GENOMICS OF BACTERIOPHAGES INFECTING SPIROPLASMA

A Thesis

by

PAULINO RAMIREZ

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

MASTER OF SCIENCE

Chair of Committee,	Mariana Mateos
Committee Members,	Jason Gill
	Paul de Figueiredo
Head of Department,	David Caldwell

May 2019

Major Subject: Wildlife and Fisherie Sciences

ABSTRACT

Viruses are in a unique position to affect the ecology of many organisms, due to their sheer numbers and diversity. Although traditionally considered pathogens, viruses can engage in more complex and mutualistic associations with their hosts. Given the large diversity and abundance of prokaryotes, viruses infecting bacteria (referred to as bacteriophages), are particularly diverse, and have been shown to engage in interesting relationships with their host. Bacteriophages found in endosymbiotic bacteria (i.e. those that live inside another organism), represent interesting tri-partite interactions that have not been recently investigated in the mollicute *Spiroplasma* bacterium, as compared to other endosymbionts such as *Wolbachia*. Due to the many phenotypes associated with the presence of bacteriophage in endosymbionts, *Spiroplasma* bacteriophages can potentially be of physiological importance to the *Spiroplasma*-host interactions. To shed light on this neglected research area, the first chapter of this thesis is a review focused on double-stranded DNA bacteriophages of *Spiroplasma*. The second chapter used a bioinformatics approach to seek potential interactions between *Spiroplasma* and bacteriophages, by searching for putative prophage sequences within the publicly available *Spiroplasma* genome sequences. Of the available genomes, only *S. citri*, *S. kunkelli* and *S. Chic* contained putative prophage regions in their genomes. The putative regions had high homology to each other. The third chapter attempted to break a several-decade hiatus of research on bacteriophage of *Drosophila*-associated *Spiroplasma*. Putative phage particles were isolated from the male-killing *Spiroplasma poulsonii*

NSRO bacterium hosted in *Drosophila melanogaster* flies. DNA of such particles was extracted and sequenced with Nanopore (long read) technology. Two putative phage draft contigs, NSROP-1 and NSROP-2, were assembled and bioinformatically annotated. The draft contigs were further characterized using homology found in a sister *Spiroplasma* strain.

ACKNOWLEDGEMENTS

I would like to thank my committee chain Dr. Mateos, and my committee members Dr. Gill, Dr. de Figueiredo for their guidance and support throughout the course of the research.

Thanks also to Dr. Justin Leavitt from Dr. Gill's lab for help with the design and technical implementation of techniques used in this paper.

Finally, thanks to undergraduates Mateos lab, Anika Stankov, Sarah Gheida, Nick Ramirez, Emile Fierro Morel, Savannah Bogard and Diandra Lombard for their assistance in lab projects and maintenance.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Dr. Mariana Mateos (advisor) of the Department of Wildlife and Fisheries, Dr. Jason Gill of the Department of Animal Sciences, and Dr. Paul de Figueiredo of Microbial Pathogenesis and Immunology.

Dr. Justin Leavitt primarily performed the electron microscopy and ultracentrifugation extractions techniques done in chapter 3.

Funding Sources

Graduate study and work were supported by Texas A&M University. The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of the Texas A&M University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION: REVIEW ON SPIROPLASMA-ASSOCIATED PHAGES

The vast presence of viruses across biomes makes them the most abundant biological element on Earth. Viruses are not limited to parasitism by lysis/killing their host; they can also engage in mutualistic and commensalistic relationships with it (Roossinck 2015). Viruses that infect bacteria (specifically referred to as bacteriophage) have been shown to have a profound impact on microbial communities. Bacteriophages are widely known for their pathogenic interactions with bacteria, namely their lytic and lysogenic replication cycles. Thus, they are generally associated with a negative impact on the fitness of bacteria. Some bacteriophages, however, are not as inherently antagonistic as others. For example, many bacteriophages code for genes that enhance the fitness of their host. The well-studied bacteriophages WO of *Wolbachia*, lambda of *Escherichia coli*, and CTX-phi of *Vibrio cholera* all contain genes that increase the virulence and transmission of their bacterial host in their respective eukaryotic host, albeit in different ways (reviewed in Bondy-Denomy and Davidson 2014). Despite the abundance of bacteriophages and their ecological relevance, they remain an understudied component of many biological model systems. One reason for this is that most phage investigation approaches rely on the ability to culture its bacterial host, which is not possible for many bacteria. In-silico techniques are helpful, but can only identify viruses and bacteriophages similar to those already sequenced; which is a drawback given the diversity of bacteriophages. Identifying and describing bacteriophages in novel organisms such as endosymbionts,

can facilitate progress in the discovery of new bacteriophages and new molecular mechanisms.

The nature of the association between bacteria and their hosts likely affects the ecological and evolutionary dynamics of phages (bacteriophages) and their bacterial hosts (Mavrich and Hatfull 2017). For example, the horizontal transmission dynamics of phages associated with vertically transmitted endosymbiotic bacteria of eukaryotes, where the fitness of the bacterium is tightly linked to that of the eukaryote, likely differ from those of phages associated with free-living and horizontally transmitted bacteria (Bordenstein and Reznikoff 2005). Therefore, phages from vertically-transmitted endosymbionts are expected to face different selective pressures to their non-heritable counterparts, and might exhibit different traits. Furthermore, heritable bacteria and the nature of their associations with hosts are diverse, ranging from reproductive parasitism to several forms of mutualisms (e.g. nutritional and defensive). How these eukaryote-bacteria associations affect phage-bacteria interactions has not been adequately examined. Few phages from heritable bacteria have been studied to date, such that general principles governing these tripartite interactions cannot be inferred at present.

Heritable bacterial endosymbionts (both intra- and extra-cellular) are widespread and diverse in arthropods, particularly insects (reviewed in Duron and Hurst 2013). These can be generally classified as obligate (primary), where eukaryote and bacterium are completely dependent on each other for survival and/or reproduction (e.g. *Buchnera* and

aphids), or facultative (secondary), where the eukaryote does not depend on the bacterium, but it may benefit from it in certain contexts (reviewed in Duron and Hurst 2013). Heritable facultative bacteria typically have imperfect vertical transmission, and are thus expected to be lost from host lineages unless mechanisms countering their loss are in place, such as reproductive manipulation or fitness benefits to the eukaryote (Werren and O'Neill 1997). Recent studies have revealed a crucial role of phages in the few interactions of eukaryotes and heritable bacteria examined to date, which are restricted to the Proteobacteria (Alpha- and Gamma-). The diversity and role of phages associated with other heritable bacteria (e.g. Mollicutes) are unknown.

Wolbachia (WO) phage has been shown to be responsible for the cytoplasmic incompatibility (CI) phenotype observed in many *Wolbachia* strains (Beckmann, Ronau et al. 2017, LePage, Metcalf et al. 2017). CI is a form of reproductive manipulation where females that are not infected with *Wolbachia* fail to produce successful offspring in matings with *Wolbachia*-infected males (Shropshire, On et al. 2018). Consequently, *Wolbachia*-infected females have a higher chance of successfully producing offspring than their *Wolbachia*-free counterparts, which leads to an increase in the number of individuals infected with *Wolbachia* in the population. Essentially, WO phage carry genes (*cifA* and *cifB*) that encode proteins responsible for damage and rescue of male sperm, such that the presence in the egg of a *Wolbachia* strain harboring a compatible phage-encoded gene is required to rescue the sperm defect upon fertilization (LePage, Metcalf et al. 2017). Phage are also involved in a defensive mutualism. The

bacteriophage ASPE2 (*Acyrthosiphon pisum* secondary endosymbiont) carried by the endosymbiont *Hamiltonella defensa*, carries toxin-encoding genes implicated in the protection of their aphid host from two parasitoid wasps (Oliver, Russell et al. 2003, Brandt, Chevignon et al. 2017). APSE phages also appear to influence the aphid-*H. defensa* interaction by preventing overproliferation of *H. defensa*. In the absence of wasp, bacteriophage-free *H. defensa* titers increased two-fold compared to their phage-carrying counterparts resulting in decreased aphid fitness (Weldon, Strand et al. 2012). Decreased fitness of aphids harboring the phage-free *H. defensa* suggests that the phage contribute to reducing fitness costs associated with *H. defensa* (Weldon, Strand et al. 2012). The above examples illustrate the influential role that phage can have on heritable facultative symbiont interactions, but how widespread these phenomena are and whether they differ between phage from heritable vs. non-heritable bacteria is unclear. The bacterial genus *Spiroplasma* (Class Mollicutes) offers a unique system to begin to answer such questions.

Members of the genus *Spiroplasma* infect a diverse range of arthropods and plants, where they engage in commensalistic, pathogenic, and mutualistic relationships (reviewed by Bolaños, Servín-Garcidueñas et al. 2015). *Spiroplasma* cells are generally motile, lack a cell wall and have a helical shape, at least under certain conditions (e.g. in extracellular environment and in exponential growth stages) (Gasparich, Whitcomb et al. 2004). *Spiroplasma* genomes are small (ca. 1–2 Mbp) and have low GC content. The transmission mode of *Spiroplasma* (predominantly vertical vs. horizontal) varies according to species/strain, but has evolved repeatedly; several maternally-transmitted

lineages have sister lineages that are horizontally transmitted. Horizontally-transmitted *Spiroplasma* strains are notable agents of pathogenic diseases, such as stunted growth in plants, and trembling syndrome in crustaceans (Whitcomb, Chen et al. 1986, Saillard, Vignault et al. 1987, Bové, Renaudin et al. 2003, Wang, Gu et al. 2011).

Among the vertically-transmitted *Spiroplasma*, several strains/species are associated with members of the fly genus *Drosophila* (Mateos, Castrezana et al. 2006) (Figure 1). *Drosophila*-associated strains are grouped into four clades (poulsonii, citri, ixodetis, tenebrosa) that represent independent invasions of *Spiroplasma* into *Drosophila* (Haselkorn 2010). One species of *Drosophila*, *D. hydei*, harbors strains from two different clades (poulsonii and citri; Mateos, Castrezana et al. 2006). Some of these heritable *Spiroplasma* strains, like other endosymbionts, have evolved reproductive manipulation in the form of male-killing, where males born to a *Spiroplasma*-infected female die at the embryonic stage (Montenegro, Petherwick et al. 2006). Due to a predominant (or exclusive) maternal transmission, males constitute an evolutionary dead end for *Spiroplasma.* Male-killing may enhance *Spiroplasma* persistence if *Spiroplasma*-infected females face lower larval competition as a result of the death of their male siblings (i.e., the resource release hypothesis; Hurst and Majerus 1993). Stemming from the highly female biased sex ratios caused by these strains, male-killing strains have been traditionally labeled "Sex Ratio Organisms" (SRO), usually preceded by the initial of *Drosophila* host species (e.g. "MSRO" for *D. melanogaster* sex ratio organism).

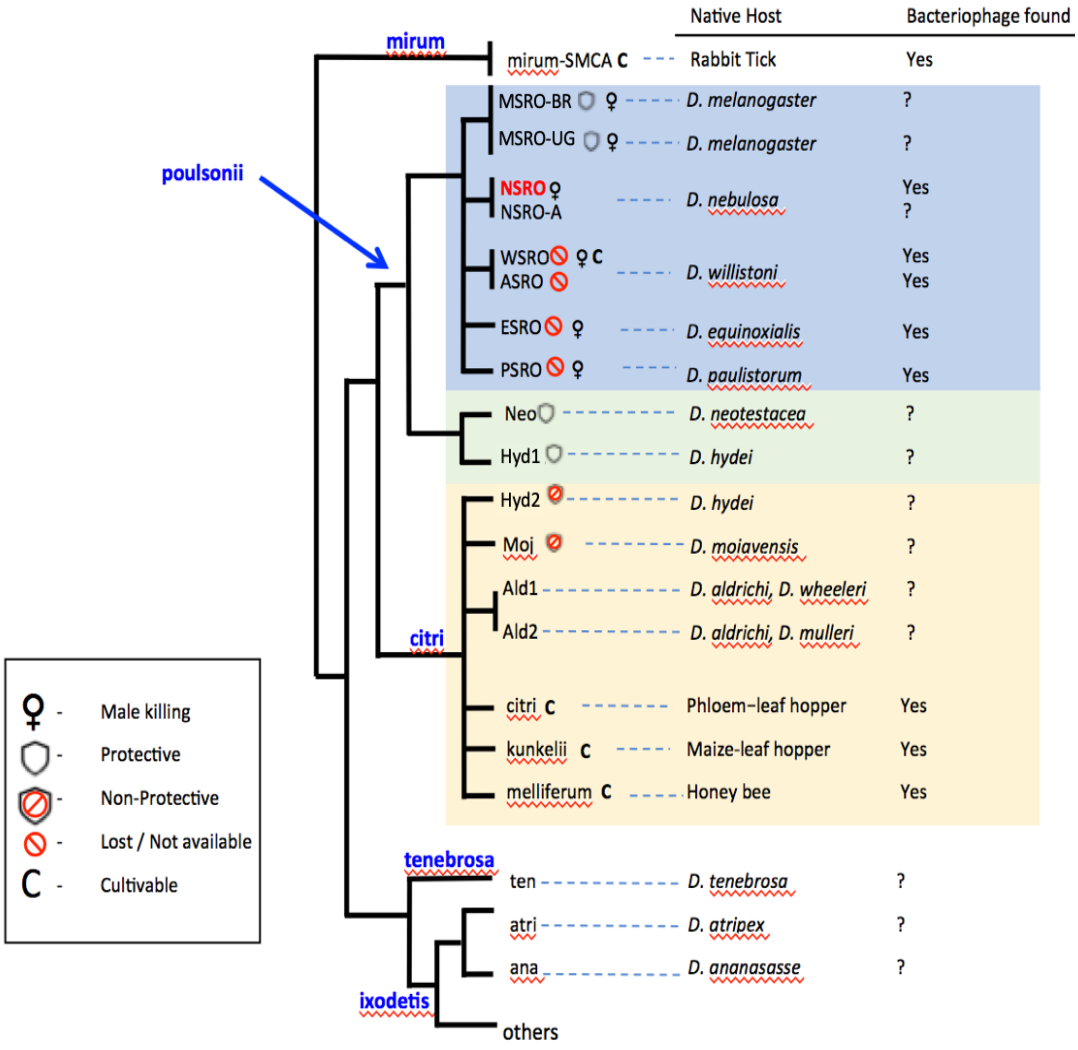

Figure 1.1 - Phylogenetic Tree of Relevant *Spiroplasma* (Adapted from Haselkorn 2010).

Phylogenetic tree of *Spiroplasma* species/strains associated with *Drosophila* and some of their close relatives (modified from Haselkorn 2010), showing known traits/phenotypes/phage, host species (*D.* = *Drosophila*). The four *Spiroplasma* clades associated with *Drosophila* are labeled with blue font. Blue background = subclade of poulsonii containing the known male-killing strains of *Drosophila*; green background = subclade of poulsonii that does not contain male-killing strains; yellow background = citri clade. The Red labeled strain is the host from which bacteriophage were studied in chapter 3.

Despite the occurrence of male-killing *Spiroplasma*, several *Drosophila*-associated strains are considered mutualistic, at least in the context of defense against natural enemies. *Spiroplasma* strains of the poulsonii clade associated with *D. melanogaster*, *D. hydei*, and *D. neotestacea* confer protection to the host fly against at least one species of parasitic wasp (and against a parasitic nematode in the case of *D. neotestacea*) (reviewed in Mateos et al. 2016). These examples suggest that fitness benefits under consumer pressure may be a common theme of *Drosophila*-*Spiroplasma* interactions. Whether *Spiroplasma* phage influence *Drosophila*-*Spiroplasma* interactions, such as male killing and protection against natural enemies, has not been studied.

Presence of phage has been documented for both horizontally-transmitted and heritable *Spiroplasma* (specifically those of *Drosophila*), with one or more of the following traits described: morphology; functional characterization (ability to lyse or affect host phenotype); genome size/structure; and genome sequence (only for phage from cultivable *Spiroplasma*; see Table 1.1). Described bacteriophage from the readily cultivable strains of *Spiroplasma* range in shape and include: filamentous rods, icosahedral heads consisting of 12 knob like capsomers (Dickinson and Townsend 1984); and Podovirus-like isometric heads with short tails (Maniloff and Dybvig 1988). Sequenced *Spiroplasma* phage genomes range in size from 4.42 to 8.27 kb, and fall within the families Microviridae (ssDNA, icosahedral; no tails) and Inoviridae (ssDNA; filamentous or rod-shaped). Thus, genome sequences of dsDNA Podovirus-like *Spiroplasma* phage have not been reported, despite being found across multiple *Spiroplasma* strains. The

chromosomes of multiple *Spiroplasma* strains, including several of the *Drosophila*-associated *Spiroplasma*, contain remnants of plectroviruses (Inoviridae) interspersed throughout the chromosome (Ku, Lo et al. 2013). Nonetheless, particles with the morphology of Inoviridae have not been observed in *Drosophila*-associated *Spiroplasma* (reviewed below), so it is unclear whether these sequences of plectroviral origin code for active viral particles.

Table 1.1 - *Spiroplasma*-phage Genome Sequences Available in the NCBI Database.

Bacteriophage	Taxonomy	Accession No.	Genome size (nt), and structure	Proteins encoded	Phage Host (Phenotype of host on eukaryote)
Spiroplasma phage 1-C74	Vespertiliovirus, Inoviridae	NC_003793	7768, ssDNA circular	13	*S. citri* (pathogenic)
Spiroplasma phage 1-R8A2B	Vespertiliovirus, Inoviridae	NC_001365	8273, ssDNA circular	12	*S. citri* (pathogenic)
Spiroplasma phage SVGII3	Unclassified Plectrovirus, Inoviridae	AJ969242	7878, ssDNA circular	11	*S. citri* (pathogenic)
Spiroplasma virus SkV1CR23x	Vespertiliovirus, Inoviridae	NC_009987	7870, ssDNA circular	13	*S. kunkelii* (pathogenic)
Spiroplasma phage 4	Spiromicrovirus, Gokushovirinae, Microviridae	NC_003438	4421, ssDNA circular	9	*S. melliferum* (pathogenic)
Spiroplasma phage SVTS2	"plectrovirus" unassigned; Inoviridae	NC_001270	6825 ssDNA circular	13	*S. melliferum* (pathogenic)

Most of the present knowledge on phage from *Drosophila*-associated *Spiroplasma* was generated prior to 30 years ago, and based on (axenic)culture-independent approaches, due to the fastidious nature of these *Spiroplasma* strains. Isolation of *Spiroplasma* putative phage particles was achieved via two general extraction approaches performed on whole flies that harbored heritable *Spiroplasma* either naturally or following artificial transfer (via fly hemolymph transfer). The "first degree" extraction was achieved by homogenizing adult flies in buffer, centrifuging to remove large debris, heating at 60ºC (on the assumption that such treatment would kill *Spiroplasma* cells but not phage particles), and centrifuged again to pellet (and remove) *Spiroplasma* cells and retain the supernatant (i.e., the putative phage extract). This "first-degree" extraction yielded particles resembling Podoviridae (isometric heads 35–45 nm in diameter) for two *Spiroplasma* strains (NSRO and "Hydei"), but not for four strains (WSRO, ESRO, PSRO, ARSO; see Table 1.2). For these latter strains, a "second-degree" extraction was required to yield particles (also resembling Podoviridae). The second-degree extraction encompassed injection of "first-degree" extract (e.g. from NSRO-infected flies) into flies infected with a different strain of *Spiroplasma* (e.g. WSRO, ESRO, PSRO) to induce/activate their putative native phage, followed by a similar extraction procedure to the first-degree extraction. Transmission Electron Microscopy (TEM) of NSRO suggests that viral particles "bud out" of the *Spiroplasma* cell (Williamson, Oishi et al. 1977). TEM of hemolymph directly obtained from flies infected with NSRO also revealed podovirus-like particles.

Table 1.2 - Genomic and Other Features of Phages Extracted from, *Drosophila*-Associated *Spiroplasma*.
Genomes sizes were determined by 0.7% agarose electrophoresis. Different phages with identical genome sizes were distinguished by distinct restriction digestion patterns, confirmed by southern blot hybridization to a labeled 6.2 kpb EcoRI fragment of NSRO-virus cloned in PBR325 (Cohen, Williamson et al. 1987). Extraction approaches are described in text.

Drosophila*-Associated *Spiroplasma	**Size of individual phage genomes (kbp)**	**Notes**	**Extraction Approach**	**Viewed via TEM**
NSRO	21.8	Circularly-permuted linear genome	First-degree	Yes[b]; 35-45 nm isometric head with a short tail
	21.8	Circularly-permuted linear genome	First-degree	Yes[b]
Hydei	21.8	Circularly-permuted linear genome	First-degree	Yes
ESRO	17	Possibly integrated into *Spiroplasma* chromosome[c]	Second-degree	Yes
PSRO	17	Possibly integrated into *Spiroplasma* chromosome[c]	Second-degree	Yes[b]
	21.8		Second-degree	Yes[b]
WSRO[d]	17	Possibly integrated into *Spiroplasma* chromosome[c]	Second-degree	Yes[b]
	>30		Second-degree	Yes[b]
ARSO[a]	17	"Strongly hybridizes with 17-kbp viruses of ESRO and PSRO, but not to the 21.8- or the >30-kbp viruses"	Unknown	Yes

[a] non-male-killing mutant of WSRO (a.k.a. $R_3 \leftarrow 3g$)

[b] because only one type of virus morphology has been observed, it is unclear whether it represents one or both viruses

[c] based on "strong hybridization" of $R_3 \leftarrow 3g$ viral DNA (used as probe) to undigested total DNA from the *Spiroplasma* strains containing the 17-kbp virus but not to strains NSRO or Hydei; akin to the 16-kbp virus strain of SPV3 of *S. citri* (Dickinson, Townsend et al. 1984)

[d] Radioactive thymidine replication assays suggested that the WSRO bacteriophage replicated in the WSRO-infected flies that had been injected with the NSRO-virus

Table 1.3 - Bacteriophage of other *Spiroplasma.*

Relevant Double-Stranded DNA Bacteriophages of *Spiroplasma* that do not have publicly available genome sequences (references in text).

Spiroplasma phage	**Size of individual phage genomes (kbp)**	**Notes**	**Replication**	**Viewed via TEM**
Citri 608: SVC3-608	21.0	Circularly-permuted linear genome : CsCl causes partial dissociation of virions	Non-lytic cytocidal infectious	Yes; 35-44 nm isometric head with a 13- 18 nm short tail
mirum–SMCA: SVC3-SMCA	21.0	SMCA- Suckling mouse cataract association: CsCl causes partial dissociation of virons: Circularly-permuted linear genome	Non-lytic cytocidal infectious	Yes; 43-54 nm isometric head with a 14 nm short tail
Citri ASP-3: AV9/3	21.0	Circularly-permuted linear genome	Non-lytic cytocidal infectious	Yes; 43-54 nm isometric head with a 14 nm short tail
Citri SP-V3: ag	21.0	Circularly-permuted linear genome	Non-lytic cytocidal infectious	Yes; 43-54 nm isometric head with a 14 nm short tail
Citri SP-V3: ai	16	Linear with cohesive ends, circular, concatemers: Lysogenic and spontaneous release of virons: plaquing properties similar to temperate phages	Non-lytic cytocidal infectious; lysogenic	Yes; 43-54 nm isometric head with a 14 nm short tail
Acholeplasma virus L3	39	Circularly-permuted linear genome: redundant ends	Non-lytic cytocidal infectious	Yes: isometric head(60nm), collar (8nm thick and 16nm wide), tail (10 nm wide and 20 nm long), and fibers attached to the tail similar to T7 phage group

Based on DNA isolation from the above putative phage extracts, followed by gel electrophoresis in combination with restriction digestion, southern blot, and/or DNA-DNA hybridization assays, it was determined that these extracts contained the genomes of one or two DNA phages with sizes shown in Table 1.2: 17; 21.8; and 30 kbp (Cohen, Williamson et al. 1987). The 17 kbp phage appears to be integrated into the host chromosome, akin to a 16 kbp phage of *S. citri* (Dickinson, Townsend et al. 1984). The 21.8 kbp phage of strains NSRO and Hydei appear to have a circularly-permuted linear genome.

SRO bacteriophages appear to be part of a larger group of *Spiroplasma* and *Acholeplasma* tailed-phage that possess a cytocidal non-lytic process of membrane budding to escape the host cell (Table 1.3). EM studies show phages (individually or in small groups) encased in a intermediate membrane vesicle post budding, which is lost before attachment to potential host *S. citri* (Alivizatos, Townsend et al. 1982). The *Spiroplasma* (*S. citri* and *S. mirum*) exclusive C3 group has an icosahedral head (40nm), a baseplate, and a tail (13-18nm). The genomes consist of linear circularly permutable double stranded DNA with sizes ranging from 13-21 kbp. Additionally, the bacteriophage *ai* of *S. citri* has plaque formation properties and genome cohesive ends similar to temperate phages such as lambda (Dickinson, Townsend et al. 1984). Lysogenesis of bacteriophages into the host genome has been suggested to account for the observation that *ai* bacteriophages can be found in *S. citri* following long-term remission from symptoms of an *ai* infection (Alivizatos, Townsend et al. 1982). *Ai*

phage-active *S. citri* produced milder symptoms in the plant than phage-inactive *S. citri*. *Spiroplasma* Virus C2 bacteriophage-like particles have been isolated from S. *citri* with no observable biological activity. The morphology of the C2 bacteriophage is a isometric head (50-55nm), a tail (6-8 nm wide and 75-83 long), and a base plate (Maniloff and Dybvig 1988).

Evidence that the putative phage extracts from *Drosophila* described above contained at least one type of active *Spiroplasma* phage (rather than virus from *Drosophila* or phage from other *Drosophila*-associated bacteria) came from *in vivo* assays that revealed their ability to cause lysis in at least one *Spiroplasma* strain derived from a different *Drosophila* species (or geographic region), as well as their ability to alter the male-killing phenotype of some strains (summarized in Table 1.4). The *in vivo* assays generally involved injecting the putative phage extracts (in serial dilutions) into flies infected with *Spiroplasma*. The effects of this treatment on *Spiroplasma* cell lysis (based on dark field microscopy observations of hemolymph made at different time points) and on sex ratio of offspring were recorded. Bacteriophage derived from *Drosophila*-associated *Spiroplasma* did not appear to induce lysis of cultivable *Spiroplasma*, as evidenced by an attempted plaque assay on 25 strains (Cohen, Williamson et al. 1987). It is important to note that at the time the above *Spiroplasma* phage studies were performed, the existence of *Wolbachia* as a common endosymbiont of *Drosophila* was unknown. Both *D. willistoni* and *D. paulistorum*, but not *D. hydei* (Mateos, Castrezana et al. 2006), harbor *Wolbachia* (in *D. paulistorum* early reports of hybrid male sterility

among races were attributed to intracellular *Mycoplasma*-like-organisms (Kernaghan and Ehrman 1970), which were later determined to be *Wolbachia* (Miller, Ehrman et al. 2010). It is therefore possible, that one or more of the phage attributed to *Spiroplasma* on the basis of morphology and genomic features could represent *Wolbachia* phage (WO) instead; which are reported to have isometric heads ca. 40 nm in diameter containing linear dsDNA ca. 20 kbp; (Fujii, Kubo et al. 2004, Bordenstein, Marshall et al. 2006). Nonetheless, the ability of putative phage extracts to lyse *Spiroplasma* cells, suggests that at least one *Spiroplasma* phage occurs in each of these strains.

Table 1.4 - Effects of *Drosophila*-Associated *Spiroplasma* Phage on Several Spiroplasma Strains.
The first column lists the viruses and their corresponding host strain (and geographic origin). The first row lists the *Spiroplasma* strains that were subjected to each phage (via injection of phage extract to *Drosophila* host containing corresponding *Spiroplasma* strain; in vivo assay). N = no observable lysis; L = lysis of Spiroplasma cells detected. ♀ = Only (predominantly) female progeny observed (as expected with viable/functional infections by SRO strains). ♀♂ male and female progeny observed; ? = has not been tested (Williamson, Oishi et al. 1977). Green cells indicate lysis and loss of male-killing phenotype. Yellow cells indicate no lysis and no alteration of sex ratio phenotype. Blue cells indicate lysis but not alteration of the sex ration phenotype. Outcomes of grey cells have not been reported.

***Spiroplasma* strain** ➔ **Virus strain** ➔	**NSRO**	**WSRO**	**ESRO-A**	**ESRO-B**	**PSRO**	**Hydei**[a]
NSRO (Haiti) - spv-1	**N** ♀	**L** ♀♂	**L** ♀♂	**L** ♀♂	**L** ♀♂	**L** ♀♂
WSRO (Jamaica) - spv-2	**L** ♀	**N** ♀	**? ?**	**? ?**	**? ?**	**? ?**
ESRO-A (Puerto Rico) - spv-4	**L** ♀	**N** ♀	**N** ♀	**? ?**	**N** ♀	**? ?**
ESRO-A (Santo Domingo) - spv-4	**L** ♀♂	**N** ♀	**N** ♀	**? ?**	**? ?**	**? ?**
ESRO-A (Jamaica) - spv-4	**L** ♀	**? ?**	**N** ♀	**? ?**	**? ?**	**? ?**
ESRO-B (Amazons - Brazil) - spv-5	**L** ♀♂	**L** ♀♂	**L** ♀♂	**N** ♀	**L** ♀♂	**? ?**
PSRO (North-East Brazil) - spv-5	**L** ♀♂	**L** ♀♂	**L** ♀♂	**? ?**	**N** ♀	**? ?**
PSRO (North Brazil) - spv-5	**L ?**	**L** ♀	**? ?**	**? ?**	**? ?**	**? ?**
PSRO (Grenada) - spv-6	**L** ♀♂	**? ?**	**? ?**	**? ?**	**? ?**	**? ?**
Hydei [a] (non-male killing) virus	**L** ♀♂	**? ?**	**? ?**	**? ?**	**? ?**	**N** ♀♂

[a] Likely Hyd1 (poulsonii clade) due to its geographic origin (i.e. Japan; see text).

In addition to causing lysis and disrupting the male-killing ability of some *Spiroplasma* strains, phage appeared to influence another *Spiroplasma* phenotype known as clumping or aggregation. When drops of hemolymph from flies infected with *Spiroplasma* from different sources (e.g. naturally found in different *Drosophila* species) were mixed under immersion oil and visualized with dark field, an aggregation or clumping of *Spiroplasma* cells would occur (Oishi and Poulson 1970). The clumping reaction was absent if hemolymph drops came from individuals infected with a *Spiroplasma* strain originally found in the same *Drosophila* species (and geographic region). Therefore, before the advent of PCR and DNA sequencing, the clumping reaction was used to distinguish different *Spiroplasma* strains (Williamson, Oishi et al. 1977). One particular study suggests a potential connection between the aggregation phenotype and *Spiroplasma* bacteriophage. Oishi, Poulson et al. (1984) exposed *Spiroplasma* NSRO cells (*in vivo*) to extracts of Hydei phage. Most of the NSRO cells subjected to Hydei phage were lysed, but a few survived. Despite this lysis, over a period of 16 days no appearance of males in the offspring was detected. Subsequent combination of the surviving "exposed" NSRO cells with "naïve" NSRO cells (i.e., those originating from flies that were not injected with the Hydei phage extract) resulted in clumping. It is unclear whether the Hydei strain used by Oishi et al., which was collected in Japan, was "Hyd1" (poulsonii clade) or "Hyd2" (citri clade). Nonetheless, it was likely Hyd1, because it is currently the more widespread of the two, and the only one reported in Japan (Kageyama, Anbutsu et al. 2006, Haselkorn, Markow et al. 2009).

CHAPTER II

BIOINFORMATIC SEARCH FOR PHAGE-LIKE SEQUENCES IN PUBLISHED GENOMES AND METAGENOMES OF *SPIROPLASMA*

To identify sequences of potential phage in available genomes of *Spiroplasma*, the following analyses were performed. All publicly available genomes of *Spiroplasma* species were submitted to the program Phaster (Arndt, Grant et al. 2016), for detection of prophage regions in chromosomal sequences and phage-like contigs in published assemblies. Regions highlighted by the programs were then manually inspected for genes that had homology to any known virion-coding genes. In addition, regions found in the immediate vicinity of the putative prophage region were examined for the presence of genes coding for polymerases, toxins, proteases, restriction modification enzymes, and unique genes not found in other *Spiroplasma* species. These regions were then searched in other strains of *Spiroplasma* to identify potential regions of homology.

Results

Using the prophage-finding program Phaster, several sites of potential prophage/phage were found for the genome assemblies of *Spiroplasma kunkelli CR2-3x, Spiroplasma citri R8-A2*, and the metagenomic assembly of *S. sp.* chic (re-assembled; see Appendix). Phaster highlighted an interesting ~ 10Kbp region in the each of these genomes. Further analysis showed the presence of tail, capsid, portal, and terminase coding genes. Given that Caudoviridae-like phages of ca. 21kbp have been reported in *S. kunkelli* and *S. citri*, there is premise that the10kbp segments could be of bacteriophage origin. The regions

were found as prophage (embedded in the intact bacterial chromosome) in both *S. kunkelli CR2 -3x* and *S. citri R8-A2* (both have complete chromosomes) but found as a metagenomic contig in *S. sp.* chic. All three species share a homologous 10kbp region consisting of major tail, minor tail, capsid, portal, and three hypothetical protein-coding genes (Figure 2.1, Table 2.1). In addition, the three strains share a terminase-coding gene except that in *S. sp chic* it seems to be truncated or misassembled. A baseplate coding gene is present in *S. citri R8-A2* and *S. sp chic* but missing in *S. kunkelli CR2 -3x*. The areas around this this phage virion element in all three species appear similar to each other. For example, both *S. kunkelli* and *S. citri* share two genes, a restriction enzyme-like and DNA binding-like coding genes, at the 3'end region, reminiscent of a restriction modification system, which is typically found in phages (Figure 2.1, Table 2.1) (Roberts 2005).

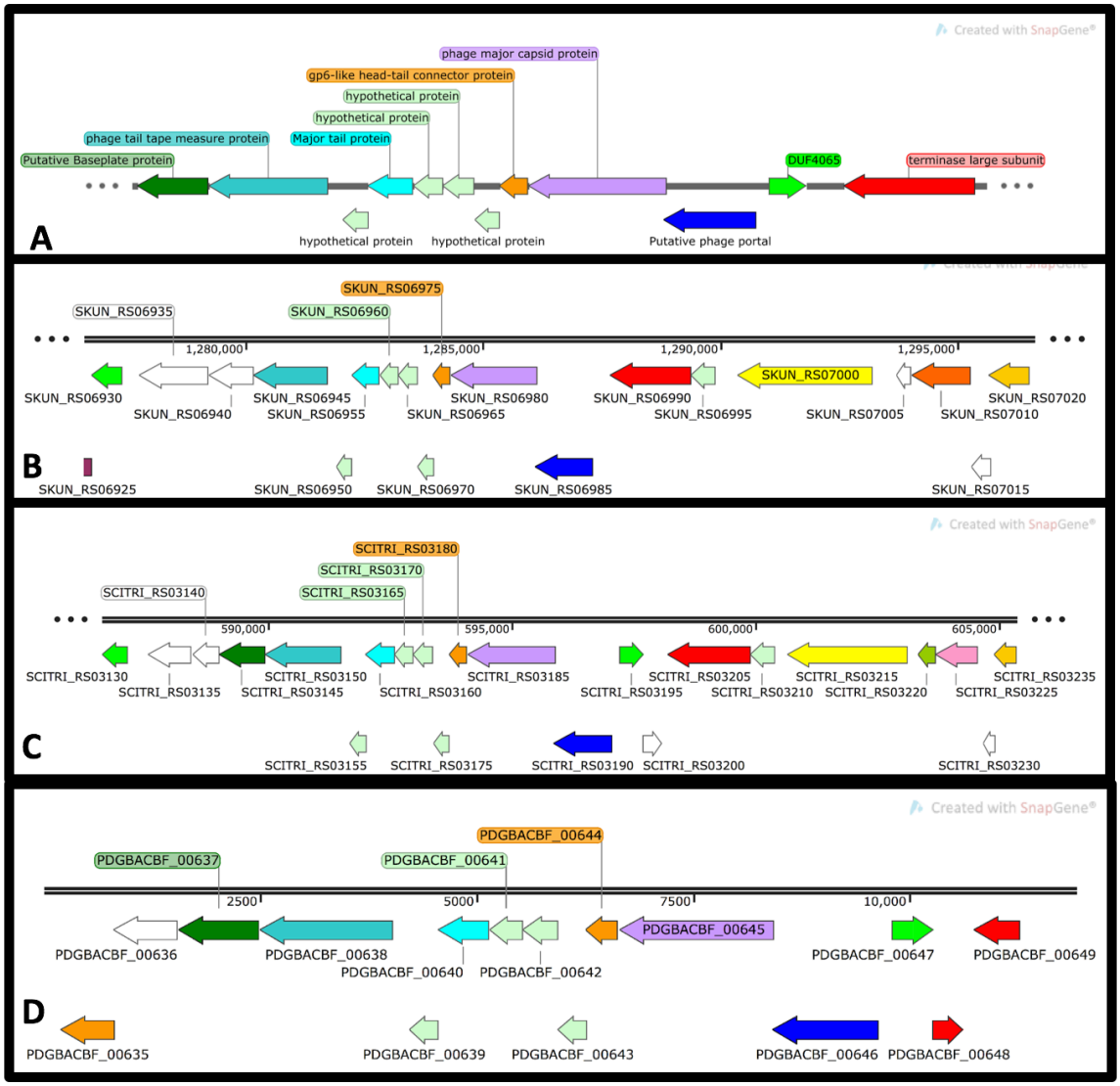

Figure 2.1 - Genomic Regions of Putative *Spiroplasma* Prophage.
A) Main homologous 10kbp region found in all three *Spiroplasma* strains. However, the putative Baseplate protein, DUF4065, is only found in *S. citri* and *S.* sp. chic. 5' end on the right and 3' end region on left. B) *Spiroplasma kunkelli* CR2-3X, highlighted prophage like region of 20kbp. C) *Spiroplasma citri* R8-2A, highlighted prophage like region of 19kbp. D) *Spiroplasma* sp chic., prophage like meta-contig of 12kbp. All arrows represent a gene. NCBI locus tags are placed below or above the respective genes except for the *S.* chic contig, which has protein IDs instead. All colors represent a homologous gene shared between two or more of the putative prophages (See table 2.1). White arrows are unique genes not shared between any putative prophage.

Table 2.1 - Description of Putative Prophage Gene Products.
NCBI gene IDs matching to their respective gene product for each of the three genomes examined. Putative virion protein and notable coding genes were analyzed bioinformatically with HHMER (EBI) to determine PFAM domains and other protein features. Color names correspond to colors in annotated figures.

	S. kunkelli	***S. citri***	***S.* chic**	**PFAM Domains**	**Coiled-Coiled**	**Transmembrane**
Baseplate (Dark Green)	NA	SCIRTI_RS03145	00637	Phage-related minor tail protein (PF10145)	Y	Y
Tail tape measure (Aqua)	SKUN_RS06945	SCIRTI_RS03150	00638	NA	N	N
Tail major (Light Blue)	SKUN_RS06955	SCIRTI_RS03160	00640	NA	Y	N
Capsid (Light Purple)	SKUN_RS06980	SCIRTI_RS03160	00645	Caudovirus prohead serine protease (PF04586), Phage capsid family (PF05065)	Y	N
Portal (Dark blue)	SKUN_RS06985	SCIRTI_RS03190	00646	Portal Protein (PF04860)	N	N
Terminase (Red)	SKUN_RS06990	SCIRTI_RS03205	00649,00648 (trunc)	PF03354	N	N
gp6-like head-tail connector (Orange)	SKUN_RS06975	SCIRTI_RS03180	00644,00635	NA	N	N
Restriction Modification (Yellow)	SKUN_RS07000	SCIRTI_RS03215	NA	***S. citri*** N-6 DNA Methylase	***S. citri*** - N	***S. citri*** - N
				S. kunkelli -Type I restriction enzyme R protein N terminus (HSDR_N) -N-6 DNA Methylase -Type I restriction modification DNA specificity domain (2)	***S. kunkelli*** -Y	***S. kunkelli*** - N
DNA binding Protein (Light orange)	SKUN_RS07020	SCIRTI_RS03235	NA	***S. citri*** -Single-strand binding protein	***S. citri*** - N	***S. citri*** - N
				S. kunkelli -RecT	***S. kunkelli*** - N	***S. kunkelli*** - N

Discussion

As mentioned in Chapter 1, five *Spiroplasma* Caudovirales-like bacteriophages were described to infect *S. citri*, but due to the time period at which they were isolated, no genetic sequences were obtained (Table 1.6). However, one of the phage strains Citri SP-V3: ai was considered to be lysogenic to *S. citri*. Therefore, it is possible that related *Spiroplasma* prophage or cryptic phages could be present in published genomes.

Spiroplasma genomes (32 publicly available; see Appendix) were searched for putative prophage, but only three genomes *S. citri* (complete), *S. kunkelli* (complete) and *S. sp* chic (draft metagenome) had possible matches. Each of these *Spiroplasma* genomes shared a prophage-like region with a similar main unit gene pattern of virion proteins and a terminase. However, the putative prophage region outside the main virion protein gene region differed among the three species. No integrase genes were detected within, or in the vicinity of, the prophage region, and thus might be cryptic (Wang and Wood 2016). The presence of intact virion genes among all three regions (Table 2.1) suggests that the putative prophage has Caudovirales-like morphology. Previously, Caudovirales-like bacteriophage had only been reported in the *Spiroplasma citri*, *Spiroplasma mirum*, and *Spiroplasma* spp. infecting *Drosophila* (Williamson, Oishi et al. 1977). Nonetheless, the present study identified two similar putative prophage regions in *S. kunkelli* (complete) and *S.* chic. The placement of a modification enzyme that modifies specific nucleotides to be recognized by the host and a restriction nuclease that degrades foreign DNA (lacking modification) near the main virion region is interesting (Meselson and

Heywood 1972). It suggests that these phage overcame the *Spiroplasma* host restriction modification system by encoding their own N-6 like modification proteins (Murphy, Mahony et al. 2013), akin to T4-phage (Kossykh, Schlagman et al. 1995).

The finding of putative Caudovirales prophage in the plant pathogens *S. citri*, *S. kunkelli*, and the recently sequenced *S. chic* offers new possibilities towards studying these organisms. Further studies should be done to test the possible induction of the putative prophage, and whether they are capable of lysing other *Spiroplasma* species. If successful in lysing foreign *Spiroplasma*, the possibility of phage-based pathogen control could be tested (Jones, Jackson et al. 2007).

CHAPTER III

PRELIMINARY CHARACTERIZATION OF BACTERIOPHAGE FROM A MALE-KILLING MOLLICUTE (THE ENDOSYMBIONT *SPIROPLASMA*)

Introduction

Viruses are the most abundant and diverse organisms on the planet. They are ubiquitously present wherever life is found, ranging from the ocean to the soil. Their presence across ecosystems has a significant effect not just on their host, but also on the surrounding community. Although viruses have commonly been viewed as pathogens, examples abound of beneficial and long-lasting effects of viruses on a diversity of hosts, including bacteria, plants, fungi, and animals (Roossinck and Bazán 2017). For example, a green algae virus carries genes with the potential to change its host metabolism, which can have substantial consequences on the surrounding ecosystem through modulating nitrogen levels in their habitat (Monier, Chambouvet et al. 2017, Schvarcz and Steward 2018).

Viruses that associate with bacteria, referred to as bacteriophages or phages, also exhibit significant effects on the microbial community and on their host's invasion success (Chatterjee and Duerkop 2018). Many phages carry genes that encode toxins that are responsible for the virulence of their host bacteria. For example, the causative agent of the severe diarrheal disease cholera is a strain of *Vibrio cholerae* infected with a phage

that encodes a toxin (Waldor 1996). Phages also influence the interactions between arthropods and their inherited bacteria.

Heritable associations between bacteria and arthropods are ubiquitous, diverse, and influential (Duron and Hurst 2013). These associations can be classified as obligate or primary (i.e., the bacterium is absolutely necessary for host survival/reproduction, vertical transmission is perfect, and infection is effectively fixed in the host population), or as facultative or secondary (i.e., the bacterium is not required for host survival, vertical transmission is usually imperfect, and infection is variable in the host population). Therefore, to persist in host populations, facultative endosymbionts tend to manipulate host reproduction to their own benefit, or confer a conditional fitness advantage to hosts (e.g. defense against natural enemies or environmental stressors) (Montllor, Maxmen et al. 2002, Ford and King 2016). Recent studies of the mechanistic bases of reproductive manipulation and defensive mutualisms have implicated phages. For example, members of the Alphaprotebacterium genus *Wolbachia* (the most common facultative symbiont of arthropods; cause several reproductive phenotypes including cytoplasmic incompatibility (CI), a type of reproductive manipulation that causes embryo lethality in the offspring of *Wolbachia*-infected males and uninfected females (Yen and Barr 1971, Werren, Baldo et al. 2008). The *Wolbachia*-encoded genes responsible for CI (*cifA* and *cifB*) were recently discovered and are actually encoded in a temperate phage (WO phage) (Fujii, Kubo et al. 2004, Harrison and Brockhurst 2017, LePage, Metcalf et al. 2017, Shropshire, On et al. 2018). Strains of the

Gammaproteobacterium *Hamiltonella defensa* infected with a particular phage (APSE3) produce a soluble factor that prevents successful development of a parasitic wasp of the aphid (Brandt, Chevignon et al. 2017). The possible influence of phages on the heritable associations of the genus *Spiroplasma* (class Mollicutes) and the fly genus *Drosophila* has not been examined. The *Drosophila*-*Spiroplasma* interactions include reproductive parasitism (in the form of male killing; Montenegro, Petherwick et al. 2006) and protective mutualism against parasitic wasps and nematodes (reviewed in Mateos, Winter et al. 2016). Although genes involved in these mechanisms have been discovered and are encoded in the bacterial chromosome and/or possibly its extrachromosomal elements (Ballinger and Perlman 2017, Harumoto and Lemaitre 2018), the role of phage has not been ruled out.

The genus *Spiroplasma* includes a diversity of endosymbionts of many eukaryotic hosts, particularly arthropods and plants, inhabiting both intra- and extra-cellular environments (e.g. phloem and hemolymph). *Spiroplasma*-host interactions range from parasitic to mutualistic, and transmission modes are either vertical (as in *Drosophila*) or horizontal (i.e., from the environment or via a vector; (Regassa and Gasparich 2006)). Several stains of *S. poulsonii* are male killers (referred to as Sex Ratio Organisms; SROs), whereby male *Drosophila* die during early embryonic development (Montenegro, Petherwick et al. 2006); have been shown to confer protection against parasitoid wasps (reviewed in Mateos, Winter et al. 2016); and are generally present in high titers

compared to other *Spiroplasma* strains (Anbutsu and Fukatsu 2003, reviewed in Haselkorn 2010).

Spiroplasma phage research was particularly active in the 1970's, with several types of phage described from *Drosophila*-associated *Spiroplasma* on the basis of particle morphology, nucleic acid content type and size, and/or phenotypic effects based on extract transfers among *Spiroplasma*-infected flies (Table 1.2 & Table 1.3) (Williamson, Oishi et al. 1977). The first *Spiroplasma* phage was extracted directly from *Spiroplasma*-infected *Drosophila* (i.e., without the need for "induction"; referred to as "first-degree" extraction in Chapter 1). The second phage type was extracted after injection of host flies with the first type of phage (referred to as "second-degree" extraction in Chapter 1); thus, resembling properties of a lysogenic phage where induction is necessary for successful isolation (Casjens 2003). Both phage types had an icosahedral podoviridae-like morphology and a genome comprised of double-stranded DNA of ca. 22kbp (first phage) and 17kbp (second phage; reviewed in Chapter 1). Genome sequences of such phages were never generated. The well-studied *S. poulsonii*, which is now cultivable in liquid cell-free media (Masson, Calderon Copete et al. 2018), and whose hosts include the model organism *Drosophila melanogaster*, is an obvious candidate for re-launching *Spiroplasma* bacteriophage research after a ~3–4 decade hiatus. Herein, we performed isolation and sequencing of phage from *Spiroplasma poulsonii* strain NSRO, which was originally found in *Drosophila nebulosa*, but transferred and maintained in *D.*

melanogaster. This strain was chosen because in previous studies, phage particles from NSRO were more readily isolated than those from other strains (Chapter 1).

Material and Methods

Drosophila Fly Rearing

A *Drosophila melanogaster* Oregon R fly lab strain harboring *Spiroplasma poulsonii* NSRO was maintained on Banana-Opuntia food at 25ºC (12:12 light:dark cycle). The NSRO-infected fly strain was maintained by mating to *Spiroplasma*-free Oregon R males. This *Spiroplasma*-free Oregon R fly strain served as a negative control in the phage isolation procedures. Adult flies were aged 1–2 weeks before phage extraction. Absence of *Wolbachia* infection was confirmed by PCR with *Wolbachia*-specific *wsp* primers (Braig, Zhou et al. 1998, Zhou, Rousset et al. 1998).

Phage Isolation

To allow for the use of a large number of *Drosophila* individuals (i.e., > 3 grams), the PEG-based precipitation approach of (van der Wilk, Dullemans et al. 1999) was adapted as follows. Whole flies (4.8g) were homogenized with a mortar and pestle in 50ml of SM buffer [50 mM Tris–HCl, pH 7.5, 0.1 M NaCl, 10 mM $MgSO_4$-$7H_2O$, 1% (w/v) gelatin and 1µg/ml RNase A]. The homogenate was incubated at room temperature for 30 min and then centrifuged at 2000xg to remove large debris. The supernatant was removed and passed through a 70µm filter (28143-312;Millipore) and incubated at 4ºC overnight. Our preliminary experiments indicated that thicker centrifugation bands were obtained if the incubation step was performed. Solid NaCl was added to final

concentration of 1.0M and incubated on ice for 1h. Solid PEG-6000 was then added to the solution to a final concentration of 10% (w/v) and gently mixed to dissolve. The mixture was incubated on ice for 1h and then centrifuged at 11,000xg for 10min at 4ºC. The supernatant was decanted and discarded, whereas the pellet was resuspended in 8ml of SM buffer by mixing for 10–20 min. An equal amount of chloroform was added to remove PEG. After gently mixing, the sample was centrifuged at 3,000xg for 15 min at 4ºC. The aqueous layer was collected and passed through a 0.45µm filter ((28143-312;Aerodisk). The sample was then incubated at 4ºC until ultra-centrifugation (i.e., for 24 to 48h). The sample was then subjected to an Optiprep (60% iodixanol) step gradient. Each sample was layered over a prepared step iodixanol gradient of 10%,30%,35%,40% and centrifuged at 209,500xg in a Beckman Coulter sw41 rotor for 2h (Figure A2). Phage-like bands were then extracted by with a needle and stored at 4ºC for further use (i.e., TEM and DNA isolation).

Transmission Electron Microscopy (TEM)

To visualize phage particles present in the above "phage" extracts, we performed TEM following Uranyl Acetate staining as follows. To collect phage particles (if present), a carbon film strip was immersed in aliquots of the extract for ~1 min (i.e., until the carbon side of the film detached). The detached film was then placed on top of a droplet of 2% uranyl acetate for staining. A dry copper grid (GT200T/Electron Microscopy Sciences) previously cleaned by immersion in isopropanol was used to retrieve the carbon film from the staining solution. Excess staining solution was gently dabbed off with filter paper (Whatman filter paper, grade 1) and placed on a grid map (Electron

Microscopy Sciences) to dry. The copper grid was stored in a desiccator until use. The grids were imaged at a magnification of 25K using a JEOL 1200 EX electron microscope.

DNA Extraction and Sequencing

DNA was extracted with the Norgen Biotek Phage DNA isolation kit in 50µl of Norgen EB buffer. DNA was quantified with a PicoGreen assay on a plate reader with a concentration of 100ng/µl. DNA was then cleaned with 0.9x AMPureXP beads (Agencourt) to remove any residual EDTA, and resuspended in 24µl Qiagen EB buffer (Tris HCl pH=8.5). The extracted DNA from the sample was subsequently run on a 1.5% Agarose gel (Figure A2). Because the DNA amount was insufficient for library preparation for Nanopore sequencing, the DNA was combined with a larger DNA sample from an unrelated source (i.e., the tephritid fruit fly *Anastrepha obliqua*, tested for absence of *Spiroplasma*) to obtain ca. 10µg of gDNA. The Nanopore SQK-LSK109 sequencing kit was used following the manufacturer's recommendations (selected options: no DNA fragmentation; and use of HSB buffer to enrich for reads with a length greater than 3kbp).

The sequencing run consisted of a Spot-on flow cell Mk 1 R9 (FLO-MIN106), MinION Mk1B sequencer and utilizing the Nanopore sequencing software Minknow v2.1 (Oxford Nanopore). Basecalling was performed with Albacore v2.3.1 (Oxford Nanopore). Resulting reads were processed with PoreChop (https://github.com/rrwick/Porechop) to remove adapter sequences. The adapter-free

fastq reads were assembled with CANU v1.7 (Koren, Walenz et al. 2017) pipeline (genomeSize=2.0m, correctedErrorRate=0.105, corMinCoverage=0, corMaxEvidenceErate=0.15, corOutCoverage=1000; to maximize the number of reads utilized). The assembly was corrected with Nanopolish (available at https://github.com/jts/nanopolish) and then blasted using blastn-discontiguous parameters to the NCBI nt database (January 2018) to identify putative *Spiroplasma* contigs. Putative *Spiroplasma* contigs were then re-corrected with a round of Racon v.1.31(Vaser 2017) to obtain a consensus. One of the originally assembled putative *Spiroplasma* contigs was ~ 31kbp, with flanking repeat regions, and a pattern of coverage that is typically found in circularly permuted linear genomes (see Results) (Casjens and Gilcrease 2009). To account for this, one of the repeat regions was removed and the resulting in a "single copy" contig, which was used as a reference to re-map the original reads to generate a new consensus. To determine if the original template molecules used to generate a contig were circularly permuted, we examined the pattern of read coverage in an artificial concatemer of the "single copy" contig (i.e., the "single copy" contig was tandemly repeated to generate a three-copy concatemer). The raw reads were then mapped to the artificial contigs to visualize mapping patterns. A gap in coverage implies that all the template molecules share the same end(s), and thus, do not represent circularly permuted elements.

Gene Prediction and Annotation

Genes from the assemblies were initially predicted and annotated using Glimmer (Delcher, Bratke et al. 2007) and Prodigal (Hyatt, Chen et al. 2010), within the Prokka

(Seemann 2014) pipeline with genetic code 4 (i.e., for *Mycoplasma*/*Spiroplasma*), which utilized a database comprised of all available NCBI *Spiroplasma* genomes (June 2018). Predicted open reading frames (ORFs) at least 100bp long were retained, allowing for overlapping ORFs (McNair, Aziz et al. 2018). Non-coding RNA regions were annotated with Aragorn (Laslett and Canback 2004). Assemblies were queried for potential prophage sequences using phaster (Arndt, Grant et al. 2016). Additionally, translated ORFs were annotated with RAST and the eggNOG (Huerta-Cepas, Szklarczyk et al. 2015) pipeline to aid in the identification of putative gene functions. Repeat regions were identified with the Geneious v.11.1.2 (Biomatters Ltd) repeat finder tool (at least 50bp and 0% mismatches allowed). Signal Peptide and Transmembrane annotations were assigned based on positive hits to both the SignalP (Nielsen 2017) and Phobius (Käll, Krogh et al. 2007) databases. Similarly, putative transmembrane regions were annotated if positive hits were observed from both the TMHMM (Krogh, Larsson et al. 2001) and Transmembrane database. IsFinder (Siguier, Pérochon et al. 2006) was used to identify possible transposable elements in the phage proteome with their blastp database; a threshold e-value of –40 was used to identify transposase families. Additionally, annotated genes, protein product names, and Interpro motifs names were visually inspected for terms associated with common toxins and virulence factors.

Results

Extraction

The Optiprep step gradient resulted in a 35–40% Iodixanol band for the *Spiroplasma* NSRO sample (Figure A1). No band was visible in the *Spiroplasma*-free control sample.

TEM of an aliquot of the NSRO band revealed phage-like particles with an icosahedral morphology measuring ca. 40nm across (Fig. 3.1). Extractions generally faired better and gave a more visible centrifugation band when flies were aged for at least two weeks, and the filtered (0.45µm) homogenate was allowed to incubate for at least 24h (results not shown). Several bands of extracted DNA from an extracted sample were visible at 1-1.5kbp, 2kbp, 3-4kbp, and 5kbp. No differences were observed among the samples subjected to EcoRI, NdeI, and no digestion enzyme (Figure A2).

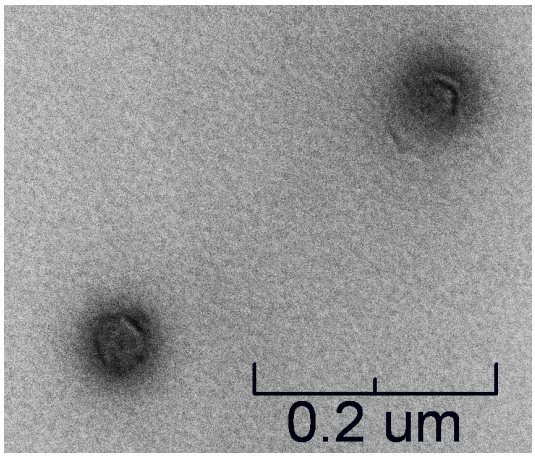

Figure 3.1 - Electron Microscopy.
TEM of negatively stained particles isolated from *Spiroplasma* strain NSRO. The particles are about 45nm across.

Assembly of Bacteriophage-like Contigs

Three contigs likely derived from the *Spiroplasma* NSRO extraction (i.e., not the *A. obliqua* DNA sample) were identified from the de-novo assembly by discontiguous-

blastn of all contigs to the NCBI nt database. One of the three contigs with blast hits to *Spiroplasma* was discarded due to low read coverage. Based on the original assembly, the sizes of the two remaining contigs were ca. 31kbp and 21kbp. However, the presence of terminal repeats at the ends of both contigs and the raw read mapping pattern observed (Figures 3.2 A and B), suggested that these original contigs contained assembly artifacts. These contigs were therefore adjusted to where a single repeat region was removed from each, leaving "single copy" contigs. The length for each "single-copy" contig was 18.9kbp and 19.5kbp; hereafter referred to as NSRO-Phage(P)-1 and NSRO-Phage(P)-2 (Figure 3.3.A and C), respectively. The average coverage for NSROP-1 and NSROP-2 was ca. 50x and 25x, respectively. Both had a similar GC (~ 24%); comparable to the GC% content of *Spiroplasma*.

Size Distribution of Reads Mapped to Putative Phage Contigs

401 reads mapped to NSROP-1, of which 14 were at least 17kbp, and none were greater than 22kbp (Figure 3.4.A), suggesting the phage genome is less than 22kbp (note: reads longer than 22kbp were obtained from this sequencing library, but could be attributed to *A. obliqua* or its associated microbes; not shown). 266 reads mapped to NSROP-2, of which 13 were > 16kbp but < 21kbp (Figure 3.4.B). The manner in which the raw reads mapped to the artificial concatemer versions of both phages, with no gaps in coverage, is consistent with a circularly permutated genome (Figures 3.5.A & 3.5.B)

Annotations

The final 18.9kbp contig of NSROP-1 contained 63 ORFs (size range 37–248 aa) of which only five could be assigned a putative function (Figure 3.3A). Notably, a putative

ssDNA binding protein (light blue in Figure 3.3A) and a Bacteriophage Terminase protein (dark green) were identified. Additionally, several Probable adhesion P58 genes were annotated, which may be involved in insect interactions (Comer, Fletcher et al. 2007). NSROP-2 contained 52 ORF (size range 25–432 aa) of which seven could be assigned a putative function (Figure 3.3D). The prophage-associated RecT ssDNA binding protein (PF03837; red in Figure 3.3D), Terminase (PF04466; dark green) and UPF0236 domain protein (PF06782; purple) were annotated.

A genome assembly of NSRO is not currently available to search for evidence of the presence of these putative phage genomes within the bacterial chromosome, and thus, whether they are likely lysogenic. Nonetheless, we mapped the two contigs to the genome of *S. poulsonii* MSRO-UG strain, which is closely related to NSRO based on several genes (Anbutsu, Goto et al. 2008). Both NSROP-1 and NSROP-2 mapped with high homology (98.9% to region 380,731–399,891 and 97.6% to region 105,846–125,415, respectively) to the main chromosomal contig of the H99 MSRO-UG assembly. NSROP-2 mapped to a region bounded by DNA polymerase C (SMH99_01220), ERA (SMH99_01230), recO (SMH99_01240) on one end, and a Glycine-tRNA ligase (SMH99_01500), DNA primase RNA polymerase (SMH99_01510), and sigma factor SigA (SMH99_01520) on the other end. NSROP-1 mapped to a region bounded by hypothetical protein genes (SMH99_04810, SMH99_04820) on one end, and Phenyllactate dehydrogenase (SMH99_05060), and Membrane transport protein (SMH99_05070) on the other.

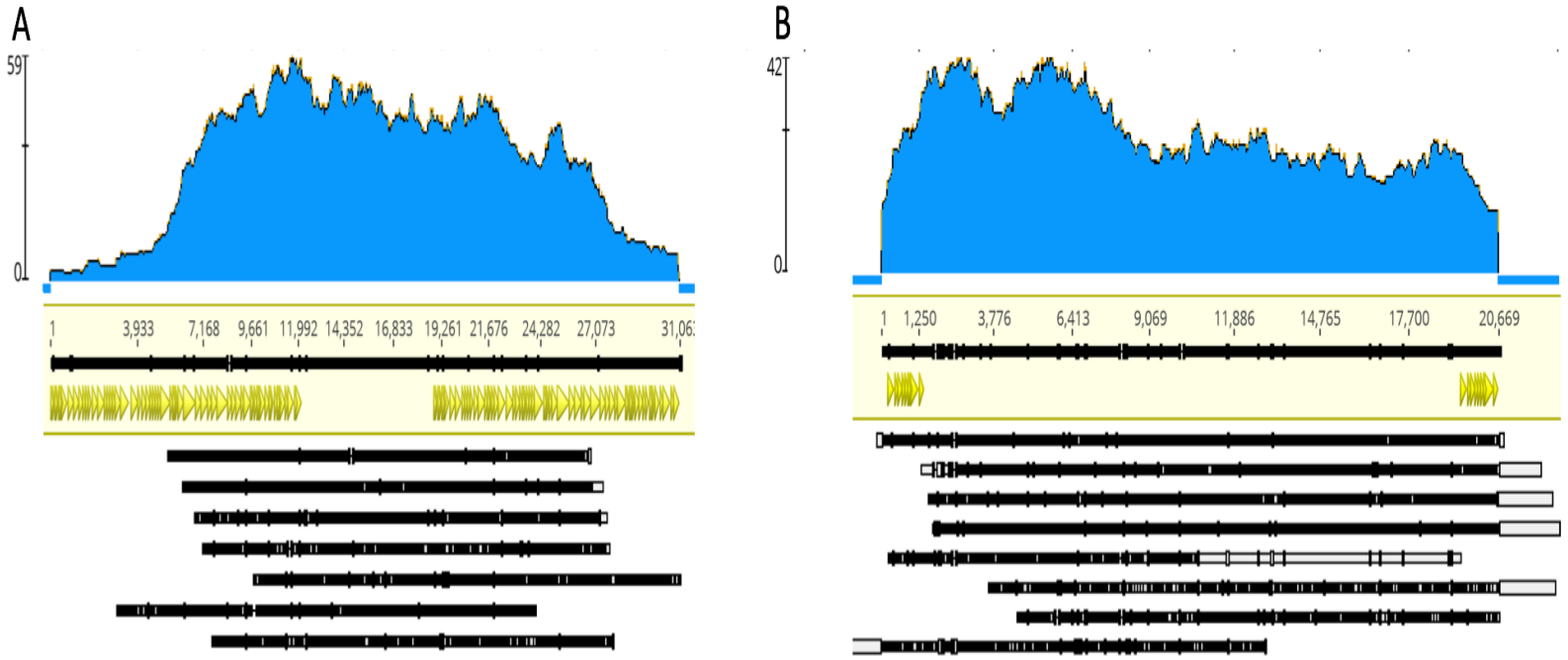

Figure 3.2 - NSROP-1 and NSROP-2 Original Assembly.
NSROP-1 (A) and NSROP-2 (B) original assembly contigs with raw Nanopore reads mapped via MiniMap2 (shown below yellow box). Only a small sample of reads mapped is shown. Yellow arrows represent direct repeats. Gray outline boxes in reads represent trimmed regions. Black vertical lines represent small (1-4bp) insertions. White vertical regions on reads represent mismatches. Top blue graph depicts coverage.

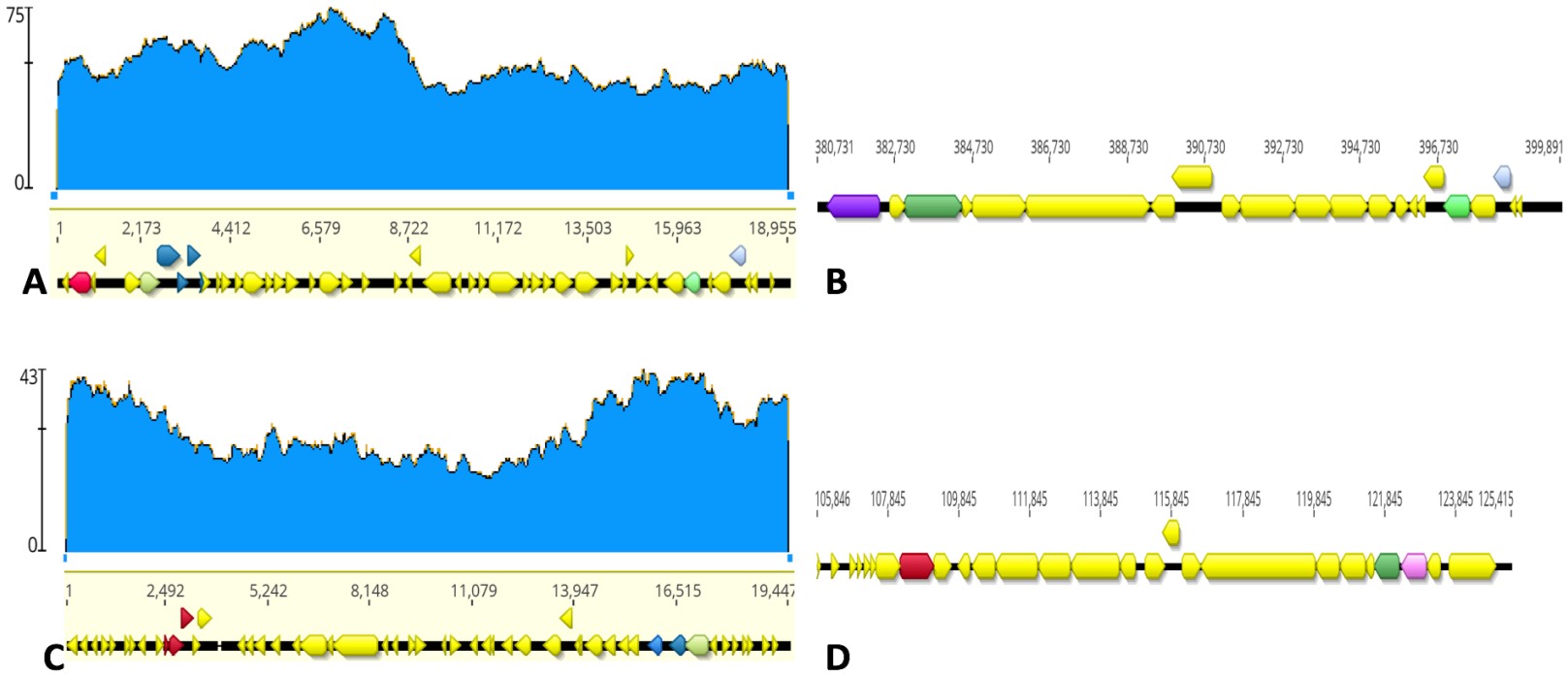

Figure 3.3 - Coverage of Final NSROP-1 and NSROP-2 Contigs With Annotations.
Final NSROP-1 (A) and NSROP-2 (C) contigs depicting coverage (blue graph) and annotations (block arrows). C Region of Spiroplasma MSRO-UG assembly H99 with substantial homology to NSROP-1 (B) and NSROP-2(D). Numbers in figure B and D indicate coordinates in Spiroplasma MSRO-UG assembly H99. Light blue arrows are single stranded DNA binding annotated genes, dark Green arrows are Terminase genes, blue arrows are adhesion protein genes, purple arrows are UPF0236 protein CDS, light green arrows are ERF superfamily gene. Red arrows are RecT annotated genes, Dark Green arrows are Terminase genes, and blue arrows are adhesion protein genes. The pink arrow is a large subunit terminase. NSROP-1 and NSROP-2 contig fasta files and annotations are available in Supplement Media Folder.

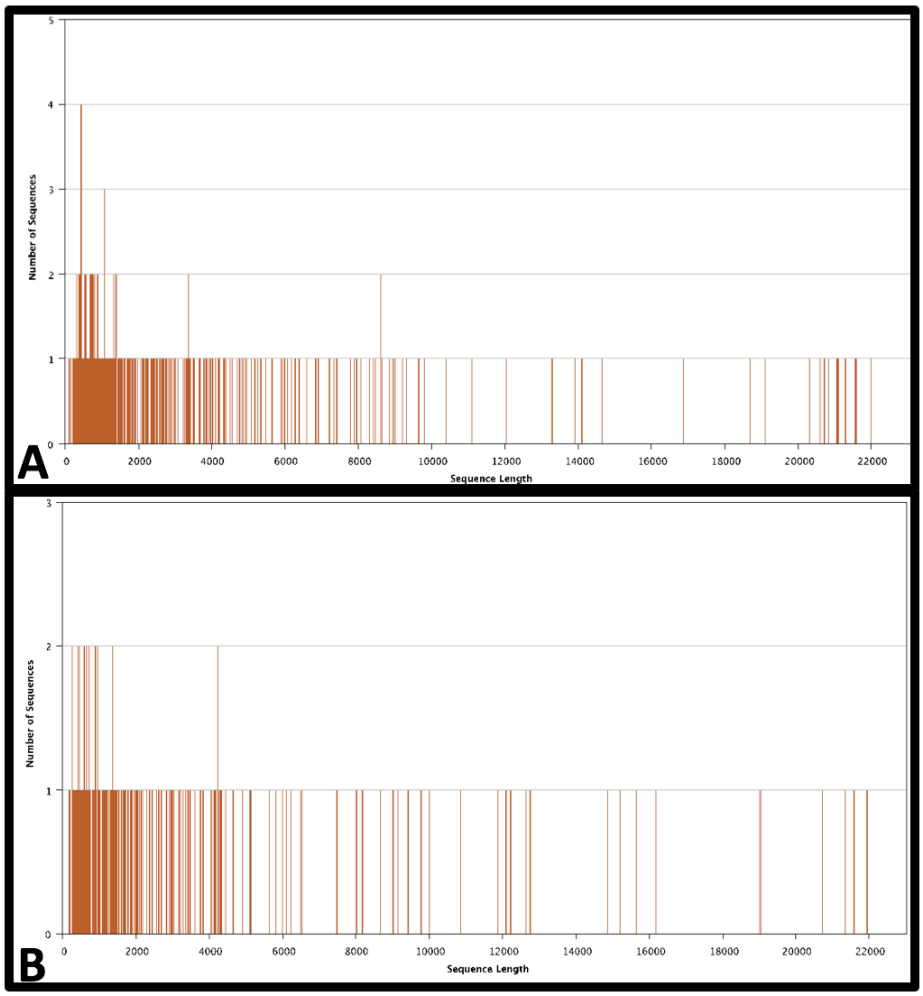

Figure 3.4 - Size Distribution of Reads.
Size distribution of raw Nanopore reads mapped to NSROP-1 (A) and NSROP-2 (B).

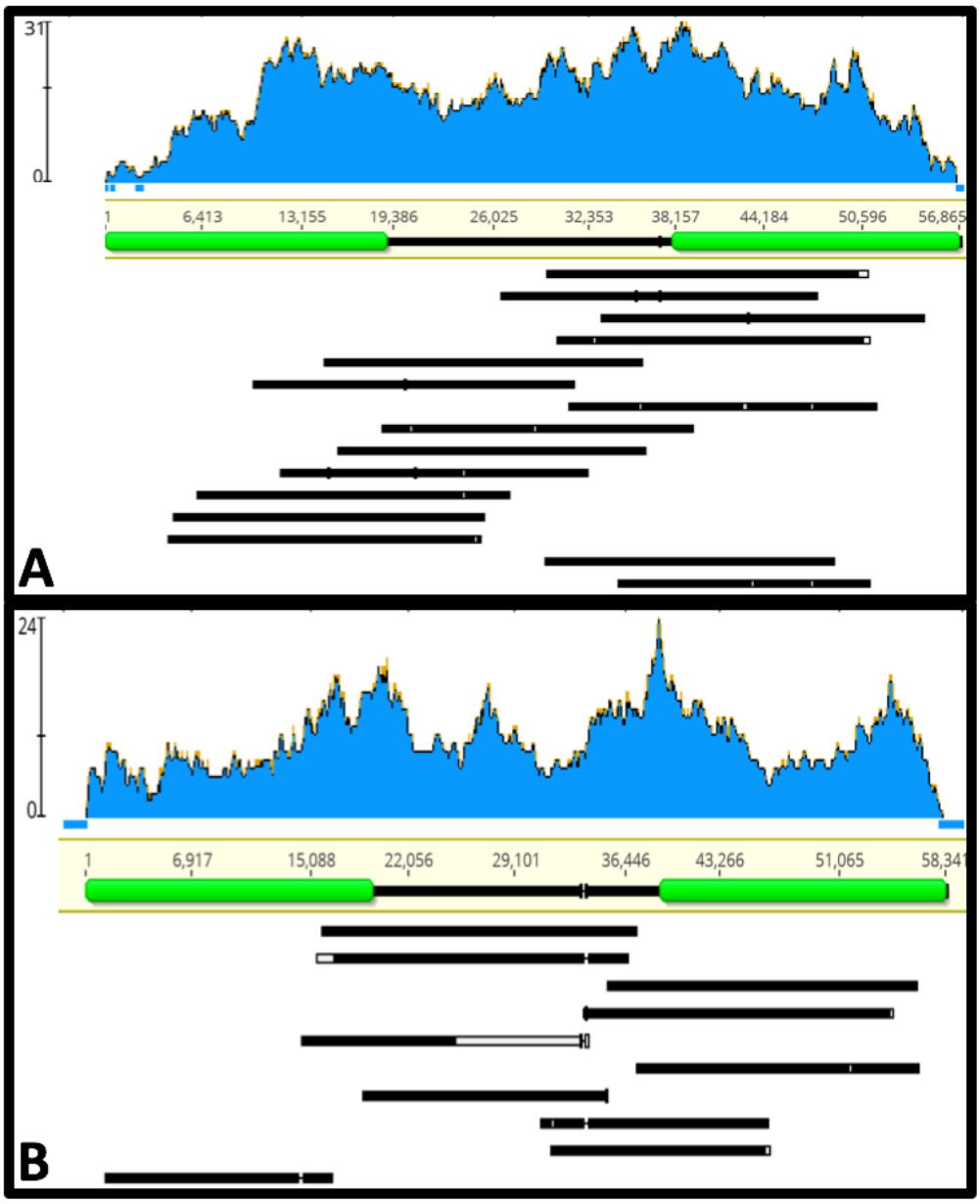

Figure 3.5 - Concatemer Construct.
The raw reads were mapped to separate concatemer construct of NSROP-1 (A) and NSROP-2 (B) to support that is replicates by Circular permutation. The green regions on the artificial concatemer each represent one whole genome bordering the black region genome in the middle. Gray outline boxes in reads represent trimmed regions. Black vertical lines represent small (1–4bp) insertions. White vertical regions on reads represent mismatches. Top blue graph depicts coverage.

Additionally, GC% shifts in the MSRO -UG H99 genome regions with homology to the phage contigs were analyzed, but no significant shifts were detected (Figure A4 and A5). Appropriate Genbank format files are included in the accompanying appendix.

Discussion

TEM of our extractions aimed at phage isolation from whole *D. melanogaster* harboring *S. poulsonii* NSRO revealed the presence of one type of particle: ca. 40nm icosahedron with short tail. Comparable particles were described following similar extraction protocols (i.e., without "induction" = "first degree" extraction in Table 1.2; Chapter 1) applied to *D. nebulosa* and *D. hydei* infected with their native (male-killing and non-male-killing, respectively) *Spiroplasma* strains (Oishi, Poulson et al. 1984, Cohen, Williamson et al. 1987). A similar morphology has been reported for particles obtained following a "second-degree extraction" protocol (i.e., induced by injecting to particles from "first degree" extractions into flies harboring *Spiroplasma*) from four sex ratio strains (ESRO, PSRO, WSRO, ASRO; see Table 1.2; Chapter 1). One caveat of the previous studies is that at that time, the existence of *Wolbachia* as a common endosymbiont of *Drosophila* was not known and therefore not tested. At least *D. willistoni* and *D. paulistorum* (native hosts of WRSO and PSRO, respectively) are known to harbor *Wolbachia* (Mateos, Castrezana et al. 2006, Miller, Ehrman et al. 2010). It is therefore possible, that one or more of the phage attributed to *Spiroplasma* on the basis of morphology and genomic features by previous studies could instead represent *Wolbachia* phage (WO); which are reported to have isometric heads ca.40 nm in diameter containing linear dsDNA ca. 20kbp (Bordenstein and Wernegreen 2004,

Fujii, Kubo et al. 2004). The present study, which used flies confirmed to be *Wolbachia*-free, rules out the possibility that any particles recovered and sequenced represent *Wolbachia* phage.

The assemblies of Nanopore reads derived from our NSRO phage DNA extracts revealed two distinct types of candidate *Spiroplasma* phage contigs with adequate read coverage, which we refer to as NSROP-1 and NSROP-2. Excluding one of their respective terminally-repeated ends, their sizes were 18.9kbp and 19.5kbp, with the longest read mapped to each just under 22kbp. Absence of longer reads that map to these contigs (despite longer reads derived from the same sequencing library; see Results) implies a maximum genome size of ca. 22kbp, which is comparable to the 21.8kbp phage genome sizes reported for *Spiroplasma* strains NSRO and Hydei following a "first-degree" extraction, and for strain PSRO (from *D. paulistorum*) following a "second-degree" extraction (Table 1.2; Chapter 1). The 21.8kbp phages of NSRO and Hydei and the 21kbp phages of the non-*Drosophila*-associated strains, *S. citri* and *S. mirum*, are all reportedly circularly permuted and terminally redundant (see Table 1.3; Chapter 1); i.e., likely possessing a headful packaging replication mechanism (Fujisawa and Morita 1997). The mapping of reads to NSROP-1 is consistent to with a circularly-permuted terminally-redundant linear genome. The NSROP-1 genome contains a phage-like large terminase, ssDNA binding protein, and an ERF DNA binding superfamily protein. The large terminase could aid in the packing of viral dsDNA (Gual, Camacho et al. 2000), whereas the ssDNA binding protein would most likely be involved in phage

DNA replication (Meyer and Laine 1990). Additionally, an ERF (Essential recombination Factor) is found in the lysogenic headful packing *Salmonella* phage P22 (Bacteriophage P22,) where it responsible for circularization of the dsDNA through recombination (Murphy 2000, Birge 2013). NSROP-2 possesses a terminase (with high homology to the terminase present in NSROP-1) and an ssDNA binding domain protein. For NSROP-2, there are several genomic features that have prophage-like characteristics. The ssDNA binding gene has homology to the prophage-associated RecT family (annotated by RAST), which is found in prophages such as *Listeria monocytogenes* phage A118 (Loessner, Inman et al. 2000). The genome itself is also identified as a prophage contig by the software phiSpy (Akhter, Aziz et al. 2012). The high (albeit imperfect) homology detected upon mapping both NSROP-1 and NSROP-2 to the H99 MSRO-UG chromosome, suggests that both might be lysogenic. The lack of homology of the predicted Terminase and UPF0236 gene regions may indicate that these genes are involved in recognition of the host.

The presence of two distinct types of phages raises interesting questions about their coexistence. Co-infection of bacteria by multiple bacteriophages is often dependent on the host environment and on the interactions among bacteriophages (Diaz-Munoz 2017). Such factors might influence whether NSROP-1 and NSROP-2 co-infect a single *Spiroplasma* cell or different cells within the same population (e.g. same host body or tissue). NSROP-1 and NSROP-2 could have an antagonistic association, as prophage have been shown to suppress co-infection by other phages to prevent reduction in host

fitness (Berngruber, Weissing et al. 2010). This would likely result in co-infection only at a population level. Alternatively, NSROP-1 and NSROP-2 might have a mutualistic interaction; e.g. if each phage encoded a factor that increased *Spiroplasma* fitness in an additive or synergistic manner. Under such scenario, both phages might con-infect the same cell.

There are several ways in which one or both of the putative phage of NSRO could enhance their host fitness. Firstly, phage might contribute to the male-killing ability of NSRO, similar to the *Wolbachia* phage that encode genes associated with cytoplasmic incompatibility (Lindsey, Rice et al. 2018). It is important to note, however, that a gene associated with make-killing in MSRO has been identified (i.e., *Spaid*; Harumoto and Lemaitre 2018)), and it appears to be located on an extra-chromosomal element; most likely a plasmid. We found no evidence of homology to *Spaid* in either of the NSRO putative phage contigs. Nonetheless, it is possible that other *Spiroplasma* (or phage)-derived factors contribute to the male-killing phenotype. Secondly, NSRO phage may enhance *Spiroplasma* fitness by enhancing the fitness of *Drosophila*, on which *Spiroplasma* is dependent for transmission. Although NSRO has not been characterized for defensive abilities against natural enemies of *Drosophila*, all the *poulsonii*-clade strains examined to date (i.e., MSRO, hyd1, and neo) have the ability to interfere with the development of parasitic wasps (reviewed in Mateos, Winter et al. 2016). One type of *Spiroplasma*-encoded factor (i.e., Ribosome inactivating proteins; RIPs) appears to be involved in the protection mechanism against parasitic wasps and nematodes (Hamilton,

Peng et al. 2016, Ballinger and Perlman 2017). None of the ORFs in the NSRO putative phage contigs have homology with RIPs (or other toxins), but other factors encoded by the phage might contribute to defense against these or other natural enemies. Lastly, it is also possible that the NSRO phage actually keep *Spiroplasma* densities in check to limit pathogenicity to the fly, akin to the phenomenon observed with ASPE phage in *Hamiltonella*-infected aphids. ASPE phage restricts over proliferation of *Hamiltonella defensa*, and in the absence of ASPE phage, *H. defensa* lowers the fitness of the aphid host (Weldon, Strand et al. 2012). High densities of *Spiroplasma* (MSRO-UG) in old *D. melanogaster* adults have been associated with reduced longevity and motor defects (Herren, Paredes et al. 2014).

CHAPTER IV

CONCLUSION

Our study provides the first genome sequences of putative phage of *Drosophila*-associated *Spiroplasma*, breaking a 3–4 decade research hiatus. Clearly, numerous aspects of these putative phage remain to be clarified or confirmed, such as: (1) their genomic size and features, for which gel electrophoresis-based assays, as well as PCR and Sanger sequencing, would be informative; (2) whether they are inserted into the host genome, for which the NSRO genome assembly is needed; and (3) whether they are able to lyse NSRO or other *Spiroplasma* strains. The recent breakthrough in cell-free liquid cultivation of MSRO (Masson, Calderon, Copete et al. 2018) is expected to facilitate phage research, but adaptation to solid media for plaque assays would be helpful. Comparative studies of *Spiroplasma* strains with different phenotypes should contribute to identify associations between phenotypes and phage presence/absence. Such studies could be followed up with functional studies to dissect the role of phage or their genes (e.g. transformed into *Spiroplasma* or *Drosophila*) on phenotypes.

Additionally, the study describes putative prophage sequences in *Spiroplasma* strains not associated with *Drosophila*. This reveals the possibility of Caudovirales double stranded DNA bacteriophage that have yet to be isolated and individually sequenced. To confirm whether the putative phage sequences are indeed from active prophage or just remnants of a cryptic phage, further work will need to be done. Follow up studies could include

isolation of phage particles, DNA sequencing and description of interactions with the host *Spiroplasma* strain.

REFERENCES

Akhter, S., R. K. Aziz and R. A. Edwards (2012). "PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity-and composition-based strategies." Nucleic Acids Research **40**(16): e126-e126.

Alivizatos, A., R. Townsend and P. Markham (1982). "Effects of Infection with a Spiroplasma Virus on the Symptoms Produced by Spiroplasma citri." Annals of Applied Biology **101**(1): 85-91.

Anbutsu, H. and T. Fukatsu (2003). "Population dynamics of male-killing and non-male-killing Spiroplasmas in *Drosophila melanogaster*." Applied and Environmental Microbiology **69**(3): 1428–1434.

Anbutsu, H., S. Goto and T. Fukatsu (2008). "High and low temperatures differently affect infection density and vertical transmission of male-killing *Spiroplasma* symbionts in *Drosophila* hosts." Applied and Environmental Microbiology **74**(19): 6053-6059.

Arndt, D., J. R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang and D. S. Wishart (2016). "PHASTER: a better, faster version of the PHAST phage search tool." Nucleic Acids Research **44**(W1): W16-W21.

Ballinger, M. J. and S. J. Perlman (2017). "Generality of toxins in defensive symbiosis: Ribosome-inactivating proteins and defense against parasitic wasps in *Drosophila*." PLoS Pathog **13**(7): e1006431.

Beckmann, J. F., J. A. Ronau and M. Hochstrasser (2017). "A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility." Nature Microbiolgy **2**: 17007.

Berngruber, T. W., F. J. Weissing and S. Gandon (2010). "Inhibition of superinfection and the evolution of viral latency." Journal of Virology **84**(19): 10200-10208.

Birge, E. A. (2013). Bacterial and bacteriophage genetics, Springer Science & Business Media.

Bolaños, L. M., L. E. Servín-Garcidueñas and E. Martínez-Romero (2015). "Arthropod–Spiroplasma Relationship in the Genomic Era." FEMS Microbiology Ecology **91**(2): 1-8.

Bondy-Denomy, J. and A. R. Davidson (2014). "When a Virus is not a Parasite: the Beneficial Effects of Prophages on Bacterial Fitness." Journal of Microbiology **52**(3): 235-242.

Bordenstein, S. R., M. L. Marshall, A. J. Fry, U. Kim and J. J. Wernegreen (2006). "The tripartite associations between bacteriophage, *Wolbachia*, and arthropods." PLoS Pathogens **2**(5): e43.

Bordenstein, S. R. and W. S. Reznikoff (2005). "Mobile DNA in Obligate Intracellular Bacteria." Nature Reviews Microbiology **3**(9): 688-699.

Bordenstein, S. R. and J. J. Wernegreen (2004). "Bacteriophage flux in endosymbionts (Wolbachia): infection frequency, lateral transfer, and recombination rates." Molecular Biology and Evolution **21**(10): 1981-1991.

Bové, J. M., J. Renaudin, C. Saillard, X. Foissac and M. Garnier (2003). "Spiroplasma citri, a Plant Pathogenic Mollicute: Relationships with its Two Hosts, the Plant and the Leafhopper Vector." Annual Review of Phytopathology **41**(1): 483-500.

Braig, H. R., W. Zhou, S. L. Dobson and S. L. O'Neill (1998). "Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*." Journal of Bacteriology **180**(9): 2373-2378.

Brandt, J. W., G. Chevignon, K. M. Oliver and M. R. Strand (2017). "Culture of an aphid heritable symbiont demonstrates its direct role in defence against parasitoids." Proceedings of the Royal Society B: Biological Sciences **284**(1866): 1925.

Casjens, S. (2003). "Prophages and bacterial genomics: what have we learned so far?" Molecular microbiology **49**(2): 277-300.

Casjens, S. R. and E. B. Gilcrease (2009). Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. Bacteriophages, Springer**:** 91-111.

Chatterjee, A. and B. A. Duerkop (2018). "Beyond bacteria: Bacteriophage-eukaryotic host interactions reveal emerging paradigms of health and disease." Frontiers in Microbiology **9**.

Cohen, A., D. Williamson and K. Oishi (1987). "SpV3 Viruses of Drosophila Spiroplasmas." Israel Journal of Medical Sciences **23**(5): 429-433.

Comer, J., J. Fletcher, R. E. Davis and U. Melcher (2007). "Evolution of the spiroplasma P58 multigene family." Biochemical genetics **45**(1-2): 25-32.

Delcher, A. L., K. A. Bratke, E. C. Powers and S. L. Salzberg (2007). "Identifying bacterial genes and endosymbiont DNA with Glimmer." Bioinformatics **23**(6): 673-679.

Diaz-Munoz, S. L. (2017). "Viral coinfection is shaped by host ecology and virus–virus interactions across diverse microbial taxa and environments." Virus Evolution **3**(1).

Dickinson, M. J. and R. Townsend (1984). "Characterization of the genome of a rod-shaped virus infecting Spiroplasma citri." Journal of General Virology **65**(9): 1607-1610.

Dickinson, M. J., R. Townsend and S. J. Curson (1984). "Characterisation of a virus infecting the wall-free prokaryote Spiroplasma citri." Virology **135**(2): 524-535.

Duron, O. and G. D. Hurst (2013). "Arthropods and inherited bacteria: from counting the symbionts to understanding how symbionts count." BMC Biol **11**: 45.

Fujii, Y., T. Kubo, H. Ishikawa and T. Sasaki (2004). "Isolation and characterization of the bacteriophage WO from *Wolbachia*, an arthropod endosymbiont." Biochemical and Biophysical Research Communications **317**(4): 1183-1188.

Fujisawa, H. and M. Morita (1997). "Phage DNA packaging." Genes to Cells **2**(9): 537-545.

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**(3): 893-918.

Gual, A., A. G. Camacho and J. C. Alonso (2000). "Functional analysis of the terminase large subunit, G2P, of Bacillus subtilis bacteriophage SPP1." Journal of Biological Chemistry **275**(45): 35311-35319.

Hamilton, P. T., F. Peng, M. J. Boulanger and S. J. Perlman (2016). "A ribosome-inactivating protein in a *Drosophila* defensive symbiont." Proceedings of the National Academy of Sciences **113**(2): 350-355.

Harrison, E. and M. A. Brockhurst (2017). "Ecological and evolutionary benefits of temperate phage: what does or doesn't kill you makes you stronger." Bioessays.

Harumoto, T. and B. Lemaitre (2018). "Male-killing toxin in a bacterial symbiont of *Drosophila*." Nature.

Haselkorn, T. S. (2010). "The *Spiroplasma* heritable bacterial endosymbiont of *Drosophila*." Fly **4**(1): 80-87.

Haselkorn, T. S., T. A. Markow and N. A. Moran (2009). "Multiple introductions of the *Spiroplasma* bacterial endosymbiont into *Drosophila*." Molecular Ecology **18**(6): 1294-1305.

Herren, J. K., J. C. Paredes, F. Schüpfer, K. Arafah, P. Bulet and B. Lemaitre (2014). "Insect endosymbiont proliferation is limited by lipid availability." Elife **3**: e02964.

Huerta-Cepas, J., D. Szklarczyk, K. Forslund, H. Cook, D. Heller, M. C. Walter, T. Rattei, D. R. Mende, S. Sunagawa and M. Kuhn (2015). "eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences." Nucleic Acids Research **44**(D1): D286-D293.

Hurst, G. D. D. and M. E. N. Majerus (1993). "Why do maternally inherited microorganisms kill males?" Heredity **71**: 81–95.

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.

Jones, J., L. Jackson, B. Balogh, A. Obradovic, F. Iriarte and M. Momol (2007). "Bacteriophages for plant disease control." Annunal Review of Phytopatholgy. **45**: 245-262.

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 Enviromental Microbiolgy. **72**(10): 6667-6673.

Käll, L., A. Krogh and E. L. Sonnhammer (2007). "Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server." Nucleic Acids Research **35**(suppl_2): W429-W432.

Kernaghan, R. P. and L. Ehrman (1970). "Antimycoplasmal antibiotics and hybrid sterility in *Drosophila paulistorum*." Science **169**: 63-64.

Ford, S. A., and K. King. (2016) "Harnessing the power of defensive microbes: evolutionary implications in nature and disease control." PLoS Pathogens **12.4**: e1005465.

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman and A. M. Phillippy (2017). "Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation." Genome Research: gr. 215087.215116.

Kossykh, V. G., S. L. Schlagman and S. Hattman (1995). "Phage T4 DNA [N]-adenine6methyltransferase. Overexpression, purification, and characterization." Journal of Biological Chemistry **270**(24): 14389-14393.

Krogh, A., B. Larsson, G. Von Heijne and E. L. Sonnhammer (2001). "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes." Journal of Molecular Biology **305**(3): 567-580.

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 Biology and Evolution **5**(6): 1151–1164.

Laslett, D. and B. Canback (2004). "ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences." Nucleic Acids Research **32**(1): 11-16.

LePage, D. P., J. A. Metcalf, S. R. Bordenstein, J. On, J. I. Perlmutter, J. D. Shropshire, E. M. Layton, L. J. Funkhouser-Jones, J. F. Beckmann and S. R. Bordenstein (2017). "Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility." Nature **543**(7644): 243-247.

Lindsey, A. R., D. W. Rice, S. R. Bordenstein, A. W. Brooks, S. R. Bordenstein and I. L. Newton (2018). "Evolutionary genetics of cytoplasmic incompatibility genes cifA and cifB in prophage WO of Wolbachia." Genome Biology and Evolution **10**(2): 434-451.

Loessner, M. J., R. B. Inman, P. Lauer and R. Calendar (2000). "Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of Listeria monocytogenes: implications for phage evolution." Molecular Microbiology **35**(2): 324-340.

Maniloff, J. and K. Dybvig (1988). "Mycoplasma Viruses." CRC Critical Reviews in Microbiology **15**(4): 339-389.

Masson, F., S. Calderon Copete, F. Schüpfer, G. Garcia-Arraez and B. Lemaitre (2018). "In vitro culture of the insect endosymbiont *Spiroplasma poulsonii* highlights bacterial genes involved in host-symbiont interaction." mBio **9**(2).

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**(1): 363-376.

Mateos, M., L. Winter, C. Winter, V. Higareda-Alvear, E. Martinez-Romero and J. Xie (2016). "Independent origins of resistance or susceptibility of parasitic wasps to a defensive symbiont." Ecology and Evolution **6**(9): 2679–2687.

Mavrich, T. N. and G. F. Hatfull (2017). "Bacteriophage evolution differs by host, lifestyle and genome." Nature Microbiolgy **2**: 17112.

McNair, K., R. K. Aziz, G. D. Pusch, R. Overbeek, B. E. Dutilh and R. Edwards (2018). Phage Genome Annotation Using the RAST Pipeline. Bacteriophages, Springer**:** 231-238.

Meselson, M. Y., Robet and J. Heywood (1972). "Restriction and modification of DNA." Annual Review of Biochemistry **41**(1): 447-466.

Meyer, R. R. and P. S. Laine (1990). "The single-stranded DNA-binding protein of Escherichia coli." Microbiological Reviews **54**(4): 342-380.

Miller, W. J., L. Ehrman and D. Schneider (2010). "Infectious speciation revisited: impact of symbiont-depletion on female fitness and mating behavior of *Drosophila paulistorum*." PLoS Pathogens **6**(12): e1001214.

Monier, A., A. Chambouvet, D. S. Milner, V. Attah, R. Terrado, C. Lovejoy, H. Moreau, A. E. Santoro, É. Derelle and T. A. Richards (2017). "Host-Derived Viral Transporter Protein for Nitrogen Uptake in Infected Marine Phytoplankton." Proceedings of the National Academy of Sciences **114**(36): E7489-E7498.

Montenegro, H., A. Petherwick, G. Hurst and L. Klaczko (2006). "Fitness effects of *Wolbachia* and *Spiroplasma* in *Drosophila melanogaster*." Genetica **127**: 207–215.

Montllor, C. B., A. Maxmen and A. H. Purcell (2002). "Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress." Ecological Entomology **27**(2): 189-195.

Murphy, J., J. Mahony, S. Ainsworth, A. Nauta and D. van Sinderen (2013). "Bacteriophage orphan DNA methyltransferases: Insights from their bacterial origins, function and occurrence." Applied and Environmental Microbiology: AEM. 02229-02213.

Murphy, K. C. (2000). "Bacteriophage P22 Abc2 protein binds to RecC increases the 5′ strand nicking activity of RecBCD and together with λ Bet, promotes Chi-independent recombination1." Journal of Molecular Biology **296**(2): 385-401.

Nielsen, H. (2017). "Predicting secretory proteins with SignalP." Protein Function Prediction: Methods and Protocols: 59-73.

Oishi, K., D. Poulson and D. Williamson (1984). "Virus Mediated Change in Clumping Properties of Drosophila SR Spiroplasma." Current Microbiology **10**: 153-158.

Oishi, K. and D. F. Poulson (1970). "A virus associated with SR-Spirochetes of *Drosophila nebulosa*." Proceedings of the National Academy of Sciences of the United States of America **67**(3): 1565-1572.

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 **100**(4): 1803-1807.

Regassa, L. B. and G. E. Gasparich (2006). "Spiroplasmas: evolutionary relationships and biodiversity." Frontiers in Bioscience **11**: 2983-3002.

Roberts, R. J. (2005). "How restriction enzymes became the workhorses of molecular biology." Proceedings of the National Academy of Sciences **102**(17): 5905-5908.

Roossinck, M. J. (2015). "Move Over, Bacteria! Viruses Make Their Mark as Mutualistic Microbial Symbionts." Journal of Virology **89**(13): 6532-6535.

Roossinck, M. J. and E. R. Bazán (2017). "Symbiosis: viruses as intimate partners." Annual Review of Virology **4**(1): 123–139.

Saillard, C., J. Vignault, J. Bové, A. Raie, J. Tully, D. Williamson, A. Fos, M. Garnier, A. Gadeau and P. Carle (1987). "Spiroplasma phoeniceum sp. nov., a New Plant-Pathogenic Species From Syria." International Journal of Systematic and Evolutionary Microbiology **37**(2): 106-115.

Schvarcz, C. R. and G. F. Steward (2018). "A Giant Virus Infecting Green Algae Encodes Key Fermentation Genes." Virology **518**: 423-433.

Seemann, T. (2014). "Prokka: rapid prokaryotic genome annotation." Bioinformatics **30**(14): 2068-2069.

Shropshire, J. D., J. On, E. M. Layton, H. Zhou and S. R. Bordenstein (2018). "One prophage WO gene rescues cytoplasmic incompatibility in Drosophila melanogaster." Proceedings of the National Academy of Sciences **115**(19): 4987-4991.

Siguier, P., J. Pérochon, L. Lestrade, J. Mahillon and M. Chandler (2006). "ISfinder: the reference centre for bacterial insertion sequences." Nucleic Acids Research **34**: D32-D36.

van der Wilk, F., A. M. Dullemans, M. Verbeek and J. F. van den Heuvel (1999). "Isolation and characterization of APSE-1, a bacteriophage infecting the secondary endosymbiont of Acyrthosiphon pisum." Virology **262**(1): 104-113.

Vaser, R., Ivan Sović, N. Nagarajan, and M. Šikić (2017). "Fast and accurate de novo genome assembly from long uncorrected reads." Genome Research.

Waldor, M. K., and J. Mekalanos (1996). " Lysogenic conversion by a filamentous phage encoding cholera toxin " Science **272**(5270): 1910–1914.

Wang, W., W. Gu, G. E. Gasparich, K. Bi, J. Ou, Q. Meng, T. Liang, Q. Feng, J. Zhang and Y. Zhang (2011). "Spiroplasma eriocheiris sp. nov., Associated with Mortality in the Chinese Mitten Crab, Eriocheir sinensis." International Journal of Systematic and Evolutionary Microbiology **61**(4): 703-708.

Wang, X. and T. K. Wood (2016). "Cryptic prophages as targets for drug development." Drug Resistance Updates **27**: 30-38.

Weldon, S. R., M. R. Strand and K. M. Oliver (2012). "Phage loss and the breakdown of a defensive symbiosis in aphids." Proceedings of the Royal Society of London B: Biological Sciences **280**(1751): 20122103.

Werren, J. and S. O'Neill (1997). The evolution of heritable symbionts. Influential Passengers: Inherited Microorganisms and Arthropod Reproduction. S. O'Neill, A. Hoffmann and J. Werren. Oxford, Oxford University Press**:** 1–41.

Werren, J. H., L. Baldo and M. E. Clark (2008). "*Wolbachia*: master manipulators of invertebrate biology." Nature Review Microbiology **6**(10): 741-751.

Whitcomb, R. F., T. Chen, D. Williamson, C. Liao, J. Tully, J. Bové, C. Mouches, D. Rose, M. Coan and T. Clark (1986). "Spiroplasma kunkelii sp. nov.: Characterization of the Etiological Agent of Corn Stunt Disease." International Journal of Systematic and Evolutionary Microbiology **36**(2): 170-178.

Williamson, D., K. Oishi and D. Poulson (1977). Viruses of *Drosophila* sex ratio *Spiroplasma*. The Atlas of Insect and Plant Viruses. K. Maramorosch. New York, Academic Press**:** 465-472.

Yen, J. H. and A. R. Barr (1971). "New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens* L." Nature **232**(5313): 657-658.

Zhou, W. G., F. Rousset and S. O'Neill (1998). "Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences." Proceedings of the Royal Society B: Biological Sciences **265**: 509–515.

APPENDIX A

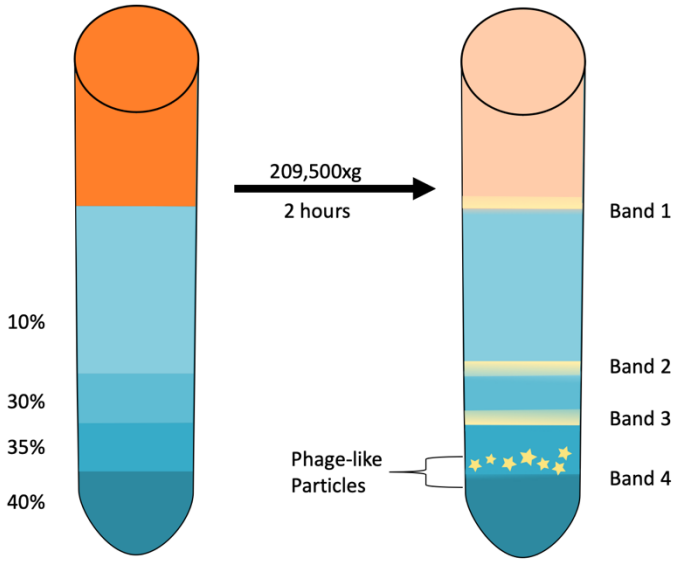

Figure A1 – Ultracentrifugation Diagram
Cartoon depicting ultracentrifugation steps that led to isolation of phage-like particles that were subjected to DNA extraction and sequencing.

The original homogenate of the sample (dark orange) was applied to an Optiprep step gradient that resulted in several bands (bands 1–4). Band 1was not checked due to the low probability of finding bacteriophage at such a low density. The contents of bands 2–4 were viewed under EM. Only band 4 (at the 35–40% part of the gradient) contained phage-like particles (represented by yellow stars). No band was visible in the *Spiroplasma*-free control sample (not shown).

Initially, to estimate the location of potential dsDNA Caudovirales bacteriophage, a purified phage K sample was centrifuged. The phage k sample was collected at 35-40% interface.

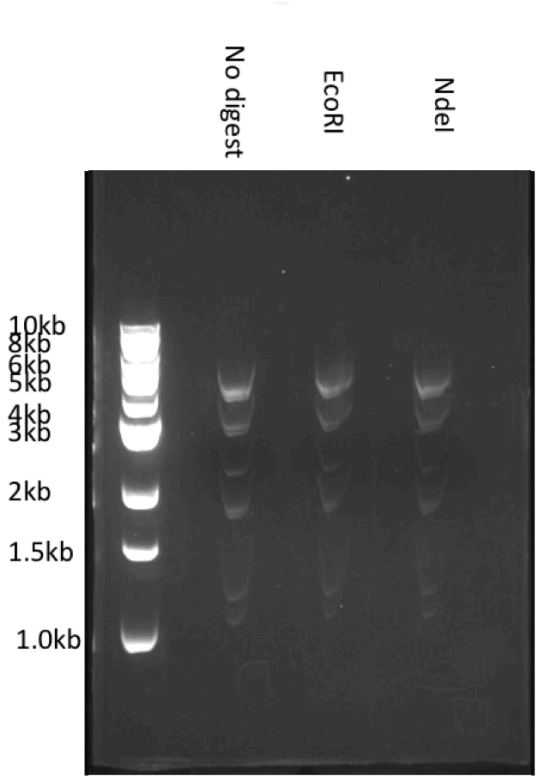

Figure A2 – NSRO Gel Electrophoresis.
Gel electrophoresis of DNA extracted from phage-like particles that was subjected to MinION DNA sequencing. *Lane 1 = 1kbp DNA ladder*; *Lane 2 = Phage DNA no*t subjected to restriction digestion; Lane 3 = subjected to EcoRI digestion; and Lane 4 = subjected to NdeI digestion. Notably, no bands were observed above 10kbp, despite the presence of long reads measuring ca. 18-20kbp that assembled into putative phage genomes NSROP-1 and NSROP-2.

In addition to *Spiroplasma* NSRO, an attempt at isolating phage like particles form MSRO-Br was performed. Using a CsCl ultracentrifugation gradient, a phage-like band was extracted from the 1.48g/ml density band. The DNA was extracted from the phage-like particles via a phenol-chloroform method, and run on a 1.5% agarose gel. A band >15 kbp was observed (lane 6; Figure S3). Unfortunately, there was not enough DNA to perform sequencing.

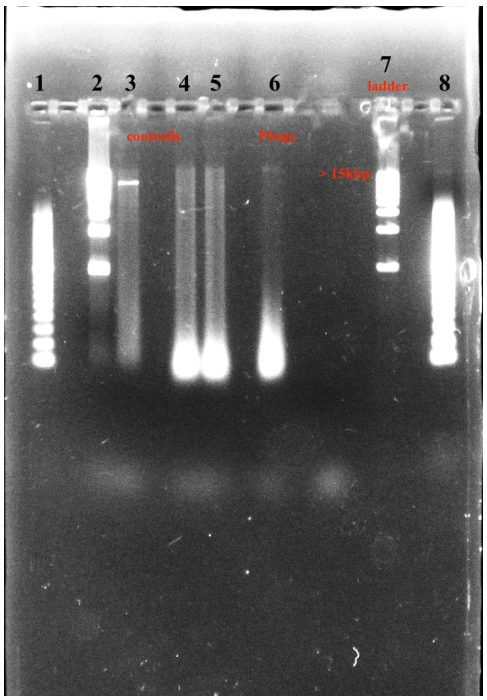

Figure A3 – MSRO Brazil Phage Extract Gel electrophoresis.
Gel electrophoresis on DNA extracted from viral-like particle band of *Spiroplasma poulsonii* MSRO-Br. *Lane 1,8 = 100bp ladder*; *lane*s *2,7 = 1kbp ladder, lane*s *3–5 = whole fly (with Spiroplasma) DNA extractions*; and *lane 6 = MSRO-BR Phage*-like particles.

GC Shift in the Prophage Region of the MSRO-Ug Genome

Both NSROP-1 and NSROP-2 partially aligned to the MSRO-UG strain H99 genome (Figure S4 and S5). There were no large changes in the GC content of the MSRO-UG genome region to which the phage contigs mapped.

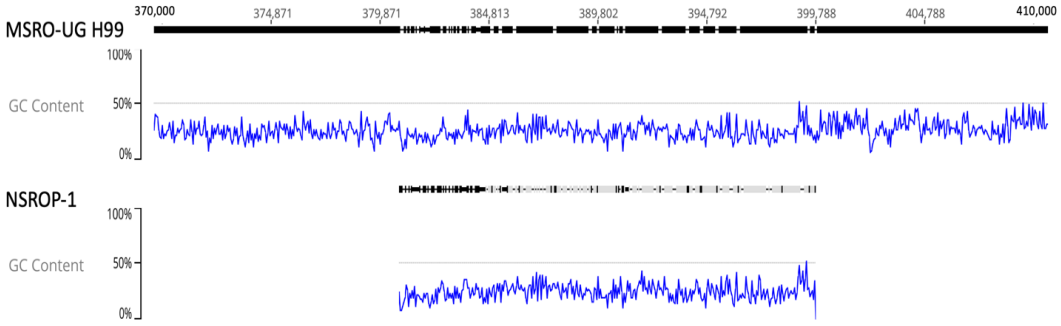

Figure A4 - GC content of the MSRO-UG genome region to which NSROP-1 Maps.
Blue line = GC Content, Black line of consensous reperesents non-homologus regions, grey lines repersent homolugus regions, empty spaces represent in-dels between the NSROP-1 draft genome and referance MSRO-UG H99 genome. Numbers above the MSRO-UG H99 line repersent regions within referance genome.

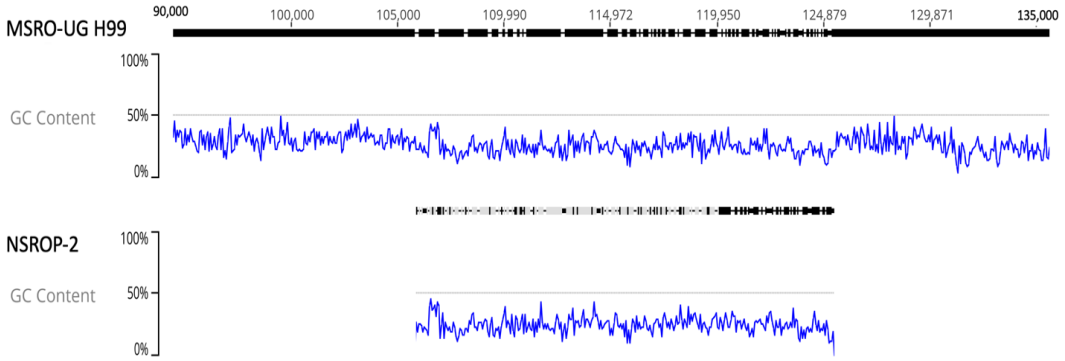

Figure A5 - GC content of the MSRO-UG genome region to which NSROP-2 Maps.
Blue line = GC Content, Black line of consensous reperesents non-homologus regions, grey lines repersent homolugus regions, empty spaces represent in-dels between the NSROP-2 draft genome and referance MSRO-UG H99 genome. Numbers above the MSRO-UG H99 line repersent regions within referance genome.

Spiroplasma Chic. SRA

The *Spiroplasma Chic*. SRA was downloaded to obtain Illumina Miseq reads. Trimmomatic was used to remove adapter sequences and filter low quality sequences. Reads were assembled using Spades v3.12. A discontiguous blastn search using the nt-database (Jan 2018) to identify *Spiroplasma* blast hits. Only contigs with *Spiroplasma* hits were accepted as the draft assembly due to the metagenomic nature of contig pool.

List of Spiroplasma genomes (RefSeq Assembly Accession Numbers) that were searched for phage

NZ_CP024870, NZ_CP011856, NZ_CP001973, NZ_CP017015, NZ_CP025057, NZ_CP013860, NZ_CP012328, NZ_CP012357, NZ_CP017658, NZ_CP018022, NZ_CP012622, NZ_CP006681, NC_022998, NZ_CP011855, NZ_CP006720, NZ_CP002082, NC_021280. NC_021284, NZ_CP006934, NC_021833, NZ_CP025543

GCF_003124115.1_ASM312411v1_genomic,

GCF_003124105.1_ASM312410v1_genomic,

GCF_002968355.1_ASM296835v1_genomic,

GCF_002868295.1_ASM286829v1_genomic,

GCF_002237575.1_ASM223757v1_genomic,

GCF_001886855.1_ASM188685v1_genomic,

GCF_001274875.1_ASM127487v1_genomic,

GCF_000820525.2_SMSRO_2016_genomic,

GCF_000439435.1_ASM43943v1_genomic,

GCF_000328865.1_SMIPMB4A_v3_genomic,

GCF_000236085.2_SpiMel2.0_genomic

Attached Files

NSRO_phage_Final.fasta - This file contains the final contigs of NSRO Phage 1 and Phage 2 based on the assembly of Nanopore reads (Fasta format).

NSRO_phage_Final.gb - This file contains the final annotated contigs of NSRO Phage 1 and Phage 2 based on the assembly of Nanopore reads (GenBank format).

NSRO_Phage_original_Canu_contigs.fasta - This file contains the original contigs of NSRO Phage 1 and Phage 2 based on the assembly of Nanopore reads (Fasta format).

Spiroplasma_Chic_assembly.fasta - This file contains the final contigs of Spiroplasma SpChic based on the assembly of Nanopore reads (Fasta format).