species have been sequenced, the male killer *S. poulsonii* MSRO ("Melanogaster Sex Ratio Organism") from *D. melanogaster* collected in Uganda (Paredes et al. 2015), the defensive strain *S. poulsonii* neo from *D. neotestacea* (S. Perlman, personal communication), and two strains from *D nebulosa*; the male-killing strain *S. poulsonii* NSRO ("Nebulosa Sex Ratio Organism") and its lab-derived non-male-killing relative NSRO-A (Anbutsu et al. 2016). Only *S. poulsonii* MSROhas been publicly released, however. The main hurdle faced by the *Drosophila*-associated *Spiroplasma* genome sequencing projects is that these strains are fastidious to culture outside of their hosts (Williamson et al. 1983; Hackett et al. 1986; Williamson et al. 1999), such that host-derived sequences must be removed through bioinformatic approaches.

2.1.2 Background on strains sequenced in this study

Drosophila hydei is a cosmopolitan species belonging to the repleta species group. Two *Spiroplasma* strains have been found to be associated with this species, the Poulsonii clade hyd1 strain and the Citri clade hyd2 (Mateos et al. 2006) (Figure 1). Of the two strains, to date only hyd1 has been found to act as defensive mutualist, protecting its host against the attack of parasitoid wasps (Xie et al. 2010). Sequencing of *Drosophila*-associated defensive strains such *S. poulsonii* hyd1 and its comparison with previously sequenced *Spiroplasma* strains could contribute to address questions related to the protection mechanism at the molecular level. To date, available evidence suggests that protection could be related to one or several cytotoxic ribosomal inactivating proteins (RIP) that are found in a few members of the genus *Spiroplasma* (Hamilton et al. 2014; Hamilton et al. 2016). Additionally recent research also proposes that competition for resources between *Spiroplasma* and parasitoid wasps may contribute to the protection mechanism (Paredes et al. 2016).

Herein we present the genome draft of two *Spiroplasma* strains: *Spiroplasma* hyd1, a defensive strain associated to *Drosophila hydei*; and *Spiroplasma* moj, harbored by *D. mojavensis*, whose effects on its host remain unknown (Haselkorn 2010b) (see Chapter 3). Comparison with other available *Spiroplasma* genomes allowed us to identify shared and unique genes and reveal interesting aspects about their metabolic capacity and potential factors involved in their phenotypes. Additionally, we performed phylogenomic analyses, which supported some of the previously hypothesized relationships, but revealed incongruences regarding the validity of certain clades.

Table 1. *Spiroplasma* genome sequencing projects to date

***Spiroplasma* species**	**Host**	**Clade**	**Genome size**	**Status**	**Accesion number**	**Ref.**
S. apis	*Apis mellifera*	Apis	1.3 Mbp	Complete	CP006682	(Ku et al. 2014)
S. atrichopogonis	Biting midges	Apis	1.1 Mbp	Complete	CP011855	(Lo et al. 2015)
S. chrysopicola	Tabanid flies	Mirum/Chrysopicola	1.3 Mbp	Complete	CP005077	(Ku et al. 2013)
S. citri	Citrus leafhoppers	Citri	1.8 Mbp	Complete		(Carle et al. 2010)
S. cantharicola	*Cantharis carolinus*		1.2 Mbp	Complete	CP012622	(Lo et al. 2015)
S. culicicola	*Aedes sollicitans*	Apis	1.3 Mbp	Complete	CP006681	(Chang et al. 2014)
S. diminutum	*Culex annulus*	Apis	1.0 Mbp	Complete	CP005076	(Lo et al. 2013b)
S. eriocheiris	Chinese mitten crab	Apis	1.3 Mbp	Complete	CP011856	(Lo et al. 2015)
S. ixodetis	*Ixodes pacificus* ticks	Ixodetis	2.2 Mbp	In progress	PRJNA184743	
S. kunkelii	*Zea mays* leafhoppers	Citri	1.5 Mbp	Complete	CP010899	(Davis et al. 2015a)
S. litorale	*Tabanus nigrovittatus*	Apis	1.2mbp	Complete	CP012357	
S. melliferum	*Apis mellifera*	Citri	1.1 Mbp	Complete	AGBZ00000000	(Alexeev et al. 2012)
S. melliferum	*Apis mellifera*	Citri	1.3 Mbp	Complete	AMGI00000000	(Lo et al. 2013a)
S. mirum	Rabbit tick	Mirum/Chrysopicola	1.3 Mbp	Complete	CP002082	(Tatusova et al. 2014)
S. mojavensis	*D. mojavensis*	Citri	1.0 Mbp	Draft	MQTY00000000	This study
S. phoeniceum	Periwinkle plants	Citri	1.8 Mbp	In progress	PRJNA184752	
S. platyhelix	Mosquitoes	Ixodetis		In progress	PRJNA184742	
S. poulsonii	*D. hydei*	Poulsonii	1.5 M	Complete	JXYY00000000	This study
S. poulsonii	*D. melanogaster*	Poulsonii	1.7 Mbp	Complete	JTLV00000000	(Paredes et al. 2015)
S. sabaudiense	Mosquitoes	Apis	1.1 Mbp	Complete	CP009634	(Tatusova et al. 2014)

* Non-cultivable strain

Table 1. Continued

Spiroplasma **species**	**Host**	**Clade**	**Genome size**	**Status**	**Accesion number**	**Ref.**
S. syrphidicola	Syrphid fly	Mirum/Chrysopicola	1.1 Mbp	Complete	CP005078	(Ku et al. 2013)
S. taiwanense	Mosquitoes	Apis	1.2 Mbp	Complete	CP005074	(Lo et al. 2013b)
S. turonicum	*Haematopota sp.*	Apis	1.2 Mbp	Complete	CP012328	(Davis et al. 2015b)

2.2 Materials and Methods

2.2.1 Spiroplasma *hyd-1 infected* Drosophila *flies*

Spiroplasma hyd-1 was obtained from a laboratory isoline TEN-104-106. The isoline was constructed from a single *Drosophila* female originally collected in Central Mexico (18.91° N; -99.61° W) in 2004 (Mateos et al. 2006). Ovaries were dissected and DNA was purified following the procedure described in the commercial kit DNEasy (Qiagen, Valencia CA). Infection status was corroborated using the *Spiroplasma*-specific primers TKSSp/63F, 16STF1/16TR1 and the universal primers 27F/1492R (Table 2).

Table 2. Universal and *Spiroplasma* specific primers used to screen TEN 104-106 flies

Locus	Primer name and sequence (5'to 3')	Target group	Annealing temp. (°C)	size (bp)
16S rDNA	27F: GAG AGT TTG ATC CTG GCT CAG[a] 1492R: GGT TAC CTT GTT ACG ACT T[a]	Most bacteria	55	~1470
16S rDNA	16STF1: GGT CTT CGG ATT GTA AAG GTC TG 16STR1: GGT GTG TAC AAG ACC CGA GAA	*Spiroplasma*	65 TD* 55	~1368
16S rDNA	63F: GCC TAA TAC ATG CAA GTC GAA C[b] TKSSp: TAG CCG TGG CTT TCT GGT AA[c]	*Spiroplasma* and several Gram-positive	55	~450

[a] Lane (1991); [b] Mateos et al. (2006); [c] Fukatsu and Nikoh (2000). *Touchdown PCR protocol.

2.2.2 Library preparation and Illumina sequencing

Spiroplasma hyd-1 cells were isolated in the laboratory through mechanical separation from hemolymph obtained by piercing with a sterile needle the mesothoracic segment from ~300 individuals belonging to the infected isoline TEN104-106. Immediately after the piercing, ~35-40 flies were placed into 0.5 ml microcentrifuge tubes previously pierced in the bottom, which was placed within a 1.5 ml microcentrifuge tube containing ~20 ul phosphate buffered saline solution 1X (PBS buffer; 137 mmol NaCl, 2.7 mmol KCl, 10 mmol Na_2HPO_4, 1.8 mmol KH_2PO_4), and centrifuged at 7000 rpm (g=4.5) for 10 sec. to collect hemolymph. DNA was recovered using a chloroform-ethanol procedure (Appendix; Supplemental protocol 1) and diluted in AE buffer (Qiagen). Presence of *Spiroplasma* was confirmed again through PCR. Prior to library preparation, dsDNA quality was examined using Picogreen fluorometric system in the Texas AgriLife Genomics and Bioinformatics Services Facility (College Station, TX).

A pair-ended library was constructed by Eureka Genomics (Hercules, CA). DNA was fragmented, end repaired, A' tagged, ligated to adaptors, size-selected, and enriched with 25 cycles of PCR during which an index was incorporated to the sample. Sample preparation was performed according to Illumina's Multiplexing Sample Preparation Guide and Eureka Genomics' proprietary method. The resulting library was subjected to Illumina HiSeq 2500 sequencing at the Texas AgriLife Genomics and Bioinformatics Services Facility (College Station, TX).

2.2.3 Library preparation and PacBio sequencing

DNA extraction of ten infected TEN-104-106 *D. hydei* flies was performed through the chloroform-ethanol procedure described above, precipitated with ethanol and diluted in AE buffer (Qiagen). Genomic DNA was sheared using g-TUBE (Covaris) to 20kb fragments. The PacBio library was constructed according to standard Pacific Biosciences protocol for 20 kb libraries. The library was size-selected using Blue Pippin (Sage Science) instrument starting at 7 kb. Finished SMRTbell library average size was 16 kb measured by Fragment analyzer (Advanced Analytical Technologies, Inc). Sequencing was performed on PacBio RSII instrument on one SMRT cell with P6-C4 chemistry magbead loading and 6 hours movie. Library preparation and sequencing were performed at University of Delaware Sequencing and Genotyping Center (Newark, DE).

2.2.4 Genome assembly

Illumina and PacBio reads were assembled separately. However, the final draft assembly was the result of a scaffold based on PacBio subreads and Illumina contigs. Galaxy Tools (Giardine et al. 2005; Afgan et al. 2016) was used to examine the quality of the Illumina reads and to discard those with low quality and sequencing artifacts. To remove reads likely belonging to the host (*Drosophila hydei*), the remaining "high quality" reads were mapped against an unreleased *D. hydei* assembly (D. Begun personal communication) using Bowtie2 v2.1.0 (Langmead and Salzberg 2012). The mapping procedure was performed with the computational resources of the Texas A&M University Whole Systems Genomics Initiative (WSGI) HPC cluster.

Assembly of the reads that did not map to the assembly of *D. hydei* (i.e., "Non-*Drosophila* reads") was performed with Velvet v1.0.0 (Zerbino and Birney 2008) in a local Galaxy environment (Giardine et al. 2005; Afgan et al. 2016), assuming paired-ends reads and stringent conditions (k-mer >75, -cov_cutoff auto). Contigs obtained were compared to the NCBI non-redundant nucleotide database using nucleotide-nucleotide BLAST (blastn) (Altschul et al. 1990; Camacho et al. 2009) to locate and remove remaining contigs that likely belonged to the host. The remaining, non-*Drosophila* contigs were scaffolded in SCARPA (Donmez and Brudno 2013), and the resulting file containing both scaffolds and contigs was compared via local blastn i.e., local ncbi-blast-2.2.27+ (Altschul et al. 1990) to the *D. hydei* assembly to discard contigs and scaffolds that could be assigned to the host genome.

The remaining contigs and scaffolds were compared via blastn to a database constructed with previously released *Spiroplasma* genomes. The output was divided into two groups: (1) sequences that were similar to *Spiroplasma* sequences, including chromosome and plasmids; and (2) sequences with no similarities to any other reported sequences to date.

The PacBio reads were quality filtered with the HGAP (Hierarchical Genome Assembly Project) Portal from the PacBio SMRT analysis pipeline v 2.2 (Pacific Biosciences). *Drosophila* contigs were filtered and removed based on identity to the *D. hydei* assembly using local blastn. The remaining contigs were also separated by identity to *Spiroplasma* (i.e., local blast described above). The Genome Finishing Module of the CLC Genomics Workbench 9 (Qiagen) was used to perform a hybrid assembly of the Illumina and PacBio data that were assigned to *Spiroplasma*, by aligning both Illumina contigs/scaffolds and PacBio subreads through the contig alignment tool. The quality of the final contigs was verified by re-aligning (with Bowtie v2.1.0) the Illumina short reads (i.e. "high-quality Non-*Drosophila*), to detect regions with insufficient read coverage, which could represent assembly artifacts.

2.2.5 Annotation and phylogenetic analyses

A local implementation of PRODIGAL v 2.60 (Hyatt et al. 2010) was used to identify Open Reading Frames (ORFs) within the scaffolds/contigs assigned to *Spiroplasma*, assuming the *Mycoplasma-Spiroplasma* translation code. Annotation of identified genes was performed on the basis of homologous proteins using the Ensembl

Bacteria database through HMMER (Eddy 2011; Finn et al. 2011), and with RPSBLAST on the cluster of orthologous groups database (COG). Biochemical pathways and additional protein descriptions were identified through the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000; Kanehisa et al. 2016a; Kanehisa et al. 2016b)

Phylogenetic relationships among newly-sequenced (*S. poulsonii* hyd1 and *S. citri* moj) and published *Spiroplasma* strains (Table 1) were inferred with PhyloPhlAn v 0.99 (Segata et al. 2013). This method first identifies and aligns amino acid sequences of ~400 proteins conserved across bacteria originally selected from 2,887 bacterial "reference" genomes. Homologs of each conserved protein are aligned individually with Muscle v 3.8.31 (Edgar 2004). PhyloPhlAn trees are constructed from a amino acid concatenated alignment introducing gaps for genomes missing specific proteins. Maximum likelihood trees are then generated using FastTree v 2.1.7 (Price et al. 2010).

In addition, we used the following phylogenetic methods on the amino acid sequence datasets generated with PhyloPhlAn: maximum-likelihood with RAxML v 8.2.9 (Stamatakis 2006); and Bayesian using MrBayes v 3.2.6 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). Several datasets composed of different taxon sampling schemes (e.g. removing particular *Spiroplasma* species or removing certain outgroup taxa) were analyzed. Several representative members of the Entomoplasmatales group along with *Spiroplasma* (i.e., *Entomoplasma*, *Mesoplasma*, *Mycoplasma*) were included in the phylogenetic analyses, as well as several outgroup

taxa including other Mollicutes (i.e., *Phytoplasma*, *Acholeplasma*, *Ureaplasma*, *Mycoplasma*) and two non-Mollicutes (i.e., *Bacillus*) were included (Table 3).

Table 3. Additional species used to reconstruct the phylogeny of *Spiroplasma*.

Species	Mollicutes Group	Size (Mb)	RefSeq
Mesoplasma photuris	Entomoplasmatales	0.78	NZ_JNJ00000000.1
Mesoplasma syrphidae	Entomoplasmatales	0.92	NZ_JMKV00000000.1
Mesoplasma seiffertii	Entomoplasmatales	0.98	NZ_JAED00000000.1
Mesoplasma florum	Entomoplasmatales	0.79	NC_006055.1
Mesoplasma chaulicicola	Entomoplasmatales	0.84	NZJAEI00000000.1
Entomoplasma somnilux	Entomoplasmatales	0.86	NZ_JAGV00000000.1
Entomopasma luminosum	Entomoplasmatales	1.02	NZ_JAGW00000000.1
Entomoplasma lucivorax	Entomoplasmatales	1.10	NZ_JADH00000000.1
Entomoplasma mealucae	Entomoplasmatales	0.82	NZ_JMKX00000000.1
Mycoplasma putrefaciens	Entomoplasmatales	0.83	NC_015946.1
Mycoplasma feriruminatoris	Entomoplasmatales	1.02	NZ_ANFU00000000.1
Mycoplasma mycoides	Entomoplasmatales	1.21	NC_005363.1
Mycoplasma leachii PG50	Entomoplasmatales	1.01	NC_014751.1
Mycoplasma capricolum	Entomoplasmatales	1.01	NC_007633.1
Mycoplasma gallicepticum	Pneumoniae	1.01	NC_004829.2
Mycoplasma pnuemoniae	Pneumoniae	0.81	NC_000912.1
Mycoplasma genitalium	Pneumoniae	0.58	NC_000908.2
Ureaplasma urealyticum	Pneumoniae	0.87	NC_011374.1
Mycoplasma hominis	Hominis	0.66	NC_013511.1
Mycoplasma hyorhinis	Hominis	0.83	NC_019552.1
Mycoplasma canis	Hominis	0.89	NZ_CP014281.1
Mycoplasma felifaucium	Hominis	0.76	NZ_JHXS00000000.1

Table 3. Continued

Species	**Mollicutes Group**	**Size (Mb)**	**RefSeq**
Mycoplasma agalactiae	Hominis	0.87	NC_009497.1
Mycoplasma bovis	Hominis	1.00	NC_014760.1
Phytoplasma mali	AAA*	0.60	NC_011047.1
Phytoplasma pruni	AAA*	0.60	NZ_LHCF00000000.1
Acholeplasma hippikon	AAA*	1.43	NZ_JNJT00000000.1
Acholeplasma laidlawii	AAA*	1.50	NC_010163.1
Bacillus cereus	Not applicable	5.41	NC_004722.1
Bacillus thuringiensis 97-27	Not applicable	5.24	NC_005957.1

*AAA group: *Acholeplasma*/*Asteroplasma*/*Anaeroplasma*

Phylogenetic methods were performed using *Spiroplasma* and all outgroup taxa, including Mollicutes and *Bacillus* (Appendix; figures S1 to S5); *Spiroplasma* only (Appendix; figures S6 to S10); *Spiroplasma* without *S. eriocheiris* (Appendix; figures S11 to S15) and removing the whole Mirum clade (Appendix; figures S16 to S20). Computational work of MrBayes and RAxML phylogenetic analysis was performed in the CIPRES server (Miller 2010).

2.2.6 Spiroplasma *core and pan-genome*

Homolog and core genes shared by all *Spiroplasma* strains included in this study were identified through clusters of homologous gene families with GET_HOMOLOGUES (Contreras-Moreira and Vinuesa 2013). GET_HOMOLOGUES uses BLAST+ (Camacho et al. 2009) and the code base of OrthoMCL (Li et al. 2003) to

perform a search for orthologous proteins using three different algorithms: 1) OrthoMCL (OMCL), which performs a BLASTP to compare all sequences and construct a matrix based on similarities, it uses a *P*-value cutoff of 1e-5. Orthologous clusters are then constructed through the Markov Cluster algorithm (Van Dongen 2000); 2) COGtriangles, which merges triangles of inter-genomic symmetrical best matches of protein sequences to build clusters of orthologous groups (COGs) (Kristensen et al. 2010); and 3) bidirectional best hit (BDBH) method, which sorts genomes by size in order to use the smallest one as a reference. Paralogues are identified and new genomes are incrementally and individually compared to the reference genome; clusters containing at least one sequence per genome are retained (Contreras-Moreira and Vinuesa 2013). In all cases GET_HOMOLOGUES performs HMMER to annotate the protein domains.

Annotation of *Spiroplasma* genomes provided two different ways to study the genus. First we estimated the pan-genome, which is defined as the number of total different genes that can be found within a genus or species (Tettelin et al. 2005). On the other hand, the set of genes ubiquitously-present in every strain within a species or genus is known as the core-genome (Figure 2). Both metrics are important in *Spiroplasma* comparative genomics; the core-genome provides information about the housekeeping genes of the genus, whereas the pan-genome reflects acquired genes that could be associated with adaptations to new environments, hosts or conditions. The size of the pan-genome is also considered an indicator of a taxon's (genus or species) degree of plasticity. Both core and pan-genome were estimated with binomial mixture models

(Snipen et al. 2009). This approach starts by comparing all genomes using BLAST, and assumes that two sequences belong to the same gene family if their alignments have at least 50% identity.

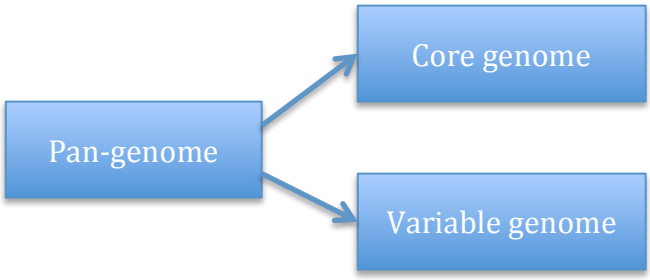


Figure 2. Pan-genome composition

2.3 Results and Discussion

2.3.1 Spiroplasma hyd1 *genome assembly*

According to the quantification performed, 46.664 ng of dsDNA diluted in buffer AE (Qiagen, Hercules CA) were used to construct the Illumina HiSeq library in Eureka Genomics. After sequencing, ~356.5 millions of reads with a length of 100 bp were obtained. 76% of the reads obtained mapped to the host genome (i.e., *D. hydei*), whereas 5.47% (a subset of the aforementioned 76%) mapped to the *Spiroplasma poulsonii* MSRO genome (Table 4). In addition, we found a small, but negligible (i.e., < 1%), number of reads that mapped to *Propiobacterium acnes*, a Gram-positive bacterium. The presence of these reads, which probably represent a contaminant, were removed and did not negatively affect the subsequent assembly. According to the data obtained and assuming that the genome size of *S. poulsonii* hyd1 ranges between 1.5 and 2.0 Mbp, the

estimated coverage was between 1139X-854X (C=LN/G; where C: coverage; G: haploid genome length; L: read length; N: number of reads. Reviewed in Illumina Technical Note: Sequencing).

Table 4. Illumina and PacBio reads generated.

	Total reads	Read metrics	QC passed	% Passed
Illumina	178,337,598 (forward) 178,337,598 (reverse)	Read length 100 bp Read length 100 bp	166,076,722 161,935,532	93.1 90.8
PacBio	115,715 (reads)	N50 read length 30,127 bp Mean read length 19,133 bp		

The PacBio library was constructed from >1 ng of DNA containing both *Spiroplasma* and *Drosophila*. Sequencing generated 115,715 long reads from 287,945 sub-reads, of which 224,599 (78%) were assigned to *D. hydei*, whereas 8,398 (2.91%) were identified as *Spiroplasma* sequences, following BLAST to the *D. hydei* assembly and the *S. poulsonii* MSRO genome, respectively.

The final *S. poulsonii* hyd1 assembly combining both Illumina reads and PacBio subreads returned 67 contigs representing 1,401,220 bp with a 27.9% G+C content (Table 5). Maximum contig length was 100,263 bp (mean 20,913 bp, N50 51,881 bp). True order and direction of *S.* hyd1 contigs cannot be determined without additional data (e.g. optical mapping; Paredes et al. 2015). To identify possible rearrangements and inversions with respect to *S.* MSRO, we used Mauve Contig Mover (Rissman et al.

2009) to iteratively order the contigs of *S.* hyd1 using *S.* MSRO as a reference (Figure 3). Surprisingly for these two closely related strains, we detected evidence of frequent inversions and rearrangements between *S.* hyd1 and *S.* MSRO genomes. Previously Alexeev et al. (2012) found similar results in the two close related strains *S. melliferum* KC3 and IPMB4A, were rearrangements occur despite 99.9% identity at the gene sequence level.

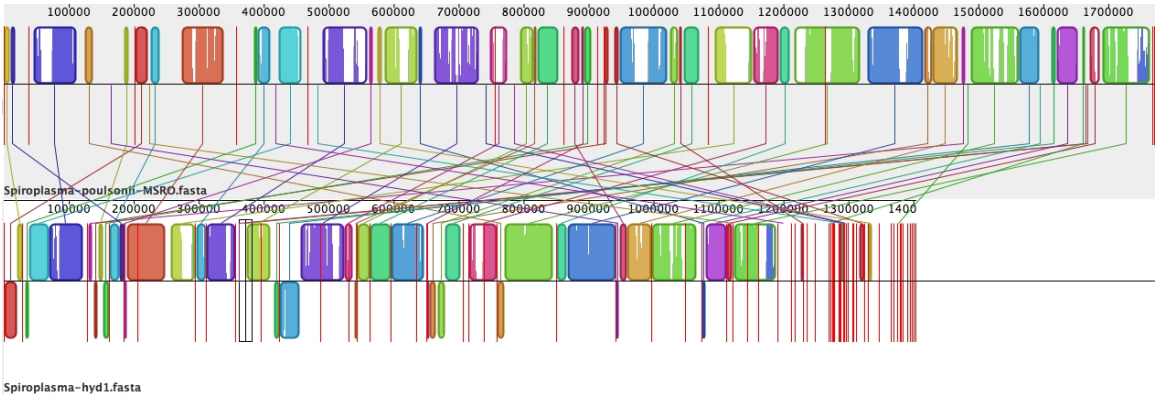


Figure 3. Whole genome alignment between *S.* MSRO (upper) and *S.* hyd1 (bottom). The lines connect homologous blocks between the genomes. Color blocks represent regions without rearrangements. Red bars indicate contig boundaries. For the hyd1 genome (bottom), those contigs below the horizontal black line represent inversions with respect to MSRO.

Sequences identified with high homology (i.e., most with an e-value ≤ 1 e-100 through local Blastn) to available sequences of *Spiroplasma citri* plasmids (e.g. pSci2, pSci4, pSci6, pBJS-O and pSKU146) were recovered. Nonetheless, circularization of such sequences could not be achieved, suggesting that they may be located on the *Spiroplasma* chromosome.

tRNAscan-SE-1.3.1 (Lowe and Eddy 1997) revealed 32 tRNA genes encoding for each of the 20 standard amino acids plus selenocysteine (Sec) (Appendix; table S1). tRNAs are distributed across seven contigs, but most of them are present in three clusters. RNAmmer revealed the presence of a complete set of rRNAs, including the large subunit (5S and 23S) and the small subunit (16S) within a single contig. To investigate the possibility of multiple rRNA operons, we used the Illumina reads that mapped to this contig to search for inconsistencies (i.e., mapping of one member of the pair to this contig and the other member of the pair to a different contig). No evidence of such inconsistencies was found suggesting that this genome contains a single ribosomal operon, which is consistent with all the *Spiroplasma* genomes sequenced to date. 1,511 protein-coding genes were identified after annotation of the *S. poulsonii* hyd1 hybrid assembly. Nevertheless, as observed in other *Spiroplasma* strains, a large number of putative, hypothetical and uncharacterized proteins were found, with such putative genes representing 55.9% of the total annotated genes in *S. poulsonii* hyd1. According to levels of protein homology, *S. poulsonii* hyd1 is closely related to *S. poulsonii* MSRO (uncorrected *p* distance=0.02) (Figure 4); 1,234 out of 1,521 genes have high homology between these two strains. 1,433 are genes found in within the *Spiroplasma* genus and 45 were identified as unique *S. poulsonii* hyd1 hypothetical genes. Annotation of predicted genes using the Kyoto Encyclopedia of Genes and Genomes (Blast Koala) showed that 545 genes have well known function of which almost half (48.8%; 266) are associated to Genetic Information Processing, whereas 11% (60) are associated to carbohydrate transport and metabolism.

Table 5. Genomic features of the assembly of *S. poulsonii* hyd1 and comparison with selected previously sequenced *Spiroplasma* strains.

	S. chrysopicola[1]	*S. mellif* KC3[2]	*S. citri* GII3-3X[3]	*S. poulsonii* MSRO[4]	*S. poulsonii* hyd1
Sequencing technique	Illumina HiSeq	SOLiD 4	miniBAC	PacBio RSII	Hybrid
Clade	Chrysopicola	Citri	Citri	Poulsonii	Poulsonii
Number of chromosomal contigs	1	4	39	16	67
Combined size of chromosomal contigs	1,123,322	1,260,174	1,525,756	1,632,994	1,401,220
Estimated chromosome size	Und	1,430,000	1,820,000	1,890,000	Und
Estimated coverage	Und	88.1	83.8	86.4	1139
G+C contents	28.8	27	25.9	26.7	27.9
Protein coding genes	1009	1,222	1,905	2,042	1,382
Plectrovirus proteins	0	132	375	307	31
Hypothetical proteins	394	485	519	1,164	622
Annotated pseudogenes	6	12	401	34	154
rRNA operon	1	1	1	1	1
tRNA genes	33	31	32	31	32
Number of plasmids	0	4	7	0	1?

[1] Ku et al. (2014); [2] Alexeev et al. (2012); [3] (Carle et al. 2010); [4] Paredes et al. (2015).

2.3.2 Spiroplasma *core and pan-genome*

19 *Spiroplasma* genomes were used to estimate the pan-genome and core genome of the genus. The average amino acid identities (i.e., Gower Distance Matrix) were obtained with Get_homologues, and the matrix is depicted as a heatmap in Figure 4. The distance values between *Spiroplasma poulsonii* MSRO and hyd1 (0.02) were very low indicating a high degree of similarity.

Using as a reference the smallest *Spiroplasma* genome (*S. diminitum*), I found 5,582 protein-coding gene clusters were identified across 19 *Spiroplasma* genomes. 4,561 of these are present in 1-2 genomes exclusively; these genes represent the "cloud" genome. 655 gene clusters were conserved across majority of the *Spiroplasma* strains included. 366 and 308 gene clusters form the soft-core and the core genome respectively (*Spiroplasma* soft-core gene list Appendix 5). The soft-core is defined as the gene clusters conserved in 95% of the analyzed taxa, (i.e., 18 *Spiroplasma* strains in this case) and the core genome includes the gene clusters conserved in every taxon analyzed. Among the core genome functions, we found the following: Amino acid transport and metabolism; Carbohydrate transport and metabolism; Cell division; Cell wall/membrane; Energy production and conversion; Lipid transport and metabolism; Transcription; recombination and repair (Appendix; table S6)

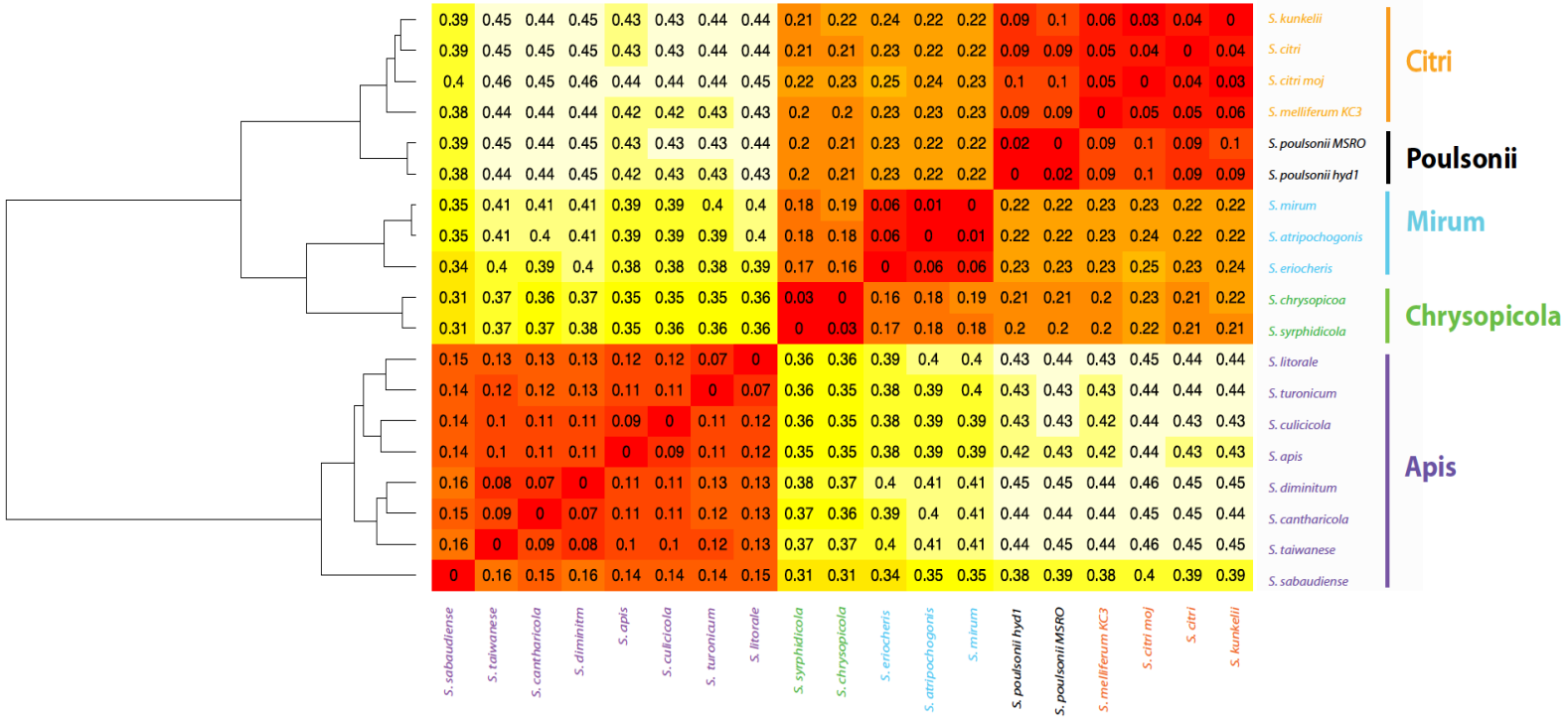


Figure 4. Heatmap of similarity within *Spiroplasma* genus constructed from the average amino acid identities. Color indicates degree of similarity among genomes; cell color closest to red (and cell value closest to 0) indicates highest similarity, whereas lighter color indicates lowest similarity. Comparisons between the 6,555 protein clusters used were performed through BLASTP.

2.3.3 DNA replication, repair, and homologous recombination

Our annotation of the *S. poulsonii* hyd1 genome revealed that most genes involved in prokaryotic DNA replication are present (e.g. *dna*B, *dna*E, *dna*G, *dna*N, *dna*X, *hol*A, *hol*B, *ssb*, *lig*A, *lig*B, *rnh*B, *rnh*C), as well as those involved in base and nucleotide excision repair (e.g. *mut*M, *nfo*, *pol*A, *MPG*, *UNG*, *uvr*A, *uvr*B, *uvr*C and *uvr*D). Although *S. poulsonii* hyd1 has genes from the mismatch repair system (*uvr*D, *exo*VII, *ssb*, *dna*E), but as reported for *S.* MSRO, it lacks critical genes such (i.e., *mut*S, *mut*L, *mut*H, *exo*I, *exo*X, *rec*J and *dam*). Also, consistent with the findings for MSRO, *S. poulsonii* hyd1 does not encode most of the necessary genes for homologous recombination (e.g. *rec*B, *rec*C, *rec*F, *rec*G, *rec*J and *ruv*C). Interestingly the bacterial recombinase A (*rec*A) gene, crucial for DNA homologous recombination and for the bacterial damage response (SOS), appears to be complete and functional in *S. poulsonii* hyd1 (table 6), whereas in *S. poulsonii* MSRO, *S. citri* and *S. melliferum,* it is a truncated non-functional gene (Carle et al. 2010; Lo et al. 2013a; Paredes et al. 2015). Pseudogenization of *rec*A in *S. poulsonii* MSRO, but not in *S. poulsonii* hyd1 strengthens the hypothesis of a recent loss event in MSRO (Paredes et al. 2015), and could explain the gene rearrangements observed between MSRO and hyd1 (Figure 3).

Table 6. Number of DNA replication and repair genes encoded by *S.* MSRO and *S.* hyd1

DNA replication and repair	***S. poulsonii* MSRO**	***S. poulsonii* hyd1**
DNA replication	13	13
Base excision repair	6	6
Nucleotide excision repair	6	6
Mismatch repair	11	11
Homologous recombination	13 (*rec*A not encoded)	14

2.3.4 Energy and metabolism

S. poulsonii hyd1 encodes the F-type ATP synthase operon (i.e., *atp*F, *atp*E, *atp*A, *atp*D *atp*H, *atp*C, *atp*G, *atp*B) and the complete prokaryotic glycolysis pathway. Regarding carbohydrate utilization, *S. poulsonii* hyd1 and *S. poulsonii* MSRO lack *tre*A or *tre*B, both of which are necessary to metabolize trehalose, the most abundant carbohydrate in insect hemolymph (Wyatt 1961). Most strains of *Spiroplasma* sequenced to date also lack the trehalose metabolism machinery (i.e., are missing one or both *tre* genes) (Figures 5 and 8). For insect-associated *Spiroplasma*, the inability to utilize trehalose might benefit the association by conferring the host better control of *Spiroplasma* densities (Paredes et al. 2015). Interestingly, three insect-associated strains, the honey-bee pathogen *S. melliferum* (Schwarz et al. 2014), the mosquito (*Culex annulus*) commensal *S. diminitum*, and insect-vectored plant pathogen *S. citri*, are capable of metabolizing trehalose (Carle et al. 2010; Alexeev et al. 2012; Lo et al.

2013b). The phylogenetic distribution of trehalose-metabolizing capability suggests repeated losses or gains of these genes in *Spiroplasma.*

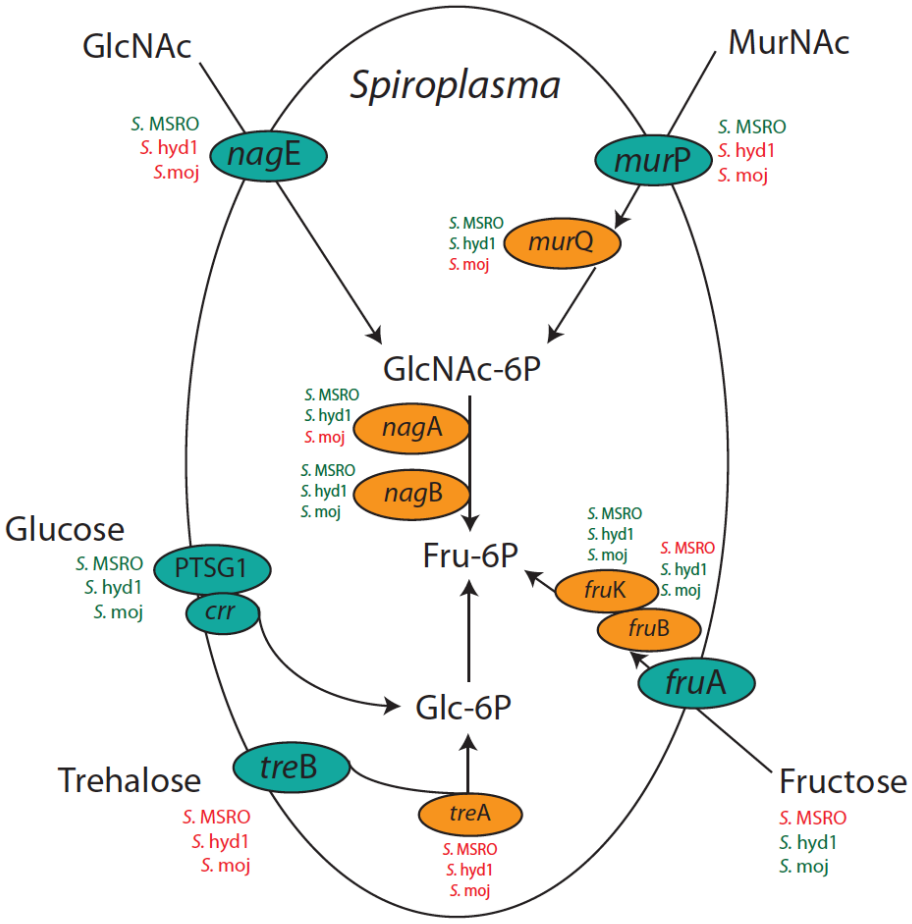


Figure 5. Comparison of carbohydrate metabolism in *Spiroplasma* strains associated to *D. melanogaster* (*S.* MSRO), *D. hydei* (*S.* hyd1) and *D. mojavensis* (*S.* moj). Blue circle indicates carbohydrate transporter enzyme. Orange circle represents intermediate enzymes. Green strain names indicate that the strain encodes a putatively functional enzyme, whereas those in red indicate that the gene is absent or pseudogenized in the corresponding strain. Abbreviations: Fru-6P, fructose-6-phosphate; Glc-6P, glucose-6-phosphate; GlcNAc, N-acetylglucosamine; GlcNAc-6P, N-acetylglucosamine-6-phosphate; MurNAc, N-acetylmuramic acid.

One other potential source of carbon for insect-associated *Spiroplasma* is chitin; an abundant biopolymer in insects, and fundamental component of the procuticle and the peritrophic matrix reviewed in Merzendorfer and Zimoch (2003). The strains *S. chrysopicola* (tabanid flies*), S. syrphidicola* (syphid flies), *S. melliferum* (bees), and *S. sabaudiense* (mosquitoes) appear to encode all the necessary machinery to use it as a source of carbon. *S. poulsonii* hyd1 encodes two type-II chitinase genes that are necessary for degradation of chitinase into smaller molecules (e.g. *N*-Acetylglucosamine;GlcNAc and N-acetylmuramic acid; MurNAc-6P). Nonetheless, absence of the transporter *nag*E in *S. poulsonii* hyd1, suggests that GlcNAc cannot be incorporated into the *Spiroplasma* cell. Similarly, although *mur*P (the transporter of MurNAc-6P) is encoded, it is interrupted by an early stop codon (residue 205/396) in *S. poulsonii* hyd1, suggesting it is not functional. Interestingly, the remaining genes, necessary for the biochemical conversion of GlcNAc-6P and MurNAc-6P to glucosamine-6-phohphate (*mur*Q and *nag*A, respectively), and subsequently into fructose-6-phosphate (*nag*B) are present in both *S. poulsonii* hyd1 and *S. poulsonii* MSRO (Figures 5 and 8). This mechanism is complete and functional in members of the clades Apis, Chrysopicola, Mirum, as well as in plant pathogens belonging to the Citri clade. Although it appears that chitin may not serve as a carbon source for *S. poulsonii* hyd1, as described below, its encoded chitinases may serve other functions.

According to the phosphotransferase systems found in *Spiroplasma*, the most common carbon source is fructose (Appendix; table S3); a monosaccharide found mostly in plants, but also in insect hemolymph at different concentrations depending of the

insect life stage (Levenbook 1947). Fructose is also a component of the *Drosophila* hemolymph and potentially a source of carbon for *Spiroplasma*. Whereas *S. poulsonii* MSRO lacks the ability to metabolize fructose (based on experimental evidence and on absence of a functional *fruB* gene; Paredes et al. 2015), *S. poulsonii* hyd1 appears to encode the necessary machinery to metabolize extracellular fructose (*fru*A and *fru*B). This ability has been reported in other arthropod-associated strains such as *S. melliferum*, *S. citri moj, S. eriocheiris* and the plant pathogens *S. kunkelii* and *S. citri* (Alexeev et al. 2012; Chang et al. 2014; Davis et al. 2015a; Paredes et al. 2015)

As reported for other sequenced *Spiroplasma* strains (Table S4), *S. poulsonii* hyd1 possesses the arginine transport and metabolism gene cluster *arc*A (arginine deiminase), *arc*B (ornithine carboamyltransferase), *arc*C (carbamate kinase) and *arc*D (Arginine/ornithine antiporter) (Figure 6), suggesting that it can use arginine as an energy source. In contrast, *S. poulsonii* MSRO has a truncated *arc*C and is unable to utilize arginine as a carbon source (Paredes et al. 2015). Given the close relationship of MSRO and hyd1, a recent pseudogenization event must have occurred within the *S. poulsonii* MSRO lineage. Other clades of *Spiroplasma* also contain members that encode the whole arginine transport and metabolism and members that do not (see Table S4), suggesting repeated losses of one or more of these genes.

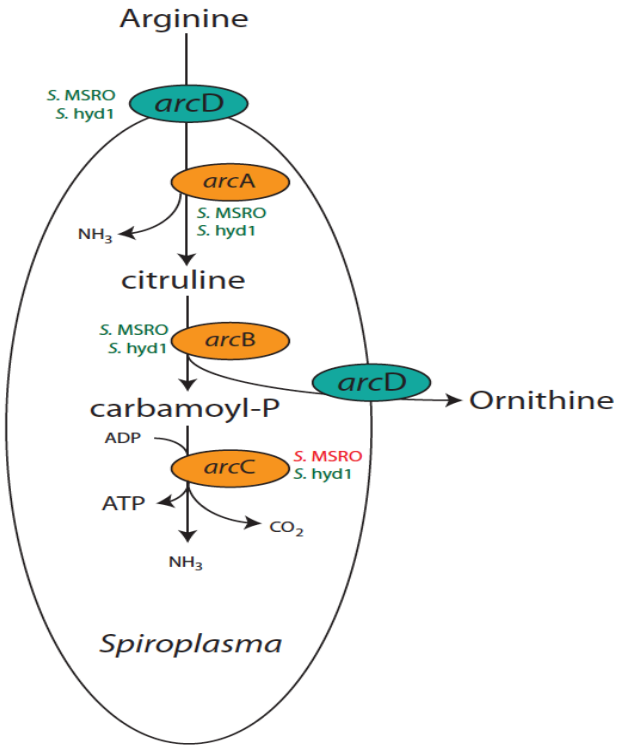


Figure 6. Comparison of arginine metabolism in *Spiroplasma* strains associated to *D. melanogaster* (*S.* MSRO), *D. hydei* (*S.* hyd1) and *D. mojavensis* (*S.* moj). Blue circle indicates carbohydrate transporter enzyme. Orange circle represents intermediate enzymes. Strains in green indicate active enzymes and strains in red indicate absent/pseudogene. Abbreviations: ADP, adenosine diphosphate; ATP, adenosine triphosphate.

Additionally to the metabolism of carbohydrates, lipid usage by *Spiroplasma* strains associated to *Drosophila* is also relevant. Previous studies concluded that growth of *Spiroplasma* within *Drosophila* is limited by the availability of lipids in the hemolymph (Herren et al. 2014). Moreover, it is suggested that competition for hemolymph lipids could also be an important factor in the protective mechanism against parasitoid wasp larvae in the *D. melanogaster*-*S. poulsonii* MSRO-*Leptopilina* system

(Herren et al. 2014; Paredes et al. 2016). Lipophorines (Lpp) from the apo-B family are the lipid transporters present in *Drosophila* hemolymph; of which diacylglycerol (DAG), phosphatidylethanolamine (PE) and sterols are the most abundant (Palm et al. 2012). *S. poulsonii* hyd1 encodes enzymes necessary for entry of extracellular glycerol into the glycolysis cycle. Although *Spiroplasma* encodes most of the enzymes required for cardiolipin biosynthesis (*pls*X, *pls*Y, *dgk*A, *cds*A, *pgs*A and *cls*A), the required gene *pgp* (phosphatidyl glycerol phosphatase) is not present. A gene encoding this enzyme has not been found in any *Spiroplasma* genome sequenced to date. The missing enzyme is required to remove the phosphate group from the phosphatidylglycerol phosphate (PGP) and transform it to phospatidylglycerol (PG) (Figure 7). However, evidence that *Spiroplasma* synthetizes cardiolipoin has been reported in *S. citri* and *S. poulsonii* MSRO (Patel et al. 1978; Herren et al. 2014). We thus hypothesize that a gene with the function of *pgp* in *Spiroplasma* is present, but highly divergent from *pgp* genes annotated in other bacterial groups. Such gene is probably among ~23 hypothetical proteins with unknown function found in the *Spiroplasma* core genome.

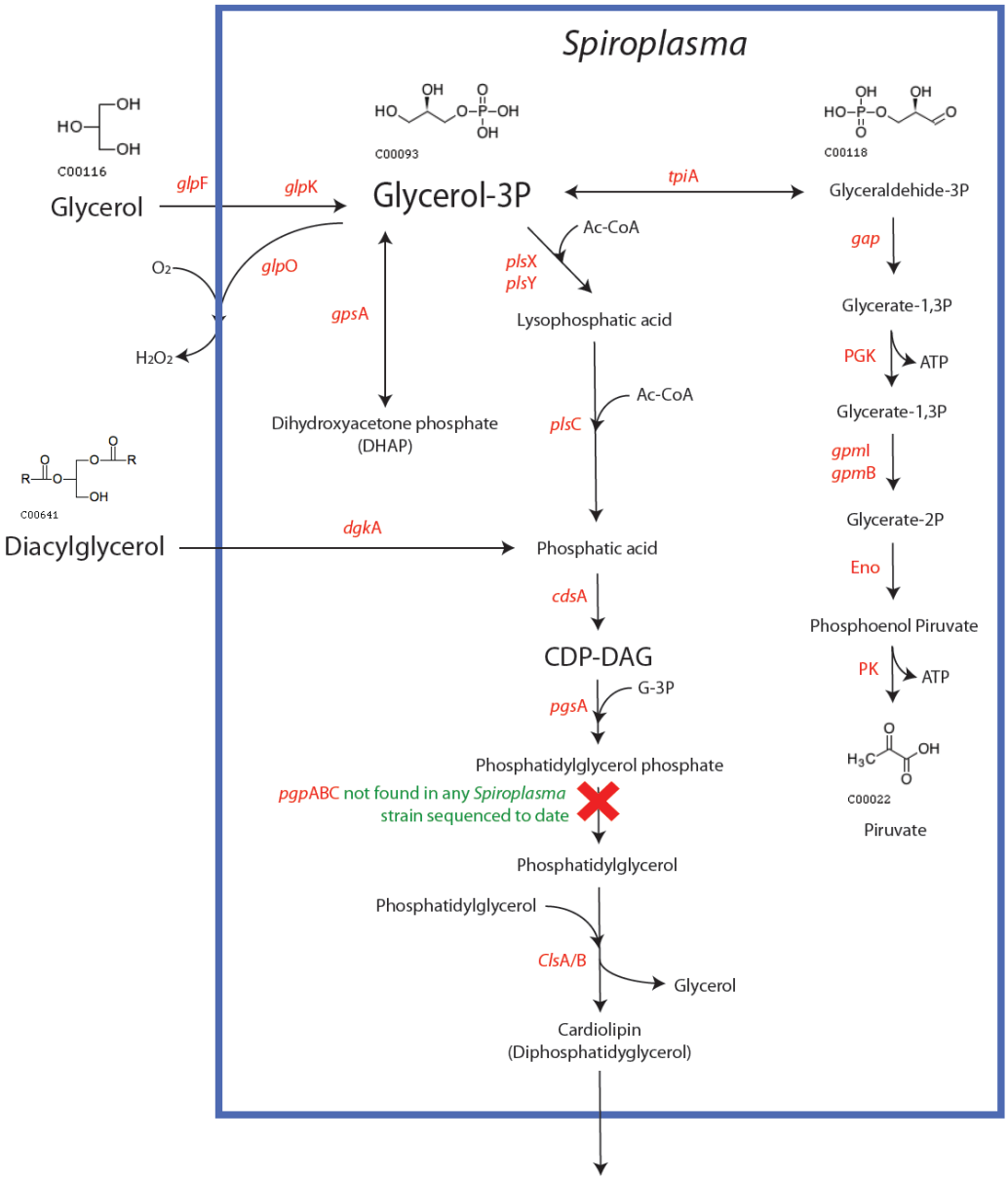


Figure 7. Glycerol and diacylglycerol metabolism in *S. poulsonii* hyd1. Intermediate products are depicted with black font. Putatively active enzymes are depicted in red. Abbreviations: ATP, adenosine triphosphate; CDP-DAG, cytidine diphosphate diacylglycerol.

Table 7. Putative Ribosome-Inactivating Proteins (RIPs) found/reported in *Spiroplasma* strains, on the basis of conserved residues found in Shiga toxin subunit A and in ricin A-chain.

			Conserved residues in RIP ricin A-chain active site				
Strain	**Location**	**Size (a.a.)**	**Y^{80}**	**Y^{123}**	**E^{177}**	**R^{180}**	**W^{211}**
S. poulsonii hyd1	Contig-25_2	474	Y	Y	E	R	W
	Contig-6_5	692	S	Y	I	D	L
S. poulsonii MSRO	WP_040093770.1	465	Y	Y	E	R	W
	WP_040093936.1	496	Y	Y	E	R	W
	WP_040094559.1	448	Y	Y	E	R	W
	WP_040092751.1	448	Y	Y	E	R	W
	WP_040093807.1	185	-	Y	G	Q	W
S. poulsonii Neo		403	Y	Y	E	R	W
S. eriocheris	AKM54484.1	192	Y	Y	E	R	W
	WP_047791682.1	190	Y	Y	E	R	W
S. sabaudiense	WP_025251437.1	268	Y	Y	E	R	W
S citri Moj	Contig-35_3	359	Y	Y	E	R	W
	Contig-93_1	321	F	L	L	N	L
	Contig-175_1	175	F	V	P	E	S
E. coli (Shiga toxin subunit A)	WP_033815585.1	318	Y	Y	E	R	W
Ricin (A-chain)	1UQ4_A	266	Y	Y	E	R	W

2.3.5 Ribosomal Inactivating Proteins (RIPs)

The genome of *S.* hyd1 contains two ORFs that exhibit high domain homology to previously reported RIP genes in other *Spiroplasma* and in plants (Table 7). RIP homologous proteins in S. hyd1 were identified through local BLASTP with *S. poulsonii MSRO, S. poulsonii Neo, S. eriocheiris, S. sabaudiense, E. coli* and Ricin. RIP protein sequences from *Drosophila*-associated *Spiroplasma* strains were aligned and conserved residues were identified. At least one RIP from each *Drosophila* strain showed all the active-site conserved residues (Table 7). However, the sequences are so divergent (overall uncorrected *p*-distance= 0.749) that adequate homology could not be determined and thus, phylogenetic inferences were not made. Presence of RIP toxins in *Drosophila*-associated *Spiroplasma* known to be protective strains strengthen the hypothesis of their role in this mechanism and suggest that protection due to RIP toxins may not be restricted to nematodes, but extend to parasitic wasps (e.g. in *S. poulsonii* MSRO and *S. poulsonii* hyd1). Interestingly, *S. moj* (citri clade), which does not appear to protect against two wasp species (see Chapter 3), also encodes what appears to be a functional RIP. Therefore, it is possible that this RIP could be used against other natural enemies.

Putative RIPs are also encoded in the genomes of other arthropod-associated *Spiroplasma* strains, such as *S. sabaudiense* (mosquitoes), *S. mirum* (ticks), *S. eriocheiris* (crab), and the *Drosophila*-associated *S. poulsonii* MSRO and *S. citri* moj. Five amino acids in the RIP active site are identified as essential for the enzyme's cytotoxic activity as reported in the RIP Saporin-6 from *Saponaria officinalis*. In the action mechanism, the target adenine of the 28S rRNA is trapped between the tyr-80 and

tyr-123 while arg-180 serve protonating N3 of the ribose ring. Glu-177 breaks the N9-C1 glycosidic bond, releasing adenine A4324 from the 28S rRNA and inactivating the ribosome (Bagga et al. 2003; Fermani et al. 2005; de Virgilio et al. 2010).

Ribosomal inactivating proteins are a well-characterized family of plant and bacterial toxins that recognize and depurinate the 28S rRNA of the eukaryotic 60s ribosome subunit (de Virgilio et al. 2010; Walsh et al. 2013). Previous experiments in *D. neotestacea* infected with its natural *Spiroplasma* symbiont revealed that when the fly was exposed to the nematode *Howardula*, *S.* Neo RIP depurinates the nematode's 28S rRNA, blocking protein translation (Hamilton et al. 2014; Hamilton et al. 2016). This suggests that RIP toxins participate in *Spiroplasma* ability to defend its host against eukaryotic parasites.

2.3.6 Potential virulence factors in Spiroplasma *hyd1*

As described above, the genome of *S.* hyd1 encodes two type II chitinases. A homologue of one of these is found in the genome of *S. poulsonii* MSRO. The presence of at least one chitinase gene in two defensive strains and its absence in the genome of *S. moj* (an apparently non-defensive strain; see Chapter 3) gives support to the hypothesis that these enzymes might aid in host defense, possibly by degrading the chitin-based cuticle of wasp larvae (Paredes et al. 2015).

The genome of *S.* hyd1 possesses a gene encoding the enzyme glycerol-3-phosphate oxidase (*glp*O), which uses glycerol-3P and molecular oxygen to produce hydrogen peroxide (Fig. 8). This enzyme plays a vital role in the pathogenicity of the

mollicutes *Mycoplasma mycoides* (Bischof et al. 2009) and *M. pneumoniae* (Hames et al. 2009). Moreover, insect-synthesized hydrogen peroxide has been found as a potential defensive factor against parasitic wasps, due to its cytotoxicity and role in the production of melanin as a part of the insect immune response against parasites (Nappi et al. 2009). Paredes et al. (2015) hypothesized that this enzyme could be part of the defensive system provided by *S. poulsonii* MSRO to its host *D. melanogaster*. Its presence in *S. poulsonii* hyd1 and absence in the non-defensive strain *S.* moj is consistent with this hypothesis.

2.3.7 Phylogenomic analyses

A total of 398 protein-coding genes were used for phylogenetic analyses. All analyses distinguished three main *Spiroplasma* clades (Figure 9; Table 8; Appendix; figures S1 to S20): the Chrysopicola-Mirum-Poulsonii-Citri clade; a clade formed by seven members of the Apis clade; and the lineage of *S. sabaudiense*. These results are generally congruent with relationships established on the basis of 16S rRNA sequences (Gasparich et al. 2004; Mateos et al. 2006; Haselkorn 2010a), and the phylogenomic analysis of Bolanos et al. (2015), but see exceptions below. Consistent with relationships based on the 16S rRNA gene alone, all analyses supported the sister relationship of *S. poulsonii* hyd1 and *S. poulsonii* MSRO. Both strains are associated with *Drosophila* and act as defensive symbionts against parasitoid wasps (Xie et al. 2010; Xie et al. 2014; Mateos et al. 2016; Paredes et al. 2016), but *S. poulsonii* MSRO is a reproductive manipulator (i.e., male killer) (Montenegro et al. 2006; Xie et al. 2014). All analyses supported a sister relationship of *S. citri* moj with *S. kunkelii* (a

phytopathogenic strain), which together with *S. melliferum* and *S. citri*, form the Citri clade. The monophyly of the Citri+Poulsonii clade was supported by all analyses and is consistent with 16S rRNA phylogenies (Montenegro et al. 2006; Haselkorn 2010a; Paredes et al. 2015). We note that previous studies include the Poulsonii clade within the Citri clade. This is compatible with our findings, but we treat them separately to facilitate comparisons.

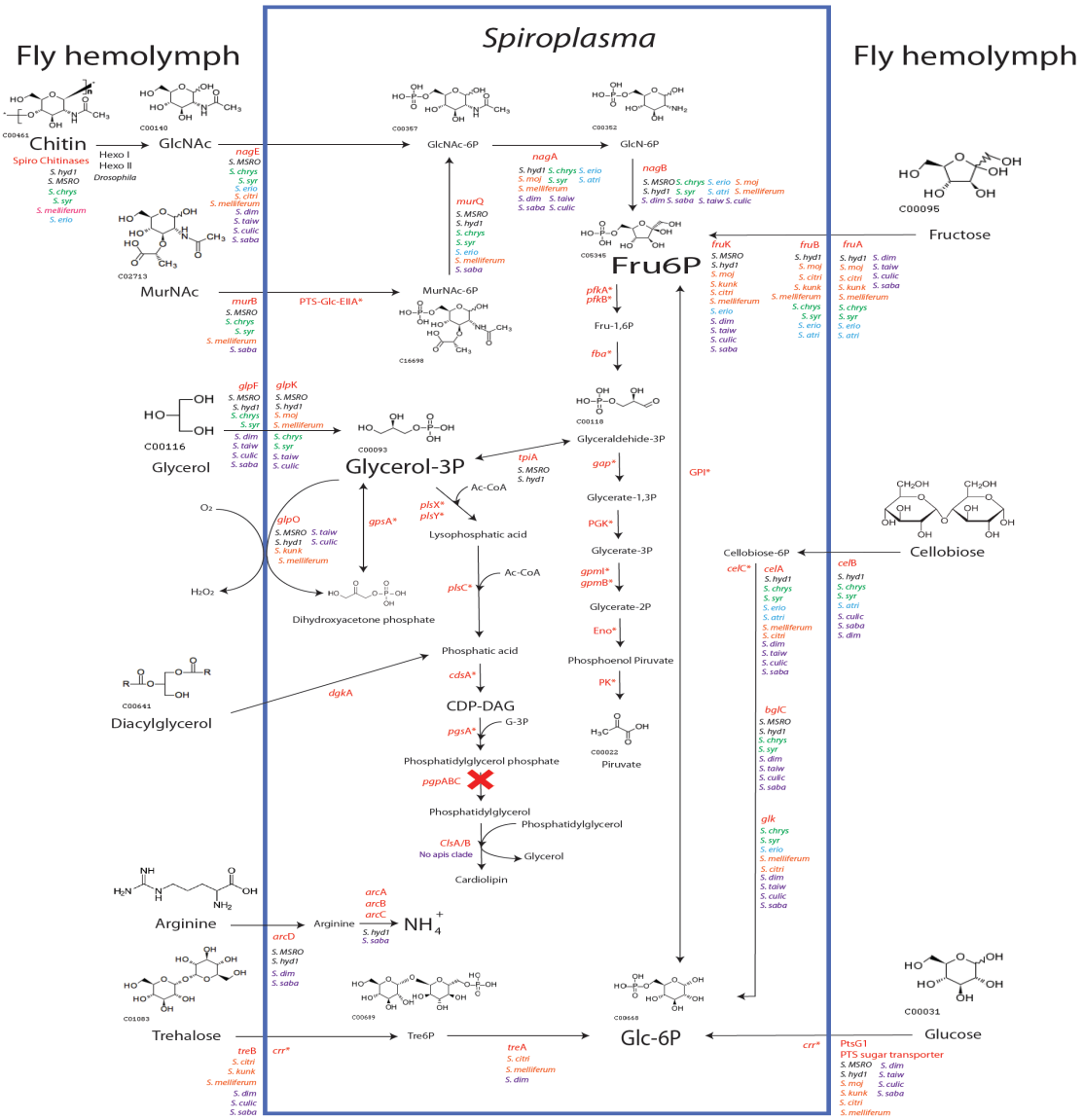


Figure 8. Selected carbohydrate and lipid metabolic pathways of genome sequenced *Spiroplasma* strains. *Spiroplasma* names are colored based on its clade; black: Poulsonii; green: Chrysopicola; blue: Mirum; orange: Citri; purple: Apis. Genes required for biochemical reactions are represented in red and *Spiroplasma* strains encoding such genes are listed. Red asterisk indicates a conserved core gene.

With few exceptions, *S. chrysopicola* and *S. syrphidicola* were recovered as sister lineages (i.e., the Chrysopicola clade; Table 8). Similarly, the monophyly of *S. eriocheiris*, *S. mirum* and *S. atripochogonis* (i.e., the Mirum clade) was supported by most analyses. Nonetheless, the relationships among the Chrysopicola, Mirum and Poulsonii-Citri clades were not clearly resolved. Whereas some analyses favored the Mirum + Poulsonii-Citri relationship, others supported the Chrysopicola + Poulsonii-Citri relationship. The latter relationship is the one previously reported on the basis of 16S rRNA, albeit with low clade support values (60% in Lo et al. 2015; 70% in Lo et al. 2013). This discrepancy could be due to the presence of horizontally transmitted genes (HTGs) in *S. eriocheiris* and *S. mirum* (Lo et al. 2015). To account for this possibility, two strategies were performed. Firstly, HGTs from *S. eriocheiris* reported in Lo et al. (2015) were removed and phylogenetic analyses were repeated (Appendix; figure S11 to S15). Secondly, the entire Mirum clade was removed, followed by new phylogenetic analyses (Table 8; Appendix; figures S16 to S20). Neither removal of *S. eriocheiris* nor removal of the entire Mirum clade resulted in more consistent support for alternative relationships. It is thus possible that there are additional horizontally acquired genes in these or the other genomes examined.

Seven out of the eight taxa assigned to the Apis clade (i.e., all except *S. sabaudiense*) consistently formed a monophyletic group in all analyses. These seven Apis-clade taxa were sister to a clade made up of members of *Entomoplasma*, *Mesoplasma* and *Mycoplasma* (i.e., *M. putrefaciens*, as well as *M. mycoides* and its close relatives). *S. sabaudiense*, considered a member of the Apis clade (Lo et al. 2013a;

Chang et al. 2014; Paredes et al. 2015; Lo et al. 2016); was in turn, sister to this latter larger clade. The paraphyly of the Apis clade had been reported by Bolanos et al. (2015), but is incongruent with phylogenies based on the 16S rRNA gene only (e.g., Lo et al. 2015). Recovery of a monophyletic Apis clade including *S. sabaudiense* in the analyses of only *Spiroplasma* taxa (Table 8) was therefore a taxon-sampling artifact. Relationships among the seven Apis-clade members that consistently formed a monophyly were not fully resolved. To reflect this, we collapsed the base of this clade into a trichotomy formed by: *S. turonicum* + *S. litorale* (clade A); *S. taiwanese* + *S. diminitum* + *S. cantharicola* (clade B); and *S. apis* + *S. culicicola* (Figure 9).

Table 8. *Spiroplasma* phylogenetic analyses. Numbers represent node support between clades/taxa. PhyloPhlAn analysis was performed through the alignment of 398 conserved proteins. RAxML and MrBayes were performed through the alignment of the same protein set used by PhyloPhlAn. Edited refers to the removal of sequences absent in more than 50% of the taxa. P=Poulsonii clade; C=Citri clade; Ch=Chrysopicola clade=*chr*+*syr*; M[a]=Mirum clade (if *S. eriocheiris* included = *atr*+*mir*+*eri*; if *S. eriocheiris* excluded = *atr*+*mir*); cladeA=*tur*+*lit*; clade B= *tws*+*dim*+*can*; *chr*=*S. chrysopicola*; *syr*=*S. syrphidicola*; *atr*=*S. atripochogonis*; *eri*=*S. eriocheiris*; *mir*=*S. mirum; tur*=*S. turonicum*; *lit*=*S. litorale*; *tws*=*S. taiwanese*; *dim*=*S. diminitum*; *sab*=*S. sabaudiense*; *cul*=*S. culicicola*; *api*=*S. apis*

	MIRUM-CHRYSOPICOLA-CITRI-POULSONII													APIS				
	Tree	Aligned characters	M[a]	Ch = *chr-syr*	P-C	M-P-C	Ch-P-C	Ch- M-P-C	*chr*-P-C	*mir-atr*-P-C-Ch	*mir*-P-C-Ch	*syr-chr*-P-C-M	*chr*-P-C-M	A-B- *api-cul*	Apis = A-B-*sab- api-cul*	A-B	*api-cul*-B	*cul*-A-B
All taxa (n = 49)																		
PhyloPhlAn	01	3704	100	90	100	0		100						100			46	
RAxML (Alignment not edited)	02	3704	66	58	100	60		98						67	97			
RAxML (Alignment edited)	03	2204		97	99		74	87						52				
MrBayes (Alignment edited)	04	2204	100	100	100		85	100						100		65		
MrBayes (Alignment no edited)	05	3704	100	100	100		87	100						100			72	
***Spiroplasma* only (n = 19)**																		
PhyloPhlAn	06	2638	100	84	100	83		100						99	100		92	
RAxML (Alignment not edited)	07	2638	88		100	84		100				100	82	77	89			
RAxML (Alignment edited)	08	2250		59	100		59	100						99	100		91	
MrBayes (Alignment edited)	09	2250	100	100	100	93		100						100	99		100	
MrBayes (Alignment not edited)	10	2638	100	100	100	79		99						99	99		98	
***S. eriocheiris* excluded (n =18)**																		
PhyloPhlAn	11	2638	100	87	100	48									100	100		71
RAxML (Alignment not edited)	12	2638	97		100			100				100	82		87	68		
RAxML (Alignment edited)	13	2310		66	100		64	100		100	61				100	100	86	
MrBayes (Alignment not edited)	14	2310	100	100	100		96	100							100		100	
MrBayes (Alignment edited)	15	2310	100	100	100		96	100							100		99	
Mirum clade excluded (n = 16)																		
PhyloPhlAn	16	2619	NA		100	NA	100	NA	60			NA	NA	99	100	98		
RAxML (Alignment no edited)	17	2619	NA		100	NA	100	NA	80			NA	NA	85	90	63		
RAxML (Alignment edited)	18	2301	NA		100	NA	100	NA	62			NA	NA	100	100		80	
MrBayes (Alignment no edited)	19	2619	NA	72	99	NA	99	NA				NA	NA	99	99			
MrBayes (Alignment edited)	20	2301	NA	67	99	NA	99	NA				NA	NA	99	99		98	

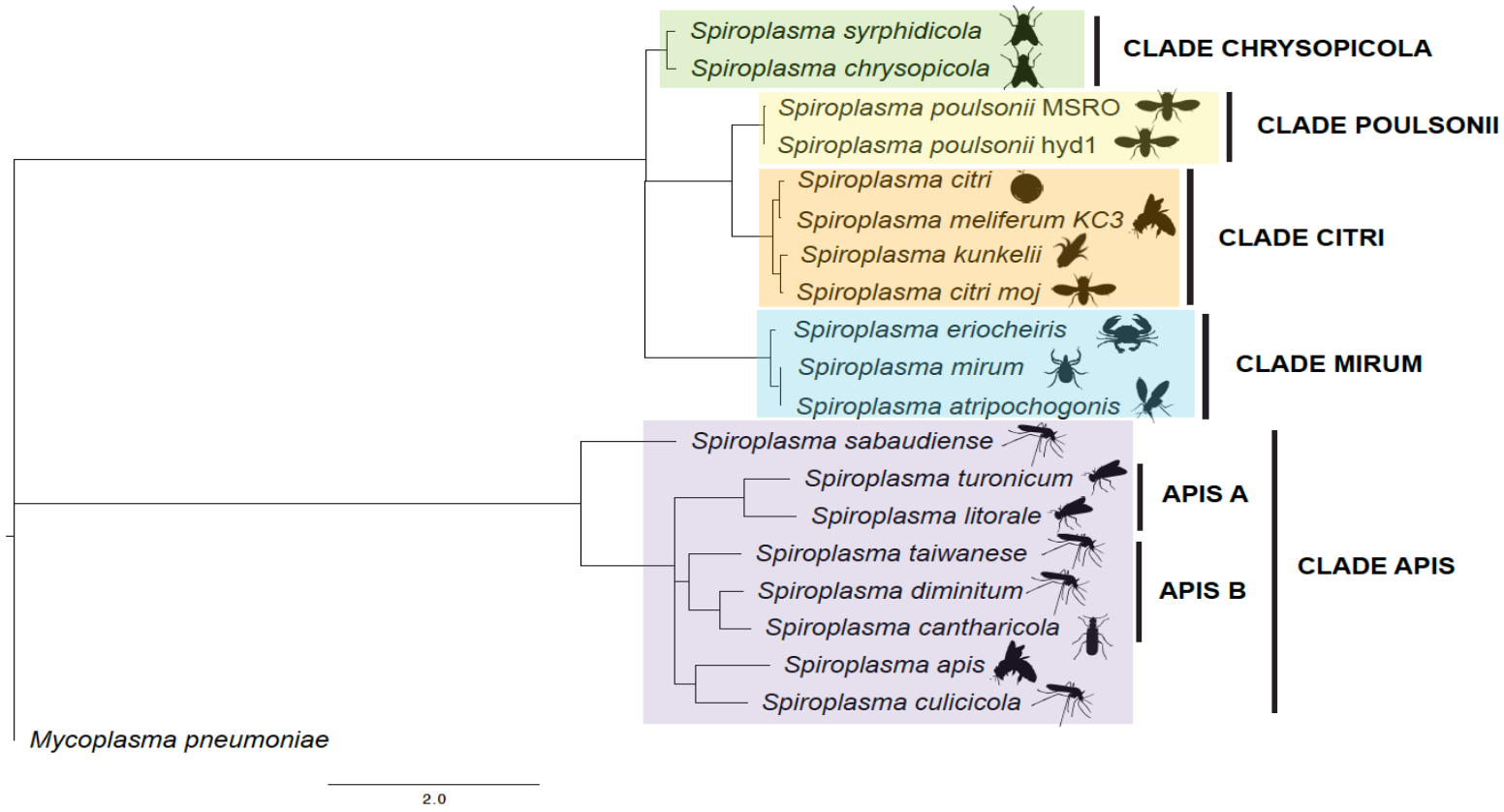


Figure 9. Consensus tree of *Spiroplasma* strains sequenced to date.

CHAPTER III

FITNESS EFFECTS OF *SPIROPLASMA CITRI* STRAINS ON ITS *DROSOPHILA HOSTS*

3.1 Introduction

The bacterial genus *Spiroplasma* includes several strains that are heritable endosymbionts of several *Drosophila* species and other insects. *Spiroplasma* are wall-less, helical, motile bacteria, phylogenetically classified as Gram-positive. *Spiroplasma* belongs to the class Mollicutes (Gasparich 2002; Gasparich et al. 2004), which contains the simplest self-replicating organisms known, and is characterized by small genomes with low GC contents (Carle et al. 1995). *Spiroplasma* strains are associated intra- and extra-cellularly with a variety of arthropods and plants (Anbutsu and Fukatsu 2011), and transmitted either vertically (in arthropods) and horizontally (e.g. in plants; vectored by sap-feeding insects). *Drosophila*-associated strains fall within four separate clades: Poulsonii; Citri; Tenebrosa; and Ixodetis (Haselkorn 2010a). Within *Drosophila*, 19 species are reported to harbor *Spiroplasma,* and in certain populations of *Drosophila* infection frequencies are relatively high (Mateos et al. 2006; Watts et al. 2009). To date, only poulsonii clade strains are known to act as defensive strains; MSRO (*D. melanogaster*), hyd1 (*D. hydei*) and neo (*D. neotestacea*) confer protection against certain parasitic wasps; and neo also confers protection against the parasitic nematode *Howardula aoronymphium* (Jaenike et al. 2010; Xie et al. 2010; Xie et al. 2014; Haselkorn and Jaenike 2015; Mateos et al. 2016).

The *citri* clade is relatively common in *Drosophila* species, but is restricted to members of the repleta species group; which includes the generalist ("human commensal") cosmopolitan *D. hydei* and numerous species that specialize on cacti. Three of these *Spiroplasma* strains reach relatively high prevalence in nature: *S. moj* reaches 60% in Catalina Island (California) (Haselkorn et al. 2009); and *S. ald2* (newly discovered) prevalence reaches 60–70% in the Austin-San Antonio area of Texas (figure 10), in both *D. mulleri* and in the eastern lineage of *D. aldrichi* (as defined by Oliveira et al. 2008). *Spiroplasma* in *D. hydei* populations reaches high frequencies in Japan, up to 66% hyd1 infection prevalence; Kageyama et al. (2006), and in North America (average ~28%; 60% in one population (Watts et al. 2009). However, the North America prevalence data do not distinguish between hyd1 (poulsonii clade) and hyd2 (citri clade). Nonetheless, they likely reflect mostly hyd1 prevalence, because hyd2 has only been reported in a handful of locations in Mexico and Western US (Mateos et al. 2006; Haselkorn et al. 2009); M. Mateos personal communication).

Current knowledge on the fitness effects of members of the citri clade on their *Drosophila* hosts is very limited. None of the strains reported to date appears to bias the sex ratio of offspring, suggesting that they do not kill males (Mateos et al. 2006; Haselkorn and Jaenike 2015) (Figure 11). The fidelity of their vertical transmission has not been formally quantified. Nonetheless, the higher (unintentional) loss of infection of citri-clade strains observed in the lab, and the lower densities reported for citri-clade strains over all life stages of the host (ref = http://dx.doi.org/10.4161/fly.25469), suggests that vertical transmission rate of citri-clade strains is lower than that of

poulsonii-clade strains. Fitness effects (survival rate, desiccation resistance, development time and body size) of *S.* moj in *D. mojavensis* isolines from Catalina Island were measured by Haselkorn (2010b). However, a clear association between infection and fitness effects (negative or positive) was not detected, as an interaction with host genetic background was apparent. No other fitness effects of citri-clade *Spiroplasma* on their *Drosophila* hosts, including defense against natural enemies and other forms of reproductive manipulation (e.g. cytoplasmic incompatibility), have been examined.

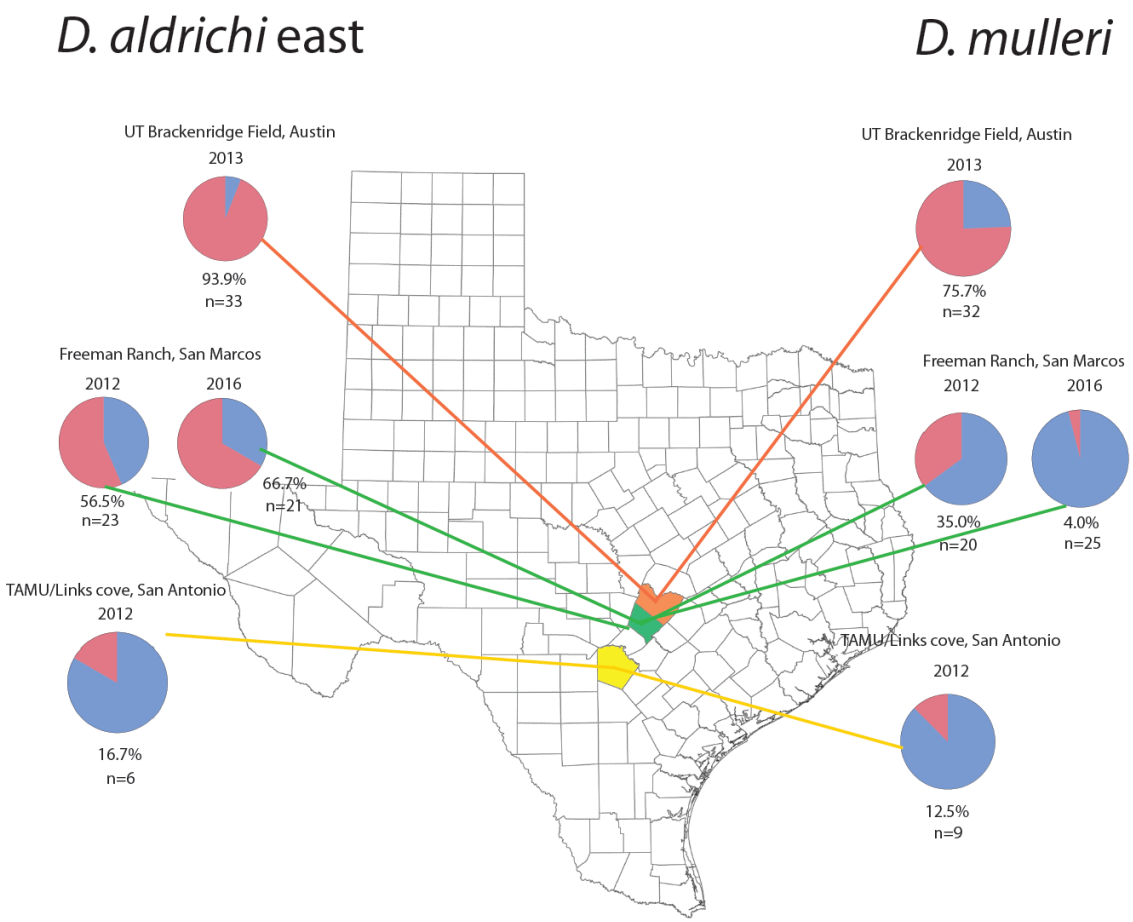


Figure 10. *Spiroplasma* ald2 frequencies in *D. aldrichi* east (left) and *D. mulleri* (right) in three regions of Central Texas. Red area and percentage in the pie chart indicate proportion of infected females per sampling/year. Collection of samples was performed from 2012 to 2016 in Texas A&M University and Cove Links at San Antonio; Texas State University's Freeman Ranch at San Marcos and the University of Texas' Brackenridge Field Station at Austin.

The present study examined whether two citri-clade strains (moj and ald2) confer protection to their respective natural host species (i.e., *D. mojavensis* and *D. aldrichi*) against two parasitic wasps: the cosmopolitan generalist *Leptopilina heterotoma* (Lh14; Figitidae); and *Asobara* sp. (Aw35; Braconidae; from Texas). In addition, we tested whether the citri-clade strains (moj and hyd2; the latter harbored by *D. hydei*) induce

reproductive phenotypes. We assessed oviposition rate in for both strains and cytoplasmic incompatibility for moj, by examining the outcome of reciprocal crosses of *Spiroplasma*-infected and *Spiroplasma*-free individuals. Cytoplasmic incompatibility (CI) is embryonic failure of the progeny from crosses between an uninfected female and an infected male, or between a female and male carrying incompatible symbiont strains (O'Neill and Karr 1990). This reproductive parasitism has been observed in *D. simulans* and *D. melanogaster* infected with *Wolbachia* (Hoffmann et al. 1986; Hoffmann et al. 1994).

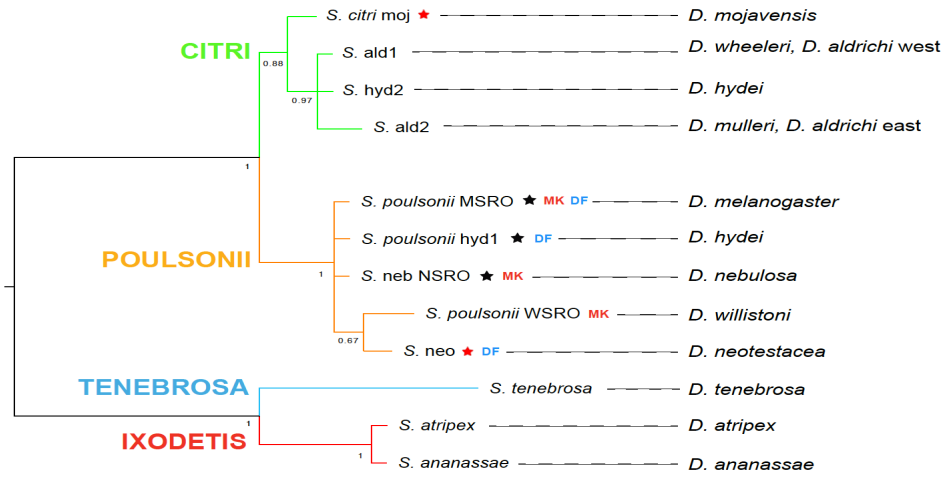


Figure 11. 16S rRNA Phylogenetic tree (Bayesian) of *Spiroplasma* strains associated to *Drosophila* flies. MK indicates male killing strains, whereas DF refers to defensive strains. Black star indicates complete genome available, red star indicates genome sequencing in process

3.2 Materials and Methods

3.2.1 Sources of flies, Spiroplasma *and wasps*

D. mojavensis isoline CI-33-15 was originally collected in Catalina Island, California in summer 2012. The isoline was maintained in the laboratory in a standard opuntia-banana diet. Infection by *Spiroplasma* was confirmed from individual DNA extracts from five female flies from the first generation in captivity (G1) with taxon-specific primers: TKSSp/63F, spoulF/spoulR and 16STF1/16STR1 (Table 1). Although *Wolbachia* has never been reported in *D. mojavensis* or any member of the *repleta* species group (Mateos et al. 2006), lack of *Wolbachia* infection was confirmed by negative results with Eubacteria universal primers (27F/1492R) and with the *Wolbachia*-specific primers *wsp*.

Spiroplasma infected and uninfected sub-isolines were generated by selection (i.e., natural loss of *Spiroplasma* infection). Virgin females were individually placed in vials and allowed to mate with CI-33-15 males. Once larvae hatched, the parental female was sacrificed and a PCR test was done to confirm its *Spiroplasma* infection status. Offspring from *Spiroplasma*-negative and *Spiroplasma*-positive females were subsequently used to establish CI-33-15 sub-isolines. Six CI-33-15 sub-isolines were constructed and maintained during two years; the *Spiroplasma*-positive sub-isolines were labeled C1, B2 and F2; whereas the *Spiroplasma*-negative sub-isolines were labeled moj ABM.

Table 9. Universal and *Spiroplasma* specific primers used to screen *Drosophila* flies.

Locus	Primer name and sequence (5'to 3')	Target group	Annealing temp. (°C)	size (bp)
16S rDNA	SpulF: GCT TAA CTC CAG TTC GCC SpulR: CCT GTC TCA ATG TTA ACC TC	*Spiroplasma*	53	~ 500
16S rDNA	16STF1: GGT CTT CGG ATT GTA AAG GTC TG 16STR1: GGT GTG TAC AAG ACC CGA GAA	*Spiroplasma*	65 TD* 55	~ 1368
16S rDNA	63F: GCC TAA TAC ATG CAA GTC GAA C TKSSp: TAG CCG TGG CTT TCT GGT AA	*Spiroplasma* and several Gram-positive	55	~ 450

*TD = touch down PCR

Table 10. Restriction enzymes and conditions to differentiate *Spiroplasma* strains hyd1 and hyd2.

Restriction enzyme	Cut site	Temp (°C)	*S.* hyd1	*S.* hyd2
*Bsa*I	5'…GGT CTC (1/5)^…3'	37	1,216 bp 161 bp	No cut
*Sac*I	5'…GAG CT/C…3'	37	No cut	818 bp 550 bp

D. aldrichi flies carrying *Spiroplasma* ald2 were collected in 2012 at Freeman Ranch (Texas State University) in San Marcos, Texas area. Infected isolines FR0512-32, FR0512-07, FR0512-02 and FR0512-66 were generated from single females in the

laboratory. FR0512-02 and FR0512-32 lost *Spiroplasma* infection after few generations. All isolines were maintained in opuntia-banana diet.

D. hydei flies were collected in 2015 in central Mexico (18.91° N; -99.61° W). Infection of females was confirmed in the lab via PCR with *Spiroplasma*-specific primer sets as described above (Table 9). Several isolines were established in the lab from single wild-caught females and non-infected subisolines were obtained originally from infected isolines that lost their infection. Although *D. hydei* can harbor two different *Spiroplasma* strains (hyd1 in the poulsonii clade, and hyd2 in the citri clade) (Mateos et al. 2006), the occurrence of dual infection within a single individual has not been reported. To confirm the presence of hyd2 and absence of hyd1, the PCR product obtained with the 16STF1-16STR1 primers was digested with restriction the enzymes *Bsa*I and *Sac*I (Table 10). The isolines used to test the effect of *Spiroplasma* hyd2 on fecundity were the infected SAG-H28 and uninfected SAG-H25.

To test whether *Spiroplasma* strains moj and ald2 confer protection to their respective natural host species against parasitic wasps (or are capable of killing wasps), we exposed *Spiroplasma*-infected and *Spiroplasma*-free isolines to the attack of two wasp species representing the two families that utilize *Drosophila* larvae as hosts: the figitid *Leptopilina* strain Lh14, which is highly virulent (Schlenke et al. 2007), but susceptible to three poulsonii-clade *Spiroplasma* strains tested to date (Xie et al. 2010; Xie et al. 2014; Haselkorn and Jaenike 2015); and the braconid *Asobara* sp. strain Aw35 collected in San Marcos, Texas (same locality where *D. aldrichi* was collected). This

line is comprised of mostly female individuals and is infected with parthenogenesis-inducing *Wolbachia* (Mateos, pers. comm).

3.2.2 Spiroplasma *protection assays*

Spiroplasma protection assays were performed following the procedures described in (Xie et al. 2010; Xie et al. 2014). Adult female flies from *Spiroplasma*-infected and *Spiroplasma*-free sub-isolines were allowed to oviposit on opuntia-banana medium for 48 hours and transferred to a new vial for a second oviposition; after which they were removed and subjected to DNA extraction and *Spiroplasma*-specific PCR with 16STF1 and 16STR1 primers to confirm infection status. For the *Spiroplasma* treatment, only vials (replicates) in which all such females were *Spiroplasma*-positive were retained. Subsequently, 30 second-instar larvae were collected and transferred to a new vial (i.e., replicate), where they were subjected to the wasp treatments described below.

Protection assays against *L. heterotoma* Lh14 and Aw35 were performed for *D. mojavensis* and *D. aldrichi.* In each assay, *Spiroplasma*-infected and *Spiroplasma*-free isolines, were exposed to either a wasp treatment (Lh14 or Aw35) or no wasp control. Each combined treatment was replicated at least three times (see Results). In each replicate vial, larvae were exposed 24 hours to the attack of Lh14 and Aw35 in a 1:6 wasp:larvae ratio. An additional 12 larvae were added to one replicate from each treatment for dissection after the parasitoid exposure, so as to estimate the wasp oviposition rate. Values of initial larvae, puparia, eclosing adult flies (including sex ratio), and eclosing adult wasps were recorded.

A generalized linear mixed model (GLIMMIX) with binomial distribution was fitted to the raw data for the following: adult flies/initial number of larva; adult flies/puparia; puparia/initial number of larvae; adult wasps/initial number of larva; adult wasps/puparia and failed pupa defined as (Total pupae-Total emerged adults)/Total pupae. The independent variables were *Spiroplasma* infection status (fixed), wasp treatment (wasp and no wasp) and fly strain or isoline (random). Statistical analyses were performed SAS statistical packages through SAS Enterprise Guide v 7.1 and graphs were generated using JMP 12 (SAS Institute Inc., Cary, NC, USA).

3.2.3 Drosophila mojavensis *and* D. hydei *reproductive assays*

Individual crosses of infected and uninfected virgin females and males (4- and 7-days-old for *D. mojavensis* and 3-10 days for *D. hydei*; respectively) were set up (n=148 mating pairs; 24–44 per cross type) in Opuntia-banana media and kept at 25^{o}C and 12:12h light:dark cycle. No other food (e.g. live yeast) was added to the vials during the experiment. Oviposition (number of eggs/embryos laid) and embryo hatching rate were recorded at 24-h-intervals over three consecutive days. Parents were then sacrificed and individually PCR-screened for *Spiroplasma* infection. For each time interval, we conducted a full factorial statistical analysis for female and male infection status in JMP12. Additionally we conducted a one-way analysis of variance of oviposition by cross type with post-hoc tests.

3.3 Results

3.3.1 Test of Spiroplasma *moj-mediated protection in* D. mojavensis *against Lh14 and Aw35*

In the absence of wasps, *Spiroplasma* infection had no effect on mean fly larva-to-adult survivorship, which was ~81.8% (Table 11). Presence of either wasp resulted in high larva-to-adult fly mortality (i.e., < 2% survivorship for all treatments exposed to wasps; Figure 13). For flies exposed to *Asobara*, *Spiroplasma* did not have a significant effect on larva-to-adult fly survivorship. For flies exposed to Lh14, however, *Spiroplasma* had a weak negative effect (P=0.0564) on larva-to-adult fly survivorship. *Spiroplasma* had no significant effect on the success of either wasp (measured as the number of emerging wasps over the initial number of fly larvae). The success of *Asobara* (62.4.%) was higher than that of Lh14 (35.3%) regardless of *Spiroplasma* infection state. Most of the fly mortality attributable to wasp occurred at the pupal stage, as the larva-to-pupa survivorship was relatively high (~83.1% for Aw35 treatments, ~84.7% for Lh14 treatments and ~92.8% for no wasp control). Overall these results indicate that *Spiroplasma* strain moj had no significant effect on the ability of the fly to fight the wasp or on the success of these wasps.

The oviposition rate of *L. heterotoma* Lh14 (proportion of *Drosophila* larvae containing at least one wasp embryo or larva) ranged from 88.89% to100% in *Spiroplasma*-infected vs. *Spiroplasma*-free treatments, respectively. The oviposition rate of *Asobara* sp. Aw35 was 66.7%, in *Spiroplasma*-infected treatments and 100%, *Spiroplasma*-free treatments. Nonetheless, I observed a significantly higher number of

Aw35 embryos per host larva in the *Spiroplasma*-free (mean ~6-7; maximum of 10 wasp embryos in one host) compared to a mean of wasp embryo per host in the *Spiroplasma*-infected treatment (Figure 12).

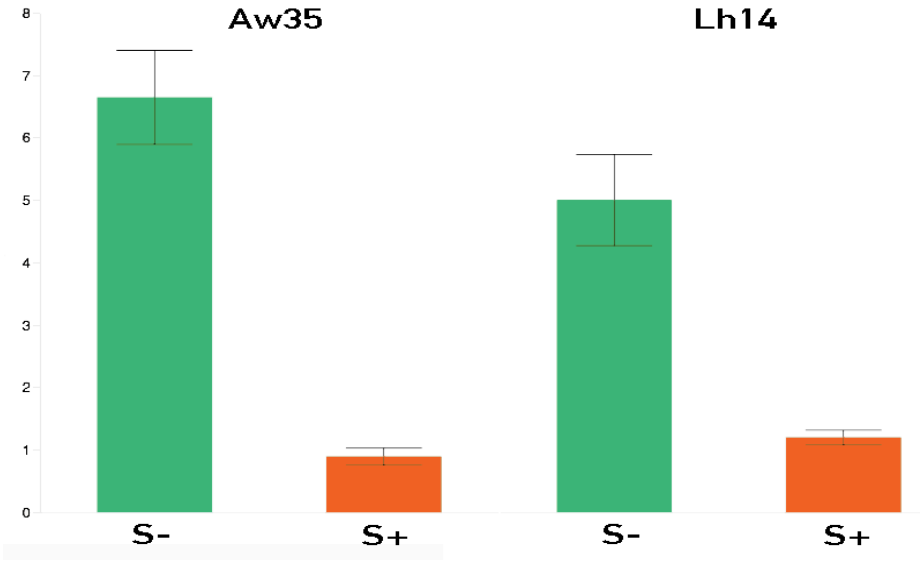


Figure 12. Mean of number of wasp embryos of *Asobara* sp. Aw35 (left) and *L. heterotoma* (right) in non-infected (green bar) and infected (red bar) *D. mojavensis*. *Spiroplasma* infected isolines used were B2, C1 and F2 while uninfected sub-isolines are ABM. Mean ± standard error

Table 11. Means of fitness effect of *D. mojavensis* parasitized by Aw35 and Lh14

	Larva-to-adult fly	Pupa-to-adult fly	Larva-to-adult wasp	Pupa-to-adult wasp	Pupal mortality
Asobara sp.	0.0181	0.0207	0.624	0.759	0.221
L. heterotoma	0.0092	0.0092	0.353	0.406	0.583
Control	0.818	0.881	NA	NA	0.119

3.3.2 Test of Spiroplasma *ald2-mediated protection in* D. aldrichi *against Lh14 and Aw35*

In the absence of wasps, larva-to-adult fly survivorship was lower, albeit not statistically significant (*P*=0.0953), in *Spiroplasma*-infected (mean ~70%) than in *Spiroplasma*-free flies (mean ~50.8%; Figure 14). *Spiroplasma* had a borderline non-significant (P=0.052) negative effect of fly survivorship in the larva-to-pupa stage (mean 73.8% vs. 55% for *Spiroplasma*-free vs. *Spiroplasma*-infected; respectively). Pupa-to-adult fly survivorship was high and not significantly affected by *Spiroplasma* infection state (>90%).

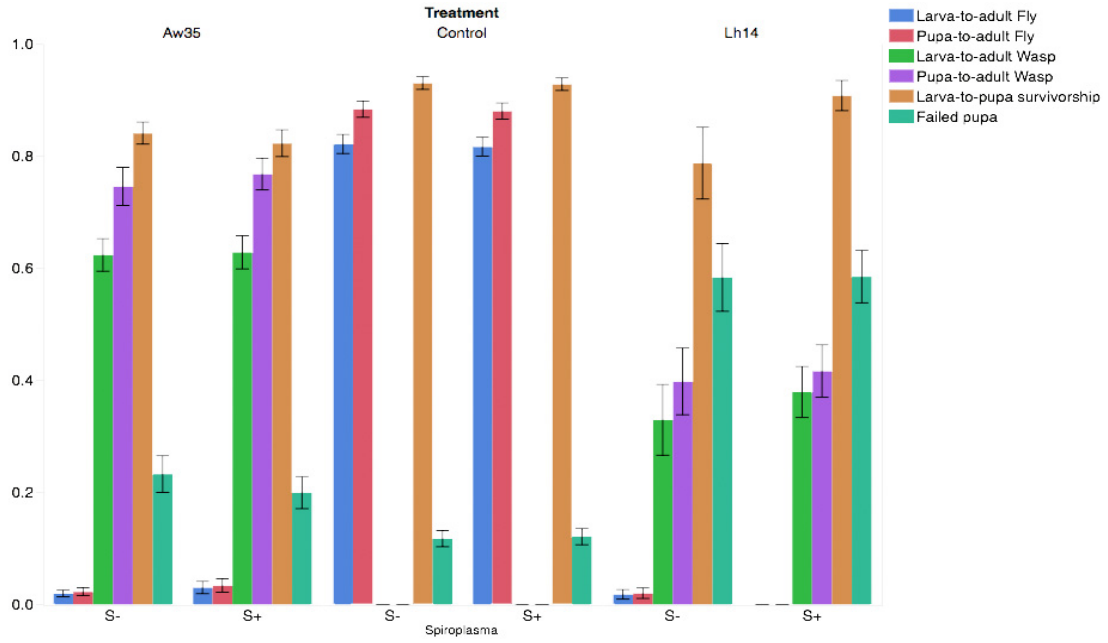


Figure 13. Fitness effects of *Spiroplasma* in *D. mojavensis* isolines exposed to *Asobara* sp, to Lh14 and unexposed controls. *S+* = *Spiroplasma* infected, *S-* = *Spiroplasma* free. S+ indicates infected replicates whereas S- indicates non-infected replicates.

Wasp oviposition of *L. heterotoma* Lh14 in *D. aldrichi* was high in both *Spiroplasma*-free and *Spiroplasma*-infected treatments (96.7% and 100%, respectively). In contrast, oviposition of Aw35 was lower in *Spiroplasma*-free (73.3%) than in *Spiroplasma*-infected (93.3%) flies (non statistically significant). Exposure to either Aw35 or Lh14 drastically reduced survival rate of *D. aldrichi* in both *Spiroplasma*-infected and *Spiroplasma*-free treatments (Figure 14).

In the presence of Aw35, *Spiroplasma* did not have a significant effect on larva-to-adult fly survival (P=0.9116). The success of Aw35 (measured both as larva-to-adult wasp and pupa-to-adult wasp survival) was negatively affected by the presence of *Spiroplasma* in *D. aldrichi*, but this effect was not statistically significant (P=0.2559 and

0.2677; respectively). The higher (albeit non-significant; P=0.1305) proportion of failed pupae in the presence of *Spiroplasma* and Aw35, suggests that most of the *Spiroplasma*-induced Aw35 mortality occurred during the pupal stage.

Spiroplasma appeared to have a negative effect on fly survival in the presence of Lh14 (mean larva-to-adult fly ~29.2% vs. ~4.7% in *Spiroplasma*-free vs. *Spiroplasma*-infected, respectively), but this effect was not statistically significant (P=0.2675). *Spiroplasma* did not have a significant effect on the success of the Lh14 wasp (*P*= 0.7026; Figure 14).

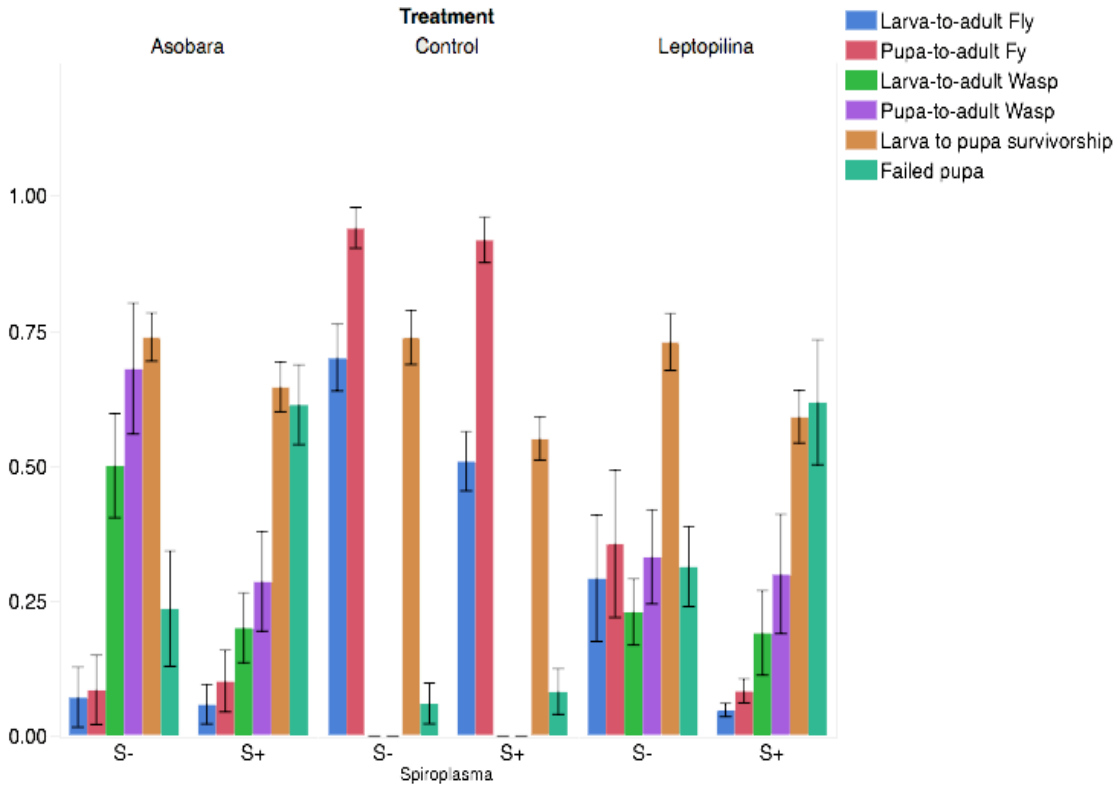


Figure 14. Fitness effects of *Spiroplasma* in *D. aldrichi* isolines exposed to *Asobara* sp, to Lh14 and unexposed controls. *S*+ = *Spiroplasma* infected, *S*- = *Spiroplasma* free.

3.3.3 Assessment of Spiroplasma*-induced reproductive phenotypes in* D. mojavensis *and* D. hydei

Our results indicate that the embryo-hatching rate of *D. mojavensis* is not influenced by the *Spiroplasma* infection state of males and females. Mean hatching rate per cross type ranged from 59.1% to 73.2%, and was not significantly different among the cross types (P=0.4252) (Figure 15). Therefore, we detected no evidence of *Spiroplasma*-induced CI.

A positive effect of *Spiroplasma* moj on oviposition of *D. mojavensis* was detected. *Spiroplasma*-infected females laid a significantly larger number of eggs (mean=16; P=0.0083) during the first 24 h, than their *Spiroplasma*-free counterparts (mean=7.45). A significant difference in oviposition rate was not detected during the second and third 24-h intervals (P=0.73 and P=0.4755, respectively), but total oviposition (i.e., over the entire 72-h) remained significant (P=0.0308). Neither male infection status nor the interaction showed significant effects on oviposition rate (Figure 16). Therefore, *Spiroplasma* moj appears to induce an early oviposition effect that results in higher fecundity over the 3-day period examined. We then tested whether the *Spiroplasma*-induced higher fecundity over this period translates into higher fecundity over a longer period of the fly's live (i.e. 15 days).

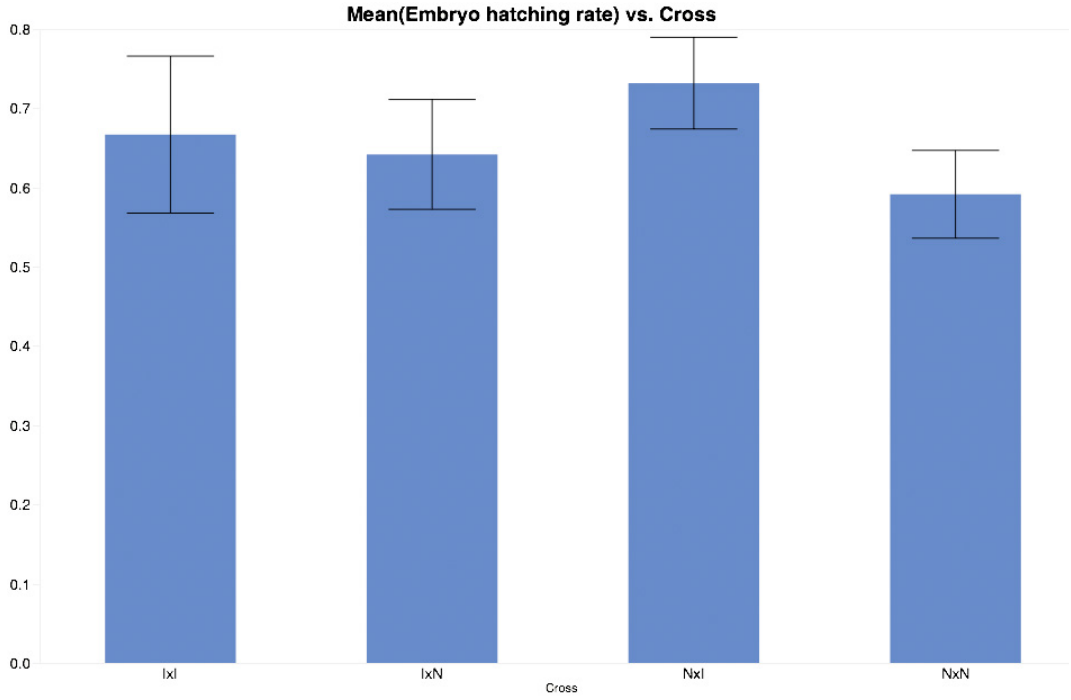


Figure 15. Embryo hatching rate on *D. mojavensis.* Replicates by cross: IxI *n*=24; IxN *n*=55; NxI *n*=55; NxN *n*=94. I: infected; N: non-infected; male and female respectively. One-way ANOVA post-hoc tests (Tukey-Kramer). Error bar constructed from the standard error of the mean.

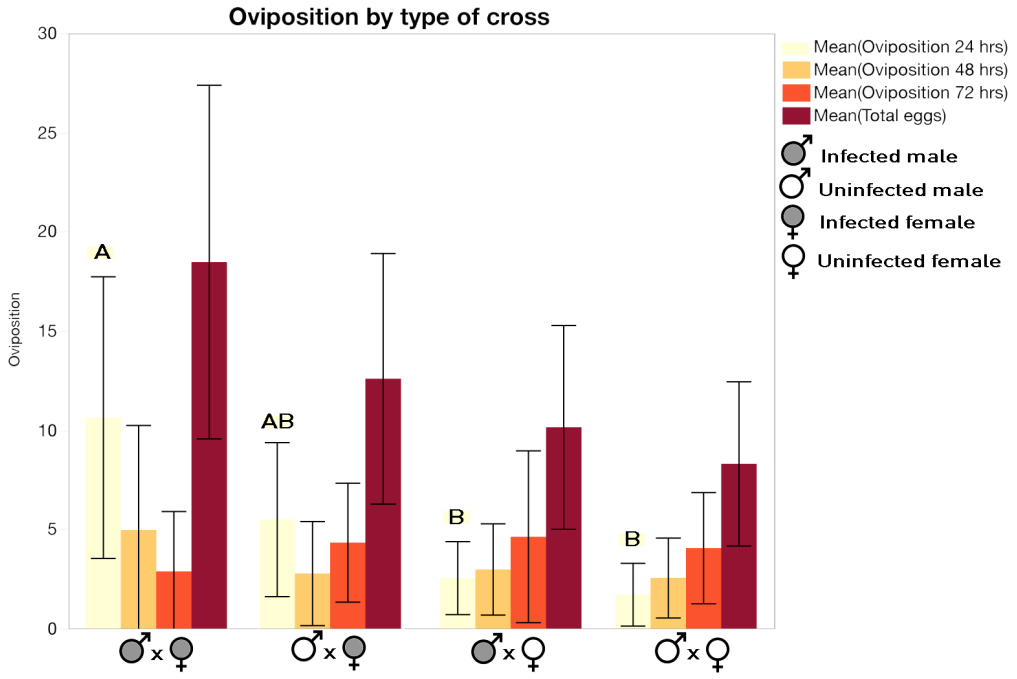


Figure 16. Oviposition mean by cross type in *D. mojavensis*. Error bar constructed using a 95% confidence interval of the mean. Different letters indicate significantly different means in one-way ANOVA post-hoc tests (Tukey-Kramer).

Virgin males and females were used to construct 67 mating pairs representing the crosses IxI, NxI, IxN and NxN (I: *Spiroplasma*-infected; N: *Spiroplasma*-free; male and female respectively). *D. mojavensis* females reached more than 50 days of age under controlled conditions in the laboratory. We observed that female oviposition decreased to zero after day 34. A one-way analysis of variance of oviposition by female infection status with post-hoc tests was performed for the first 15 days of oviposition. Consistent with the 72-hour results, *Spiroplasma*-infected females laid a significantly larger number of eggs (mean=86.8; P=<0.0001) during the first 15 days, than their *Spiroplasma*-free counterparts (mean=22.6). Neither male infection status nor the interaction showed significant effects on oviposition rate.

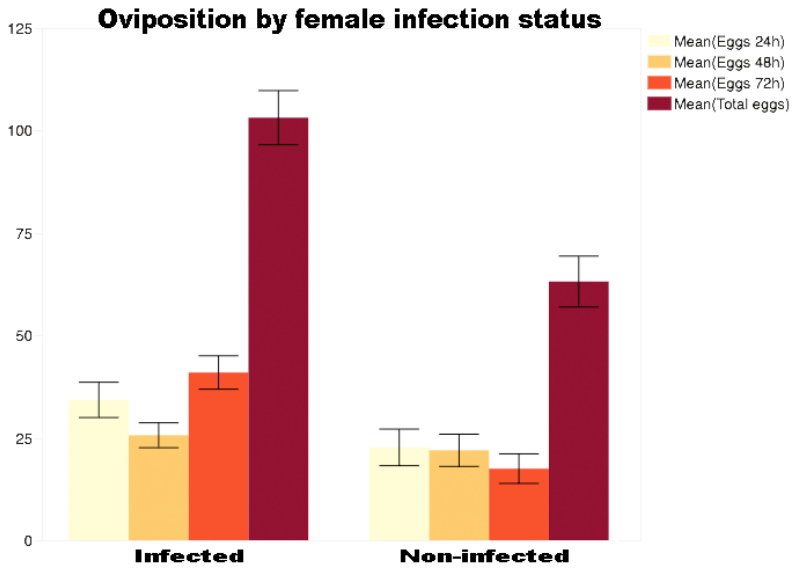


Figure 17. *Drosophila hydei* oviposition mean by female infection status. Error bar constructed using on standard error of the mean.

Similarly, to the results observed in *D. mojavensis*, *Spiroplasma*(hyd2)-infected *D. hydei* females laid a significantly larger number of eggs than their *Spiroplasma*-free counterparts (cumulative mean = 103 vs. 53; respectively; Figure 17). A significant effect of *Spiroplasma* on oviposition was detected for the cumulative 72 h-period (P=<0.0001), as well as during the first and last individual 24 h-periods.

3.4 Discussion

The persistence of heritable facultative endosymbionts in their host populations despite imperfect vertical transmission presents a conundrum, unless one or more of the following conditions are fulfilled to a degree that can counter the loss by imperfect vertical transmission: (1) the symbiont confers a fitness advantage to its host; (2) the symbiont manipulates host reproduction to its benefit (e.g. male killing under certain

circumstances and cytoplasmic incompatibility); and (3) horizontal transmission. Herein we examined aspects of fitness benefits and reproductive manipulation for members of the poorly studied, but prevalent, citri-clade of *Spiroplasma* that associate with *Drosophila* from the repleta species group.

Whereas *Spiroplasma* moj had a neutral effect on larva- and pupa-to-adult survivorship of *D. mojavensis*, *Spiroplasma* ald2 had an apparently negative effect on larva-to-adult survivorship of *D. aldrichi*, suggesting that under these experimental conditions and host genetic backgrounds, *Spiroplasma* is detrimental to *D. aldrichi*. A similar observation was reported for *D. melanogaster* infected by *S. poulsonii* MSRO (Mateos et al. 2016), but not by Xie et al. (2014) under comparable experimental conditions. Few other reports of detrimental effects (aside from male killing) of *Spiroplasma* on *Drosophila* hosts exist. Herren et al. (2014) reported that the longevity of *D. melanogaster* was negatively affected by *S. poulsonii* MSRO, but this effect was detectable only at fly old ages.

The wasp protection assay results indicate that the citri-clade strains moj and ald2 confer no protection to their hosts against the wasp *L. heterotoma* (Lh14). This was reflected by both a lack of *Spiroplasma*-mediated fly rescue and of *Spiroplasma*-mediated wasp death. This is in striking contrast to the reports that *L. heterotoma* (Lh14) is highly susceptible to three poulsonii-clade strains (hyd1, MSRO, neo) in their respective fly hosts (Haselkorn and Jaenike 2015; Paredes et al. 2016; Xie et al. 2010; Xie et al. 2014). Furthermore, the presence of Ribosome Inactivating Protein (RIP) encoding genes (which are involved in protection against nematodes; Hamilton et al.

(2016), in the genome of *Spiroplasma* moj (Chapter 2) raises the possibility that this strain may confer protection against other eukaryotic natural enemies of *D. mojavensis* (e.g. other wasps, nematodes, protozoans). A ~one-week survey of parasitoid wasps throughout Catalina Island yielded no specimens (Mateos, personal communication), but further surveys during different seasons, and for other natural enemies should be performed.

In the presence of the braconid wasp *Asobara sp*. Aw35, neither moj nor ald2 enhanced fly survivorship. Aw35 success, however, was negatively (non-significantly) affected by strain ald2. This finding is interesting because instances of *Spiroplasma*-induced wasp mortality with little to no fly rescue have been reported for poulsonii-clade *Drosophila* combinations involving other wasps (reviewed in (Mateos et al. 2006). It is possible that ald2 exerts a genuine protective effect (i.e., increases fly rescue) against Aw35 under conditions/host backgrounds yet to be tested. Furthermore, whereas Lh14 (a temperate-region cosmopolitan species) does not appear to occur in sympatry with the population of *D. aldrichi* studied herein, Aw35 does, and appears to be relatively host-specific (e.g. to date it has only been found to successfully parasitize *D. aldrichi*, *D. mulleri*, *D. mojavensis*, *D. hydei*, but not *D. melanogaster*; Mateos pers. comm.). Further studies examining a broader set of conditions and host/symbiont/parasitoid backgrounds are needed to rule out ald2-mediated wasp protection against Aw35.

The citri-clade strain tested for cytoplasmic incompatibility (moj) revealed no evidence of this reproductive manipulation phenomenon. This is not surprising given the CI has only been attributed to several members of *Wolbachia* and *Cardinium*

(Werren et al. 2008; Penz et al. 2012). Similarly, none of the citri-clade strains reported to date appear to induce male-killing. Therefore, it appears that persistence of citri-clade strains in *Drosophila* populations is not enhanced by any form of reproductive manipulation.

The oviposition assays, which were performed on moj and hyd2, suggest that citri-clade strains may confer a fitness advantage to their hosts by increasing early and lifetime oviposition. An influence of the poulsonii-clade hyd1 on *D. hydei* oviposition (under similar experimental conditions) has not been detected (Xie et al. 2011). Increased early, but not lifetime, oviposition was reported for the poulsonii-clade (male-killing) strain (WSRO) harbored by *D. willistoni* (Ebbert 1991). An early mating propensity induced by the poulsonii-clade strain (NSRO) harbored by *D. nebulosa* was also reported (Malogolowkin-Cohen and Rodrigues-Pereira 1975)

Overall, the results suggest that citri-clade strains do not utilize cytoplasmic incompatibility or male-killing as a means of enhancing their persistence in host populations. The results tentatively indicate that citri-clade strains induce higher overall fecundity, reflected particularly early in the host's life, and that this mechanism might explain their persistence in host populations. The results also indicate that citri-clade strains do not confer protection to their hosts against parasitoids, but the detrimental effect of one *Spiroplasma* strain (ald2) on one wasp species Aw35, awaits further study. One caveat is that we examined individuals from a single population of each species of fly. Another caveat is that the number of host genetic backgrounds within these populations and conditions examined were limited. It should be noted that host

background (e.g, (Haselkorn 2010b) and conditions (e.g., crowding; (Ebbert 1995) has been shown to influence *Spiroplasma*-induced phenotypes. Unfortunately, experimentation with citri-clade strains has been more challenging than that with poulsonii-clade strains (M. Mateos, pers. comm), due to the difficulty of maintaining and artificially transferring citri-clade strains between hosts (compared to poulsonii-clade strains). Such difficulties might stem from the comparatively lower titers (Haselkorn et al. 2013)and transmission rates exhibited by citri-clade strains. Future studies might benefit from tracking *Spiroplasma* frequencies over time in larger fly populations (e.g. Xie et al. 2015).

CHAPTER IV

SPIROPLASMA CITRI MOJ SEQUENCING PROJECT

4.1 Introduction

The bacterial genus *Spiroplasma* includes several strains that are heritable endosymbionts of several *Drosophila* species and other insects. *Spiroplasma* is wall-less, helical, motile bacteria, phylogenetically classified as gram-positive. *Spiroplasma* belongs to the class Mollicutes (Gasparich 2002; Gasparich et al. 2004). *Spiroplasma* strains are associated intra- and extra-cellularly with a variety of arthropods and plants (Anbutsu and Fukatsu 2011), and transmitted either vertically (in arthropods) and horizontally (mainly in plants; vectored by sap-feeding insects). *Drosophila*-associated strains fall within four separate clades: *poulsonii*; *citri*; *tenebrosa*; and *ixodetis* (Haselkorn 2010a). Within *Drosophila*, 19 species are reported to harbor *Spiroplasma*, and in certain populations of *Drosophila* infection frequencies are relatively high.

Interestingly, *Drosophila* associated species from the repleta group are infected only by *Spiroplasma citri* clade members. This group is comprised of ~100 mostly cactophilic species within the subgenus *Drosophila*, which is highly divergent from the subgenus *Sophophora*, which contains *D. melanogaster* (Pelandakis et al. (1991); O'Grady and Kidwell (2002).

Drosophila mojavensis is a member of the repleta group and its infection with Spiroplasma has been previously documented (Mateos et al. 2006). In nature *Spiroplasma* can reach relatively high prevalence, and has been observed that this

symbiont can reach 60% only in the Catalina Island population (California) (Haselkorn et al. 2009); To date, the fitness consequences of most strains belonging to the *citri* clade remain unknown. Although preliminary work suggests that *citri* clade strains do not kill males, and no strong evidence of protection against parasitoid wasps but is it not completely discarded. Genomic comparison of phylogenetically distant *Spiroplasma* such as S. moj (species will allow the identification of common regions/genes shared between plant and arthropod *Spiroplasma* strains as well the role its extra chromosomal elements have in the *Spiroplasma-Drosophila* interaction.

4.2 Materials and Methods

4.2.1 Spiroplasma citri Moj infected flies

Drosophila mojavensis flies were originally collected in Catalina Island, California in summer 2012. Infection by *Spiroplasma* of collected individuals was confirmed from individual DNA extracts with taxon-specific primers: TKSSp/63F, spoulF/spoulR and 16STF1/16STR1 and universal primers (27F/1492R).

Infected and non-infected subisolines were constructed by offspring selection from a single-infected female. Virgin females were individually placed in vials and allowed to mate with CI-33-15 males. Once larvae hatched, the parental female was sacrificed and a PCR test was done to confirm its *Spiroplasma* infection status. Offspring from negative and positive females was maintained then as CI-33-15 sub-isolines. Six CI-33-15 sub-isolines were constructed and maintained during two years; C1, B2 and F2 were the infected sub-isolines whereas A, B and M were the uninfected

sub-isolines. The subisolines are currently being maintained in the laboratory in a standard opuntia-banana diet.

4.2.2 Library preparation and Illumina sequencing

Spiroplasma Moj DNA was isolated in the laboratory from a single *Drosophila mojavensis* female from the infected subisoline C1. DNA was recovered through the chloroform-ethanol procedure used for *Spiroplasma* hyd1 (Appendix 1) and diluted in AE buffer (Qiagen).

Presence of *Spiroplasma* was confirmed again through PCR. Prior to library preparation, dsDNA quality was examined using Picogreen fluorometric system in the Texas AgriLife Genomics and Bioinformatics Services Facility (College Station, TX). Sample preparation was performed according to Illumina's Multiplexing Sample Preparation Guide. Pair ended library construction and sequencing procedure were performed at the Texas AgriLife Genomics and Bioinformatics Services Facility (College Station, TX).

4.2.3 Genome assembly

Illumina reads quality control was verified in Galaxy Tools to discard those with low quality and with presence of sequencing artifacts. In order to remove reads belonging to the *Drosophila mojavensis* host, remaining "high quality" reads were mapped against an *D. mojavensis* assembly (GenBank AAPU00000000.1) using Bowtie2 v2.1.0 (Langmead and Salzberg 2012). Mapping procedure was performed

under the computational resources of the Texas A&M University Whole Systems Genomics Initiative (WSGI) HPC cluster.

Non-*Drosophila* reads assembly was performed using Velvet v1.0.0 (Zerbino and Birney 2008) in local Galaxy environment (Blankenberg et al. 2010; Goecks et al. 2010), using astringent conditions. Contigs obtained were blasted to the NCBI nucleotide database to locate and remove remaining *Drosophila* sequences. Non-*Drosophila* contigs were scaffolded in Geneious 9.1.5 and the resulting file containing both scaffolds and contigs was locally blasted to *Drosophila mojavensis* to discard presence of *Drosophila* sequences.

4.2.4 Annotation and phylogenetic analyses

Spiroplasma Moj ORFs were identified directly from the identified *Spiroplasma*-like contigs using PRODIGAL (PROkaryotic DynamIc Programming Genefinding Algorithm) v 2.60 (Giardine et al. 2005; Hyatt et al. 2010) and proteins were identified through homolog proteins using the Ensembl Bacteria database through HMMER (Eddy 2011; Finn et al. 2011). Biochemical pathways were identified through the Kyoto Encyclopedia of Genes and Genomes (KEGG).

Phylogenetic comparisons of genome annotations from sequenced and available *Spiroplasma* strains were performed using a RAxML approach with PhyloPhlAn (Segata et al. 2013) and final trees were visualized in FigTree v 1.3.1.

4.3 Results and Discussion

4.3.1 S. citri moj *preliminary assembly*

After sequencing, ~165.6 millions of reads with a length of 100 bp were obtained. 93.77% of the obtained reads showed high homology to the host *D. mojavensis* genome while 0.25% mapped to the closest relative *Spiroplasma kunkelii* genome (Table 12). According to the data obtained and estimating a *S.* Moj genome size of 1.5 Mbp, the estimated coverage was ~70X.

Table 12. Illumina reads generated for the *S. citri* moj sequencing project.

	Total Reads	***D. mojavensis* (Host)**	***P. acnes* (Contaminant)**	***S. poulsonii* MSRO**	***S. kunkelii***
S. moj	165,652,448	155,336,611 (93.77%)	1,261 (0.00%)	935,108 (0.57%)	1,063,945 (0.65%)

Draft *S.* moj assembly resulted in 181 contigs representing 919,378 bp with a G+C content of 27.4%. Contig maximum length was 26,722 bp, (mean 5,079 bp, N50 7,613 bp) (Table 13). BLASTN revealed homologies with previously reported *Spiroplasma* plasmid sequences; however, there is no evidence of circularization and the plasmid sequences identified could be part of the *Spiroplasma* chromosome.

tRNA scan revealed 29 tRNA genes encoding for 18 standard aminoacids plus selenocysteine (Sec) (Appendix 2), histidine (His) and isoleucine (Ile) were not found in our assembly. tRNAs are distributed across six contigs but most of them are present in four clusters. A single operon of the ribosomal gene 16S was found in *S.* moj, which is consistent with previous findings in previously *Spiroplasma* strains.

HMMER ORFs search and Prodigal annotation returned 947 protein-coding genes. 36.4% (345) were hypothetical and/or uncharacterized genes while remaining 63.5% have a well-known function. KEGG-KOALA annotation classified 227 (23.9%) genes related to genetic information, 29 (3.06%) to energy metabolism and 47 (4.9%) to carbohydrate transport and metabolism.

Table 13. Genomic features of *Spiroplasma citri* moj and comparison with *S. poulsonii* MSRO and *S. poulsonii* hyd1.

	***S. poulsonii* MSRO[1]**	***S. poulsonii* hyd1**	***S. citri* moj**
Sequencing technique	PacBio RSII	Hybrid	Illumina
Clade	Poulsonii	Poulsonii	Citri
Number of chromosomal contigs	16	67	181
Combined size of chromosomal contigs	1,632,994	1,401,220	919,378
Estimated chromosome size	1,890,000	Und	Und
Estimated coverage	86.4	1139	70
G+C contents	26.7	27.9	27.4
Protein coding genes	2,042	1,382	903
Plectrovirus proteins	307	31	1
Hypothetical proteins	1,164	622	314
Annotated pseudogenes	34	154	42
rRNA operon	1	1	1
tRNA genes	31	32	29
Number of plasmids	0	>1?	>1?

[1] (Paredes et al. 2015)

CHAPTER V

CONCLUSIONS

The purpose of my dissertation was focused on the heritable endosymbiont *Spiroplasma* and its relationship with its *Drosophila* hosts. In the Chapter 2, I describe the genome sequencing and annotation of the protective strain *S.* hyd1. One of the objectives I had when started the genome sequencing of *Spiroplasma* was to perform comparisons between strains with different effects on their hosts in order to find regions or genes that could be associate to the phenotypic effect. To date, available evidence suggests that protection could be related to one or several cytotoxic ribosomal inactivating proteins (RIP) that are found in a few members of the genus *Spiroplasma.* I found two different RIP genes that could be related to protection of *Spiroplasma* infected *D. hydei* against parasitoids. However additional studies involving strains with functional and non-functional copies of the genes are needed to verify this. I found also several metabolic differences between *S.* hyd1 and the closest sequenced strain *S.* MSRO. It is remarkable that *S.* MSRO lacks a functional copy of the genes *fru*A and *fru*B and in consequence, appears to be incapable of metabolize fructose, while *S.* hyd1 possess a functional copy of those genes and could use fructose as an additional source of carbon. Similarly, arginine could be used as energy source by *S.* hyd1 but not by *S.* MSRO since this strain apparently is missing a functional copy of arcC. As reported previously by Patel *et al.* (1978) and Herren *et al.* (2014), the *Spiroplasma* genus appears to be capable of synthetize cardiolipin. However, our results did not find evidence of a functional copy of the enzyme phosphatidylglycerol phosphate (*pgp*) which is required

in the cardiolipin pathway, but based on previous evidence of highly divergent *pgp* genes in other bacteria grouo, we hypothesized that this gene could be among the ~23 hypothetical proteins with unknown function that are common across all the sequenced strains to date.

We used a total of 398 protein-coding genes for phylogenetic analyses using our sequenced strain and the previously sequenced strains present in the NCBI genomes database. All analyses distinguished three main *Spiroplasma* clades; the Chrysopicola-Mirum-Poulsonii-Citri clade; a clade formed by seven members of the Apis clade; and the lineage of *S. sabaudiense*. Our phylogenetic analysis of Spiroplasma revealed that with few exceptions, *S. chrysopicola* and *S. syrphidicola* were recovered as sister lineages. Similarly, the monophyly of *S. eriocheiris*, *S. mirum* and *S. atripochogonis* (i.e., the Mirum clade) was supported by most analyses. However, the relationships among the Chrysopicola, Mirum and Poulsonii-Citri clades were not clearly resolved. It is possible that horizontally acquired genes are responsible of the observed inconsistencies. On the other hand, relationships among the seven Apis-clade members that consistently formed a monophyly were not fully resolved. To reflect this, we collapsed the base of this clade into a trichotomy formed by: *S. turonicum* + *S. litorale* (clade A); *S. taiwanese* + *S. diminitum* + *S. cantharicola* (clade B); and *S. apis* + *S. culicicola*.

The persistence of heritable facultative endosymbionts in their host populations despite imperfect vertical transmission presents a conundrum, to achieve success, one or more of the following conditions should be fulfilled to a degree that can counter the loss

by imperfect vertical transmission: (1) the symbiont confers a fitness advantage to its host; (2) the symbiont manipulates host reproduction to its benefit; and (3) horizontal transmission. In the chapter 3, I examined aspects of fitness benefits and reproductive manipulation for members of the poorly studied, but prevalent, citri-clade of *Spiroplasma* that associate with *Drosophila* from the repleta species group. Specifically, my work was focused in the protection of *Spiroplasma* ald2 in *D. aldrichi*, *S.* moj in *D. mojavensis* and *S.* hyd2 in *D. hydei* against two species of parasitoid wasps (Lh14 and Aw35). Our conclusions showed that whereas *Spiroplasma* moj had a neutral effect on larva- and pupa-to-adult survivorship of *D. mojavensis*, *Spiroplasma* ald2 had an apparently negative effect on larva-to-adult survivorship of *D. aldrichi*, suggesting that under these experimental conditions and host genetic backgrounds, *Spiroplasma* is detrimental to *D. aldrichi*.

The wasp protection assay results indicate that the citri-clade strains moj and ald2 confer no protection to their hosts against the wasp *L. heterotoma* (Lh14). This was reflected by both a lack of *Spiroplasma*-mediated fly rescue and of *Spiroplasma*-mediated wasp death.

In the presence of the braconid wasp *Asobara sp*. Aw35, neither moj nor ald2 enhanced fly survivorship. Aw35 success, however, was negatively (non-significantly) affected by strain ald2. Further studies examining a broader set of conditions and host/symbiont/parasitoid backgrounds are needed to rule out ald2-mediated wasp protection against Aw35.

We conducted oviposition assays on moj and hyd2, our results suggest that citri-clade strains may confer a fitness advantage to their hosts by increasing early and lifetime oviposition.

Overall, the results suggest that citri-clade strains do not utilize cytoplasmic incompatibility or male-killing as a means of enhancing their persistence in host populations. The results tentatively indicate that citri-clade strains induce higher overall fecundity, reflected particularly early in the host's life, and that this mechanism might explain their persistence in host populations. The results also indicate that citri-clade strains do not confer protection to their hosts against parasitoids, but the detrimental effect of one *Spiroplasma* strain (ald2) on one wasp species Aw35, awaits further study. One caveat is that we examined individuals from a single population of each species of fly.

REFERENCES

Afgan, E., D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Cech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Gruning, A. Guerler, J. Hillman-Jackson, G. Von Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko, and J. Goecks. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 44:W3-W10.

Alexeev, D., E. Kostrjukova, A. Aliper, A. Popenko, N. Bazaleev, A. Tyakht, O. Selezneva, T. Akopian, E. Prichodko, I. Kondratov, M. Chukin, I. Demina, M. Galyamina, D. Kamashev, A. Vanyushkina, V. Ladygina, S. Levitskii, V. Lazarev, and V. Govorun. 2012. Application of *Spiroplasma melliferum* proteogenomic profiling for the discovery of virulence factors and pathogenicity mechanisms in host-associated spiroplasmas. J Proteome Res 11:224-236.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic Local Alignment Search Tool. Journal of Molecular Biology 215:403-410.

Ammar, E.-D. and S. A. Hogenhout. 2006. Mollicutes associated with arthropods and plants. Pp. 97-118 *in* K. Bourtzis, and T. A. Miller, eds. Insect Symbiosys CRC, Boca Raton.

Anbutsu, H. and T. Fukatsu. 2011. *Spiroplasma* as a model insect endosymbiont. Environmental Microbiology Reports 3:144-153.

Anbutsu, H., N. Shibata, N. Nikoh, K. Tanaka, T. Harumoto, T. Nishiyama, S. Shingenbou, M. Hasebe, and T. Fukatsu. 2016. Comparative genomic analysis of male-killing Spiroplasma and it's non male-killing variant in *Drosohpila.* Wolbachia 2016 in the Rain Forest, Lamington Plateau, Queensland, Australia.

Bagga, S., M. V. Hosur, and J. K. Batra. 2003. Cytotoxicity of ribosome-inactivating protein saporin is not mediated through alpha(2)-macroglobulin receptor. Febs Letters 541:16-20.

Bischof, D. F., E. M. Vilei, and J. Frey. 2009. Functional and antigenic properties of GlpO from Mycoplasma mycoides subsp. mycoides SC: characterization of a flavin adenine dinucleotide-binding site deletion mutant. Vet Res 40:35.

Blankenberg, D., G. Von Kuster, N. Coraor, G. Ananda, R. Lazarus, M. Mangan, A. Nekrutenko, and J. Taylor. 2010. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol Chapter 19:Unit 19 10 11-21.

Bolanos, L. M., L. E. Servin-Garciduenas, and E. Martinez-Romero. 2015. Arthropod-*Spiroplasma* relationship in the genomic era. FEMS Microbiol Ecol 91:1-8.

Brownlie, J. C., B. N. Cass, M. Riegler, J. J. Witsenburg, I. Iturbe-Ormaetxe, E. A. McGraw, and S. L. O'Neill. 2009. Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipientis*, during periods of nutritional stress. PLoS Pathogens 5:e1000368.

Bull, J. J. 1983. Evolution of sex determining mechanisms. Benjamin/Cummings Pub. Co.

Bull, J. J. and M. Turelli. 2013. *Wolbachia* versus dengue: Evolutionary forecasts. Evolution Medicine and Public Health 2013:197-207.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST plus : Architecture and applications. BMC Bioinformatics 10:1-9.

Carle, P., F. Laigret, J. G. Tully, and J. M. Bove. 1995. Heterogeneity of genome sizes within the genus *Spiroplasma*. International Journal of Systematic Bacteriology 45:178-181.

Carle, P., C. Saillard, N. Carrere, S. Carrere, S. Duret, S. Eveillard, P. Gaurivaud, G. Gourgues, J. Gouzy, P. Salar, E. Verdin, M. Breton, A. Blanchard, F. Laigret, J. M. Bove, J. Renaudin, and X. Foissac. 2010. Partial chromosome sequence of *Spiroplasma citri* reveals extensive viral invasion and important gene decay. Applied and Environmental Microbiology 76:3420-3426.

Chang, T. H., W. S. Lo, C. Ku, L. L. Chen, and C. H. Kuo. 2014. Molecular evolution of the substrate utilization strategies and putative virulence factors in mosquito-associated *Spiroplasma* species. Genome Biology and Evolution 6:500-509.

Chen, X. A., S. Li, and S. Aksoy. 1999. Concordant evolution of a symbiont with its host insect species: Molecular phylogeny of genus Glossina and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. Journal of Molecular Evolution 48:49-58.

Cheng, B., N. Kuppanda, J. C. Aldrich, O. S. Akbari, and P. M. Ferree. 2016. Male-Killing *Spiroplasma* alters behavior of the dosage compensation complex during *Drosophila melanogaster* Embryogenesis. Curr Biol 26:1339-1345.

Clark, E. L., A. J. Karley, and S. F. Hubbard. 2010. Insect endosymbionts: manipulators of insect herbivore trophic interactions? Protoplasma 244:25-51.

Clark, M. A., N. A. Moran, P. Baumann, and J. J. Wernegreen. 2000. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of Aphids (Uroleucon) and Pitfalls of testing for phylogenetic congruence. Evolution 54:517-525.

Clay, K. 2014. Defensive symbiosis: A microbial perspective. Functional Ecology 28:293-298.

Contreras-Moreira, B. and P. Vinuesa. 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microb 79:7696-7701.

Davis, R. E., J. Shao, E. L. Dally, Y. Zhao, G. E. Gasparich, B. J. Gaynor, J. C. Athey, N. A. Harrison, and N. Donofrio. 2015a. Complete genome sequence of *Spiroplasma kunkelii* strain CR2-3x, causal agent of corn stunt disease in *Zea mays* L. Genome Announc 3.

Davis, R. E., J. Shao, Y. Zhao, G. E. Gasparich, B. J. Gaynor, and N. Donofrio. 2015b. Complete genome sequence of *Spiroplasma turonicum* strain Tab4cT, a parasite of a Horse Fly, Haematopota sp. (Diptera: Tabanidae). Genome Announc 3.

de Virgilio, M., A. Lombardi, R. Caliandro, and M. S. Fabbrini. 2010. Ribosome-Inactivating Proteins: From Plant Defense to Tumor Attack. Toxins 2:2699-2737.

Donmez, N. and M. Brudno. 2013. SCARPA: Scaffolding reads with practical algorithms. Bioinformatics 29:428-434.

Douglas, A. E. 1989. Mycetocyte symbiosis in insects. Biological Reviews of the Cambridge Philosophical Society 64:409-434.

Douglas, A. E. 2009. The microbial dimension in insect nutritional ecology. Functional Ecology 23:38-47.

Ebbert, M. A. 1991. The Interaction phenotype in the *Drosophila Willistoni*-*Spiroplasma* symbiosis. Evolution 45:971-988.

Ebbert, M. A. 1995. Variable effects of crowding on *Drosophila* hosts of male-lethal and non-male-lethal *Spiroplasmas* in laboratory populations. Heredity 74 (Pt 3):227-240.

Eddy, S. R. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol 7:e1002195.

Edgar, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:1-19.

Fermani, S., G. Falini, A. Ripamonti, L. Polito, F. Stirpe, and A. Bolognesi. 2005. The 1.4 anstroms structure of dianthin 30 indicates a role of surface potential at the active site of type 1 ribosome inactivating proteins. J Struct Biol 149:204-212.

Ferrari, J. and F. Vavre. 2011. Bacterial symbionts in insects or the story of communities affecting communities. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 366:1389-1400.

Finn, R. D., J. Clements, and S. R. Eddy. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39:W29-37.

Fukatsu, T. and N. Nikoh. 2000. Endosymbiotic microbiota of the bamboo pseudococcid *Antonina crawii* (Insecta, Homoptera). Applied and Environmental Microbiology 66:643–650.

Gasparich, G. E. 2002. *Spiroplasmas*: Evolution, adaptation and diversity. Frontiers in Bioscience 7:D619-D640.

Gasparich, G. E., R. F. Whitcomb, D. Dodge, F. E. French, J. Glass, and D. L. Williamson. 2004. The Genus *Spiroplasma* and its non-helical descendants: phylogenetic classification, correlation with phenotype and roots of the *Mycoplasma mycoides* clade. International Journal of Systematic and Evolutionary Microbiology 54:893-918.

Giardine, B., C. Riemer, R. C. Hardison, R. Burhans, L. Elnitski, P. Shah, Y. Zhang, D. Blankenberg, I. Albert, J. Taylor, W. Miller, W. J. Kent, and A. Nekrutenko. 2005. Galaxy: A platform for interactive large-scale genome analysis. Genome Research 15:1451-1455.

Goecks, J., A. Nekrutenko, J. Taylor, and T. Galaxy. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11:R86.

Hackett, K. J., D. E. Lynn, D. L. Williamson, A. S. Ginsberg, and R. F. Whitcomb. 1986. Cultivation of the *Drosophila* sex-ratio *Spiroplasma*. Science 232:1253-1255.

Hames, C., S. Halbedel, M. Hoppert, J. Frey, and J. Stulke. 2009. Glycerol metabolism is important for cytotoxicity of Mycoplasma pneumoniae. J Bacteriol 191:747-753.

Hamilton, P. T., J. S. Leong, B. F. Koop, and S. J. Perlman. 2014. Transcriptional Responses in a *Drosophila* defensive symbiosis. Molecular Ecology 23:1558-1570.

Hamilton, P. T., F. Peng, M. J. Boulanger, and S. J. Perlman. 2016. A ribosome-inactivating protein in a *Drosophila* defensive symbiont. Proc Natl Acad Sci U S A 113:350-355.

Harris, H. L., L. J. Brennan, B. A. Keddie, and H. R. Braig. 2010. Bacterial symbionts in insects: Balancing life and death. Symbiosis 51:37-53.

Harumoto, T., H. Anbutsu, and T. Fukatsu. 2014. Male-Killing *Spiroplasma* induces sex-specific cell death via host apoptotic pathway. Plos Pathogens 10.

Haselkorn, T. S. 2010a. The *Spiroplasma* heritable bacterial endosymbiont of *Drosophila*. Fly 4:80-87.

Haselkorn, T. S. 2010b. Understanding the distribution of the *Spiroplasma* heritable bacterial endosymbiont in *Drosophila*. Biology. University of Californa, San Diego, San Diego.

Haselkorn, T. S. and J. Jaenike. 2015. Macroevolutionary persistence of heritable endosymbionts: Acquisition, retention and expression of adaptive phenotypes in *Spiroplasma.* Mol Ecol 24:3752-3765.

Haselkorn, T. S., T. A. Markow, and N. A. Moran. 2009. Multiple introductions of the *Spiroplasma* bacterial endosymbiont into *Drosophila*. Molecular Ecology 18:1294-1305.

Haselkorn, T. S., T. D. Watts, and T. A. Markow. 2013. Density dynamics of diverse *Spiroplasma* strains naturally infecting different species of *Drosophila*. Fly (Austin) 7:204-210.

Hedges, L. M., J. C. Brownlie, S. L. O'Neill, and K. N. Johnson. 2008. *Wolbachia* and virus protection in insects. Science 322:702-.

Herren, J. K., J. C. Paredes, F. Schupfer, K. Arafah, P. Bulet, and B. Lemaitre. 2014. Insect endosymbiont proliferation is limited by lipid availability. Elife 3.

Hoffmann, A. A., D. J. Clancy, and E. Merton. 1994. Cytoplasmic incompatibility in Australian populations of *Drosophila melanogaster.* Genetics 136:993-999.

Hoffmann, A. A., M. Turelli, and G. M. Simmons. 1986. Unidirectional incompatibility between populations of *Drosophila simulans.* Evolution 40:692-701.

Hosokawa, T., R. Koga, Y. Kikuchi, X. Y. Meng, and T. Fukatsu. 2010. *Wolbachia* as a bacteriocyte-associated nutritional mutualist. Proceedings of the National Academy of Sciences of the United States of America 107:769-774.

Huelsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Hussain, M., G. J. Lu, S. Torres, J. H. Edmonds, B. H. Kay, A. A. Khromykh, and S. Asgari. 2013. Effect of *Wolbachia* on replication of West Nile Virus in a mosquito cell line and adult mosquitoes. Journal of Virology 87:851-858.

Hyatt, D., G. L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer, and L. J. Hauser. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

Jaenike, J., R. Unckless, S. N. Cockburn, L. M. Boelio, and S. J. Perlman. 2010. Adaptation via symbiosis: Recent spread of a *Drosophila* defensive symbiont. Science 329:212-215.

Joshi, B. D., M. Berg, J. Rogers, J. Fletcher, and U. Melcher. 2005. Sequence comparisons of plasmids pBJS-O of *Spiroplasma citri* and pSKU146 of *Spiroplasma kunkelii*: Implications for plasmid evolution. Bmc Genomics 6.

Kageyama, D., H. Anbutsu, M. Watada, T. Hosokawa, M. Shimada, and T. Fukatsu. 2006. Prevalence of a non-male-killing *Spiroplasma* in natural populations of *Drosophila hydei*. Applied and Environmental Microbiology 72:6667-6673.

Kanehisa, M. and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28:27-30.

Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. 2016a. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research 44:D457-D462.

Kanehisa, M., Y. Sato, and K. Morishima. 2016b. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. Journal of Molecular Biology 428:726-731.

Kristensen, D. M., L. Kannan, M. K. Coleman, Y. I. Wolf, A. Sorokin, E. V. Koonin, and A. Mushegian. 2010. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. Bioinformatics 26:1481-1487.

Ku, C., W. S. Lo, L. L. Chen, and C. H. Kuo. 2013. Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in *Spiroplasma*. Genome Biol Evol 5:1151-1164.

Ku, C., W. S. Lo, L. L. Chen, and C. H. Kuo. 2014. Complete genome sequence of *Spiroplasma apis* B31T (ATCC 33834), a bacterium associated with May Disease of Honeybees (Apis mellifera). Genome Announc 2.

Kuriwada, T., T. Hosokawa, N. Kumano, K. Shiromoto, D. Haraguchi, and T. Fukatsu. 2010. Biological role of *Nardonella* endosymbiont in its Weevil Host. Plos One 5.

Lane, D. J. 1991. 16S/23S rRNA sequencing. Pp. 115-175 *in* E. Stackebrandt, and M. Goodfellow, eds. Nucleic acid techniques in bacterial systematics. John Wiley and Sons, New York, NY.

Langmead, B. and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359.

Levenbook, L. 1947. Fructose and the reducing value of insects blood. Nature 160:465-465.

Li, L., C. J. Stoeckert, and D. S. Roos. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Research 13:2178-2189.

Lo, W. S., L. L. Chen, W. C. Chung, G. E. Gasparich, and C. H. Kuo. 2013a. Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a Honeybee-Associated Bacterium. BMC Genomics 14.

Lo, W. S., G. E. Gasparich, and C. H. Kuo. 2015. Found and lost: The fates of horizontally acquired genes in arthropod-symbiotic *Spiroplasma*. Genome Biology and Evolution 7:2458-2472.

Lo, W. S., Y. Y. Huang, and C. H. Kuo. 2016. Winding paths to simplicity: genome evolution in facultative insect symbionts. FEMS Microbiol Rev.

Lo, W. S., C. Ku, L. L. Chen, T. H. Chang, and C. H. Kuo. 2013b. Comparison of metabolic capacities and inference of gene content evolution in mosquito-associated Spiroplasma diminutum and S. taiwanense. Genome Biol Evol 5:1512-1523.

Lowe, T. M. and S. R. Eddy. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25:955-964.

Malogolowkin-Cohen, C. and M. A. Q. Rodrigues-Pereira. 1975. Sexual drive of normal and SR flies of Drosophila nebulosa. Evolution 29:579-580.

Markow, T. A. and P. M. O'Grady. 2006. *Drosophila*. A guide to species identification and use. Elsevier.

Mateos, M., S. J. Castrezana, B. J. Nankivell, A. M. Estes, T. A. Markow, and N. A. Moran. 2006. Heritable endosymbionts of *Drosophila*. Genetics 174:363-376.

Mateos, M., L. Winter, C. Winter, V. M. Higareda-Alvear, E. Martinez-Romero, and J. Xie. 2016. Independent origins of resistance or susceptibility of parasitic wasps to a defensive symbiont. Ecol Evol 6:2679-2687.

Merzendorfer, H. and L. Zimoch. 2003. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. J Exp Biol 206:4393-4412.

Miller, M. A., Pfeiffer, W., Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Pp. 1-8. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA.

Montenegro, H., A. S. Petherwick, G. D. Hurst, and L. B. Klaczko. 2006. Fitness effects of *Wolbachia* and *Spiroplasma* in *Drosophila melanogaster*. Genetica 127:207-215.

Moran, N. A., J. P. McCutcheon, and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. Annual Review of Genetics 42:165-190.

Nappi, A., M. Poirié, and Y. Carton. 2009. The role of melanization and cytotoxic by-products in the cellular immune response of *Drosophila* against parasitic wasps. Pp. 99-121. Adv. Parasitol.

O'Grady, P. M. and M. G. Kidwell. 2002. Phylogeny of the subgenus sophophora (Diptera: drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. Molecular Phylogenetics and Evolution 22:442-453.

O'Neill, S. L. and T. L. Karr. 1990. Bidirectional Incompatibility Between Conspecific Populations of *Drosophila simulans*. Nature 348:178-180.

Oliver, K. M., J. A. Russell, N. A. Moran, and M. S. Hunter. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proceedings of the National Academy of Sciences of the United States of America 100:1803-1807.

Palm, W., J. L. Sampaio, M. Brankatschk, M. Carvalho, A. Mahmoud, A. Shevchenko, and S. Eaton. 2012. Lipoproteins in *Drosophila melanogaster* assembly, function, and influence on tissue lipid composition. Plos Genetics 8.

Paredes, J. C., J. K. Herren, F. Schupfer, and B. Lemaitre. 2016. The Role of lipid competition for endosymbiont-mediated protection against parasitoid wasps in *Drosophila*. MBio 7.

Paredes, J. C., J. K. Herren, F. Schupfer, R. Marin, S. Claverol, C. H. Kuo, B. Lemaitre, and L. Beven. 2015. Genome sequence of the *Drosophila melanogaster* Male-Killing *Spiroplasma* strain MSRO Endosymbiont. Mbio 6.

Patel, K. R., P. F. Smith, and W. R. Mayberry. 1978. Comparison of lipids from *Spiroplasma citri* and Corn Stunt *Spiroplasma*. Journal of Bacteriology 136:829-831.

Pelandakis, M., D. G. Higgins, and M. Solignac. 1991. Molecular phylogeny of the subgenus sophophora of *Drosophila* derived from large subunit of ribosomal RNA sequences. Genetica 84:87-94.

Penz, T., S. Schmitz-Esser, S. E. Kelly, B. N. Cass, A. Muller, T. Woyke, S. A. Malfatti, M. S. Hunter, and M. Horn. 2012. Comparative genomics suggests an independent origin of cytoplasmic incompatibility in Cardinium hertigii. PLoS Genet 8:e1003012.

Pontes, M. H. and C. Dale. 2006. Culture and manipulation of insect facultative symbionts. Trends in Microbiology 14:406-412.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2-approximately maximum-likelihood trees for large alignments. Plos One 5.

Razin, S., D. Yogev, and Y. Naot. 1998. Molecular biology and pathogenicity of mycoplasmas. Microbiol Mol Biol Rev 62:1094-1156.

Renaudin, J., M. C. Pascarel, and J. M. Bove. 1987a. Cloning and sequencing of the genome of *Spiroplasma* Virus-4. Israel Journal of Medical Sciences 23:427-428.

Renaudin, J., M. C. Pascarel, and J. M. Bove. 1987b. *Spiroplasma* Virus-4 - nucleotide-sequence of the Viral-DNA, Regulatory Signals, and Proposed Genome Organization. Journal of Bacteriology 169:4950-4961.

Rissman, A. I., B. Mau, B. S. Biehl, A. E. Darling, J. D. Glasner, and N. T. Perna. 2009. Reordering contigs of draft genomes using the Mauve Aligner. Bioinformatics 25:2071-2073.

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

Saillard, C., P. Carle, S. Duret-Nurbel, R. Henri, N. Killiny, S. Carrere, J. Gouzy, J. M. Bove, J. Renaudin, and X. Foissac. 2008. The abundant extrachromosomal DNA content of the *Spiroplasma citri* GII3-3X genome. BMC Genomics 9.

Schlenke, T. A., J. Morales, S. Govind, and A. G. Clark. 2007. Contrasting infection strategies in generalist and specialist wasp parasitoids of *Drosophila melanogaster*. Plos Pathogens 3:1486-1501.

Schwarz, R. S., E. W. Teixeira, J. P. Tauber, J. M. Birke, M. F. Martins, I. Fonseca, and J. D. Evans. 2014. Honey bee colonies act as reservoirs for two *Spiroplasma* facultative symbionts and incur complex, multiyear infection dynamics. Microbiologyopen 3:341-355.

Segata, N., D. Bornigen, X. C. Morgan, and C. Huttenhower. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nature Communications 4.

Snipen, L., T. Almoy, and D. W. Ussery. 2009. Microbial comparative pan-genomics using binomial mixture models. BMC Genomics 10:385.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Tatusova, T., S. Ciufo, B. Fedorov, K. O'Neill, and I. Tolstoy. 2014. RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Research 42:D553-D559.

Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, R. T. Deboy, T. M. Davidsen, M. Mora, M. Scarselli, I. Margarit y Ros, J. D. Peterson, C. R. Hauser, J. P. Sundaram, W. C. Nelson, R. Madupu, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, S. A. Sullivan, S. C. Daugherty, D. H. Haft, J. Selengut, M. L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K. J. O'Connor, S. Smith, T. R. Utterback, O. White, C. E. Rubens, G. Grandi, L. C. Madoff, D. L. Kasper, J. L. Telford, M. R. Wessels, R. Rappuoli, and C. M. Fraser. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A 102:13950-13955.

Van Dongen, S. 2000. Graph clustering by flow simulation. University of Utrecht, The Netherlands.

Walsh, M. J., J. E. Dodd, and G. M. Hautbergue. 2013. Ribosome-inactivating proteins Potent poisons and molecular tools. Virulence 4:774-784.

Watts, T., T. S. Haselkorn, N. A. Moran, and T. A. Markow. 2009. Variable incidence of *Spiroplasma* infections in natural populations of *Drosophila* species. Plos One 4.

Werren, J. H., L. Baldo, and M. E. Clark. 2008. *Wolbachia*: Master manipulators of invertebrate biology. Nature Reviews Microbiology 6:741-751.

Williamson, D. L., B. Sakaguchi, K. J. Hackett, R. F. Whitcomb, J. G. Tully, P. Carle, J. M. Bove, J. R. Adams, M. Konai, and R. B. Henegar. 1999. *Spiroplasma poulsonii* sp. nov., a new species associated with male-lethality in *Drosophila willistoni,* a neotropical species of fruit fly. International Journal of Systematic Bacteriology 49:611-618.

Williamson, D. L., T. Steiner, and G. J. McGarrity. 1983. *Spiroplasma* taxonomy and identification of the sex ratio organisms: can they be cultivated? Yale J Biol Med 56:583-592.

Woese, C. R. 1987. Bacterial evolution. Microbiol Rev 51:221-271.

Wyatt, G. R. 1961. Biochemistry of insect hemolymph. Annual Review of Entomology 6:75-102.

Xie, J., S. Butler, G. Sanchez, and M. Mateos. 2014. Male killing *Spiroplasma* protects Drosophila melanogaster against two parasitoid wasps. Heredity 112:399-408.

Xie, J., B. Tiner, I. Vilchez, and M. Mateos. 2011. Effect of the *Drosophila* endosymbiont *Spiroplasma* on parasitoid wasp development and on the reproductive fitness of wasp-attacked fly survivors. Evolutionary Ecology 53:1065-1079.

Xie, J. L., I. Vilchez, and M. Mateos. 2010. *Spiroplasma* bacteria enhance survival of *Drosophila hydei* attacked by the parasitic wasp *Leptopilina heterotoma*. PLoS One 5:1-7.

Ye, F. C., J. Renaudin, J. M. Bove, and F. Laigret. 1994. Cloning and sequencing of the replication origin (Oric) of the *Spiroplasma citri* chromosome and construction of autonomously replicating artificial plasmids. Current Microbiology 29:23-29.

Zabalou, S., M. Riegler, M. Theodorakopoulou, C. Stauffer, C. Savakis, and K. Bourtzis. 2004. *Wolbachia*-induced cytoplasmic incompatibility as a means for insect pest population control. Proceedings of the National Academy of Sciences of the United States of America 101:15042-15045.

Zele, F., A. Nicot, A. Berthomieu, M. Weill, O. Duron, and A. Rivero. 2014. *Wolbachia* increases susceptibility to *Plasmodium* infection in a natural system. Proceedings of the Royal Society B-Biological Sciences 281.

Zerbino, D. R. and E. Birney. 2008. Velvet: Algorithms for *de novo* short read assembly using De Bruijn graphs. Genome Research 18:821-829.

APPENDIX

Supplementary protocol 1

Chloroform-Ethanol DNA purification

Reagents:

Extraction buffer (described by volume for 1 sample)

Distilled water 0.85 mL

0.5 M EDTA 0.1 mL

10% SDS (Sodium dodecyl sulfate, vendor VWR) 0.05 mL

Potassium acetate 3M pH 4.2 (adjust pH with Acetic Acid glacial)

Chloroform

Isopropanol 100%

Ethanol 70%

Procedure

1. Add 1 mL of extraction buffer to each sample, homogenize and mix well

2. Put in 72^{o} C water bath for 12 minutes, vortex and put in another 12 minutes

Add 2 ul of RNase and incubate 5 minutes.

Spin for 1 min (max speed)

3. Transfer supernatant to a new microcentrifuge tube with 50 µl of potassium acetate and mix well

4. Incubate on ice for 12 minutes, vortex and incubate another 12 minutes on ice

5. Spin 12 minutes at top speed (>13000 rpm)

6. Transfer supernatant into microcentrifuge tube with 500 µl of chloroform

7. Vortex well

8. Spin 3 minutes at top speed

9. Transfer 750 µl of the top phase into a new eppendorf with 500 µl of 100% IsoOH (mixing by inverting six times)

10. Spin 6 minutes at top speed

11. Discard supernatant

12. Add 1 mL of 70% EtOH

13. Spin 6 minutes at top speed

14. Discard supernatant

15. Tap dry on paper towel

16. Dry in Speed Vac for 12 minutes

17. Dissolve DNA in 100 µl TE buffer

Supplementary protocol 2

Maximum likelihood and Bayesian analyses were performed through CIPRES Science Gateway portal (https://www.phylo.org)

RAxML-HPC BlackBox – Parameters

RAxML v 8.2.9
Sequence Type = Protein
Maximum Hours to Run = 0.25
Protein Substitution Matrix = JTT
Find best tree using maximum likelihood search = True
Print branch lengths = True
Do not use BFGS searching algorithm = False
Use Bootstopping = True

MrBayes on XSEDE – Parameters

MrBayes v 3.2.6
Data Type = Protein
Maximum hours to Run = 2.0
Number of generations = 100000000
Number of runs = 4
Number of chains = 4
Print Branch Lenghts = Yes
Markov chain sampled = every 1000

Table S1. tRNAs encoded by *Spiroplasma poulsonii* hyd1

Seq. Name	tRNA begin	Bounds End	tRNA type	Cove score
Contig-12	87770	87843	AsnGTT	783
Contig-12	87660	87733	GluTTC	685
Contig-12	87580	87653	ValTAC	766
Contig-12	87490	87563	ThrTGT	762
Contig-12	87406	87479	LysTTT	725
Contig-12	87303	87385	LeuTAG	595
Contig-12	72846	72920	IleGAT	816
Contig-12	72754	72827	AlaTGC	716
Contig-38	79261	79334	SeCTCA	711
Contig-38	79092	79164	TrpCCA	683
Contig-38	78986	79076	SerCGA	568
Contig-39	97972	98045	HisGTG	626
Contig-39	1570	1643	HisGTG	626
Contig-43	16198	16271	CysGCA	619
Contig-43	16312	16386	ArgACG	717
Contig-43	16412	16486	ProTGG	762
Contig-43	16504	16577	AlaTGC	710
Contig-43	16614	16688	MetCAT	732
Contig-43	16692	16766	MetCAT	845
Contig-43	16800	16890	SerTGA	605
Contig-43	16949	17022	MetCAT	727
Contig-43	17027	17100	AspGTC	733
Contig-43	17108	17181	PheGAA	757
Contig-43	10705	10787	LeuCAA	590

Table S1. Continued

Contig-44	14783	14871	SerGCT	591
Contig-48	25473	25547	ArgTCT	710
Contig-51	9129	9202	ThrTGT	765
Contig-51	9210	9291	TyrGTA	587
Contig-51	9309	9381	GlnTTG	665
Contig-51	9391	9464	LysTTT	725
Contig-51	9483	9567	LeuTAA	648
Contig-51	9647	9718	GlyTCC	734

Table S2. tRNAs encoded by *Spiroplasma citri* Moj

Seq. Name	tRNA begin	Bounds End	tRNA type	Cove score
contig_26	13354	13264	SerGCT	759
contig_26	6758	6683	SeCTCA	839
contig_26	6601	6527	TrpCCA	699
contig_26	6515	6423	SerCGA	643
contig_71	7109	7025	LeuCAA	681
contig_79	5318	5242	ArgTCT	809
contig_121	236	311	AsnGTT	902
contig_121	346	421	GluTTC	766
contig_121	426	501	ValTAC	935
contig_121	518	593	ThrTGT	871
contig_121	602	677	LysTTT	900
contig_121	695	779	LeuTAG	743
contig_131	4014	3939	CysGCA	712
contig_131	3897	3821	ArgACG	807
contig_131	3798	3722	ProTGG	879
contig_131	3705	3630	AlaTGC	940
contig_131	3598	3522	MetCAT	846
contig_131	3520	3444	MetCAT	992
contig_131	3411	3319	SerTGA	775
contig_131	3267	3193	MetCAT	843
contig_131	3190	3115	AspGTC	759
contig_131	3109	3034	PheGA	860
contig_132	1316	1391	ValTAC	792
contig_132	1404	1479	ThrTGT	882

Table S2. Continued

contig_132	1486	1569	TyrGTA	710
contig_132	1585	1659	GlnTTG	757
contig_132	1664	1739	LysTTT	842
contig_132	1756	1842	LeuTAA	753
contig_132	1929	2002	GlyTCC	758

Table S3. Carbohydrates metabolized by the *Spiroplasma* strains analyzed. Clade abbreviations, A: apis; M: mirum; Ch: chrysopicola; C: citri; P: poulsonii. Carbohydrates abbreviations, Gluc: glucose; MurNAc: N-acetylmuramic acid; GlcNAc; *N*-Acetylglucosamine; Malt: maltose; Tre: trehalose; Fru: fructose.

	Clade	Strain	Gluc	Mur NAc	GlcNAc	Malt	Tre	Fru	Cellobiose
Mosquitoes	A	*S. culicicola*			Yes			Yes	Yes
	A	*S. diminitum*		Yes	Yes		Yes	Yes	
	A	*S. sabaudiense*	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	A	*S. taiwanese*			Yes			Yes	
Flies	A	*S. litorale*			Yes		Yes	Yes	Yes
	A	*S. turonicum*			Yes		Yes	Yes	Yes
	M	*S. atripochogonis*			Yes			Yes	Inc.
	Ch	*S. chrysopicola*		Yes	Yes			Yes	Yes
	Ch	*S. syrphidicola*		Yes	Yes		Inc.	Yes	
	P	*S. MSRO*	Yes	Yes					
	P	*S. hyd1*	Yes			Yes		Yes	Inc.
	C	*S. moj*	Yes					Yes	
Bees	A	*S. apis*	Yes		Yes	Yes	Yes	Yes	Yes
	C	*S. melliferum* KC3	Yes	Yes		Yes	Yes	Yes	Inc.
Ticks	M	*S. mirum*			Yes		Inc..	Yes	Inc.
Beetles	A	*S. canthariciola*	Yes		Yes	Yes	Yes	Yes	Yes
Crabs	M	*S. eriocheiris*	Yes	Yes	Yes	Yes	Inc.	Yes	Yes
Plants	C	*S. citri*	Yes		Yes		Yes	Yes	Inc.
	C	*S. kunkelii*	Yes					Yes	

Table S4. *arc* complex enzymes in *Spiroplasma*

		arcA	arcB	arcC	arcD
POULSONII	*S.* hyd1	□	□	□	□
	S. MSRO	□	□		□
CITRI	*S. moj*	P	?	?	?
	S. kunkelii	□	□		□
	S. citri	□	□	□	□
	S. melliferum	□	□	□	□
CHRYSOPICOLA	*S. chrysopicola*	□	□	□	□
	S. syrphidicola	□	□	□	□
	S. mirum				
	S. atripochogonis	□	□	□	□
	S. eriocheiris	□	□	□	□
APIS	*S. diminitum*				□
	S. taiwanese				
	S. apis	□	□	□	
	S. turonicum	□	□	□	□
	S. litorale				
	S. cantharicola				□
	S. culicicola				
	S. sabaudiense	□	□	□	□
	S. TU-14	□	□	□	□

Table S5. Intermediate enzymes required to synthetize cardiolipin.

	Clade	Strain	*pls*	*pls*C	*cds*A	*pgs*A	*pgp*	*cls*
Mosquitoes	A	*S. culicicola*	Y	Y	Y	Y		
	A	*S. diminitum*	Y	Y	Y	Y		
	A	*S. sabaudiense*	Y	Y	Y	Y		
	A	*S. taiwanese*	Y	Y	Y	Y		
Flies	A	*S. litorale*	Y	Y	Y	Y		
	A	*S. turonicum*	Y	Y	Y	Y		
	M	*S. atripochogonis*	Y	Y	Y	Y	?	Y
	Ch	*S. chrysopicola*	Y	Y	Y	Y	?	Y
	Ch	*S. syrphidicola*	Y	Y	Y	Y	?	Y
	P	*S. MSRO*	Y	Y	Y	Y	?	Y
	P	*S. hyd1*	Y	Y	Y	Y	?	Y
	C	*S. moj*	Y	Y	Y	Y	?	Y
Bees	A	*S. apis*	Y	Y	Y	Y		
	C	*S. melliferum* KC3	Y	Y	Y	Y	**?**	Y
Ticks	M	*S. mirum*	Y	Y	Y	Y	?	Y
Beetles	A	*S. canthariciola*	Y	Y	Y	Y		
Crabs	M	*S. eriocheiris*	Y	Y	Y	Y	?	Y
Plants	C	*S. citri*	Y	Y	Y	Y	**?**	Y
	C	*S. kunkelii*	Y	Y	Y	Y	?	Y

Table S6. Gene list of *Spiroplasma* core-genome

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-1_35	coaX	K03525	coaX; type III pantothenate kinase [EC:2.7.1.33]
Contig-1_40	Hypothetical	NA	
Contig-1_43	folD	K01491	folD; methylenetetrahydrofolate dehydrogenase (NADP+) / methenyltetrahydrofolate cyclohydrolase [EC:1.5.1.5 3.5.4.9]
Contig-1_44	rbfA	K02834	rbfA; ribosome-binding factor A
Contig-4_2		K07015	uncharacterized protein
Contig-4_3	yqeH	K06948	yqeH; 30S ribosome assembly GTPase
Contig-5_3	RP-S21	K02970	RP-S21; small subunit ribosomal protein S21
Contig-5_16	nadE	K01916	nadE; NAD+ synthase [EC:6.3.1.5]
Contig-5_21	miaA	K00791	miaA; tRNA dimethylallyltransferase [EC:2.5.1.75]
Contig-7_1	truB	K03177	truB; tRNA pseudouridine55 synthase [EC:5.4.99.25]
Contig-7_7		K00615	E2.2.1.1; transketolase [EC:2.2.1.1]
Contig-7_21	xseA	K03601	xseA; exodeoxyribonuclease VII large subunit [EC:3.1.11.6]
Contig-7_24	dxs	K01662	dxs; 1-deoxy-D-xylulose-5-phosphate synthase [EC:2.2.1.7]
Contig-7_25	pepP	K01262	pepP; Xaa-Pro aminopeptidase [EC:3.4.11.9]
Contig-10_5	mntN	K01243	mtnN; adenosylhomocysteine nucleosidase [EC:3.2.2.9]
Contig-10_14	parE	K02622	parE; topoisomerase IV subunit B [EC:5.99.1.-]
Contig-10_16	parC	K02621	parC; topoisomerase IV subunit A [EC:5.99.1.-]
Contig-10_21		K01537	E3.6.3.8; Ca2+-transporting ATPase [EC:3.6.3.8]
Contig-12_7	Hypothetical	NA	
Contig-12_8	cbf	K03698	cbf; 3'-5' exoribonuclease [EC:3.1.-.-]
Contig-12_11	phoU	K02039	phoU; phosphate transport system protein
Contig-12_12	pstBC	K02036	pstBC; phosphate transport system ATP-binding protein [EC:3.6.3.27]
Contig-12_15	sms	K03529	smc; chromosome segregation protein
Contig-12_23	rbgA	K14540	rbgA; ribosome biogenesis GTPase A

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-12_25	RP-L19	K02884	RP-L19; large subunit ribosomal protein L19
Contig-12_27	rimM	K02860	rimM; 16S rRNA processing protein RimM
Contig-12_39	deoD	K03784	deoD; purine-nucleoside phosphorylase [EC:2.4.2.1]
Contig-12_42	KAE1	K01409	KAE1; N6-L-threonylcarbamoyladenine synthase [EC:2.3.1.234]
Contig-12_43	PTS-HDPR	K11189	PTS-HPR; phosphocarrier protein
Contig-12_46	uvrD	K03657	uvrD; DNA helicase II / ATP-dependent DNA helicase PcrA [EC:3.6.4.12]
Contig-12_48	Hypothetical	NA	
Contig-12_51	map	K01265	map; methionyl aminopeptidase [EC:3.4.11.18]
Contig-12_52	RP-L17	K02879	RP-L17; large subunit ribosomal protein L17
Contig-12_53	rpoA	K03040	rpoA; DNA-directed RNA polymerase subunit alpha [EC:2.7.7.6]
Contig-12_54	RP-S11	K02948	RP-S11; small subunit ribosomal protein S11
Contig-12_55	RP-S13	K02952	RP-S13; small subunit ribosomal protein S13
Contig-12_57	infA	K02518	infA; translation initiation factor IF-1
Contig-12_58	adk	K00939	adk; adenylate kinase [EC:2.7.4.3]
Contig-12_59	secY	K03076	secY; preprotein translocase subunit SecY
Contig-12_60	RP-L15	K02876	RP-L15; large subunit ribosomal protein L15
Contig-12_62	RP-L18	K02881	RP-L18; large subunit ribosomal protein L18
Contig-12_63	RP-L6	K02933	RP-L6; large subunit ribosomal protein L6
Contig-12_64	RP-S8	K02994	RP-S8; small subunit ribosomal protein S8
Contig-12_65	RP-S14	K02954	RP-S14; small subunit ribosomal protein S14
Contig-12_66	RP-L5	K02931	RP-L5; large subunit ribosomal protein L5
Contig-12_67	RP-L24	K02895	RP-L24; large subunit ribosomal protein L24
Contig-12_68	RP-L14	K02874	RP-L14; large subunit ribosomal protein L14
Contig-12_69	RP-S17	K02961	RP-S17; small subunit ribosomal protein S17
Contig-12_71	RP-L16	K02878	RP-L16; large subunit ribosomal protein L16
Contig-12_72	RP-S3	K02982	RP-S3; small subunit ribosomal protein S3

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-12_73	RP-L22	K02890	RP-L22; large subunit ribosomal protein L22
Contig-12_74	RP-S19	K02965	RP-S19; small subunit ribosomal protein S19
Contig-12_75	RP-L12	K02886	RP-L2; large subunit ribosomal protein L2
Contig-12_76	RP-L23	K02892	RP-L23; large subunit ribosomal protein L23
Contig-12_77	RP-L4	K02926	RP-L4; large subunit ribosomal protein L4
Contig-12_78	RP-L3	K02906	RP-L3; large subunit ribosomal protein L3
Contig-12_79	RP-S10	K02946	RP-S10; small subunit ribosomal protein S10
Contig-12_85	dnaB	K02314	dnaB; replicative DNA helicase [EC:3.6.4.12]
Contig-12_86	RP-L9	K02939	RP-L9; large subunit ribosomal protein L9
Contig-12_96	ppnK	K00858	ppnK; NAD+ kinase [EC:2.7.1.23]
Contig-12_100	LARS	K01869	LARS; leucyl-tRNA synthetase [EC:6.1.1.4]
Contig-12_101	Hypothetical	NA	
Contig-12_104	trxA	K03671	trxA; thioredoxin 1
Contig-12_105	Hypothetical	NA	
Contig-13_1	frr	K02838	frr; ribosome recycling factor
Contig-13_2	pyrH	K09903	pyrH; uridylate kinase [EC:2.7.4.22]
Contig-13_26	coaE	K00859	coaE; dephospho-CoA kinase [EC:2.7.1.24]
Contig-13_31	hprT	K00760	hprT; hypoxanthine phosphoribosyltransferase [EC:2.4.2.8]
Contig-13_34	hubP	K03530	hupB; DNA-binding protein HU-beta
Contig-13_35	Hypothetical	NA	
Contig-13_37	Hypothetical	NA	
Contig-13_39	ABCB-BAC	K06147	ABCB-BAC; ATP-binding cassette, subfamily B, bacterial
Contig-24_2	ABC-2.A	K01990	ABC-2.A; ABC-2 type transport system ATP-binding protein
Contig-26_2		K07124	uncharacterized protein
Contig-26_3	pgsA	K00995	pgsA; CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-27_8	rnc	K03685	rnc; ribonuclease III [EC:3.1.26.3]
Contig-27_9	plsX	K03621	plsX; glycerol-3-phosphate acyltransferase PlsX [EC:2.3.1.15]
Contig-27_11	NA	NA	
Contig-27_12	dnaJ	K03686	dnaJ; molecular chaperone DnaJ
Contig-27_13	dnaK	K04043	dnaK; molecular chaperone DnaK
Contig-27_14	GRPE	K03687	GRPE; molecular chaperone GrpE
Contig-27_15	hrcA	K03705	hrcA; heat-inducible transcriptional repressor
Contig-32_13	pepF	K08602	pepF; oligoendopeptidase F [EC:3.4.24.-]
Contig-32_14	CARP	K01255	CARP; leucyl aminopeptidase [EC:3.4.11.1]
Contig-32_16	mgtA	K01531	mgtA; Mg2+-importing ATPase [EC:3.6.3.2]
Contig-34_4	infB	K02519	infB; translation initiation factor IF-2
Contig-34_5	Hypothetical	NA	
Contig-34_6	ylxR	K07742	ylxR; uncharacterized protein
Contig-34_7	nusA	K02600	nusA; N utilization substance protein A
Contig-35_3	uvrB	K03702	uvrB; excinuclease ABC subunit B
Contig-35_4	hprK	K06023	hprK; HPr kinase/phosphorylase [EC:2.7.11.- 2.7.4.-]
Contig-35_6	trxB	K00384	trxB; thioredoxin reductase (NADPH) [EC:1.8.1.9]
Contig-35_7	PGK	K00927	PGK; phosphoglycerate kinase [EC:2.7.2.3]
Contig-35_14	NA	K09762	uncharacterized protein
Contig-35_15	TC.APA	K03294	TC.APA; basic amino acid/polyamine antiporter, APA family
Contig-35_17	TPI	K01803	TPI; triosephosphate isomerase (TIM) [EC:5.3.1.1]
Contig-35_28	uvrA	K03701	uvrA; excinuclease ABC subunit A
Contig-35_29	folC	K11754	folC; dihydrofolate synthase / folylpolyglutamate synthase [EC:6.3.2.12 6.3.2.17]
Contig-35_30	NA	NA	
Contig-35_31	dnaB	K03346	dnaB; replication initiation and membrane attachment protein

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-35_32	dnaI	K11144	dnaI; primosomal protein DnaI
Contig-35_34	GAPDH	K00134	GAPDH; glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]
Contig-35_37	gidA	K03495	gidA; tRNA uridine 5-carboxymethylaminomethyl modification enzyme
Contig-35_43	nusG	K02601	nusG; transcriptional antiterminator NusG
Contig-35_47	gidB	K03501	gidB; 16S rRNA (guanine527-N7)-methyltransferase [EC:2.1.1.170]
Contig-35_48	parA	K03496	parA; chromosome partitioning protein
Contig-35_50	ychF	K06942	ychF; ribosome-binding ATPase
Contig-35_51	ispF	K01770	ispF; 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [EC:4.6.1.12]
Contig-35_52	gltX	K09698	gltX; nondiscriminating glutamyl-tRNA synthetase [EC:6.1.1.24]
Contig-35_53		K06885	uncharacterized protein
Contig-35_55	pyrG	K01937	pyrG; CTP synthase [EC:6.3.4.2]
Contig-35_56	purA	K01939	purA; adenylosuccinate synthase [EC:6.3.4.4]
Contig-35_57	purB	K01756	purB; adenylosuccinate lyase [EC:4.3.2.2]
Contig-35_58	PTH1	K01056	PTH1; peptidyl-tRNA hydrolase, PTH1 family [EC:3.1.1.29]
Contig-35_59	rsml	K07056	rsmI; 16S rRNA (cytidine1402-2'-O)-methyltransferase [EC:2.1.1.198]
Contig-35_71	Hypothetical	NA	
Contig-35_76		K00954	E2.7.7.3A; pantetheine-phosphate adenylyltransferase [EC:2.7.7.3]
Contig-35_77	rnj	K12574	rnj; ribonuclease J [EC:3.1.-.-]
Contig-35_78	ftsK	K03466	ftsK; DNA segregation ATPase FtsK/SpoIIIE, S-DNA-T family
Contig-35_80	ecfA2	K16787	ecfA2; energy-coupling factor transport system ATP-binding protein [EC:3.6.3.-]
Contig-35_81	ecfT	K16785	ecfT; energy-coupling factor transport system permease protein
Contig-35_82	truA	K06173	truA; tRNA pseudouridine38-40 synthase [EC:5.4.99.12]
Contig-37_02		K07043	uncharacterized protein
Contig-37_09	FARSA	K01889	FARSA; phenylalanyl-tRNA synthetase alpha chain [EC:6.1.1.20]
Contig-37_10	FARSB	K01890	FARSB; phenylalanyl-tRNA synthetase beta chain [EC:6.1.1.20]

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-37_12	NA	K07040	uncharacterized protein
Contig-37_13	RP-L32	K02911	RP-L32; large subunit ribosomal protein L32
Contig-37_14	mraZ	K03925	mraZ; MraZ protein
Contig-37_15	mraW	K03438	mraW; 16S rRNA (cytosine1402-N4)-methyltransferase [EC:2.1.1.199]
Contig-37_16	ftsA	K03590	ftsA; cell division protein FtsA
Contig-37_17	ftsZ	K03531	ftsZ; cell division protein FtsZ
Contig-37_21	IARS	K01870	IARS; isoleucyl-tRNA synthetase [EC:6.1.1.5]
Contig-37_23	rluD	K06180	rluD; 23S rRNA pseudouridine1911/1915/1917 synthase [EC:5.4.99.23]
Contig-37_24	Hypothetical	NA	
Contig-37_25	comEB	K01493	comEB; dCMP deaminase [EC:3.5.4.12]
Contig-37_29	acpS	K00997	acpS; holo-[acyl-carrier protein] synthase [EC:2.7.8.7]
Contig-37_32	tsf	K02357	tsf; elongation factor Ts
Contig-38_7	UNG	K03648	UNG; uracil-DNA glycosylase [EC:3.2.2.27]
Contig-38_8	NA	NA	
Contig-38_10	RP-L13	K02871	RP-L13; large subunit ribosomal protein L13
Contig-38_11	RP-S9	K02996	RP-S9; small subunit ribosomal protein S9
Contig-38_12	Hypothetical	NA	
Contig-38_14	Hypothetical	NA	
Contig-38_23	Hypothetical	NA	
Contig-38_24	RP-L31	K02909	RP-L31; large subunit ribosomal protein L31
Contig-38_28	nrnA	K06881	nrnA; bifunctional oligoribonuclease and PAP phosphatase NrnA [EC:3.1.3.7 3.1.13.3]
Contig-38_29	tdk	K00857	tdk; thymidine kinase [EC:2.7.1.21]
Contig-38_31	prfA	K02835	prfA; peptide chain release factor 1
Contig-38_32	hemK	K02493	hemK; release factor glutamine methyltransferase [EC:2.1.1.297]
Contig-38_33	Hypothetical	NA	

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-38_34	tsaC	K07566	tsaC; L-threonylcarbamoyladenylate synthase [EC:2.7.7.87]
Contig-38_52	Hypothetical	NA	
Contig-38_57	Hypothetical	NA	
Contig-38_58	Hypothetical	NA	
Contig-38_59	sufC	K09013	sufC; Fe-S cluster assembly ATP-binding protein
Contig-38_60	sufD	K09015	sufD; Fe-S cluster assembly protein SufD
Contig-38_61	sufS	K11717	sufS; cysteine desulfurase / selenocysteine lyase [EC:2.8.1.7 4.4.1.16]
Contig-38_62	iscU	K04488	iscU; nitrogen fixation protein NifU and related proteins
Contig-38_63	sufB	K09014	sufB; Fe-S cluster assembly protein SufB
Contig-38_66	YARS	K01866	YARS; tyrosyl-tRNA synthetase [EC:6.1.1.1]
Contig-38_67	NA	NA	
Contig-38_70	pncB	K00763	pncB; nicotinate phosphoribosyltransferase [EC:6.3.4.21]
Contig-38_73	Hypothetical	NA	
Contig-38_74	Hypothetical	NA	
Contig-38_78	mgtA	K01531	mgtA; Mg2+-importing ATPase [EC:3.6.3.2]
Contig-39_08	ispE	K00919	ispE; 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase [EC:2.7.1.148]
Contig-39_09	ksgA	K02528	ksgA; 16S rRNA (adenine1518-N6/adenine1519-N6)-dimethyltransferase [EC:2.1.1.182]
Contig-39_10	rnmV	K05985	rnmV; ribonuclease M5 [EC:3.1.26.8]
Contig-39_11	mreB	K03569	mreB; rod shape-determining protein MreB and related proteins
Contig-39_16	bmpA	K07335	bmpA; basic membrane protein A and related proteins
Contig-39_17		K02056	ABC.SS.A; simple sugar transport system ATP-binding protein [EC:3.6.3.17]
Contig-39_18		K02057	ABC.SS.P; simple sugar transport system permease protein
Contig-39_19		K02057	ABC.SS.P; simple sugar transport system permease protein
Contig-39_24		K03046	rpoC; DNA-directed RNA polymerase subunit beta' [EC:2.7.7.6]

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-39_25		K02950	RP-S12; small subunit ribosomal protein S12
Contig-39_26		K02992	RP-S7; small subunit ribosomal protein S7
Contig-39_27		K02355	fusA; elongation factor G
Contig-39_28		K02358	tuf; elongation factor Tu
Contig-39_42	upp	K00761	upp; uracil phosphoribosyltransferase [EC:2.4.2.9]
Contig-39_44		K02108	ATPF0A; F-type H+-transporting ATPase subunit a
Contig-39_45		K02110	ATPF0C; F-type H+-transporting ATPase subunit c
Contig-39_46		K02109	ATPF0B; F-type H+-transporting ATPase subunit b
Contig-39_47		K02113	ATPF1D; F-type H+-transporting ATPase subunit delta
Contig-39_48		K02111	ATPF1A; F-type H+-transporting ATPase subunit alpha [EC:3.6.3.14]
Contig-39_49		K02115	ATPF1G; F-type H+-transporting ATPase subunit gamma
Contig-39_50		K02112	ATPF1B; F-type H+-transporting ATPase subunit beta [EC:3.6.3.14]
Contig-39_51		K02114	ATPF1E; F-type H+-transporting ATPase subunit epsilon
Contig-39_54	Hypothetical	NA	
Contig-39_55	rnr	K12573	rnr; ribonuclease R [EC:3.1.-.-]
Contig-39_56	smpB	K03664	smpB; SsrA-binding protein
Contig-39_59	ftsY	K03110	ftsY; fused signal recognition particle receptor
Contig-39_60		K09787	uncharacterized protein
Contig-39_61		K09769	uncharacterized protein
Contig-39_63	metK	K00789	metK; S-adenosylmethionine synthetase [EC:2.5.1.6]
Contig-39_64	Hypothetical	NA	
Contig-39_66	potA	K11072	potA; spermidine/putrescine transport system ATP-binding protein [EC:3.6.3.31]
Contig-39_67	potB	K11071	potB; spermidine/putrescine transport system permease protein
Contig-39_68	potC	K11070	potC; spermidine/putrescine transport system permease protein
Contig-39_75	DPO3D1	K02340	DPO3D1; DNA polymerase III subunit delta [EC:2.7.7.7]

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-39_76	Hypothetical	NA	
Contig-39_77	RP-S-20	K02968	RP-S20; small subunit ribosomal protein S20
Contig-39_79	rpiB	K01808	rpiB; ribose 5-phosphate isomerase B [EC:5.3.1.6]
Contig-39_82	Hypothetical	NA	
Contig-39_83	PRPS	K00948	PRPS; ribose-phosphate pyrophosphokinase [EC:2.7.6.1]
Contig-40_1		K02003	ABC.CD.A; putative ABC transport system ATP-binding protein
Contig-40_2	hslO	K04083	hslO; molecular chaperone Hsp33
Contig-40_3	ftsH	K03798	ftsH; cell division protease FtsH [EC:3.4.24.-]
Contig-40_4	tilS	K04075	tilS; tRNA(Ile)-lysidine synthase [EC:6.3.4.19]
Contig-40_7	tmk	K00943	tmk; dTMP kinase [EC:2.7.4.9]
Contig-40_9	DPO3G	K02343	DPO3G; DNA polymerase III subunit gamma/tau [EC:2.7.7.7]
Contig-40_10	tadA	K11991	tadA; tRNA(adenine34) deaminase [EC:3.5.4.33]
Contig-40_11	SARS	K01875	SARS; seryl-tRNA synthetase [EC:6.1.1.11]
Contig-40_13	gyrB	K02470	gyrB; DNA gyrase subunit B [EC:5.99.1.3]
Contig-40_14	DPO3B	K02338	DPO3B; DNA polymerase III subunit beta [EC:2.7.7.7]
Contig-40_15	dnaA	K02313	dnaA; chromosomal replication initiator protein
Contig-40_19	yidC	K03217	yidC; YidC/Oxa1 family membrane protein insertase
Contig-40_21	mnmE	K03650	mnmE; tRNA modification GTPase [EC:3.6.-.-]
Contig-40_22	tatD	K03424	tatD; TatD DNase family protein [EC:3.1.21.-]
Contig-40_23	mreB	K03569	mreB; rod shape-determining protein MreB and related proteins
Contig-40_24	mreB	K03569	mreB; rod shape-determining protein MreB and related proteins
Contig-41_5	nusB	K03625	nusB; N utilization substance protein B
Contig-41_14	ruvX	K07447	ruvX; putative holliday junction resolvase [EC:3.1.-.-]
Contig-41_15	AARS	K01872	AARS; alanyl-tRNA synthetase [EC:6.1.1.7]
Contig-41_17	relA	K00951	relA; GTP pyrophosphokinase [EC:2.7.6.5]
Contig-41_18	APRT	K00759	APRT; adenine phosphoribosyltransferase [EC:2.4.2.7]

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-41_30	lepA	K03596	lepA; GTP-binding protein LepA
Contig-41_33	MTFMT	K00604	MTFMT; methionyl-tRNA formyltransferase [EC:2.1.2.9]
Contig-41_34	Hypothetical	NA	
Contig-41_35	Hypothetical	NA	
Contig-41_43	Hypothetical	NA	
Contig-41_48	tsaB	K14742	tsaB; tRNA threonylcarbamoyladenosine biosynthesis protein TsaB
Contig-41_49	tsaE	K06925	tsaE; tRNA threonylcarbamoyladenosine biosynthesis protein TsaE
Contig-41_50	trxB	K00384	trxB; thioredoxin reductase (NADPH) [EC:1.8.1.9]
Contig-41_51	Hypothetical	NA	
Contig-42_6	oppF	K10823	oppF; oligopeptide transport system ATP-binding protein
Contig-42_14	greA	K03624	greA; transcription elongation factor GreA
Contig-42_21	Hypothetical	NA	
Contig-42_23	bmpA	K07335	bmpA; basic membrane protein A and related proteins
Contig-42_28	acpD	K01118	acpD; FMN-dependent NADH-azoreductase [EC:1.7.-.-]
Contig-42_33	tlyA	K06442	tlyA; 23S rRNA (cytidine1920-2'-O)/16S rRNA (cytidine1409-2'-O)-methyltransferase [EC:2.1.1.226 2.1.1.227]
Contig-43_9	prpC	K20074	prpC; PPM family protein phosphatase [EC:3.1.3.16]
Contig-43_10	prkC	K12132	prkC; eukaryotic-like serine/threonine-protein kinase [EC:2.7.11.1]
Contig-43_11	rsgA	K06949	rsgA; ribosome biogenesis GTPase [EC:3.6.1.-]
Contig-43_12	rpe	K01783	rpe; ribulose-phosphate 3-epimerase [EC:5.1.3.1]
Contig-43_13	thiN	K00949	thiN; thiamine pyrophosphokinase [EC:2.7.6.2]
Contig-43_14	RP-L28	K02902	RP-L28; large subunit ribosomal protein L28
Contig-43_19	Hypothetical	NA	
Contig-43_20		K07030	uncharacterized protein
Contig-43_21	Hypothetical	NA	
Contig-43_28	ezrA	K06286	ezrA; septation ring formation regulator

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-43_29	thiI	K03151	thiI; thiamine biosynthesis protein ThiI
Contig-43_31	DPO3A1	K02337	DPO3A1; DNA polymerase III subunit alpha [EC:2.7.7.7]
Contig-44_4	PTS-EI	K08483	PTS-EI.PTSI; phosphotransferase system, enzyme I, PtsI [EC:2.7.3.9]
Contig-44_6	PTS-Glc	K02777	PTS-Glc-EIIA; PTS system, sugar-specific IIA component [EC:2.7.1.-]
Contig-44_7	Hypothetical	NA	
Contig-44_8		K01972	E6.5.1.2; DNA ligase (NAD+) [EC:6.5.1.2]
Contig-44_9	gatC	K02435	gatC; aspartyl-tRNA(Asn)/glutamyl-tRNA(Gln) amidotransferase subunit C [EC:6.3.5.6 6.3.5.7]
Contig-44_10	gatA	K02433	gatA; aspartyl-tRNA(Asn)/glutamyl-tRNA(Gln) amidotransferase subunit A [EC:6.3.5.6 6.3.5.7]
Contig-44_11	gatB	K02434	gatB; aspartyl-tRNA(Asn)/glutamyl-tRNA(Gln) amidotransferase subunit B [EC:6.3.5.6 6.3.5.7]
Contig-44_12	trkA	K03499	trkA; trk system potassium uptake protein
Contig-44_15	Hypothetical	NA	
Contig-44_16	PARS	K01881	PARS; prolyl-tRNA synthetase [EC:6.1.1.15]
Contig-44_19	acpP	K02078	acpP; acyl carrier protein
Contig-44_20	perR	K09825	perR; Fur family transcriptional regulator, peroxide stress response regulator
Contig-45_8	cshB	K18692	cshB; ATP-dependent RNA helicase CshB [EC:3.6.4.13]
Contig-45_11	ispH	K03527	ispH; 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase [EC:1.17.7.4]
Contig-45_17	trmK	K06967	trmK; tRNA (adenine22-N1)-methyltransferase [EC:2.1.1.217]
Contig-45_18	SIG1	K03086	SIG1; RNA polymerase primary sigma factor
Contig-45_19	dnaG	K02316	dnaG; DNA primase [EC:2.7.7.-]
Contig-45_20	GARS	K01880	GARS; glycyl-tRNA synthetase [EC:6.1.1.14]
Contig-45_22		K03595	era; GTPase

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-45_23	DPO3A2	K03763	DPO3A2; DNA polymerase III subunit alpha, Gram-positive type [EC:2.7.7.7]
Contig-45_24	rseP	K11749	rseP; regulator of sigma E protease [EC:3.4.24.-]
Contig-45_25	dxr	K00099	dxr; 1-deoxy-D-xylulose-5-phosphate reductoisomerase [EC:1.1.1.267]
Contig-45_26	cdsA	K00981	E2.7.7.41; phosphatidate cytidylyltransferase [EC:2.7.7.41]
Contig-46_7	HARS	K01892	HARS; histidyl-tRNA synthetase [EC:6.1.1.21]
Contig-46_8	DARS	K01876	DARS; aspartyl-tRNA synthetase [EC:6.1.1.12]
Contig-46_10	NA	K09976	uncharacterized protein
Contig-47_4	spoU	K03437	spoU; RNA methyltransferase, TrmH family
Contig-47_10	hemN	K02495	hemN; oxygen-independent coproporphyrinogen III oxidase [EC:1.3.98.3]
Contig-47_11	Hypothetical	NA	
Contig-47_16	RARS	K01887	RARS; arginyl-tRNA synthetase [EC:6.1.1.19]
Contig-47_17	rsmB	K03500	rsmB; 16S rRNA (cytosine967-C5)-methyltransferase [EC:2.1.1.176]
Contig-47_19	Hypothetical	NA	
Contig-47_20	PDF	K01462	PDF; peptide deformylase [EC:3.5.1.88]
Contig-47_21	rnj	K12574	rnj; ribonuclease J [EC:3.1.-.-]
Contig-47_28		K07052	uncharacterized protein
Contig-47_29	gpml	K15633	gpmI; 2,3-bisphosphoglycerate-independent phosphoglycerate mutase [EC:5.4.2.12]
Contig-47_33	gpsA	K00057	gpsA; glycerol-3-phosphate dehydrogenase (NAD(P)+) [EC:1.1.1.94]
Contig-47_34	engA	K03977	engA; GTPase
Contig-47_35	cmk	K00945	cmk; cytidylate kinase [EC:2.7.4.14]
Contig-47_36	Hypothetical	NA	
Contig-47_39	Hypothetical	NA	
Contig-47_47	Hypothetical	NA	
Contig-47_57	rny	K18682	rny; ribonucrease Y [EC:3.1.-.-]

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-47_58	SRP54	K03106	SRP54; signal recognition particle subunit SRP54 [EC:3.6.5.4]
Contig-48_13	mreB	K03569	mreB; rod shape-determining protein MreB and related proteins
Contig-48_20	tig	K03545	tig; trigger factor
Contig-48_26	Hypothetical	NA	
Contig-48_27	Hypothetical	NA	
Contig-48_32	ppa	K01507	ppa; inorganic pyrophosphatase [EC:3.6.1.1]
Contig-48_33	Hypothetical	NA	
Contig-48_35	scpB	K06024	scpB; segregation and condensation protein B
Contig-48_36	scpA	K05896	scpA; segregation and condensation protein A
Contig-48_37	pdp	K00756	pdp; pyrimidine-nucleoside phosphorylase [EC:2.4.2.2]
Contig-48_38	plcC	K00655	plsC; 1-acyl-sn-glycerol-3-phosphate acyltransferase [EC:2.3.1.51]
Contig-48_39	folA	K00287	folA; dihydrofolate reductase [EC:1.5.1.3]
Contig-48_40	thyA	K00560	thyA; thymidylate synthase [EC:2.1.1.45]
Contig-48_48	gapN	K00131	gapN; glyceraldehyde-3-phosphate dehydrogenase (NADP+) [EC:1.2.1.9]
Contig-48_50	VARS	K01873	VARS; valyl-tRNA synthetase [EC:6.1.1.9]
Contig-49_20	FBA	K01624	FBA; fructose-bisphosphate aldolase, class II [EC:4.1.2.13]
Contig-49_21	mutM	K10563	mutM; formamidopyrimidine-DNA glycosylase [EC:3.2.2.23 4.2.99.18]
Contig-49_27	DPO1	K02335	DPO1; DNA polymerase I [EC:2.7.7.7]
Contig-50_3	WARS	K01867	WARS; tryptophanyl-tRNA synthetase [EC:6.1.1.2]
Contig-50_4	ENO	K01689	ENO; enolase [EC:4.2.1.11]
Contig-51_2	infC	K02520	infC; translation initiation factor IF-3
Contig-51_3	RP-L35	K02916	RP-L35; large subunit ribosomal protein L35
Contig-51_4	RP-L20	K02887	RP-L20; large subunit ribosomal protein L20
Contig-52_7	ispD	K00991	ispD; 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase [EC:2.7.7.60]
Contig-52_9	ssb	K03111	ssb; single-strand DNA-binding protein

Table S6. Continued

S. hyd1 location	Gene	KEGG Number	KEGG Description
Contig-52_10	RP-S18	K02963	RP-S18; small subunit ribosomal protein S18
Contig-52_12	Hypothetical	NA	
Contig-52_18	RP-L7	K02935	RP-L7; large subunit ribosomal protein L7/L12
Contig-52_19	RP-L10	K02864	RP-L10; large subunit ribosomal protein L10
Contig-53_7	Hypothetical	NA	
Contig-53_13	thiJ	K03152	thiJ; 4-methyl-5(b-hydroxyethyl)-thiazole monophosphate biosynthesis
Contig-53_14	Hypothetical	NA	
Contig-54_3	mnmA	K00566	mnmA; tRNA-specific 2-thiouridylase [EC:2.8.1.-]
Contig-54_4	recD	K03581	recD; exodeoxyribonuclease V alpha subunit [EC:3.1.11.5]
Contig-54_8	deoD	K03784	deoD; purine-nucleoside phosphorylase [EC:2.4.2.1]
Contig-54_9	Hypothetical	NA	
Contig-55_4	pfkA	K00850	pfkA; 6-phosphofructokinase 1 [EC:2.7.1.11]
Contig-55_5	PK	K00873	PK; pyruvate kinase [EC:2.7.1.40]
Contig-55_6	TARS	K01868	TARS; threonyl-tRNA synthetase [EC:6.1.1.3]
Contig-56_13	Hypothetical	NA	

Table S7. *D. mojavensis* protection experiment data.

Isoline	*Spiroplasma* infection status	Treatment	Initial larvae	Pupa	Total adult flies	Total adult wasps
C1	S+	Aw35	30	29	23	0
C1	S+	Aw35	30	29	21	0
C1	S+	Aw35	30	20	20	0
C1	S+	Aw35	30	23	17	0
C1	S+	Aw35	30	27	22	0
C1	S+	Aw35	30	27	23	0
C1	S+	Aw35	30	27	21	0
C1	S+	Aw35	30	28	17	0
C1	S+	Aw35	30	27	23	0
C1	S+	Aw35	30	24	18	0
B2	S+	Aw35	30	30	26	0
B2	S+	Aw35	31	31	28	0
B2	S+	Aw35	30	30	27	0
B2	S+	Aw35	30	30	26	0
B2	S+	Aw35	32	32	30	0
B2	S+	Aw35	30	27	22	0
B2	S+	Aw35	30	29	26	0
B2	S+	Aw35	30	29	26	0
B2	S+	Control	30	28	28	0
B2	S+	Control	30	27	26	0
F2	S+	Control	30	28	24	0
F2	S+	Control	30	30	24	0

Table S7. Continued

Isoline	*Spiroplasma* infection status	Treatment	Initial larvae	Pupa	Total adult flies	Total adult wasps
F2	S+	Control	30	25	24	0
F2	S+	Control	30	26	25	0
F2	S+	Control	30	30	26	0
F2	S+	Control	30	30	27	0
F2	S+	Control	30	30	25	0
F2	S+	Control	30	29	29	0
F2	S+	Control	30	27	27	0
F2	S+	Control	31	31	24	0
B2	S+	Control	30	27	0	23
B2	S+	Control	30	29	0	19
B2	S+	Control	30	30	0	16
B2	S+	Aw35	30	27	0	24
B2	S+	Aw35	42	31	0	25
B2	S+	Aw35	30	28	1	14
B2	S+	Aw35	30	22	2	11
B2	S+	Control	30	21	0	18
B2	S+	Control	41	41	0	36
B2	S+	Control	30	27	0	14
B2	S+	Control	30	23	0	16
B2	S+	Aw35	30	15	0	12
B2	S+	Aw35	30	23	0	14
B2	S+	Aw35	30	25	0	15

Table S7. Continued

Isoline	*Spiroplasma* infection status	Treatment	Initial larvae	Pupa	Total adult flies	Total adult wasps
C1	S+	Aw35	30	22	2	15
C1	S+	Aw35	30	16	0	13
C1	S+	Aw35	30	22	0	18
C1	S+	Aw35	30	21	0	20
C1	S+	Aw35	42	12	0	11
C1	S+	Aw35	30	27	0	25
C1	S+	Control	30	20	0	17
C1	S+	Control	30	26	0	15
C1	S+	Control	42	34	0	34
F2	S+	Control	30	26	1	18
F2	S+	Control	30	29	0	26
F2	S+	Control	30	27	8	18
ABM	S-	Control	30	27	26	0
ABM	S-	Control	30	30	28	0
ABM	S-	Control	32	32	29	0
ABM	S-	Control	30	22	18	0
ABM	S-	Control	30	29	21	0
ABM	S-	Control	30	29	25	0
ABM	S-	Control	30	29	27	0
ABM	S-	Control	30	30	23	0
ABM	S-	Control	30	27	22	0
ABM	S-	Lh14	30	23	23	0

Table S7. Continued

Isoline	*Spiroplasma* infection status	Treatment	Initial larvae	Pupa	Total adult flies	Total adult wasps
ABM	S-	Lh14	30	26	22	0
ABM	S-	Lh14	30	26	22	0
ABM	S-	Lh14	30	26	21	0
ABM	S-	Lh14	30	28	27	0
ABM	S-	Lh14	30	29	29	0
ABM	S-	Lh14	31	31	0	27
ABM	S-	Lh14	42	37	2	24
ABM	S-	Lh14	30	28	0	19
ABM	S-	Lh14	42	36	1	22
ABM	S-	Lh14	30	28	26	0
ABM	S-	Lh14	30	29	27	0
ABM	S-	Lh14	30	27	24	0
ABM	S-	Lh14	30	28	20	0
ABM	S-	Lh14	30	24	0	12
ABM	S-	Aw35	30	25	0	16
ABM	S-	Aw35	42	29	0	23
ABM	S-	Aw35	30	28	3	18
ABM	S-	Aw35	32	32	0	19
ABM	S-	Aw35	30	28	1	21
ABM	S-	Aw35	30	25	0	22
ABM	S-	Aw35	30	28	0	18
ABM	S-	Aw35	30	21	2	8

Table S7. Continued

Isoline	*Spiroplasma* infection status	Treatment	Initial larvae	Pupa	Total adult flies	Total adult wasps
ABM	S-	Aw35	30	26	24	0
ABM	S-	Aw35	30	30	27	0
ABM	S-	Aw35	30	29	25	0
ABM	S-	Aw35	30	29	28	0
ABM	S-	Aw35	30	26	22	0
ABM	S-	Aw35	30	27	25	0
ABM	S-	Aw35	30	30	29	0
ABM	S-	Control	30	29	29	0
ABM	S-	Control	30	30	29	0
ABM	S-	Control	30	27	27	0
ABM	S-	Control	30	30	22	0
ABM	S-	Control	30	26	21	0
ABM	S-	Control	30	29	26	0
ABM	S-	Control	30	29	23	0
ABM	S-	Control	30	28	22	0
ABM	S-	Control	30	29	0	4
ABM	S-	Control	30	18	0	8
ABM	S-	Control	30	17	0	6
ABM	S-	Control	30	22	0	3
ABM	S-	Control	30	3	0	1
ABM	S-	Control	30	26	1	16
ABM	S-	Control	30	30	0	21

Table S7. Continued

Isoline	*Spiroplasma* infection status	Treatment	Initial larvae	Pupa	Total adult flies	Total adult wasps
ABM	S-	Lh14	30	28	0	10
ABM	S-	Lh14	30	29	0	18
ABM	S-	Lh14	30	22	2	2
ABM	S-	Lh14	30	30	2	8
ABM	S-	Lh14	30	30	3	21
ABM	S-	Lh14	30	26	0	20
ABM	S-	Lh14	30	27	0	7
ABM	S-	Lh14	30	17	0	3
ABM	S-	Lh14	30	24	1	24
ABM	S-	Lh14	30	29	1	16
ABM	S-	Lh14	30	24	3	18
ABM	S-	Lh14	30	24	0	22
ABM	S-	Lh14	30	21	0	20
ABM	S-	Lh14	30	25	1	24
ABM	S-	Aw35	30	17	0	13
ABM	S-	Aw35	30	25	0	25
ABM	S-	Aw35	30	26	0	20
ABM	S-	Aw35	30	23	0	23
ABM	S-	Aw35	30	27	2	16
ABM	S-	Aw35	30	22	0	20
ABM	S-	Aw35	30	23	0	13
ABM	S-	Aw35	30	28	0	28

Table S7. Continued

Isoline	*Spiroplasma* infection status	Treatment	Initial larvae	Pupa	Total adult flies	Total adult wasps
ABM	S-	Aw35	30	28	0	14
F2	S+	Aw35	30	25	24	0
F2	S+	Aw35	30	29	28	0
F2	S+	Aw35	30	30	29	0
F2	S+	Aw35	30	30	28	0
F2	S+	Aw35	30	30	29	0
B2	S+	Aw35	30	27	21	0
B2	S+	Aw35	30	26	25	0
B2	S+	Aw35	30	29	29	0
B2	S+	Aw35	30	25	19	0
B2	S+	Aw35	30	27	26	0
C1	S+	Aw35	30	25	24	0
C1	S+	Aw35	30	28	26	0
C1	S+	Aw35	30	27	18	0
C1	S+	Aw35	30	28	25	0
C1	S+	Aw35	30	30	29	0
F2	S+	Aw35	30	30	0	4
F2	S+	Aw35	30	27	0	13
F2	S+	Aw35	30	25	0	14
F2	S+	Aw35	30	27	0	16
B2	S+	Aw35	30	25	0	5
B2	S+	Aw35	30	30	0	13

Table S7. Continued

Isoline	*Spiroplasma* infection status	Treatment	Initial larvae	Pupa	Total adult flies	Total adult wasps
B2	S+	Aw35	30	34	0	17
B2	S+	Control	30	28	0	21
B2	S+	Control	30	27	0	14
C1	S+	Control	30	21	0	8
C1	S+	Control	30	28	0	13
C1	S+	Control	30	26	0	7
C1	S+	Control	30	28	0	5
C1	S+	Control	30	25	0	9
F2	S+	Control	30	30	1	28
F2	S+	Control	30	26	0	18
F2	S+	Control	30	27	0	26
F2	S+	Control	30	28	7	19
F2	S+	Control	30	28	0	24
B2	S+	Control	30	26	1	23
B2	S+	Control	30	27	7	18
B2	S+	Control	30	22	0	13
B2	S+	Aw35	30	25	0	19
C1	S+	Aw35	30	22	0	21
C1	S+	Aw35	30	27	0	10
C1	S+	Aw35	30	27	1	34
C1	S+	Control	30	29	4	24

Table S8. *D. aldrichi* protection experiment data.

Isoline	Spiroplasma infection status	Treatment	Initial larvae	Pupa	Total adult flies	Total adult wasps	Oviposition rate
FR0512-32	S-	Lh14	30	15	1	5	0.7
FR0512-32	S-	Lh14	30	21	4	7	0.7
FR0512-32	S-	Lh14	30	22	0	10	0.8
FR0512-32	S-	Aw35	30	19	0	15	0.9
FR0512-32	S-	Aw35	30	25	0	25	1
FR0512-32	S-	Aw35	30	23	0	23	1
FR0512-32	S-	Control	30	16	11	0	
FR0512-32	S-	Control	30	25	24	0	
FR0512-32	S-	Control	30	26	26	0	
FR0512-07	S+	Lh14	30	23	3	12	0.9
FR0512-07	S+	Lh14	30	17	0	9	1
FR0512-07	S+	Lh14	30	21	2	15	0.9
FR0512-07	S+	Aw35	30	22	0	13	1
FR0512-07	S+	Aw35	30	20	0	15	1
FR0512-07	S+	Aw35	30	24	0	8	
FR0512-07	S+	Control	30	18	17	0	
FR0512-07	S+	Control	30	14	13	0	
FR0512-07	S+	Control	30	19	19	0	
FR0512-02	S-	Aw35	30	26	12	12	
FR0512-02	S-	Aw35	30	24	1	16	
FR0512-02	S-	Aw35	30	22	2	2	

Table S8. Continued.

FR0512-02	S-	Aw35	30	16	0	12	
FR0512-66	S+	Aw35	30	11	3	0	
FR0512-66	S+	Aw35	30	17	1	2	
FR0512-66	S+	Aw35	30	20	1	3	
FR0512-66	S+	Aw35	30	20	0	5	
FR0512-66	S+	Aw35	30	21	9	2	
FR0512-02	S-	Control	30	20	20	0	
FR0512-02	S-	Control	30	26	25	0	
FR0512-02	S-	Control	30	23	21	0	
FR0512-02	S-	Control	30	25	25	0	
FR0512-02	S-	Control	30	16	16	0	
FR0512-66	S+	Control	30	15	12	0	
FR0512-02	S-	Lh14	30	29	21	6	
FR0512-02	S-	Lh14	30	26	21	1	
FR0512-02	S-	Lh14	30	23	2	16	
FR0512-02	S-	Lh14	30	22	20	0	
FR0512-02	S-	Lh14	30	17	1	10	
FR0512-66	S+	Lh14	30	17	1	1	
FR0512-66	S+	Lh14	30	16	1	0	
FR0512-66	S+	Lh14	30	11	2	3	
FR0512-66	S+	Lh14	30	19	1	0	
FR0512-32	S-	Lh14	30	15	1	5	

Table S9. Statistical models and results for infected and non-infected *D. mojavensis* on Aw35 and Lh14 survivorship

		Fly over larva			Fly over pupa			Wasp over larva			Wasp over pupa			Failed pupa		
Treat.	**Model**	**DF[2]**	**F-value[1]**	**P-value**	**DF[2]**	**F-value[1]**	**P-value**	**DF[2]**	**F-value**	**P-value**	**DF[2]**	**F-value**	**P-value**	**DF[2]**	**F-value**	**P-value**
Aw35	Logistic/Firth	*1*	*0.3039*	*0.5814*	*1*	*0.1717*	*0.6786*									
	GLIMMIX (random residual and isoline)	*1,56*	*0.02*	*0.8819*	*1,56*	*0.00*	*0.9117*	*1,56*	*0.07*	*0.7870*	*1,56*	*0.00*	*0.9510*	*1,56*	*0.01*	*0.9342*
	(random residual)	*1,60*	*0.04*	*0.8349*	*1,60*	*0.03*	*0.8621*	*1,60*	*0.45*	*0.5064*	*1,60*	*0.00*	*0.9941*	*1,60*	*0.00*	*0.9540*
	(random isoline)	*1,56*	*0.00*	*0.9455*	*1,56*	*0.00*	*0.9519*	*1,56*	*0.05*	*0.8277*	*1,56*	*0.00*	*0.9977*	*1,56*	*0.00*	*0.9820*
Control	Logistic/Firth	*1*	*0.884*	*0.7672*	*1*	*0.0554*	*0.8140*									
	GLIMMIX (random residual and isoline)	*1,73*	*0.01*	*0.9397*	*1,73*	*0.01*	*0.9412*							*1,73*	*0.01*	*0.9412*
	(random residual)	*1,77*	*0.04*	*0.8416*	*1,77*	*0.03*	*0.8698*	*1,77*	*0.00*	*1.0000*	*1,77*	*0.00*	*1.0000*	*1,77*	*0.03*	*0.8698*
	(random isoline)	*1,73*	*0.00*	*0.9557*	*1,73*	*0.00*	*0.9630*							*1,73*	*0.00*	*0.9630*
Lh14	Logistic/Firth	*1*	*3.6387*	***0.0564***	*1*	*4.0313*	***0.0447***									
	GLIMMIX (random residual and isoline)							*1,23*	*0.23*	*0.6372*	*1,23*	*0.06*	*0.8119*	*1,23*	*0.00*	*0.9921*
	(random residual)	*1,27*	*0.02*	*0.8812*	*1,27*	*0.02*	*0.8848*	*1,27*	*0.39*	*0.5369*	*1,27*	*0.06*	*0.8107*	*1,27*	*0.00*	*0.9873*
	(random isoline)							*1,23*	*0.08*	*0.7823*	*1,23*	*0.01*	*0.9202*	*1,23*	*0.00*	*0.9947*

[1] Chi-square for Logistic/Firth
[2] Degrees of freedom. For F-value: numerator, denominator

Table S10. Statistical models and results for infected and non-infected *D. aldrichi* on Aw35 and Lh14 survivorship

		Fly over larva			Fly over pupa			Wasp over larva			Wasp over pupa			Failed pupa		
Treat.	**Model**	**DF[2]**	**F-value[1]**	**P-value**	**DF[2]**	**F-value[1]**	**P-value**	**DF[2]**	**F-value**	**P-value**	**DF[2]**	**F-value**	**P-value**	**DF[2]**	**F-value**	**P-value**
Aw35	Logistic/Firth	*1*	*0.3184*	*0.5726*	*1*	*0.0366*	*0.8482*									
	GLIMMIX (random residual and isoline)	*1,11*	*0.01*	*0.9116*	*1,11*	*0.00*	*0.9799*	*1,11*	*1.44*	*0.2559*	*1,11*	*1.36*	*0.2677*	*1,11*	*2.67*	*0.1305*
	(random residual)	*1,13*	*0.04*	*0.8419*	*1,13*	*0.01*	*0.9427*	*1,13*	*6.08*	***0.0283***	*1,13*	*4.90*	***0.0454***	*1,13*	*6.68*	***0.0227***
	(random isoline)	*1,11*	*0.00*	*0.9559*	*1,11*	*0.00*	*0.9889*	*1,11*	*1.48*	*0.2498*	*1,11*	*1.48*	*0.2499*	*1,11*	*2.03*	*0.1819*
Control	Logistic/Firth	*1*	*12.345*	***0.0004***	*1*	*0.6802*	*0.4095*									
	GLIMMIX (random residual and isoline)	*1,8*	*3.58*	*0.0953*	*1,8*	*0.18*	*0.6862*							*1,8*	*0.16*	*0.7017*
	(random residual)	*1,10*	*3.58*	*0.0879*	*1,10*	*0.16*	*0.6941*	*1,10*	*0.00*	*1.0000*	*1,10*	*0.16*	*0.7008*	*1,10*	*0.12*	*0.7361*
	(random isoline)	*1,8*	*12.46*	***0.0077***	*1,8*	*0.33*	*0.5827*							*1,8*	*0.02*	*0.8897*
Lh14	Logistic/Firth	*1*	*35.334*	***<.0001***	*1*	*30.671*	***<.0001***									
	GLIMMIX (random residual and isoline)	*1,11*	*1.36*	*0.2675*	*1,11*	*1.25*	*0.2880*	*1,11*	*0.12*	*0.7324*	*1,11*	*0.02*	*0.8794*	*1,11*	*0.64*	*0.4400*
	(random residual)	*1,13*	*3.99*	*0.0672*	*1,13*	*3.96*	*0.0680*	*1,13*	*0.15*	*0.7026*	*1,13*	*0.00*	*0.9546*	*1,13*	*4.59*	***0.0517***
	(random isoline)	*1,11*	*1.28*	*0.2818*	*1,11*	*1.16*	*0.3040*	*1,11*	*0.19*	*0.6753*	*1,11*	*0.07*	*0.8012*	*1,11*	*0.79*	*0.3945*

[1] Chi-square for Logistic/Firth
[2] Degrees of freedom. For F-value: numerator, denominator

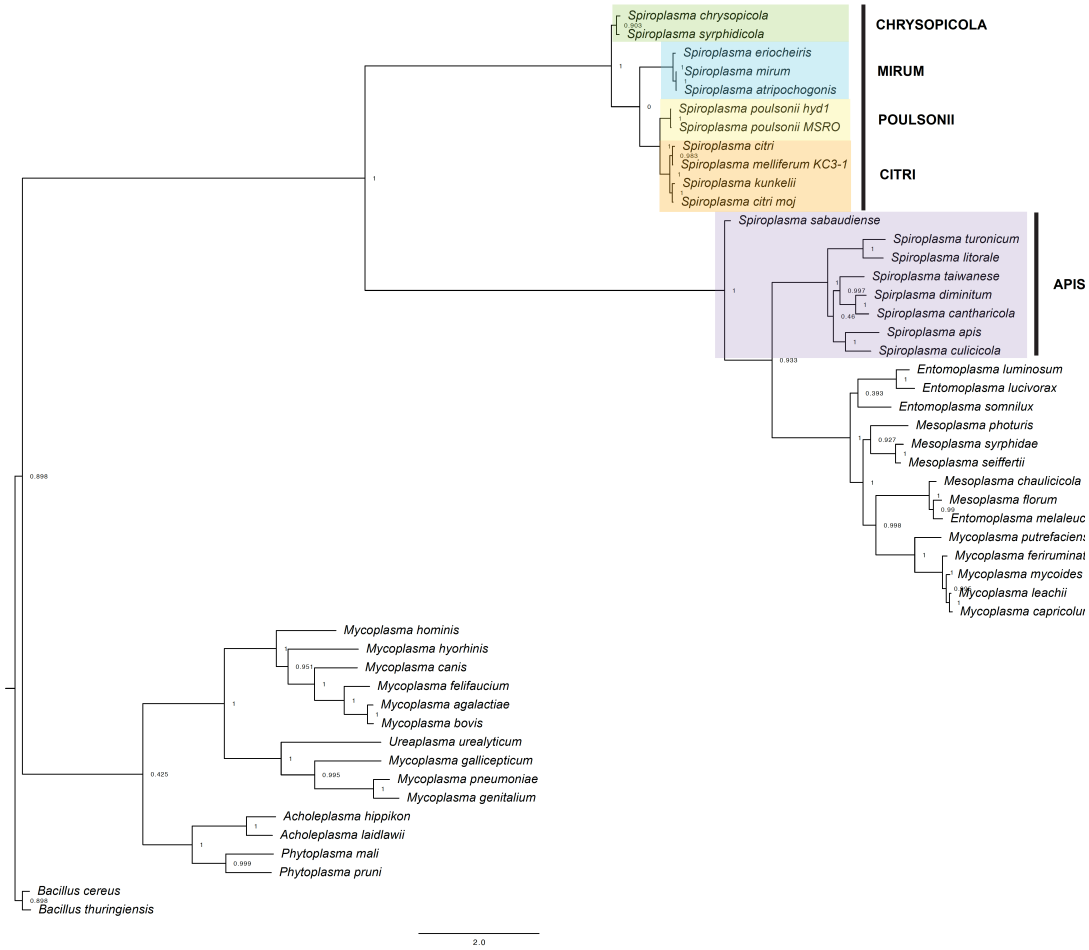


Figure S1. Phylogeny of *Spiroplasma* within Mollicutes. Tree was reconstructed using PhyloPhlAn through the alignment of 399 conserved proteins. Alignment w constructed using MUSCLE with 3704 variable characters across taxa. Identical regions in all taxa were removed. Highlighted *Spiroplasma* clades.

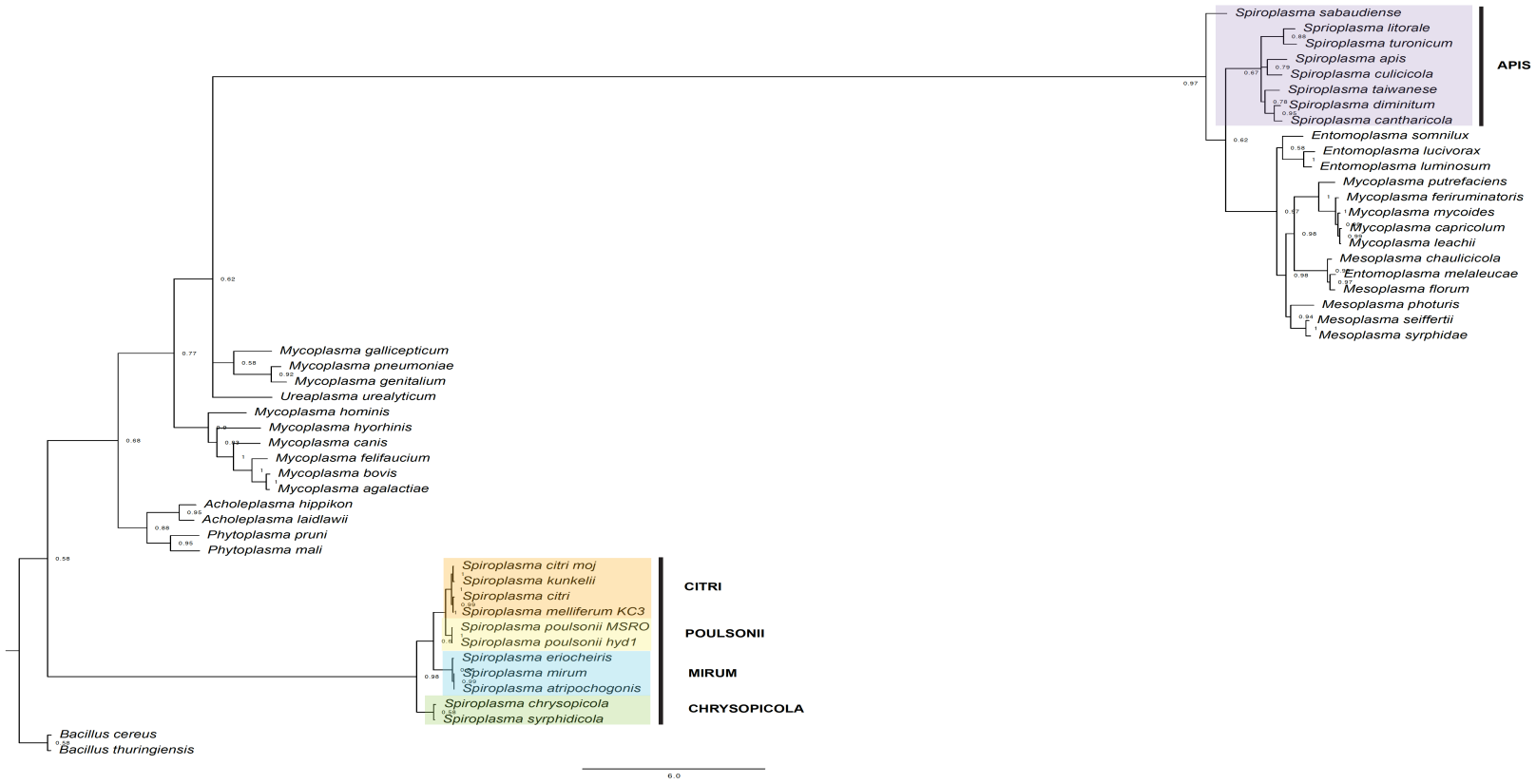


Figure S2. Randomized accelerated maximum likelihood analysis (RAxML) of *Spiroplasma* within Mollicutes. Alignment comprises 3704 variable characters using MUSCLE; PhyloPhlAn removed identical regions across taxa before the phylogenetic analysis. Analysis was performed using RAxML-HPC v 8.2.9 in CIPRES. Highlighted *Spiroplasma* clades.

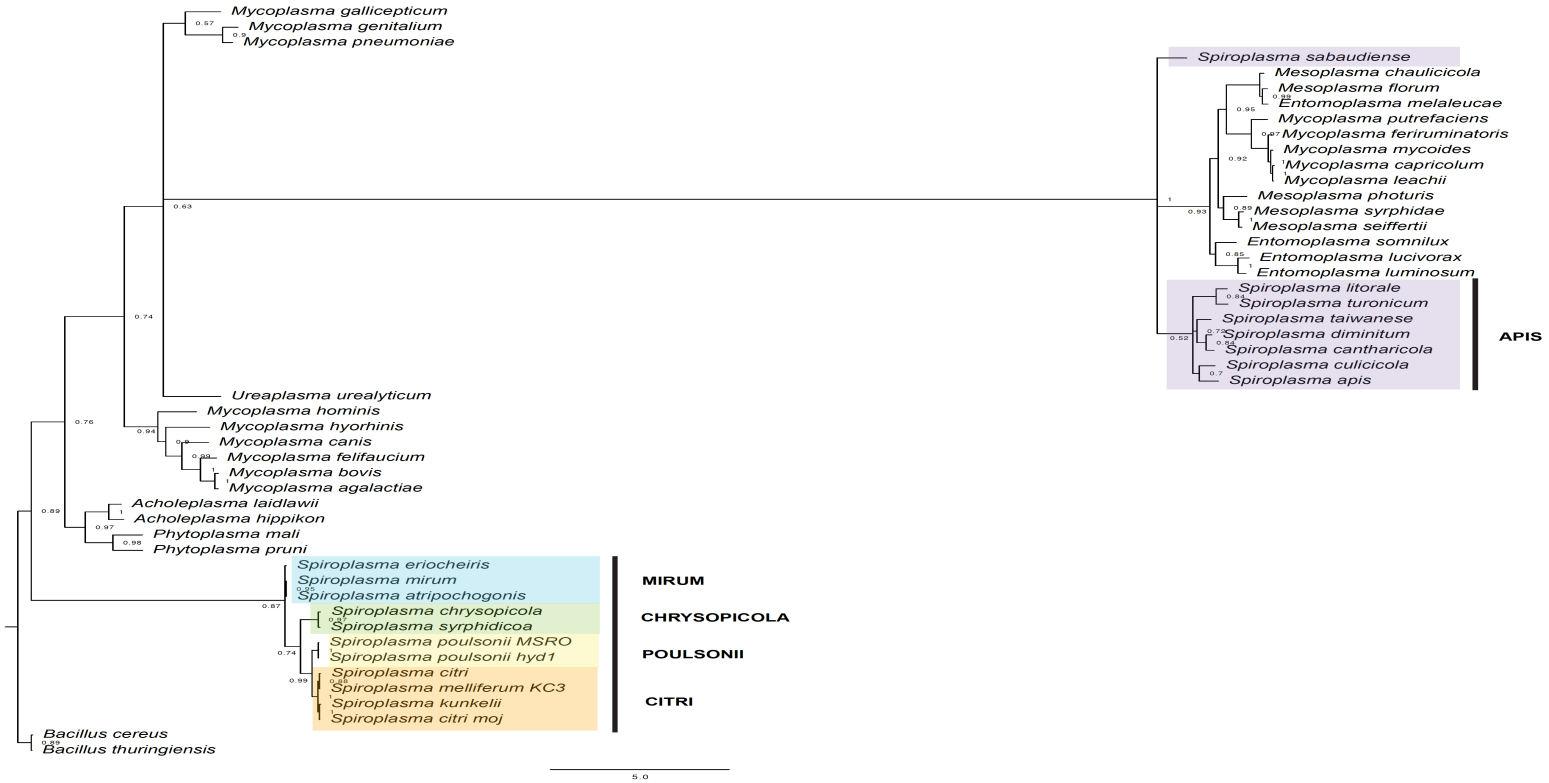


Figure S3. Randomized accelerated maximum likelihood analysis (RAxML) of *Spiroplasma* within Mollicutes. Alignment comprises 2204 variable characters using MUSCLE; Regions present in less than 60% of all taxa were removed. PhyloPhlAn removed identical regions across taxa before the phylogenetic analysis. Analysis was performed using RAxML-HPC v 8.2.9 in CIPRES. Highlighted *Spiroplasma* clades.

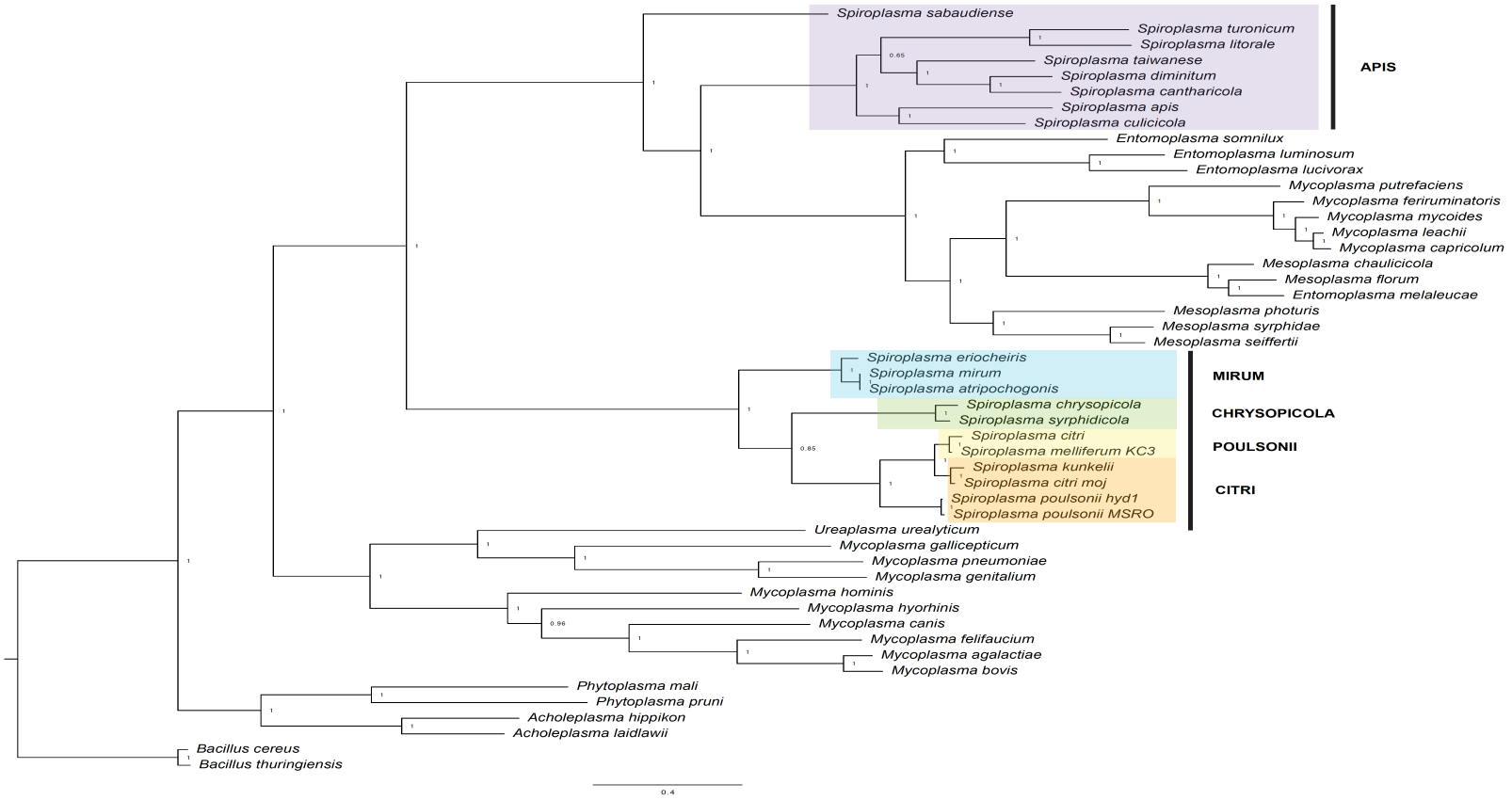


Figure S4. Bayesian analysis (MrBayes) of *Spiroplasma* within Mollicutes. Alignment comprises 2204 variable characters using MUSCLE; PhyloPhlAn removed identical regions across taxa before the phylogenetic analysis. Regions present in less than 60% of all taxa were removed. Analysis was performed using MrBayes v 3.2.6 in CIPRES. Highlighted *Spiroplasma* clades.

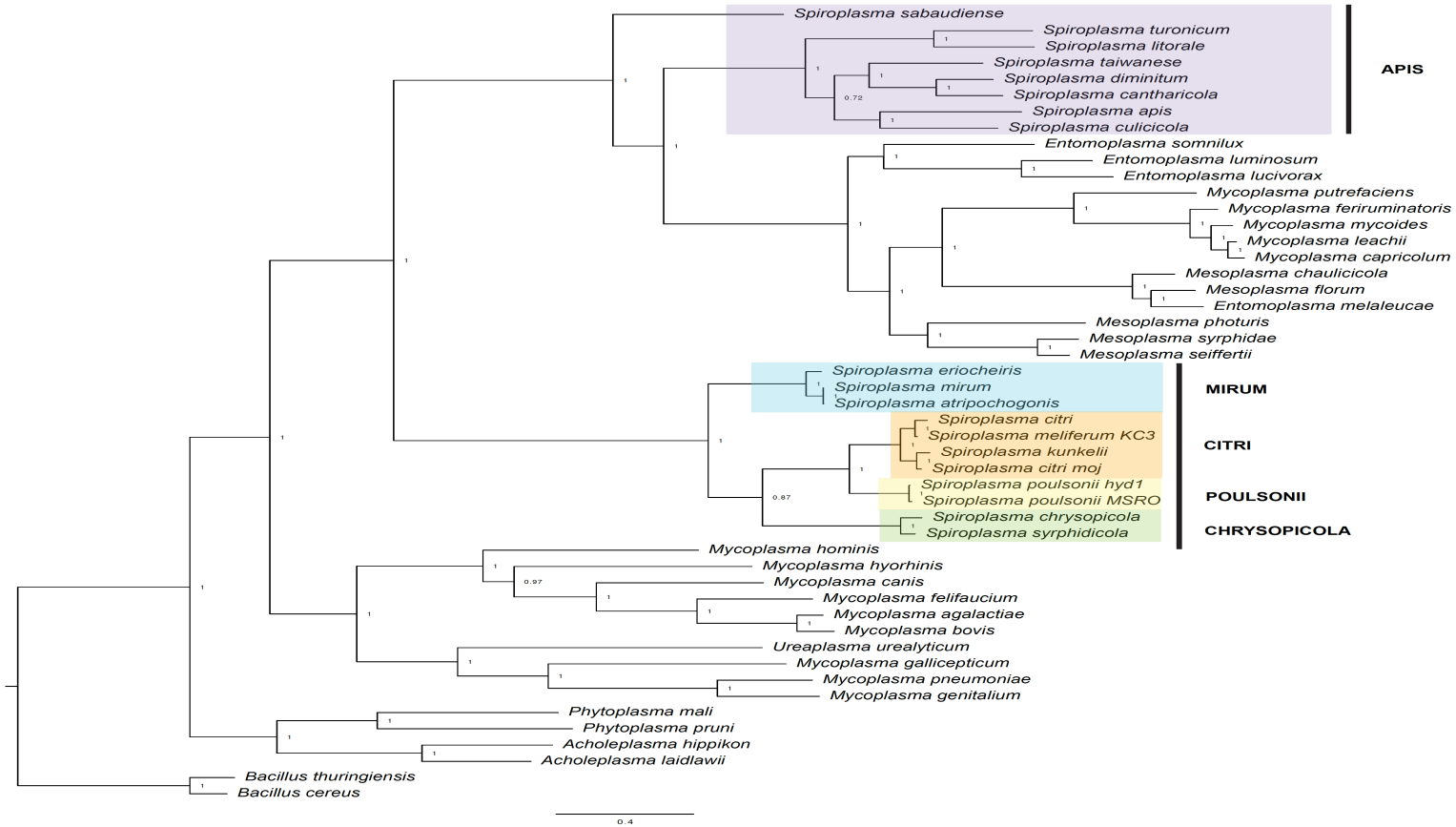


Figure S5. Bayesian analysis (MrBayes) of *Spiroplasma* within Mollicutes. Alignment comprises 3704 variable characters using MUSCLE; PhyloPhlAn removed identical regions across taxa before the phylogenetic analysis. Analysis was performed using MrBayes v 3.2.6 in CIPRES. Highlighted *Spiroplasma* clades.

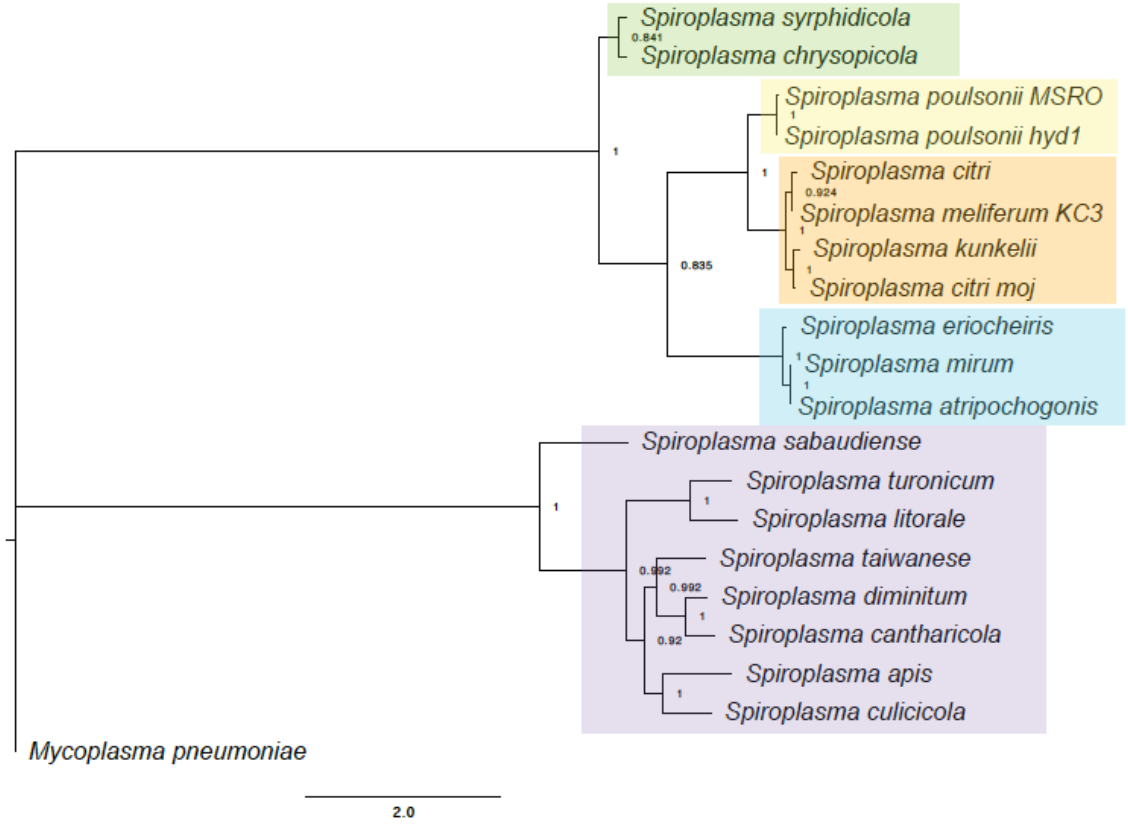


Figure S6. Phylogeny of *Spiroplasma*. Tree was reconstructed using PhyloPhlAn through the alignment of 399 conserved proteins. Alignment was constructed using MUSCLE with 2638 variable characters across taxa. Identical regions across all taxa were removed.

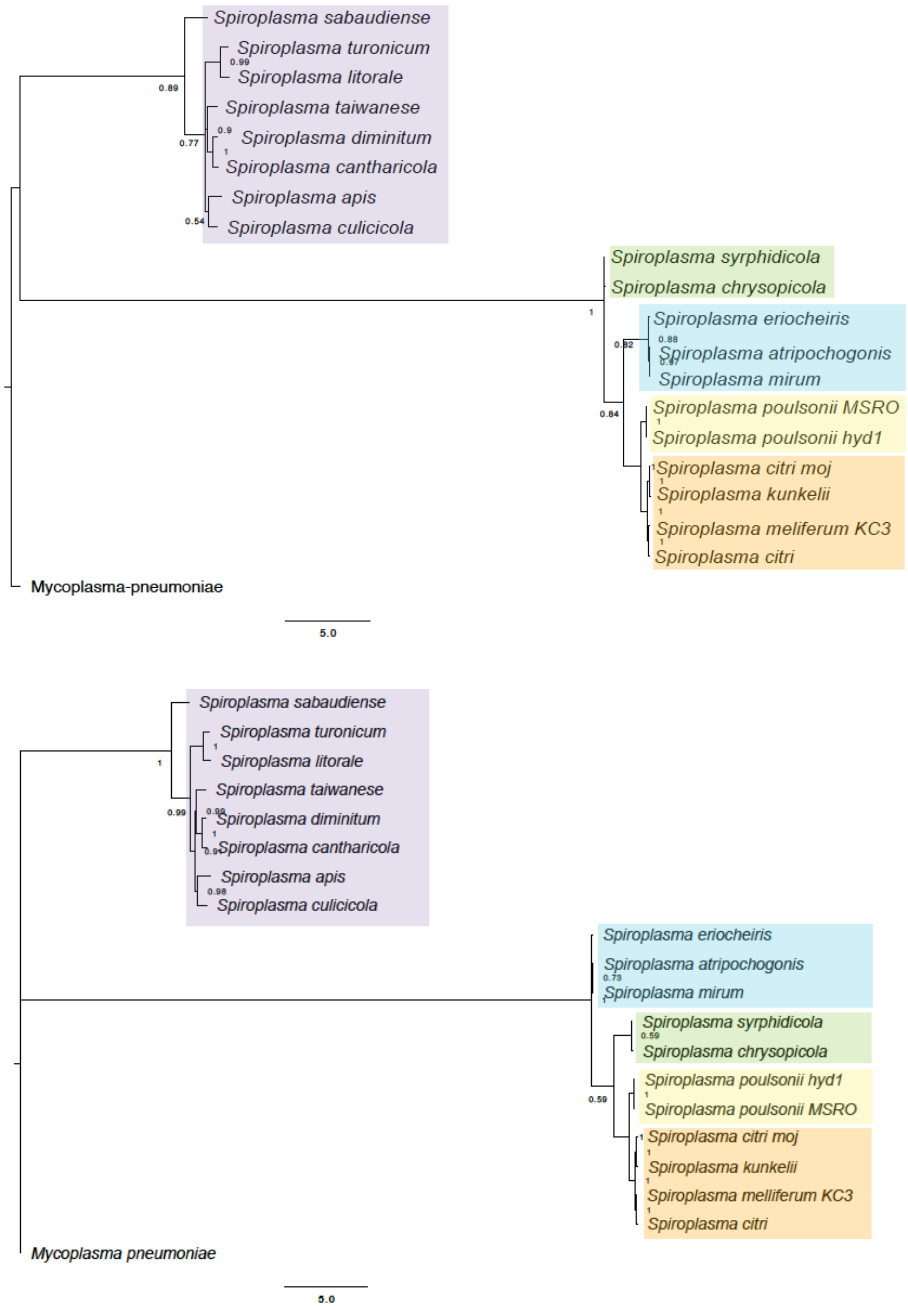


Figure S7 (upper). RAxML phylogeny of *Spiroplasma*. Alignment contains 2638 variable characters using MUSCLE; PhyloPhlAn removed identical regions across taxa before the phylogenetic analysis. **Figure S8 (lower).** RAxML phylogeny of *Spiroplasma*. Regions present in less than 60% of all taxa were removed Alignment contains 2250 variable characters. Analysis was performed using RAxML-HPC v 8.2.9 in CIPRES.

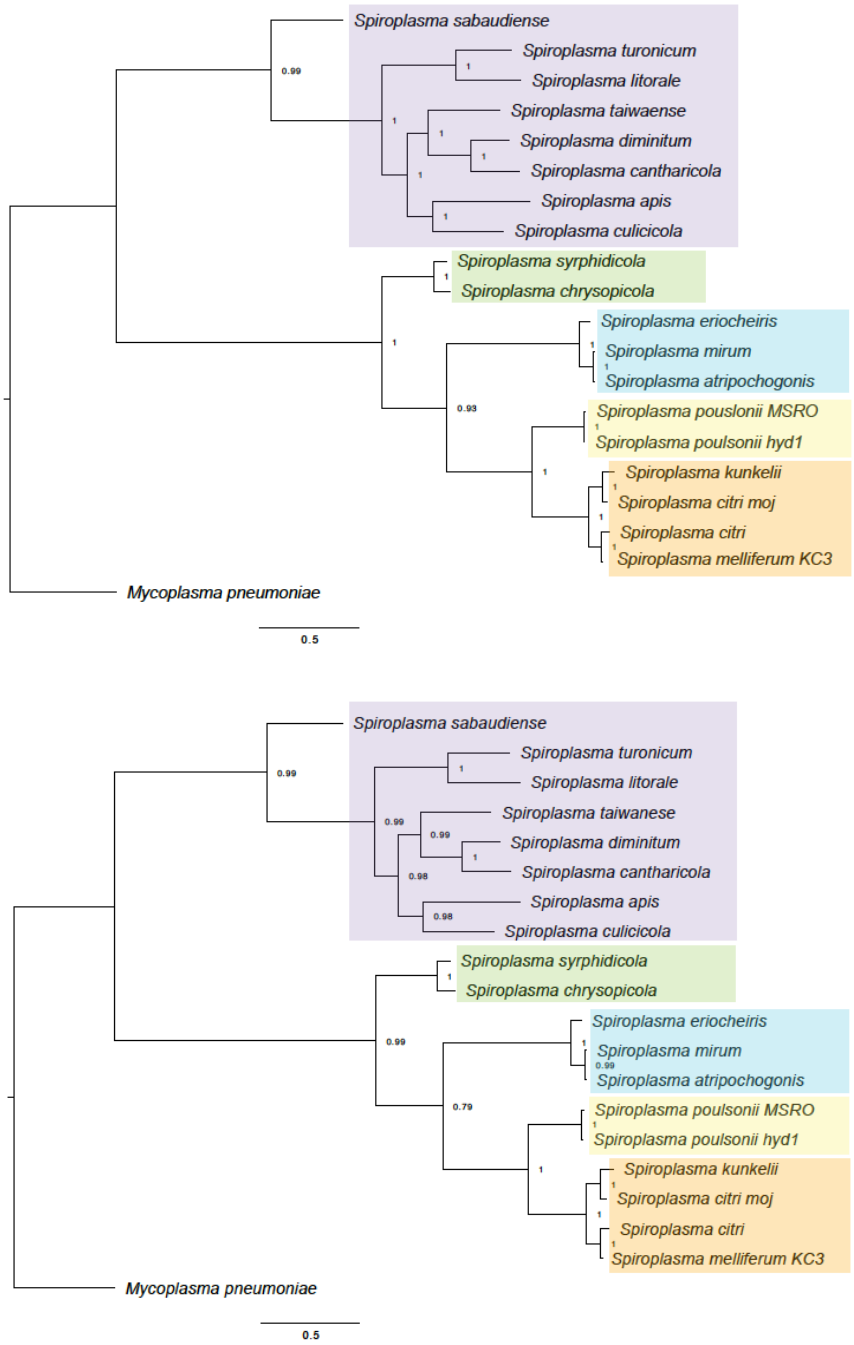


Figure S9 (upper). Bayesian phylogeny of *Spiroplasma*. Alignment contains 2250 variable characters; PhyloPhlAn removed identical regions across taxa before the phylogenetic analysis. Regions present in less than 60% of all taxa were removed **Figure S10 (lower).** Bayesian phylogeny of *Spiroplasma*. Alignment contains 2638 variable characters. In both trees analysis was performed through MrBayes v 3.2.6 in CIPRES

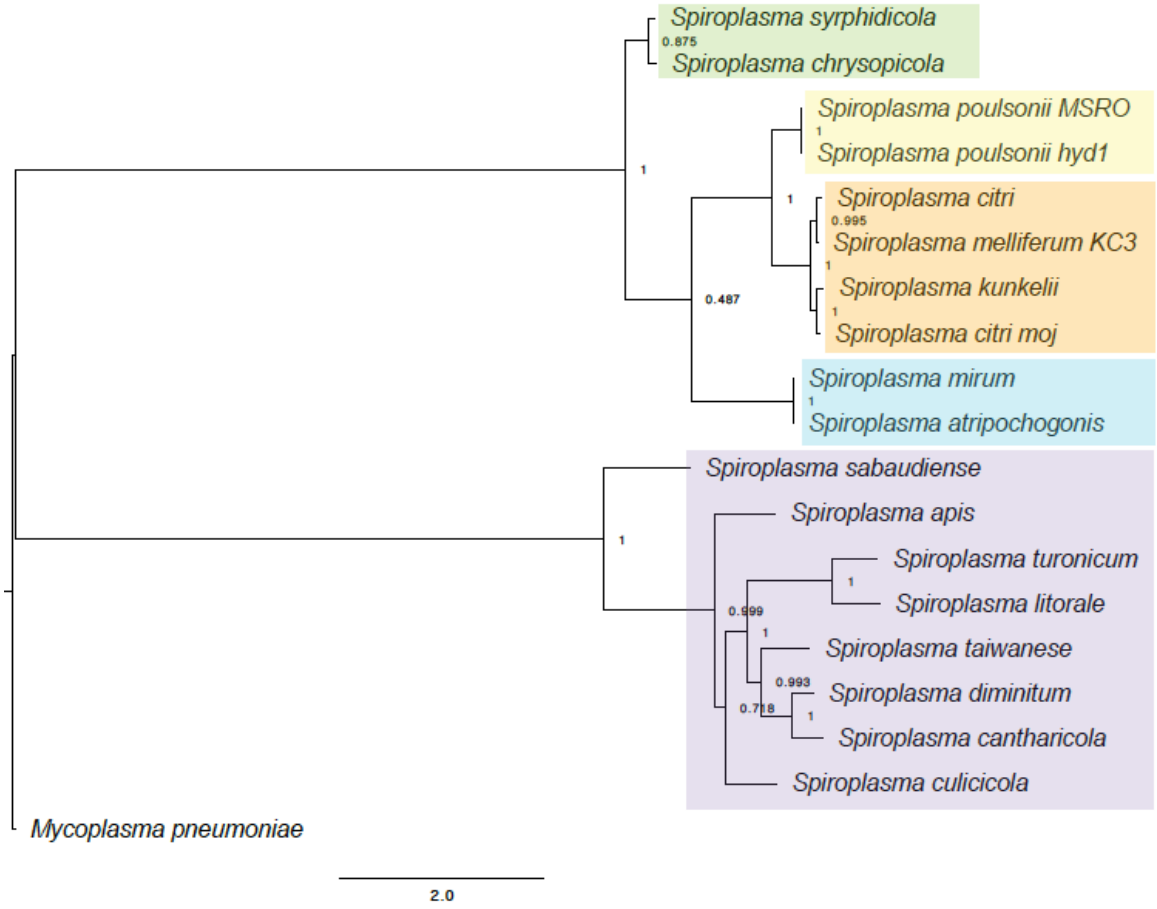


Figure S11. Phylogeny of *Spiroplasma*. Tree was reconstructed using PhyloPhlAn through the alignment of 399 conserved proteins. Alignment was constructed using MUSCLE with 2638 variable characters across taxa. Identical regions across all taxa were removed. *Spiroplasma eriocheiris* was excluded from the analysis

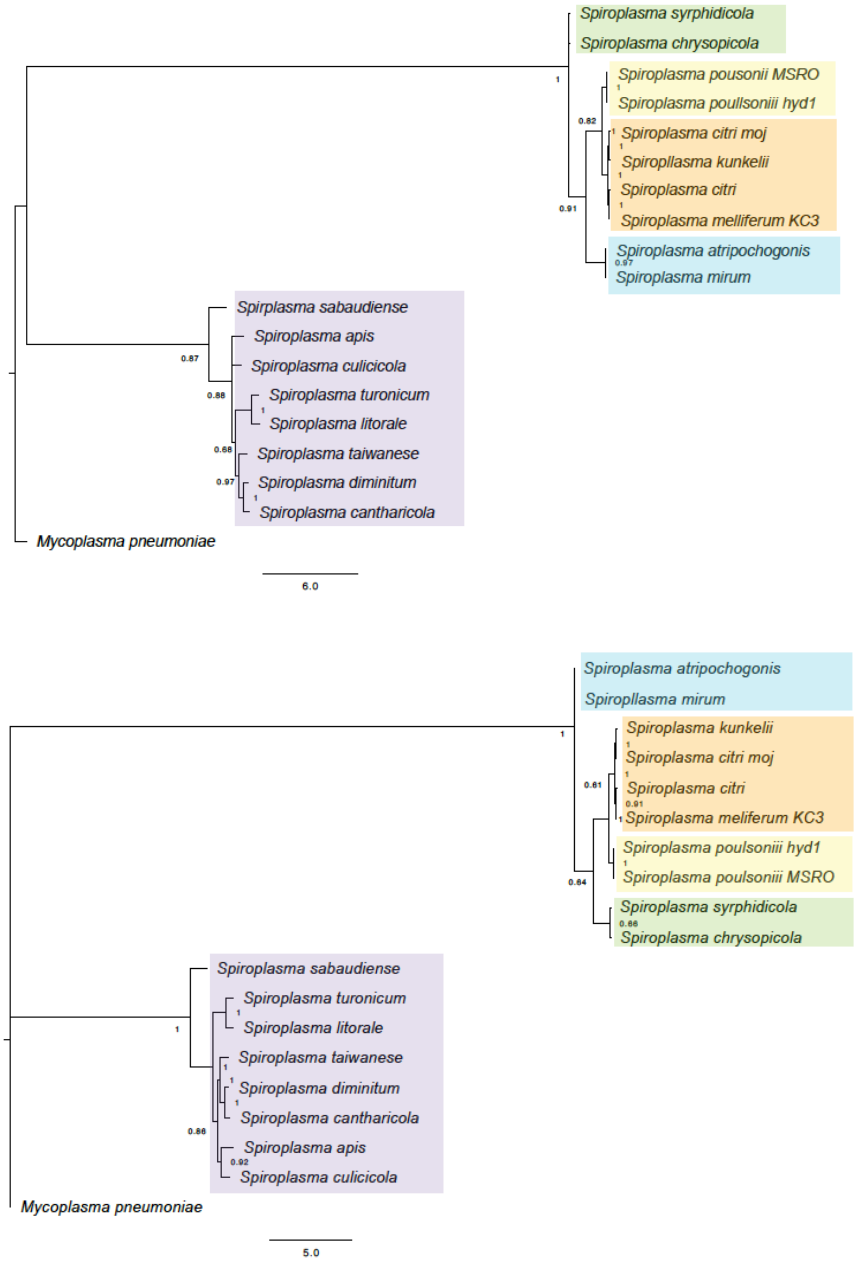
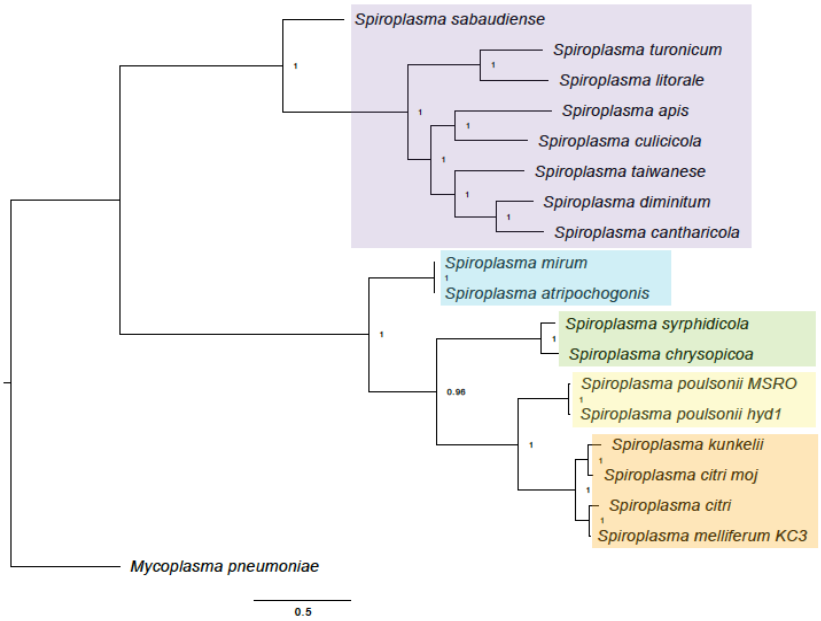


Figure S12 (upper). RAxML phylogeny of *Spiroplasma*. Alignment contains 2638 variable characters using MUSCLE; PhyloPhlAn removed identical regions across taxa before the phylogenetic analysis. **Figure S13 (lower).** RAxML phylogeny of *Spiroplasma*. Regions present in less than 60% of all taxa were removed Alignment contains 2310 variable characters. In both trees *Spiroplasma eriocheiris* was excluded from the analysis. Analysis was performed using RAxML-HPC v 8.2.9 in CIPRES.



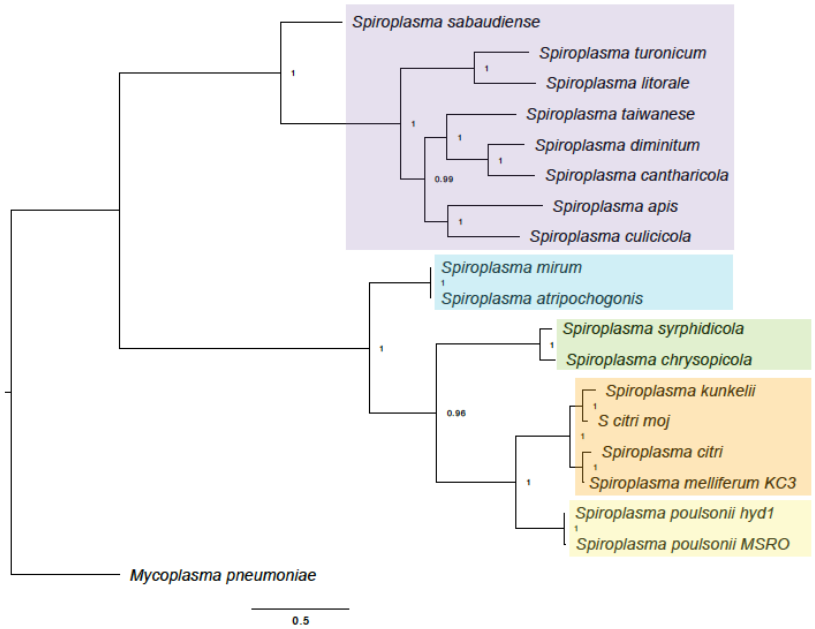


Figure S14 (upper). Bayesian phylogeny of *Spiroplasma*. Alignment contains 2638 variable characters; PhyloPhlAn removed identical regions across taxa before the phylogenetic analysis. **Figure S15 (lower).** Bayesian phylogeny of *Spiroplasma*. Alignment contains 2310 variable characters, regions present in less than 60% of all taxa were removed. In both trees analysis was performed through MrBayes v 3.2.6 in CIPRES and *Spiroplasma eriocheiris* was excluded

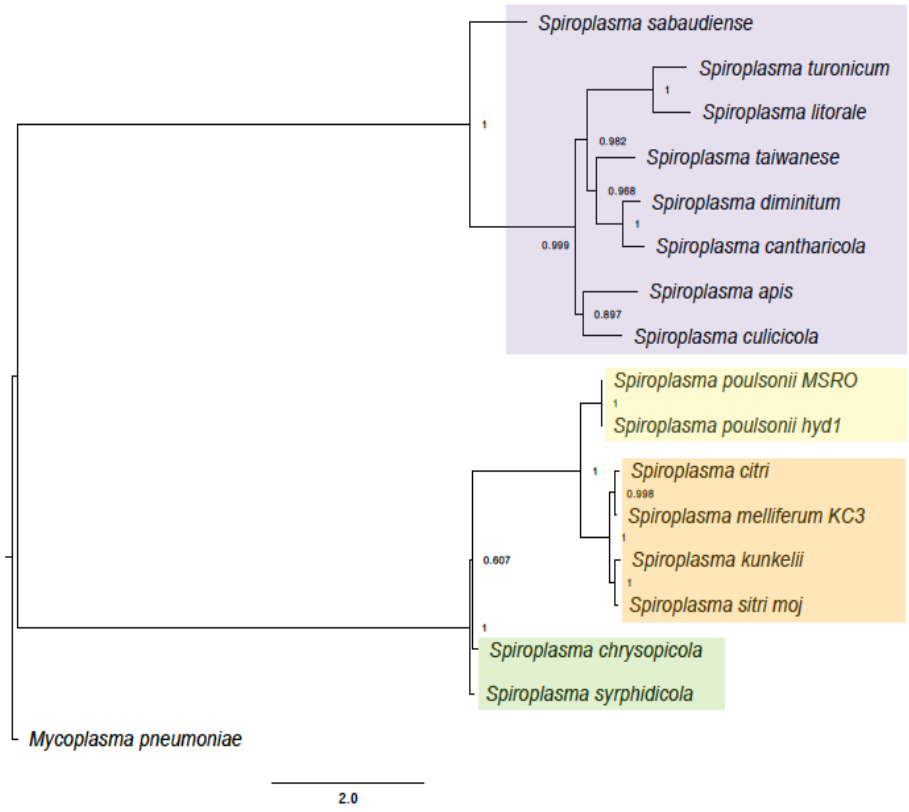


Figure S16. Phylogeny of *Spiroplasma*. Tree was reconstructed using PhyloPhlAn through the alignment of 399 conserved proteins. Alignment was constructed using MUSCLE with 2619 variable characters across taxa. Identical regions across all taxa were removed. Mirum clade was excluded from the analysis.

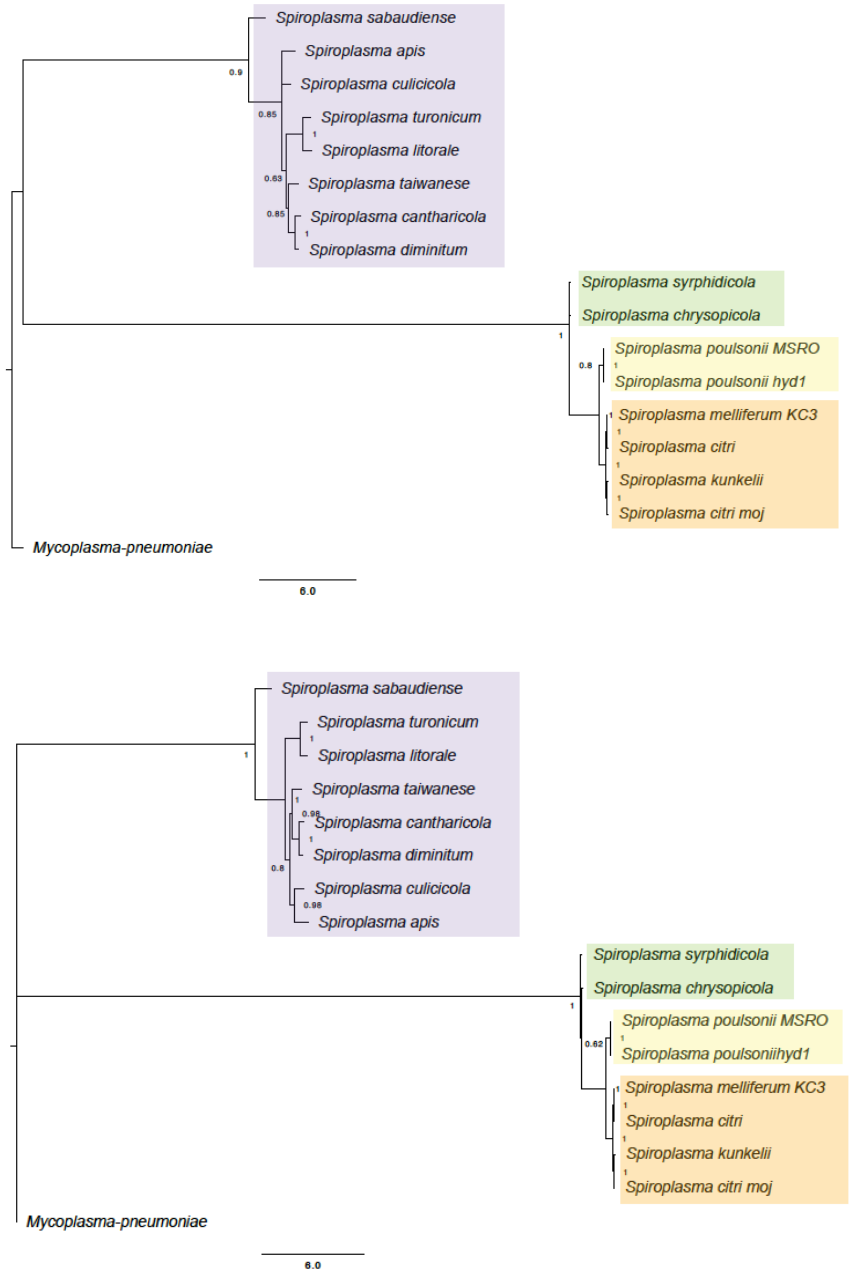
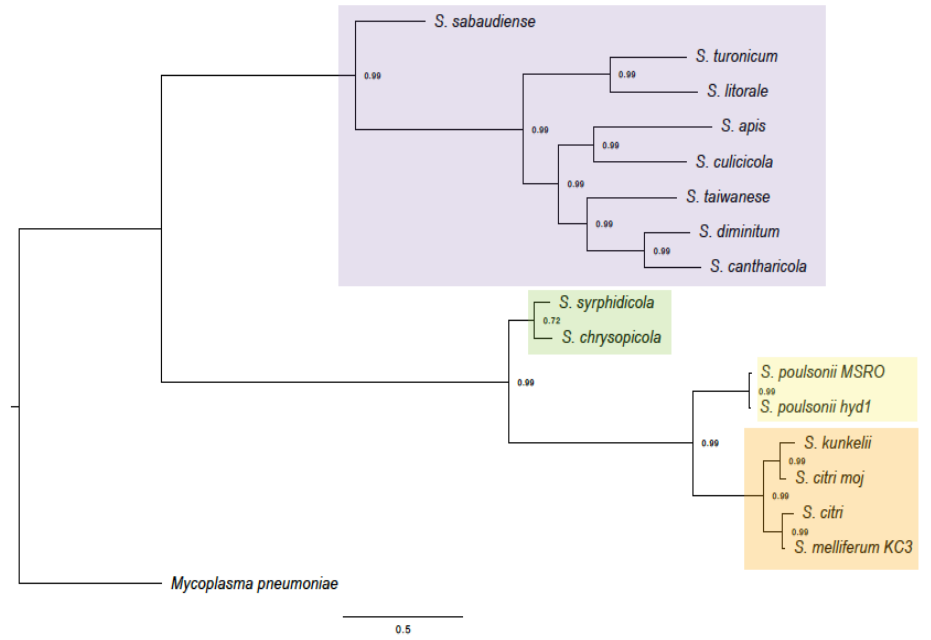


Figure S17 (upper). RAxML phylogeny of *Spiroplasma*. Alignment contains 2619 variable characters using MUSCLE; PhyloPhlAn removed identical regions across taxa before the phylogenetic analysis. **Figure S18 (lower).** RAxML phylogeny of *Spiroplasma*. Regions present in less than 60% of all taxa were removed Alignment contains 2301 variable characters. In both trees the mirum clade was excluded from the analysis. Analysis was performed using RAxML-HPC v 8.2.9 in CIPRES.



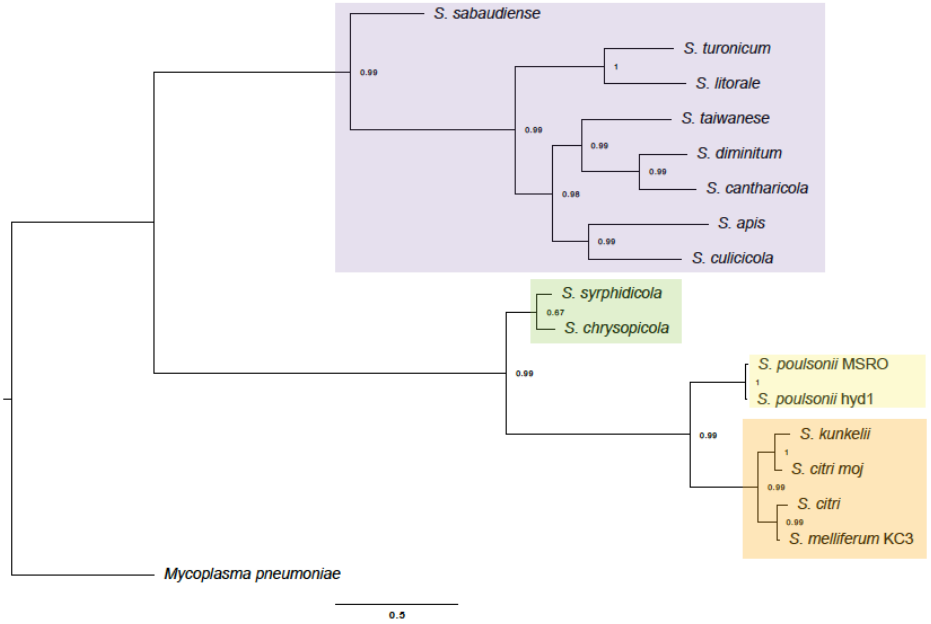


Figure S19 (upper). Bayesian phylogeny of *Spiroplasma*. Alignment contains 2619 variable characters using MUSCLE; PhyloPhlAn removed identical regions across taxa before the phylogenetic analysis. **Figure S20 (lower).** Bayesian phylogeny of *Spiroplasma*. Regions present in less than 60% of all taxa were removed Alignment contains 2301 variable characters. In both trees the mirum clade was excluded from the analysis. Analysis was performed through MrBayes v 3.2.6 in CIPRES.