Preliminary Characterization of Phage-like Particles from the Male-Killing Mollicute *Spiroplasma poulsonii* (an Endosymbiont of *Drosophila*)

Paulino Ramirez[1,2], Justin C. Leavitt[3], Jason J. Gill[3], Mariana Mateos[1,4*]

[1]Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, TX, USA

[3]Department of Animal Science, Texas A&M University, College Station, TX, USA

[4]Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX, USA

[2]present address: University of Texas Health Science Center-San Antonio, San Antonio, TX, USA

*corresponding author: mmateos@tamu.edu

ORCID IDs:

Paulino Ramirez: 0000-0001-8532-9590

Justin C. Leavitt: 0000-0003-4544-0483

Jason J. Gill: 0000-0002-9494-6053

Mariana Mateos: 0000-0001-5738-0145

Author contributions:

Conceived and designed the analysis: PR, JCL, JJG, MM

Collected the data: PR, JCL, JJG

Contributed materials, data or analysis tools: PR, JCL, JJG, MM

Performed the analyses and interpreted the results: PR, JCL, JJG, MM,

Wrote the paper: PR, MM

Edited the paper: PR, MM, JCL, JJG

Acknowledgments

Anika Stankov for technical assistance. Paul de Figueiredo provided feedback on the research and early versions of the manuscript. Portions of this research were conducted with high-performance research computing resources provided by Texas A&M University (**https://hprc.tamu.edu**). This research was conducted in partial fulfillment of the Master's of Science degree requirements of Paulino Ramirez.

Data Availability Statement

The raw sequence data used for the present study are available at NCBI under Project Number PRJNA545743; BioSamples SAMN11919470, SAMN11919470, SAMN23459927; and SRA SRR17050036, SRR17063333, SRR17065466. Assembled and annotated contigs are available at NCBI under GenBank Accession Nos. OL689226-OL689230 and OL778852.

Keywords: bacteriophage, Spiroplasma, phage-like particles, gene transfer agents, Nanopore, Illumina

Abstract

Viruses are vastly abundant and influential in all ecosystems, and are generally regarded as pathogens. Viruses of prokaryotes (themselves highly diverse and abundant) are known as bacteriophages or phages. Phages engage in diverse associations with their hosts, and contribute to regulation of biogeochemical processes, horizontal movement of genes, and control of bacterial populations. Recent studies have revealed the influential role of phage in the association of arthropods and their heritable endosymbiotic bacteria (e.g. the Proteobacteria genera *Wolbachia* and *Hamiltonella*). Despite prior studies (~30 years ago) documenting presence of phage in the mollicute *Spiroplasma* infecting *Drosophila*, genomic sequences of such phage are lacking, and their effects on the *Spiroplasma-Drosophila* interaction have not been comprehensively characterized. The present work isolated phage-like particles from the male-killing *Spiroplasma poulsonii* (strains NSRO and MSRO-Br) harbored by *Drosophila melanogaster*. Isolated particles were subjected to DNA sequencing, assembly, and annotation. Our results recovered three ~19 kb phage-like contigs (two in NSRO and one in MSRO-Br), and two smaller non-phage-like contigs encoding a known *Spiroplasma* toxin and an insertion element. Whole or parts of the particle-derived contigs were found in the genome assemblies of members of the *Spiroplasma poulsonii* clade. Although our results do not allow us to distinguish whether the contigs obtained represent infective phage-like particles capable of transmitting their DNA to new hosts, their encoding of several typical phage genes suggests that they are at least remnants of functional phage. We discuss potential implications of our findings and suggest future directions.

Introduction

Viruses are the most abundant and diverse biological entities 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 only on their host, but also on the surrounding community, and influences biogeochemical processes, contributes to horizontal gene transfer, and controls their host populations (reviewed in Luque et al. 2020). 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).

Viruses that associate with bacteria, referred to as bacteriophages or phages, also exhibit significant effects on microbial communities and on their host's ability to invade and proliferate in eukaryotes (Chatterjee and Duerkop 2018; Dion et al. 2020; Stone et al. 2019). Many phages carry genes that encode toxins that are responsible for the virulence of their host bacteria (Brüssow et al. 2004; Łoś et al. 2021). For example, the causative agent of the severe diarrheal disease cholera is a strain of *Vibrio cholerae* that carries a prophage-encoded toxin (Waldor and Mekalanos 1996). Phages also influence the interactions between inherited bacteria and their arthropod hosts.

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 required 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. tolerance to environmental stressors; Montllor et al. 2002) or defense against natural enemies (reviewed in Hajek et al. 2019; Mateos et al. 2016). Recent studies of the mechanistic bases of arthropod-bacteria reproductive manipulation and defensive mutualisms have implicated phages. For example, members of the Alphaprotebacterium genus *Wolbachia* (the most common facultative heritable symbiont of arthropods), cause several reproductive phenotypes including cytoplasmic incompatibility (CI). CI is a type of reproductive manipulation that leads to embryo lethality in the offspring of *Wolbachia*-infected males and uninfected females (Werren et al. 2008; Yen and Barr 1971). The *Wolbachia*-encoded genes responsible for CI were recently discovered and are encoded by a temperate phage known as "WO" (Fujii et al. 2004; Harrison and Brockhurst 2017; LePage et al. 2017; Shropshire et al. 2021; Shropshire et al. 2018; Shropshire et al. 2020b; Shropshire et al. 2020a). Similarly, the male-killing phenotype of certain *Wolbachia* strains also appears to be encoded by a phage (Perlmutter et al. 2019).

Phages have also been implicated in the defense phenotype conferred by certain bacteria to their arthropod hosts. Strains of the Gammaproteobacterium *Hamiltonella defensa* that are infected with a particular phage (APSE3) produce a soluble factor that prevents successful development of a parasitic wasp of the aphid (Brandt et al. 2017). Therefore, such phage-infected *H. defensa* strains are defensive mutualists of their aphid host. Several other bacterial lineages confer protection to their insect host, including members of the genus *Spiroplasma* (class Mollicutes).

The possible influence of phages on the heritable associations of the genus *Spiroplasma* and arthropods, including the fly genus *Drosophila*, has not been examined. *Drosophila-Spiroplasma*

interactions include reproductive parasitism (in the form of male-killing; Montenegro et al. 2006) and protective mutualism against parasitic wasps and nematodes (reviewed in Mateos 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; Garcia-Arraez et al. 2019; Harumoto and Lemaitre 2018), the role of phages 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). Four clades of *Spiroplasma* are associated with members of the genus *Drosophila*; each clade represents an independent invasion of *Drosophila* (Haselkorn et al. 2009). The most well studied *Drosophila*-associated *Spiroplasma* clade is *Spiroplasma poulsonii*, which includes strains that associate naturally with members of both the subgenus *Drosophila* (*D. neotestacea and D. hydei*) and the subgenus *Sophophora* (*D. melanogaster*, *D. simulans*, *D. willistoni*, and *D. nebulosa*) (reviewed in (Haselkorn 2010; Jaenike et al. 2010). Several strains of *S. poulsonii* are male killers (referred to as Sex Ratio Organisms, SROs), whereby male *Drosophila* die during early embryonic development (Montenegro et al. 2006). In addition, all strains of *S. poulsonii* examined to date have been shown to confer protection against parasitoid wasps (Mateos et al. 2016); and they are generally present in high titers compared to members of other *Spiroplasma* lineages (Haselkorn et al. 2013).

Spiroplasma phage research was particularly active in the 1970's, when several types of phage were 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 (Williamson et al. 1977). The first *Spiroplasma* phage was extracted directly from *Spiroplasma*-infected *Drosophila* (i.e., without the need for "induction" by an outside source; hereafter "first-degree extraction"). The second type of phage was extracted after injection of the "first-degree extraction" into *Drosophila* harboring a "sister" strain of *Spiroplasma* (hereafter referred to as "second-degree extraction"). Therefore, the phage from the "second-degree extraction" has properties that resemble lysogenic phage, where induction is necessary for successful isolation (Casjens 2003). Both types of phage had an icosahedral podoviridae-like morphology and a genome comprised of double-stranded DNA of ca. 22kbp (first phage) and 17kbp (second phage) (Oishi and Poulson 1970). Nonetheless, genome sequences of such phages were never generated. The well-studied *S. poulsonii*, which is now cultivable in liquid cell-free media (Masson et al. 2018), and whose hosts include the model organism *Drosophila melanogaster*, is an obvious candidate for re-launching *Spiroplasma* bacteriophage research after a ~4 decade hiatus. Herein, we performed isolation and DNA sequencing (short and long reads) of putative phage-like particles from *Spiroplasma poulsonii* strain NSRO. This strain was originally found in *Drosophila nebulosa*, but transferred and maintained in *D. melanogaster*. The NSRO strain was chosen because in previous studies, phage particles were more readily isolated from NSRO (i.e., with a "first-degree extraction") than from other *Drosophila*-associated strains (Oishi and Poulson 1970; Williamson et al. 1977). In addition, for comparison, we performed isolation and sequencing (short reads) of putative phage-like particles from "MSRO-Br"; a closely related *Spiroplasma poulsonii* strain originally derived from *D. melanogaster* in Brazil (Red-42; Montenegro et al. 2005).

Material and Methods

Source of Biological Material

A *Drosophila melanogaster* Oregon R fly lab strain harboring *Spiroplasma poulsonii* NSRO or MSRO-Br was maintained on Banana-Opuntia food at 25ºC (12:12 light:dark cycle). Both infected fly strains were 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 et al. 1998; Zhou et al. 1998).

Isolation of Phage-Like Particles

To enable the use of a large number of *Drosophila* individuals (i.e., > 3 grams), the PEG-based precipitation approach of van der Wilk et al. (1999) was adapted as follows. Whole flies 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 2000 x g to remove large debris. The supernatant was removed and passed through a 70 µm filter (28143-312; Millipore) and incubated at 4ºC overnight. Our preliminary experiments indicated that thicker centrifugation bands were obtained if this incubation step was performed. Solid NaCl was added to final concentration of 1.0 M 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 1 h and then centrifuged at 11,000 x g for 10 min at 4ºC. The supernatant was decanted and discarded, and the pellet was resuspended in 8 ml of SM buffer by mixing for 10–20 min. An equal amount of chloroform was added to remove residual PEG. After gently mixing, the sample was centrifuged at 3,000 x g for 15 min at 4 ºC. The aqueous layer was

collected and passed through a 0.45 µm filter (Cat. #28143-312; Acrodisk). The sample was stored at 4ºC until ultra-centrifugation (i.e., for 24 to 48 h). The sample was subjected to an Optiprep (60% iodixanol) step gradient as previously described (Takemura-Uchiyama et al. 2013). Each sample was layered over a prepared step iodixanol gradient of 10%, 30%, 35%, 40% and centrifuged at 209,500 x g in a Beckman Coulter SW41 rotor for 2 h (Fig. 1A). A phage-like band (see Results) was then extracted with a needle and stored at 4ºC until further use (i.e., TEM and/or DNA isolation). These procedures were performed separately on NSRO-infected, MSRO-Br-infected, or uninfected flies (to serve as a *Spiroplasma*-free control).

Transmission Electron Microscopy (TEM)

To visualize particles present in putative phage NSRO extracts (MSRO-Br extract was not visualized due to lack of sufficient sample), samples were examined by negative-stain TEM following preparation by the Valentine method (Valentine et al. 1968); samples were adsorbed to a carbon film, stained with 2% (w/v) uranyl acetate and placed on a copper grid (GT200T/Electron Microscopy Sciences). The grids were imaged at a magnification of 25,000x in a JEOL 1200 EX electron microscope (100 keV accelerating voltage) at the Texas A&M Microscopy and Imaging Center.

DNA Isolation from Particle Extracts

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).

Library Preparation and Sequencing

Because the DNA amount was insufficient for the library preparation protocols for MinIon Nanopore (NSRO only) and Illumina MiSeq sequencing (NSRO and MSRO-Br), the DNA was combined with a larger DNA sample from an unrelated source. For the NSRO Nanopore library, we used DNA from the tephritid fruit fly *Anastrepha obliqua* to obtain ca. 10 µg of gDNA. Absence of *Spiroplasma* in the sample from *A. obliqua* was verified by PCR with *Spiroplasma*-specific primers targeting the 16S rRNA (16STF1 and 16STR1; Haselkorn et al. 2009). 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 Nanopore sequencing run consisted of a Spot-on flow cell Mk 1 R9 (FLO-MIN106), MinION Mk1B sequencer, and utilized 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. For the Illumina libraries, we used DNA from known unrelated phages (see Results). The Illumina libraries were prepared with TruSeq Nano low-throughput (LT) kit (Illumina) generating inserts with an average size of 580 bp. The Illumina libraries were sequenced with the MiSeq Platform (paired-end 150 bp). Hereafter, the total reads from the MSRO-Br Illumina sequencing will be referred to as the "MSRO-Br read set", and likewise, the NSRO particle-derived reads will be referred to as "Illumina NSRO read set", "Nanopore NSRO read set", or "NSRO read set" to refer to both.

Assembly of NSRO particle-derived contigs

The Nanopore NSRO read set was used alone for a preliminary assembly aimed at extracting the reads derived from the NSRO putative phage-like particles (as opposed to the DNA derived from the *Anastrepha obliqua* sample). The adapter-free fastq Nanopore reads were assembled

with Canu v.1.7 (Koren et al. 2017) pipeline (genomeSize=2.0m, correctedErrorRate=0.105, corMinCoverage=0, corMaxEvidenceErate=0.15, corOutCoverage=1000; to maximize the number of reads utilized), then searched using blastn-discontiguous parameters to the NCBI nt database (January 2018) to identify putative *Spiroplasma* contigs based on an e-value of zero. One of the originally assembled putative *Spiroplasma* contigs was ~ 31kbp, with flanking repeat regions (see Supporting Figure S1); a pattern that is typically found in circularly permuted linear genomes (Casjens and Gilcrease 2009). To generate a new consensus, we removed one of the repeat regions to generate a "single copy" contig, which was used as a reference to re-map the original reads. In addition, 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 concatemer to visualize mapping patterns. A gap in coverage would imply that all the template molecules share the same end(s), and thus, do not represent circularly permuted elements. The same process was repeated with a second contig that was originally 21kbp (see Supporting Figure S2). To correct for commonly encountered errors of Nanopore-only assemblies (i.e., insertions and deletions), the Illumina NSRO read set was aligned (10 iterations using the "map to reference" tool of Geneious v.11.1.2; Biomatters, Ltd.) to the two single-copy-Nanopore resulting contigs, and consensus contigs were generated (hereafter "Illumina-corrected-Canu" assembly).

The Nanopore NSRO read set was then remapped with Minimap2 (-ax map-ont --secondary=no; Li, 2018) to the putative single-copy Nanopore *Spiroplasma* contigs to extract "*Spiroplasma*-only" long reads. These "*Spiroplasma*-only" Nanopore reads and all of the quality-filtered Illumina reads were subjected to a SPAdes v.3.12 hybrid assembly (33,55,77,107 kmer and default settings). The assembled contigs were then subjected to a Blast search (Altschul et

al. 1990; Camacho et al. 2009), using blastn-discontiguous parameters to the NCBI nt database (January 2018) to identify and retain putative *Spiroplasma* contigs with a maximum e-value of 1e-50 (hereafter, the “Hybrid Spades” assembly).

To obtain the most accurate assembly possible, the “Illumina-corrected-Canu” and the “Hybrid Spades” assemblies were combined. This was done by aligning the “Hybrid Spades” assembly (which contained smaller contigs) to the “Illumina-corrected-Canu” assembly. “Hybrid Spades” contigs that aligned to the Canu reference with high identity (>99%), were considered to be part of the reference (“Illumina-corrected-Canu”) contig. The remaining “Hybrid Spades” contigs plus the all of “Illumina-corrected-Canu” contigs were considered as the final assembly. For the two largest contigs, which exhibited evidence of potential circular permutation, the origin was set so that they aligned to highly similar respective regions in the published *Spiroplasma* MSRO genomes (see Results).

Assembly of MSRO-Br particle-derived contigs (short reads only)

The MSRO-Br read set containing MSRO-Br phage like particle DNA was assembled with SPAdes v.3.12 assembly (33,55,77,107 kmer and default settings). Akin to the procedures with the NSRO assembly, to identify the potential source of the assembled contigs, a blastn-discontiguous search (Altschul et al. 1990; Camacho et al. 2009) was performed using the NCBI nt database (January 2018). Contigs with hits to *Spiroplasma* strains having an e-value of 1e-50 or less were considered to be representative of the MSRO-Br phage like particles.

ORF Annotation

Open reading frames (ORFs) from the NSRO and MSRO-Br assemblies were initially predicted and annotated using Prodigal (Hyatt et al. 2010), within the Prokka v.1.13 (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 ORFs at least 100bp long were retained, allowing for overlapping ORFs (McNair et al. 2018). Non-coding RNA regions were annotated with Aragorn v.1.2.38 (Laslett and Canback 2004) . Assemblies were queried for potential prophage sequences using Phaster (Arndt et al. 2016). Additionally, translated ORFs were annotated with RASTtk v.2.0 (Brettin et al. 2015; accessed June 2019) and searched manually by HHPred (Zimmermann et al. 2018) and CD-Search (Yang et al. 2020) to assign functional annotations. Repeat regions were identified with the repeat finder tool of Geneious (parameters: at least 50bp long and 0% mismatches allowed). Signal peptide and transmembrane domains were assigned based on positive hits to the SignalP (Nielsen 2017) and Phobius (Kall et al. 2007) databases. Similarly, putative transmembrane regions were annotated if positive hits were observed from both the TMHMM (Krogh et al. 2001) and PHOBIUS (Kall et al. 2004) databases. IsFinder (Siguier et al. 2006; accessed 27 June 2019) was used to identify possible transposable-element-associated proteins in the translated sequences with the IsFinder 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 or virulence factors. The Geneious “Repeat finder” tool (parameters: at least 10bp long and 0% mismatches allowed) was used to manually identify inverted repeats flanking ORFs in assembled contigs (see Results).

Comparison of particle-derived contigs to sequenced *Spiroplasma* genomes and phage

Ten genomes from *Spiroplasma poulsonii* and from the *citri* clade were analyzed for endogenous presence of our particle-derived contigs or parts of them, using a homology-based search with RepeatMasker v.4.0.9 (Smit et al. 2013-2015). Detected regions were filtered based

on a minimum query coverage of 50% compared to the query contig, and a Smith-Waterman (SW) score of less than 4.0%. Individual ORFs of our phage-like contigs were also searched with RepeatMasker in MSRO-Ug and MSRO-Br genomic assemblies to identify potential multicopy ORFs in the respective genomes (same minimum scores were used as for contig search).

To identify potential homology, we used Muscle v. 3.8.425 (Edgar 2004a, b) to attempt alignment of our NSRO/MSRO-Br particle-derived contigs and previously sequenced *Spiroplasma* phages (NC_009987, NC_003793, NC_003438, NC_001365, NC_001270, AJ969242).

Assessment of expression of particle derived contigs with available RNA sequencing datasets

To determine whether regions assembled in the particle-derived contigs show evidence of expression, we used publicly available RNA-seq data. An RNA sequencing dataset corresponding to *Spiroplasma* strain MSRO-UG cultured in vitro and extracted from fly hemolymph were downloaded for NCBI GEO (GSE112290). Reads were trimmed and filtered with FASTP (Chen et al. 2018). Processed reads were then aligned with the Geneious RNA aligner to either the NSRO particle- or MSRO-Br particle-derived contigs. Alignments were then visually inspected with Geneious to qualitatively determine which of the ORFs in the particle-derived contigs were being expressed. Expression levels of entire contigs were computed using Reads per Kilobase Million (RPKM) method, except that gene length was replaced with the corresponding contig length. Expression was estimated by taking the number of reads aligned a contigs and dividing by the total number of sequenced reads. Then the resulting number was multiplied by a million to obtain a scaling factor to normalize for sequencing depth. The total number of reads aligned to a contig was then divided by both scaling factor and contig length in

kilobases to obtain an RPKM-like value. Geneious was used to calculate raw reads and RPKMs of ORFs of NSRO particle- or MSRO-Br particle-derived contigs.

Results

Isolation of phage-like particles

The iodixanol step gradient of the NSRO sample resulted in a visible band at the interface of the 35–40% layers (Fig. 1), consistent with the migration position for *Caudovirales* phages (Takemura-Uchiyama et al. 2013). Such a band was barely visible in the MSRO-Br sample, and it was not visible in the control samples of *Spiroplasma*-free fly homogenates. Extractions generally yielded a more visible band in the Optiprep gradient when flies were aged for at least two weeks, and the filtered (0.45 µm) homogenate was allowed to incubate for at least 24 h (results not shown). TEM of an aliquot of the NSRO band revealed tail-less, icosahedral particles measuring ca. 40 nm across (Fig. 2). DNA extracted from this interface from the NSRO sample produced multiple bands when run on an agarose gel, visible at approximately 1–1.5 kbp, 2 kbp, 3–4 kbp, and 5 kbp (Fig. 3). TEM and gel electrophoresis were not performed on the MSRO-Br sample due to its low yield, which was subjected to Illumina DNA sequencing only.

Assembly of NSRO and MSRO-Br particle-derived contigs

From the NSRO sample, three contigs with blast hits to *Spiroplasma* were assembled with Canu, but one was discarded due to low read coverage (7X coverage; only one read > 15kbp) and close sequence similarity to a higher coverage contig, resulting in two remaining contigs of size 31 kbp and 21 kbp. However, the presence of 12 kbp- and 1.3 kbp-long terminal repeats (oriented in the same direction) at their ends, respectively, and the raw read mapping pattern observed (Supporting Figs. S1 and S2), 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, resulting in two contigs ~19 kbp-long each. These contigs were then corrected with Illumina reads. The resulting “Illumina-corrected-Canu” contigs were then reconciled with the SPAdes Hybrid assembly.

Four contigs from the reconciliation assembly were recovered. The two aforementioned ~19 kbp contigs (hereafter referred to as “NSRO-P1” and “NSRO-P2”) were not evident in the gel of DNA isolated from the particle extract (Fig. 3). The sizes of the remaining two contigs (Table 1) were consistent with the DNA bands observed in the gel. The two other contigs, hereafter Contig 3 and “NSRO-IS5”, were 3.8 and 1.2 kbp, respectively. The GC content of the five contigs ranged from 21.6 to 24.8%.

From the MSRO-Br sample for which only Illumina reads were generated, we obtained six contigs with high homology to *Spiroplasma* based on blastn, of which only two had adequate coverage: a 19.7 kbp contig (hereafter referred to as “MSRO-Br-P1”) along with a 1.1 kbp contig (hereafter “MSRO-Br-IS5”; Table 1). The other four contigs were of lengths shorter than 500 bp and considerably lower coverage (See Supporting Dataset S1). MSRO-Br-P1 and the MSRO-Br-IS5 had GC% content levels of 24.9% and 20.7%, respectively.

<u>Coverage and size distribution of reads mapped to the particle-derived contigs</u>

400 Nanopore reads mapped to NSRO-P1 (not shown), of which 15 were at least 16 kbp, and none were greater than 22 kbp (Fig. 4A), suggesting that if this represents a phage genome, its size is less than 22 kbp (reads longer than 22 kbp were obtained from this sequencing library, but could be attributed to *A. obliqua* or its associated microbes; not shown). The average, maximum and minimum coverage of the Nanopore reads was respectively 52, 72, and 32X. The manner in which the longest Nanopore raw reads mapped to the artificial three-copy

concatemer version of NSRO-P1, with no gaps in coverage for the middle copy, or in its 5' and 3' junctions with the adjacent copies (Fig. 5A), is consistent with a circularly permuted genome and a headful packaging mechanism (Merrill et al. 2016). 1,795 Illumina reads aligned to the NSRO-P1 single-copy contig (13.7X coverage; Table 1), of which 18 pairs had the following pattern consistent with circular permutation: one member of the pair mapped to one end of the contig whereas the other mapped to the other end (not shown).

268 Nanopore reads mapped to NSRO-P2 (not shown) with an average, maximum and minimum coverage of 28.2, 42, 16X respectively. Seven Nanopore reads were > 14 kbp long and their mapping to the single-copy contig is shown in Fig. 4B. Mapping of these seven Nanopore reads to the three-copy concatemer (Fig. 5B), also suggests circular permutation and headful packaging. 1,163 Illumina reads aligned to the NSRO-P2 genome (8.679X coverage), of which 10 pairs had each member mapped to opposite ends of the contig (not shown), consistent with circular permutation.

Read coverage for the Contigs 3 and NSRO-IS5 was lower (Table 1). Two Nanopore and 121 Illumina reads mapped to Contig 3 (Illumina coverage = 4.33X). Zero Nanopore and 55 Illumina reads mapped to NSRO-IS5 (Illumina coverage = 6X).

Coverage of MSRO-Br-P1 ranged from 1–31 Illumina reads with an average of 9 reads for a total of 1,298 reads (Table 1). Coverage of MSRO-Br-IS5 ranged from 1-10 reads with an average of 5 reads for a total of 106 reads.

Functional annotation of particle-derived contigs

The final 18.8 kbp contig of NSRO-P1 contained 23 predicted protein-coding genes ranging from 41 to 1067 aa (Supporting Dataset S2; Fig. 6A). The contig resembles a partial phage genome, with large and small terminase subunits, portal, head-tail connector and capsid proteins identifiable by conserved domain or HHPred searches. In addition to these structural proteins, genes encoding phage-like DNA replication functions were also identified, including a single-stranded DNA binding protein, a potential replicative helicase/nuclease, a RecT-like recombinase and a Holliday junction resolvase. Two ORFs, peg.18 and peg.21 coded for hypothetical proteins with a bacterial signal motif and a eukaryotic signal motif, respectively. Additionally, three ORFs were predicted to contain transmembrane domains, peg.16 with three transmembrane domains, and peg.15 and peg.22 each with one transmembrane domain. This contig does not appear to be a complete phage genome, however, as it is notably missing recognizable regulatory switches, tail structural proteins, lysis genes, and an integrase that would be responsible for insertion and excision of the element. The NSRO-P1 sequence encodes recognizable capsid and DNA packaging machinery components, and the presence of this DNA in an ultracentrifuge gradient position consistent with phage, suggests it is capable of being packaged into a phage-like particle.

NSRO-P2, 19.1 kbp in length, contained 24 ORFs (size range 38–1073 aa; Fig. 6B) of which one could be assigned a putative function: a prophage-associated RecT ssDNA binding protein (peg.35; Uniprot ID: PF03837). Two ORFs coded for transmembrane proteins motifs, peg.38 (three motifs) and peg.39 (one motif), which is identical to peg.15 and peg.16 in NSRO-P1. The entirety of Contig 3 (3.8 kbp) mapped with 98.7% similarity to a region (1,436,327–1,440,042) of the *Spiroplasma* MSRO-UG assembly GCF_000820525.2_SMSRO_2016_genomic. This region contains three ORFs, of which one (peg.3 in our contig) corresponds the MSRO-UG's

Ribosome Inactivating Protein 2 (RIP2) gene containing a signal peptide, whose RIP-function has been confirmed experimentally (Garcia-Arraez et al. 2019). Another ORF (peg.2 in our contig) also has a signal peptide motif and is found in NSRO-P1 and NSRO-P2. NSRO-IS5 contained two ORFs (peg.4 and peg.5). The former has a DDE superfamily endonuclease, and the latter has a helix-turn-helix DNA-binding motif that is common among transposable elements (based on interpro: IPR027805). IsFinder indicated that both peg.4 and peg.5 had homology to transposable elements of the family IS5, group IS2. Additionally, using the Geneious repeat finder, we identified flanking inverted repeats (12bp long, 5'-AAAATATAATTT-'3) in NSRO-IS5. According to the NCBI Reference Sequence (WP_105628893.1), peg.4 and peg.5 undergo ribosomal slippage to produce a larger combined single protein composed of both a DDE superfamily endonuclease and a helix-turn-helix DNA-binding motif. A transmembrane motif also exists in the protein sequence.

MSRO-Br-P1 has 22 predicted ORFs (ranging from 45-1,071 aa) with 7 ORFs containing protein with known motifs according to InterPro, Phobius or SignalP (Fig. 6C). The contig contains genes with large and small terminase subunits, portal, head-tail connector and capsid proteins identifiable by conserved domain or HHPred searches. In addition, a RecT-like recombinase and a Holliday junction resolvase were identified (Supporting Dataset S2). MSRO-Br-IS5 was identical in annotation to NSRO-IS5 consisting of two ORFs and the observed inverted repeats (Supporting Dataset S2).

Similarity between the particle-derived contigs from NSRO and MSRO-Br

MSRO-Br-P1 was aligned to both NSRO-P1 and NSRO-P2 with the Geneious Aligner (Global alignment with free end gaps, Cost Matrix: 65% similarity, Gap Open Penalty: 12, Gap Extension Penalty: 3). MSRO-Br-P1 shares 78% and 81% identity with NSRO-P1 and NSRO-

P2, respectively (Supporting Fig. S3A). Interestingly, the recT-like recombinase is similar to the NSRO-P2 annotated recT gene (94.8%). Unlike NSRO-P1, a single-stranded DNA-binding protein or DNA helicase/nuclease protein was not annotated in MSRO-Br-P1. MSRO-Br-IS5 is identical to the NSRO-IS5 except on the terminal ends of the contigs (Supporting Fig. S3B).

Comparison of particle-derived contigs to available sequenced genomes of *Spiroplasma* phage

Attempts to align the NSRO particle- and MSRO-Br particle-derived contigs to previously described plectroviral phage and other phages from the *citri* and *melliferum* clades of *Spiroplasma* revealed no significant homology.

Presence of the particle-derived contigs in genome assemblies of the *poulsonii* clade, but not of the *citri* clade

A genome assembly of *Spiroplasma poulsonii* strains NSRO is not currently available to search for evidence of the presence of our particle-derived contigs within the bacterial chromosome, and thus confirm them as originating from the *Spiroplasma* strain (e.g. if they are lysogenic). Nonetheless, we searched for regions homologous to the *Spiroplasma* particle-derived contigs in the main chromosomal contigs and associated plasmids of the currently available genomes of the members of the *poulsonii* and *citri* clades of *Spiroplasma* (Table 2). For the *poulsonii* clade, we included *S. poulsonii* Neo (associated with *D. neotestacea*), and four sub-strains of MSRO, three of which are derived from “MSRO-UG”; a sub-strain of MSRO originally found in *D. melanogaster* from Uganda (Pool et al. 2006). The fourth substrain is MSRO-Br, from which MSRO-P1 was extracted. Based on sequences from several genes, *S. poulsonii* MSRO and NSRO are each other’s closest relatives (Anbutsu et al. 2008). We found no regions of homology in the currently available *citri* clade genome assemblies; none of these *citri* clade strains are associated with *Drosophila*.

Contig NSRO-P1 mapped with high homology to 2–3 regions in the main chromosomal contigs of the three MSRO-UG genomes (Table 2; Supporting Dataset S6). Similarly, NSRO-P2 mapped with high homology to 3–5 regions of the three MSRO-UG genomes. The RIP2-containing Contig 3 mapped with high homology (i.e., 98.7% similarity) to a single location in the chromosome of two MSRO substrains (MSRO-Masson GCA_000820525.2 and MSRO-SE GCA_003124105.1), and at a single location of a contig deemed an extrachromosomal element (or assembly artifact) of MSRO-H99 (GCA_003124115.1). Contig 3 also mapped to a single contig of the MSRO-Br assembly (contig VIGU01000005) with 98.6% similarity. As mentioned above, the RIP2-encoding region of Contig 3 had 100% match to the corresponding regions in the four MSRO assemblies. NSRO-P1, NSRO-P2, and Contig 3 had no matches fulfilling our criteria in the other six *Spiroplasma* genomes examined.

NSRO-IS5 (and MSRO-Br-IS5; its identical counterpart in MSRO), or its remnants, were found in numerous locations of all four of the *poulsonii*-clade genomes examined (Table 2): found at 174–237 chromosomal and 8–19 plasmid locations of the three MSRO-Ug genomes; and at 265 locations in the genome assembly of *S. poulsonii* Neo. The IS5 contig was only found at 3 locations in the MSRO-Br draft genome, which is highly fragmented.

Incomplete parts of the MSRO-Br-P1 contig were found in the three MSRO assemblies (Table 2). The full MSRO-Br-P1 contig was found in the MSRO-Br whole genome assembly with 85% percent identity to the circular VIGU01000004 (18.9 kbp) contig. Additionally, MSRO-Br-P1 was found in 5–7 different regions of each MSRO-UG assembly, with the largest regions being ~12,000bp long. Interestingly, parts of the MSRO-Br-P1 contig were found in 13 regions

belonging to putative assembly artifacts in the MSRO-JTV (GCA_000820525.2) assembly, and 6 regions in putative assembly artifacts in MSRO-H99 assembly (GCA_003124105.1).

Assessment of contamination by DNA from sources other than *Spiroplasma*

Due to the low amount of DNA obtained from the extractions targeting phage-like particles, our DNA extracts were combined with ("spiked into") a different sample that contained a much larger amount of DNA. In Supporting Dataset S3, we report the top blast hits of all contigs obtained as a result of our assemblies. For the Illumina MiSeq run (Index 4), which contained DNA from *Staphylococcus* and *Klebsiella* phages with our NSRO particle extraction, we obtained large contigs whose top blast hits were the known phage. Numerous contigs whose top blast hit was *Klebsiella* or other Proteobacteria were also recovered.

For the Nanopore run, the only contigs whose top blast hit was *Spiroplasma* were the ones we assigned to *Spiroplasma* NSRO particles. As expected from a library prep containing a predominance of DNA from *Wolbachia*-infected *A. obliqua*, numerous other contigs had blast hits to *Anastrepha*, *Wolbachia*, and Proteobacteria (not reported).

For the Illumina MiSeq run (Index 1), which contained DNA from *Vibrio, Staphylococcus* and *Klebsiella* phages, combined with our MSRO-Br particle extraction, we obtained large contigs whose top blast hits were these phage. In addition to the contigs that we assigned to *Spiroplasma* MSRO particles, we only detected four short contigs (259 to 468 bp) whose top blast hit was *Spiroplasma* (very low coverage). A few contigs whose top blast hit was Proteobacteria were also recovered, as well as a few short contigs (<400bp) whose top blast hit was *Homo sapiens*.

Representation in RNA-seq datasets of the particle-derived contigs

To determine if there is evidence of expression (at the mRNA level) of the ORFs annotated in the particle-derived contigs, we used currently available RNA sequencing datasets from *Spiroplasma poulsonii* MSRO-UG. These datasets were derived from fly hemolymph and from in vitro culture (NCBI PRJNA445479; GEO: GSE112290). We decided to exclude the RNAseq dataset of MSRO-BR (Higareda Alvear et al. 2021), because it was based on whole insect extractions and the yield of *Spiroplasma* reads was relatively low. Based on Reads per Kilobase Million (RPKM), at the contig level, NSRO-P1 had the lowest expression level (average 16–17; Table 3; Supporting Dataset S4). At the ORF level, NSRO-P1 also had the least representation with only 18 out of the 23 ORFs having at least 5 reads per replicate (in culture; Table 3; Supporting Dataset S5). Higher RPKMs were observed at the contig level for NSRO-P2 (50 in culture; 31 in hemolymph) and MSRO-P1 (57 in culture; 42 in hemolymph). At the ORF level, at least 21 out of 24 ORFs of NSRO-P2 had at least 5 reads mapped per replicate, but small regions of ORFs lacked coverage. For MSRO-P1, at least 21 out of 22 ORFs had coverage of at least 5 reads. Regions lacking coverage by RNA-seq reads may reflect absence of template DNA (chromosomal or extrachromosomal) in MSRO-UG, or lack of expression despite presence of template DNA (e.g. because it is not a functional gene or because expression of the functional gene does not occur under those experimental conditions). At the contig level, IS5 had the highest RPKM (294 in culture; 119 in hemolymph), and both ORFs were covered by at least 5 reads. Finally, NSRO Contig 3 (the one containing the RIP2 gene) had higher RPKM in hemolymph (43.5) than in culture (38.7) of MSRO-UG, as reported by Masson et al (Masson et al. 2018).

Discussion

TEM of our extraction aimed at isolation of phage-like particles from whole *D. melanogaster* homogenates harboring *S. poulsonii* NSRO revealed the presence of one type of icosahedral particle ca. 40 nm in diameter. Particles were not observed in the negative control samples (i.e., same *D. melanogaster* genetic background, but lacking *Spiroplasma* and *Wolbachia*). Previous extractions of *Spiroplasma poulsonii* NSRO also contained 40 nm icosahedral particles, but these also possessed a short tail (Oishi and Poulson 1970). The discrepancy between the presence and absence of a tail could be the result of differences in the genetic content of these particles. Comparable tailed phage particles were described following similar extraction protocols (i.e., without “induction” = “first degree extraction”; see Introduction) applied to *D. nebulosa* and *D. hydei* infected with their native (male-killing and non-male-killing, respectively) *Spiroplasma* strains (Cohen et al. 1987; Oishi et al. 1984). One caveat of the previous studies is that at that time, the existence of *Wolbachia* as a common endosymbiont of *Drosophila* (and many arthropods) was not known and therefore not tested. At least *D. willistoni* and *D. paulistorum* (native hosts to *Spiroplasma* strains known as WRSO and PSRO, respectively), but not *D. hydei*, are known to harbor *Wolbachia* (Mateos et al. 2006; Miller et al. 2010). Therefore, it is 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 (reported to have isometric heads ca. 40nm in diameter containing linear dsDNA ca. 20kbp; Bordenstein and Wernegreen 2004; Fujii et al. 2004). The present study, which used flies confirmed to be *Wolbachia*-free, rules out the possibility that any particles recovered or sequenced represent *Wolbachia* phage.

The reconciliation of the SPAdes hybrid and Illumina-corrected-Canu assemblies from NSRO particle-derived reads revealed two distinct types of *Spiroplasma* phage-like contigs with

adequate read coverage, to which we refer as NSRO-P1 and NSRO-P2. We consider the other two NSRO contigs (Contig 3 and NSRO-IS5) too small and lacking key genes to represent a phage genome. Assembly of only the long reads revealed that NSRO-P1 and NSRO-P2 were terminally redundant, which along with the alignment of long and short reads to NSRO-P1 and NSRO-P2, is consistent with a circularly-permuted terminally-redundant linear genome, which is a common DNA packaging mechanism in phages. To verify circular permutation, future studies should document electrophoretic patterns of NSRO particle DNA subjected to restriction digestion (e.g. Just and Klotz 1990). Excluding one of their respective terminally-repeated ends, their sizes were 18.936 kbp and 19.148 kbp, respectively. Such band sizes were not evident in the agarose gel, where the largest fragment was estimated at ~5 kbp. It is possible that ~19 kbp fragments were present, but at concentrations lower than detectable in the gel (the Nanopore library preparation was aimed at enriching fragments > 5 kbp). It is also possible that the conformation of the DNA run on the gels was different from double-stranded and linear, such as supercoiled DNA, which tends to run faster in the agarose gel matrix (Levene and Zimm 1987; Zivanovic et al. 1986).

The size of NSRO-P1 and NSRO-P2 contigs is similar, albeit somewhat smaller, to the 21.8 kbp 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 (Oishi et al. 1984; Williamson et al. 1977). The 21.8 kbp phages of NSRO and Hydei and the 21 kbp phages of two non-*Drosophila*-associated strains (*S. citri* and *S. mirum*), are all reportedly circularly permuted and terminally redundant (Alivizatos et al. 1982; Maniloff and Dybvig 1988), and thus, likely possess a headful packaging replication mechanism (Fujisawa and Morita 1997). The large phage-like terminase annotated in NSRO-P1 and NSRO-P2 could

aid in the packing of viral dsDNA (Gual et al. 2000), whereas the ssDNA binding protein and recombinases would most likely be involved in phage DNA replication (Meyer and Laine 1990).

NSRO-P2 also possesses an ssDNA binding domain protein. The ssDNA binding gene has homology to the temperate-associated RecT family (Lopes et al. 2010), which is found in prophages such as *Listeria monocytogenes* phage A118 (Loessner et al. 2000). RecT is a ssDNA binding protein that promotes strand exchange among homologous DNA. Despite belonging to a superfamily of single stranded DNA binding proteins, ERF and RecT are evolutionarily unrelated. The presence of both types of temperate recombinases, as well as the high (albeit imperfect) homology detected upon mapping NSRO-P1 and NSRO-P2 to the main chromosome of the three MSRO-UG substrains with available draft genomes, suggests that both might be lysogenic.

Compared to the NSRO sample, the extraction procedure of MSRO-Br resulted in a much lower yield of particles, and thus of DNA and reads. The assembly from the MSRO-Br particle-derived Illumina reads produced two contigs (MSRO-Br-P1 and MSRO-Br-IS5) that were similar in genetic content and size to NSRO-P2 and NSRO-IS5, respectively. The available draft genome assembly for strain MSRO-Br (GCA_019295385) allowed us to search for the presence of the two MSRO-Br particle-derived contigs in what is effectively its host's genome. The two MSRO-Br particle-derived contigs were found in the genome assembly. MSRO-Br-P1 showed high similarity to a circular contig from the whole genome draft assembly. Nonetheless, the particle-derived contig showed more complete ORFs compared to the genome assembly contig. This may be the result of errors in MSRO-Br whole genome assembly, stemming from the repetitive nature of regions the MSRO-Br-P1 contig. It is also possible that the differences between MSRO-Br-P1 and the circular contig from the whole genome assembly reflect mutations

accrued between the collection time points of the two samples, as *Spiroplasma poulsonii* strains evolve rapidly (Gerth et al. 2021).

The present data do not allow us to distinguish whether one or more of the three ~19kbp phage-like contigs obtained (two in NSRO and one in MSRO-Br) represent infective phage-like particles capable of transmitting their DNA to new hosts. Nonetheless, their encoding of several typical phage genes suggests that at the very least they are remnants of functional phage. Our RNAseq analyses revealed that most ORFs are transcribed. The DNA sequences reported here were recovered from a density gradient designed to isolate phage-like particles, which only produced visible bands in *Spiroplasma*-positive samples and in which particles were observed (only NSRO was subjected to TEM), all suggesting they are packaged into virions (Scenario 1; Fig. 7), but further work is needed (e.g. DNAse treatment of particles prior to DNA isolation) to fully rule out that they represent non-packaged extra-chromosomal (Scenario 2; Fig. 7) or non-packaged chromosomal DNA fragments (Scenario 3). If the particle-derived contigs we obtained indeed are the result of DNA present in particles (i.e., Scenario 1), such DNA-containing particles could be released via lysis or budding (i.e., a lysis-free release mechanism reported in the mollicute Acholeplasma; Maniloff and Dybvig 2006). Evidence consistent with a “budding” extrusion mechanism was also reported in NSRO (Williamson et al. 1977). Even if Scenario 1 were true, the particle-derived phage-like contigs we recovered may not represent functional phage; instead, they may represent elements that are more similar to Gene Transfer Agents (GTA; see discussion below).

Implications if the phage-like contigs represent functional phage

There are several ways in which phages of *Spiroplasma* SRO could enhance their host fitness. Firstly, phage might contribute to the male-killing ability of SRO strains, akin to the *Wolbachia* phage that encode genes associated with cytoplasmic incompatibility and male-killing (Lindsey et al. 2018; Bordenstein and Bordenstein 2021; Perlmutter et al. 2019). Secondly, SRO 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 et al. 2016). Nonetheless, NSRO-P1, NSRO-P2 or MSRO-Br-P1 do not encode any of the genes that have been previously associated with the male killing and/or protective phenotypes of *Spiroplasma*: *Spaid* (Harumoto and Lemaitre 2018); and *RIP* (Ballinger et al. 2019; Ballinger and Perlman 2017; Hamilton et al. 2016; Garcia-Arraez et al. 2019) . Lastly, it is also possible that the SRO phage 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 et al. 2013). High densities of *Spiroplasma* (MSRO-UG) in old *D. melanogaster* adults have been associated with reduced longevity and motor defects (Herren et al. 2014).

The presence of two distinct types of potential phages in NSRO 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 (Díaz-Muñoz 2017). Such factors might influence whether NSRO-P1 and NSRO-P2 co-infect a single *Spiroplasma* cell or different *Spiroplasma* cells within the same population (e.g. same insect body or tissue).

NSRO-P1 and NSRO-P2 could have an antagonistic association, as prophage have been shown to suppress co-infection by other phages to prevent reduction in host fitness (Berngruber et al. 2010). This would likely result in co-infection only at a population level (i.e., each *Spiroplasma* cell infected by at most one type of phage). Alternatively, NSRO-P1 and NSRO-P2 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 a scenario, both phages might co-infect the same cell. However, it is possible that NSRO-P1 and NSRO-P2 represent genetic variants of the same phage-like entity, or that one of them is an assembly artifact.

Implications of the non-phage-like small contigs

The recovery of the non-phage-like small contigs (the RIP2-containing Contig 3 of NSRO, and the NSRO-IS5 and MSRO-Br-IS5) from extractions aimed at phage particles raises a question as to whether segments of the *Spiroplasma* chromosome (not adjacent to prophage) are being packaged into phage-like particles, akin to the gene transfer agents (GTA) that are particularly common in α-proteobacteria, where random ~5–15 kbp fragments of chromosomal DNA are packaged into phage-like particles via headful packaging and released via lysis (Christensen and Serbus 2020; Lang and Beatty 2000; Lang et al. 2017; Lang et al. 2012; Yen et al. 1979). Unless Scenario 3 (i.e., chromosomal contamination from multicopy genes) holds true for the non-phage-like contigs of NSRO and MSRO, our recovery of these contigs from a particle extraction suggests that they are not random fragments of the chromosome. Packaging of non-random regions of the chromosome into GTAs has been reported in *Bartonella bacilliformis*, where run-off replication appears to preferentially amplify a particular region of the genome, which is then packaged into particles (Berglund et al. 2009). It is thus possible that the particle-derived contigs we recovered from *Spiroplasma* NSRO and MSRO particle extractions represent preferentially amplified and packaged regions of the chromosome.

The recovery of RIP2 (from NSRO contig 3) in our particle-derived contigs is particularly intriguing given the function and known expression patterns of this gene in MSRO strains. This gene is one of five RIP genes encoded in the MSRO genome (Gerth et al. 2021; Paredes et al. 2015). The *Spiroplasma*-mediated defense against parasitic wasps and nematodes of *Drosophila* appears to involve RIP-induced damage (depurination) to the ribosomes of these natural enemies (Ballinger et al. 2019; Ballinger and Perlman 2017; Hamilton et al. 2016). A direct connection between RIP2 specifically and wasp death has not been demonstrated yet. Nonetheless, based on transgenic expression of RIP2 in *D. melanogaster*, RIP2 has been demonstrated to cause *Drosophila* ribosome depurination, as well as to contribute to male embryo death and reduced fly longevity (Garcia-Arraez et al. 2019). Assuming that RIP2 has a similar role in NSRO, what is the potential benefit of having RIP2 DNA copies in virions? Could this lead to higher RIP2 transcription within the *Spiroplasma* cell where the virions are being formed? Or could such particles infect neighboring *Spiroplasma* cells, where the additional RIP2 copies could serve as template for transcription? Interestingly, RIP2 is among the most highly transcribed genes of MSRO when inside its native host *D. melanogaster* (Higareda Alvear et al. 2021), but RIP2 expression is significantly lower when in culture (Masson et al. 2018). It is possible that the difference in RIP2 mRNA levels between the two environments is a result of differential abundance in extra-chromosomal copies; e.g. high abundance in fly vs. low abundance in culture. Isolation of particles from in vitro cultured *Spiroplasma* should help address this question. Moreover, upon comparison of coverage by Illumina reads (from WGS datasets SAMN10488582 and SAMN15339386; i.e., not aimed at particle enrichment; Table 4), we estimate that RIP2 has a coverage consistent with a single-copy gene in MSRO-Ug, but not in MSRO-Br, where RIP2 has ~14x copies compared to the dnaA gene (presumably single-copy). In contrast, RIP1 exhibited coverage consistent with a single-copy gene in both MSRO-

Ug and MSRO-Br. The reason behind the difference in RIP2 inferred copy number is unknown. The source material for the two datasets differed (whole fly extraction for MSRO-Br vs. cell-free culture for MSRO-Ug). Thus, the higher RIP2 copy numbers detected in MSRO-Br could reflect a higher number of chromosomal copies, or a higher abundance of extrachromosomal copies, which could be condition-dependent. If RIP2 gene copy numbers differ between MSRO-Br and MSRO-Ug, they could be related to the higher degree of overall protection against certain wasps reported for MSRO-Br (Jones and Hurst 2020).

The recovery of NSRO-IS5 and MSRO-Br-IS5 contigs in NSRO and MSRO-Br phage particle isolations is also interesting. Both contigs showed high homology to the IS5 family of Insertion Sequences (Humayun et al. 2017). Insertion Sequence elements are considered to be the shortest mobile elements capable of autonomous transmission, and generally contain an ORF coding for a transposase that is required for transposition (Siguier et al. 2015). IS5 elements contain consecutive reading frames coding for Aspartate-Aspartate-Glutamine (DDE) transposases. Given the numerous copies of IS5 contigs found in the genome assemblies of *S. poulsonii*, it appears that transposition of this element is frequent. Whether the potential presence of IS5 contigs in virions has any consequence is unclear. For example, the IS5 fragments contained in virions could potentially infect neighboring *Spiroplasma* cells and thereby further contribute to expansion of this element. Ultimately, transposition of elements such as IS5 can lead to genome instability, genome reduction, and alteration of expression (Cerveau et al. 2011; Glansdorff et al. 1981; Moran and Plague 2004). This could contribute to the rapid evolution seen in *Spiroplasma poulsonii* genomes (Gerth et al. 2021).

Conclusions and Outlook

The generation of bacteriophage-like genome sequences from *Drosophila*-associated *Spiroplasma* represents an initial step to explore a poorly understood aspect of the *Drosophila*-*Spiroplasma* endosymbiosis. Given the influential role of bacteriophages in the pathogenicity and survival of many endosymbiotic bacteria, interesting avenues of research include the role of bacteriophages on the negative (e.g. male-killing and over-proliferation) and positive (defense against natural enemies) phenotypes that *Spiroplasma* exerts on *Drosophila*. Phage research in this system would benefit from the development of solid media culture of *Spiroplasma* (i.e., for phage plaque assays) and elucidation of the function of the proteins encoded in these phage genomes. Although transformation is not yet feasible in *S. poulsonii*, its host *D. melanogaster* is highly amenable to genetic manipulation.

Several assays, which can be performed in the absence of (solid-media) cultivation, should contribute to verify that the contigs described herein represent DNA contained in particles. Achieving this will require substantially higher quantities of particles than those produced in the present study, which could be achieved by utilizing published culturing techniques for *Spiroplasma* MSRO, combined with induction techniques such as UV light exposure or Mitomycin-C to induce phage production (Broudy et al. 2001). A thorough treatment of particles with DNAse prior to DNA extraction must be undertaken to ensure that only particle-encapsulated DNA is examined (Owen et al. 2021). The particle-encapsulated DNA should be subjected to gel electrophoresis comparing intact DNA and DNA treated with restriction enzymes to assess DNA conformation (e.g. circular, linear, supercoiled) and verify patterns consistent with circular replication. Much deeper sequencing with High Fidelity (HiFi) long reads (or with a combination of short reads and low fidelity long reads) of particle-derived DNA, as well as “total bacterial DNA” from the same *Spiroplasma* “population”, should enable determination

of number and location of copies on the chromosome, verify terminal repetition of particle-derived contigs, and assess evidence of biased amplification (versus random fragments of the chromosome incorporated into particles). Protein expression of the putative phage genes encoded by the particle-derived contigs should be assessed; if the phage-like contigs represent functional phage, the structural proteins they encode should be detectable in particle isolates. If treatments/conditions that differ in particle abundance exist, protein and transcript levels of the putative phage-encoded genes should be higher in treatments where particles are more abundant. Combining the above assays with *in vivo* exposure to isolated particles (e.g. expose flies infected with one strain of Spiroplasma to particles derived from a different strain of Spiroplasma, as in Oishi 1971), may help elucidate whether these particles contribute to horizontal gene transfer (HGT), inter-strain antagonism, control of *Spiroplasma* titers within the fly, or influence other *Spiroplasma*-mediated phenotypes, such as male killing or protection against natural enemies.

Figure legends

Figure 1. Cartoon of the banding patterns obtained after gradient centrifugation of the NSRO extract aimed at isolating phage particles. The Optiprep step gradient resulted in a 35–40% Iodixanol band for the *Spiroplasma* NSRO sample. Additionally, other bands corresponding to the other gradient interfaces were obtained as well. The contents of bands 2, 3, and 4 were viewed under TEM to determine the composition of the band. Figure 2A shows particles from band 4. No viral-like particles were observed in bands 2 or 3. No band was visible in the *Spiroplasma*-free control sample (not shown).

Figure 2. **TEM image of putative NSRO phage particles. A)** TEM of negatively stained particles isolated with Iodixanol ultracentrifugation from *Spiroplasma* strain NSRO (band 4; Figure 3). B) TEM of negatively stained particles isolated with Cesium Chloride based gradient at a density of at 1.48g/ml. The particles are ca. 45nm across.

Figure 3. **Gel electrophoresis of DNA extracted from the NSRO phage-like particles.** An aliquot of this DNA isolation was subjected to Nanopore sequencing. Left Lane = DNA ladder (band sizes indicated in Kb); Right Lane = DNA sample not subjected to restriction digestion.

Figure 4. Mapping results of the trimmed long reads (black) to the NSRO-P1 and NSRO-P2 contigs. A) The 14 longest reads (16 kbp to 21.9 kbp; avg 20 kbp) mapped to NSRO-P1 are shown. B) The 7 longest reads mapped to NSRO-P2 (14.852 kbp to 21.9 kbp; avg 19 kbp) are shown. Light gray boxes in reads represent trimmed regions. Thin lines represent gaps in alignment of the reads to the reference. Annotated ORFs (colored arrows) are shown for reference and described in detail in Figure 6.

Figure 5. Mapping patterns of the longest Nanopore reads to artificial three-copy concatemers of NSRO-P1 (A) and NSRO-P2 (B). The raw long reads are shown in black. Gray outline boxes in reads represent trimmed regions (i.e., different from reference).

Figure 6. Annotation of NSRO-P1 (A), NSRO-P2 (B) and MSRO-Br-P1 genomes (C). Open reading frames (ORFs) are indicated with colored block arrows. Yellow = hypothetical CDS with no homology to known proteins with known functions. Blue graphs depict the GC content (grey line = 50%) at a15 bp sliding window. **A)** NSRO-P1, **B)** NSRO-P2, **C**) MSRO-Br-P1 contained annotated ORFs, Blue = transposase proteins. Green = capsid structure proteins. Orange = DNA binding/recombination proteins.

Figure 7. Possible scenarios to explain overrepresentation of a particular locus in a Whole Genome Shotgun sequencing dataset. Cartoon depicting three scenarios that could result in an overrepresentation of a particular DNA sequence (the blue double strand) in an assembly of reads obtained from an isolation aimed at viral or viral-like particles from a sample containing cells of *Spiroplasma*.

Supporting Figure Legends

Figure S1. Original Canu assembly of NSRO phage like contigs. Grey arrows in the reference are imperfect terminal repeats in the contig. A) NSRO-P1, B) NSRO-P2

Figure S2. Alignment of Nanopore reads to the original Canu assembly contigs. Grey arrows in the reference are imperfect terminal repeats in the contig. The raw long reads are shown in black. Gray outline boxes in reads represent trimmed regions (i.e., different from reference). A) NSRO-P1, B) NSRO-P2

Figure S3. Geneious-based alignments with annotated ORFs of (A) NSRO-P1, NSRO-P2 and MSRO-P1 phage-like contigs, and (B) Insertion element identified in the phage particle assembly from NSRO and MSRO-Br. Sites identical bases in all positions are depicted green identity bar. In (B) two ORFs were identified along with inverted repeat sequences (grey). The two sequences have 100% homology, except at terminal ends of the contigs. The yellow ORFs (possible translational frameshift) code for a DDE superfamily endonuclease and mobile element like protein respectively, which can form an Insertion element**.**

Tables

Table 1. Characteristics of the contigs derived from the putative phage particles with sufficient read coverage.

Contig Name	Length bp	GC content %	Number of ORFs	Illumina Miseq reads mapped (coverage)	Nanopore Minion reads mapped (coverage)
NSRO					
NSRO-P1	18,836	24.0	23	1,795 (13.7)	400 (51.4)
NSRO-P2	19,169	24.2	24	1,163 (8.79)	268 (28.2)
3	3,805	24.8	3	121 (4.33)	2 (0.55)
NSRO-IS5	1266	21.6	2	55 (6.060)	0 (0)
MSRO-BR					
MSRO-BR-P1	19,700	24.9	22	1,298 (9)	NA
MSRO-BR_IS5	1,106	20.7	2	106 (5)	NA

Table 2. Summary of the results from the search for regions with similarity to the NSRO and MSRO-Br particle-derived contigs in *Spiroplasma* genome assemblies from the *poulsonii* and *citri* clades. Only regions with high homology (SW <4.0 and > 50% length of original contig) were counted as complete regions or remnants of the NSRO putative phage contigs. Detailed output available in Supporting Dataset S6. All listed genomes are considered "complete" genomes except for Neo and MSRO-Br. MK = male killing. *S. poulsonii* belong to the *poulsonii* clade, whereas all other belong to the *citri* clade.

GenBank assembly accession number	Phenotype	Species, strain, and chromosome, contig, or plasmid ID [size in kbp]	NSRO-P1 (18.8 kpb)	NSRO-P2 (19.2 kbp)	NSRO Contig 3 (3.8 kbp)	IS5 (1.26 kbp)	MSRO-Br-P1 (19.71 kbp)
GCA_000820525.2	MK,Cultured	*S. poulsonii* MSRO-Masson (JTLV02000001) [1,800]	3	5	1	236	5
		S. poulsonii MSRO-Masson Plasmid (JTLV02000002) [36.2]	0	0	0	8	0
GCA_003124105.1	Non-MK	*S. poulsonii* MSRO-SE (NZ_PETF01000001.1) [1,800]	3	6	1	237	7
		S. poulsonii MSRO-SE Plasmid (NZ_PETF01000002) [55]	0	0	0	8	0
GCA_003124115.1	MK	*S. poulsonii* MSRO-H99 (NZ_PETG01000001.1) [1,400]	2	4	0	174	7
		S. poulsonii MSRO-H99 Plasmid (NZ_PETG01000004) [87]	0	0	0	19	0
		S. poulsonii MSRO-H99 Artifact (NZ_PETG01000003) [257]	0	4	1	0	2
GCA_019295385.1	MK	*S. poulsonii* MSRO-Br (Red 42)	0	0	1	3	1
GCA_003989055.1	Non-MK	*S. poulsonii* Neo	0	0	0	265	0

GCA_003339775.1	Plant pathogen	*S. phoeniceum* strain:p40	0	0	0	0	0
GCF_000236085.2	Bee pathogen	*S. melliferum*	0	0	0	0	0
GCF_000328865.1	Bee pathogen	*S. melliferum*	0	0	0	0	0
GCF_001274875.1	Corn pathogen	*S kunkelii* CR2-3x	0	0	0	0	0
GCF_001886855.1	Plant pathogen	*S. citri*	0	0	0	0	0

Table 3. Analysis of mRNA expression levels of particle-derived contigs, by published *Spiroplasma poulsonii* MSRO-UG RNA-seq datasets. RNA-seq reads were aligned to the five particle-derived contigs (from both NSRO and MSRO) to estimate expression at the contig level (Reads per Kilobase Million; RPKM). The numerator in "ORFs expressed" rows indicates ORFs that had at least five reads mapped per replicate. Detailed output by contig and ORF is provided in Supporting Dataset S4.

	NSRO-P1 (18.8 kpb)	**NSRO-P2 (19.2 kbp)**	**NSRO Contig 3 (3.8 kbp)**	**IS5 (1.26 kbp)**	**MSRO-P1 (19.71 kbp)**
Culture	16.06	49.90	35.95	294.16	56.67
Hemolymph	17.22	30.66	55.06	118.60	42.04
Culture: ORFs expressed	18/23	24/24	3/3	2/2	22/22
Hemolymph: ORFs expressed	23/23	21/24	3/3	2/2	21/22

Table 4. Comparison on inferred copy numbers of RIP1 and RIP2 genes in *Spiroplasma poulsonii* substrains MSRO-Br (NCBI Acc. Nos.: PRJNA507275; SAMN10488582) and MSRO-Ug (NCBI Acc. Nos.: PRJNA640980; SAMN15339386). Mean coverage of RIP1, RIP2, and dnaA (presumably a single-copy gene). Locus tags in MSRO-Br genome (VIGU00000000.1) are shown in parenthesis. Mean coverage was estimated in Geneious by examining the output of reads mapped to the three genes by Bowtie2 (Langmead and Salzberg 2012) with default parameters. Within each substrain, mean coverages of RIP1 and RIP2 were divided by mean coverage of dnaA to estimate their copy numbers.

	MSRO-Br		**MSRO-UG**	
Gene	**Mean Coverage**	**estimated copy number (coverage standardized by dnaA)**	**Mean Coverage**	**estimated copy number (coverage standardized by dnaA)**
RIP1 (EIB84_03210)	305	0.90	22	0.60
RIP2 (EIB84_07200)	4725	13.98	32.9	0.89
dnaA (EIB84_04815)	338		36.9	

Supporting Dataset Descriptions

Supporting Dataset S1. Four contigs from the MSRO-BR particle extraction that had blast hits to *Spiroplasma*, but whose coverage was extremely low.

Supporting Dataset S2. Annotations of the particle-derived phage-like contigs (NSRO-P1, NSRO-P2, Contig3, IS5, and MSRO-P1). Each contig is in a separate sheet. ORFs were predicted and annotated using RastK software. ORFs were then manually annotated with HHPred to elucidate potential function of ORF products. HHPred predictions are described in notes with probability, accession ID and name.

Supporting Dataset S3. Contamination Report based on Blast Hits.

Top blast hit for each of the contigs recovered from the three sequencing runs: Nanopore and Illumina MiSeq Index 4 for NSRO particle derived DNA (sheets Blast_Index_4_NSRO and Index_4_NSRO_unique_hits); and Illumina MiSeq Index 1 for MSRO particle DNA (sheets Blast_Index_1_MSRO-Br and Index_1_MSRO-Br_unique_hits). The first worksheet contains metadata for each run consisting of information related to potential sources of contamination. Blast output of SPAdes assembled contigs and unique blast hits to organisms are included.

Supporting Dataset S4. Coverage of the NSROP-1, NSROP-2, Contig3, IS5 and MSROP-1 contigs by published RNAseq reads. RNAseq read coverage, raw and RPKM, for NSRO and MSRO ORFs are shown in separate sheets.

Supporting Dataset S5. Repeat detection at the ORF level of particle-derived contings. Results of RepeatMasker at the Open Reading Frame (ORF) level, for the particle-derived contigs (NSROP-1, NSROP-2, Contig3, IS5 and MSROP-1) against the MSRO-UG (GCA_000820525.2) and MSRO-Br (GCA_019295385.1) genome assemblies. The first sheet ("Totals") shows the number of times that each ORF (peg) of the particle-derived contigs was found by RepeatMasker in the genome assemblies. Raw output for each analysis is in a separate sheet.

Supporting Dataset S6. Repeat detection at the contig level of the particle-derived contigs. Results of RepeatMasker for the particle-derived contigs (NSROP-1, NSROP-2, Contig3, IS5 and MSROP-1) against ten *Spiroplasma* genome assemblies listed in Table 3. The first sheet "Legend" describes the header of each column. Raw output for each contig is in a separate sheet.

Reference list

Alivizatos A, Townsend R, Markham P (1982) Effects of infection with a spiroplasma virus on the symptoms produced by *Spiroplasma citri*. Ann Appl Biol 101:85-91. https://doi.org/10.1111/j.1744-7348.1982.tb00803.x

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410

Anbutsu H, Goto S, Fukatsu T (2008) High and low temperatures differently affect infection density and vertical transmission of male-killing *Spiroplasma* symbionts in *Drosophila* hosts. Appl Environ Microbiol 74:6053-6059. https://doi.org/10.1128/aem.01503-08

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16-21. https://doi.org/10.1093/nar/gkw387

Ballinger MJ, Gawryluk RMR, Perlman SJ (2019) Toxin and genome evolution in a *Drosophila* defensive symbiosis. Genome Biol Evol 11:253-262. https://doi.org/10.1093/gbe/evy272

Ballinger MJ, Perlman SJ (2017) Generality of toxins in defensive symbiosis: Ribosome-inactivating proteins and defense against parasitic wasps in *Drosophila*. PLoS Pathog 13:e1006431. https://doi.org/10.1371/journal.ppat.1006431

Berglund EC, Frank AC, Calteau A, Vinnere Pettersson O, Granberg F, Eriksson AS, Naslund K, Holmberg M, Lindroos H, Andersson SG (2009) Run-off replication of host-adaptability genes is associated with gene transfer agents in the genome of mouse-infecting *Bartonella grahamii*. PLoS Genet 5:e1000546. https://doi.org/10.1371/journal.pgen.1000546

Berngruber TW, Weissing FJ, Gandon S (2010) Inhibition of superinfection and the evolution of viral latency. J Virol 84:10200-10208. https://doi.org/10.1128/JVI.00865-10

Bordenstein S, Bordenstein S (2021) The most widespread phage in animals: Genomics and taxonomic classification of Phage WO. bioRxiv. https://doi.org/10.1101/2021.11.04.467296

Bordenstein SR, Wernegreen JJ (2004) Bacteriophage flux in endosymbionts (Wolbachia): infection frequency, lateral transfer, and recombination rates. Mol Biol Evol 21:1981-1991. https://doi.org/10.1093/molbev/msh211

Braig HR, Zhou W, Dobson SL, O'Neill SL (1998) Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. J Bacteriol 180:2373-2378. https://doi.org/10.1128/JB.180.9.2373-2378.1998

Brandt JW, Chevignon G, Oliver KM, Strand MR (2017) Culture of an aphid heritable symbiont demonstrates its direct role in defence against parasitoids. Proc Biol Sci 284. https://doi.org/10.1098/rspb.2017.1925

Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD (2015) RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5:8365. https://doi.org/10.1038/srep08365

Broudy TB, Pancholi V, Fischetti VA (2001) Induction of lysogenic bacteriophage and phage-associated toxin from group a streptococci during coculture with human pharyngeal cells. Infect Immun 69:1440-1443. https://doi.org/10.1128/IAI.69.3.1440-1443.2001

Brüssow H, Canchaya C, Hardt W-D (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev 68:560-602. https://doi.org/10.1128/MMBR.68.3.560-602.2004

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

Casjens S (2003) Prophages and bacterial genomics: what have we learned so far? Mol Microbiol 49:277-300. https://doi.org/10.1046/j.1365-2958.2003.03580.x

Casjens SR, Gilcrease EB (2009) Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. Methods Mol Biol 502:91-111. https://doi.org/10.1007/978-1-60327-565-1_7

Cerveau N, Leclercq S, Leroy E, Bouchon D, Cordaux R (2011) Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from *Wolbachia* endosymbionts. Genome Biol Evol 3:1175-1186. https://doi.org/10.1093/gbe/evr096

Chatterjee A, Duerkop BA (2018) Beyond Bacteria: Bacteriophage-Eukaryotic Host Interactions Reveal Emerging Paradigms of Health and Disease. Front Microbiol 9:1394. https://doi.org/10.3389/fmicb.2018.01394

Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884-i890. https://doi.org/10.1093/bioinformatics/bty560

Christensen S, Serbus LR (2020) Gene Transfer Agents in Symbiotic Microbes. In: Kloc M (ed) Symbiosis: Cellular, Molecular, Medical and Evolutionary Aspects. Springer International Publishing, Cham, pp 25-76. https://doi.org/10.1007/978-3-030-51849-3_2

Cohen A, Williamson D, Oishi K (1987) SpV3 viruses of *Drosophila* spiroplasmas. Isr J Med Sci 23:429-433

Díaz-Muñoz SL (2017) Viral coinfection is shaped by host ecology and virus–virus interactions across diverse microbial taxa and environments. Virus evolution 3. https://doi.org/10.1093/ve/vex011

Dion MB, Oechslin F, Moineau S (2020) Phage diversity, genomics and phylogeny. Nat Rev Microbiol 18:125-138. https://doi.org/10.1038/s41579-019-0311-5

Duron O, Hurst GD (2013) Arthropods and inherited bacteria: from counting the symbionts to understanding how symbionts count. BMC Biol 11:45. https://doi.org/10.1186/1741-7007-11-45

Edgar RC (2004a) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113. https://doi.org/10.1186/1471-2105-5-113

Edgar RC (2004b) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797. https://doi.org/10.1093/nar/gkh340

Fujii Y, Kubo T, Ishikawa H, Sasaki T (2004) Isolation and characterization of the bacteriophage WO from *Wolbachia*, an arthropod endosymbiont. Biochem Biophys Res Commun 317:1183-1188. https://doi.org/10.1016/j.bbrc.2004.03.164

Fujisawa H, Morita M (1997) Phage DNA packaging. Genes Cells 2:537-545. https://doi.org/10.1046/j.1365-2443.1997.1450343.x

Garcia-Arraez MG, Masson F, Escobar JCP, Lemaitre B (2019) Functional analysis of RIP toxins from the *Drosophila* endosymbiont *Spiroplasma poulsonii*. BMC Microbiol 19:46. https://doi.org/10.1186/s12866-019-1410-1

Gerth M, Martinez-Montoya H, Ramirez P, Masson F, Griffin JS, Aramayo R, Siozios S, Lemaitre B, Mateos M, Hurst GDD (2021) Rapid molecular evolution of *Spiroplasma* symbionts of *Drosophila*. Microbial Genomics 7. https://doi.org/10.1099/mgen.0.000503

Glansdorff N, Charlier D, Zafarullah M (1981) Activation of gene expression by IS2 and IS3. Cold Spring Harb Symp Quant Biol 45 Pt 1:153-156. https://doi.org/10.1101/SQB.1981.045.01.024

Gual A, Camacho AG, Alonso JC (2000) Functional analysis of the terminase large subunit, G2P, of Bacillus subtilis bacteriophage SPP1. J Biol Chem 275:35311-35319. https://doi.org/10.1074/jbc.M004309200

Hajek AE, Morris EE, Hendry TA (2019) Context dependent interactions of insects and defensive symbionts: insights from a novel system in siricid woodwasps. Curr Opin Insect Sci. https://doi.org/10.1016/j.cois.2019.03.006

Hamilton PT, Peng F, Boulanger MJ, Perlman SJ (2016) A ribosome-inactivating protein in a *Drosophila* defensive symbiont. Proc Natl Acad Sci U S A 113:350-355. https://doi.org/10.1073/pnas.1518648113

Harrison E, Brockhurst MA (2017) Ecological and evolutionary benefits of temperate phage: what does or doesn't kill you makes you stronger. BioEssays. https://doi.org/10.1002/bies.201700112

Harumoto T, Lemaitre B (2018) Male-killing toxin in a bacterial symbiont of *Drosophila*. Nature. https://doi.org/10.1038/s41586-018-0086-2

Haselkorn TS (2010) The *Spiroplasma* heritable bacterial endosymbiont of *Drosophila*. Fly (Austin) 4:80-87. https://doi.org/10.4161/fly.4.1.10883

Haselkorn TS, Markow TA, Moran NA (2009) Multiple introductions of the *Spiroplasma* bacterial endosymbiont into *Drosophila*. Mol Ecol 18:1294-1305. https://doi.org/10.1111/j.1365-294X.2009.04085.x

Haselkorn TS, Watts TD, Markow TA (2013) Density dynamics of diverse *Spiroplasma* strains naturally infecting different species of *Drosophila*. Fly (Austin) 7:204-210. https://doi.org/10.4161/fly.25469
Herren JK, Paredes JC, Schüpfer F, Arafah K, Bulet P, Lemaitre B (2014) Insect endosymbiont proliferation is limited by lipid availability. Elife 3:e02964. https://doi.org/10.7554/eLife.02964
Higareda Alvear VM, Mateos M, Cortez D, Tamborindeguy C, Martinez-Romero E (2021) Differential gene expression in a tripartite interaction: *Drosophila*, *Spiroplasma* and parasitic wasps. PeerJ 9:e11020. https://doi.org/10.7717/peerj.11020
Humayun MZ, Zhang Z, Butcher AM, Moshayedi A, Saier Jr MH (2017) Hopping into a hot seat: Role of DNA structural features on IS5-mediated gene activation and inactivation under stress. PLoS ONE 12:e0180156. https://doi.org/10.1371/journal.pone.0180156
Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119
Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ (2010) Adaptation via symbiosis: recent spread of a *Drosophila* defensive symbiont. Science 329:212–215. https://doi.org/10.1126/science.1188235
Jones JE, Hurst GD (2020) Symbiont-mediated fly survival is independent of defensive symbiont genotype in the Drosophila melanogaster–Spiroplasma–wasp interaction. J Evol Biol 33:1625-1633. https://doi.org/10.1111/jeb.13702
Just W, Klotz G (1990) Terminal redundancy and circular permutation of mycoplasma virus L3 DNA. J Gen Virol 71 (Pt 9):2157-2162. https://doi.org/10.1099/0022-1317-71-9-2157
Kall L, Krogh A, Sonnhammer EL (2004) A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338:1027-1036. https://doi.org/10.1016/j.jmb.2004.03.016
Kall L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res 35:W429-432. https://doi.org/10.1093/nar/gkm256
Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722-736. https://doi.org/10.1101/gr.215087.116
Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567-580. https://doi.org/10.1006/jmbi.2000.4315
Lang AS, Beatty JT (2000) Genetic analysis of a bacterial genetic exchange element: the gene transfer agent of Rhodobacter capsulatus. Proc Natl Acad Sci U S A 97:859-864. https://doi.org/10.1073/pnas.97.2.859
Lang AS, Westbye AB, Beatty JT (2017) The distribution, evolution, and roles of Gene Transfer Agents in Prokaryotic genetic exchange. Annu Rev Virol 4:87-104. https://doi.org/10.1146/annurev-virology-101416-041624
Lang AS, Zhaxybayeva O, Beatty JT (2012) Gene transfer agents: phage-like elements of genetic exchange. Nat Rev Microbiol 10:472-482. https://doi.org/10.1038/nrmicro2802
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359. https://doi.org/10.1038/nmeth.1923
Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11-16. https://doi.org/10.1093/nar/gkh152
LePage DP, Metcalf JA, Bordenstein SR, On J, Perlmutter JI, Shropshire JD, Layton EM, Funkhouser-Jones LJ, Beckmann JF, Bordenstein SR (2017) Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. Nature 543:243-247. https://doi.org/10.1038/nature21391
Levene SD, Zimm BH (1987) Separations of open-circular DNA using pulsed-field electrophoresis. Proc Natl Acad Sci U S A 84:4054-4057. https://doi.org/10.1073/pnas.84.12.4054
Lindsey ARI, Rice DW, Bordenstein SR, Brooks AW, Bordenstein SR, Newton ILG (2018) Evolutionary genetics of cytoplasmic incompatibility genes *cifA* and *cifB* in prophage WO of *Wolbachia*. Genome Biol Evol 10:434-451. https://doi.org/10.1093/gbe/evy012

Loessner MJ, Inman RB, Lauer P, Calendar R (2000) Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of Listeria monocytogenes: implications for phage evolution. Mol Microbiol 35:324-340. https://doi.org/10.1046/j.1365-2958.2000.01720.x
Lopes A, Amarir-Bouhram J, Faure G, Petit MA, Guerois R (2010) Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. Nucleic Acids Res 38:3952-3962. https://doi.org/10.1093/nar/gkq096
Łoś J, Zielińska S, Krajewska A, Michalina Z, Małachowska A, Kwaśnicka K, Łoś M (2021) Temperate phages, prophages, and lysogeny. Bacteriophages: Biology, Technology, Therapy:119-150. https://doi.org/10.1007/978-3-319-41986-2_3
Luque A, Benler S, Lee DY, Brown C, White S (2020) The missing tailed phages: prediction of small capsid candidates. Microorganisms 8. https://doi.org/10.3390/microorganisms8121944
Maniloff J, Dybvig K (1988) *Mycoplasma* viruses. CRC Crit Rev Microbiol 15:339-389
Maniloff J, Dybvig K (2006) Mycoplasma phages. In: Calendar R, Abedon S (eds) The bacteriophages. Oxford University Press, Oxford, pp 636-652
Masson F, Calderon Copete S, Schüpfer F, Garcia-Arraez G, Lemaitre B (2018) In vitro culture of the insect endosymbiont *Spiroplasma poulsonii* highlights bacterial genes involved in host-symbiont interaction. mBio 9. https://doi.org/10.1128/mBio.00024-18
Mateos M, Castrezana S, Nankivell B, Estes A, Markow TA, Moran NA (2006) Heritable endosymbionts of *Drosophila*. Genetics 174:363–376. https://doi.org/10.1534/genetics.106.058818
Mateos M, Winter L, Winter C, Higareda-Alvear V, Martinez-Romero E, Xie J (2016) Independent origins of resistance or susceptibility of parasitic wasps to a defensive symbiont. Ecol Evol 6:2679–2687. https://doi.org/10.1002/ece3.2085
McNair K, Aziz RK, Pusch GD, Overbeek R, Dutilh BE, Edwards R (2018) Phage Genome Annotation Using the RAST Pipeline. Methods Mol Biol 1681:231-238. https://doi.org/10.1007/978-1-4939-7343-9_17
Merrill BD, Ward AT, Grose JH, Hope S (2016) Software-based analysis of bacteriophage genomes, physical ends, and packaging strategies. BMC Genomics 17:679. https://doi.org/10.1186/s12864-016-3018-2
Meyer RR, Laine PS (1990) The single-stranded DNA-binding protein of Escherichia coli. Microbiol Rev 54:342-380. https://doi.org/10.1128/mr.54.4.342-380.1990
Miller WJ, Ehrman L, Schneider D (2010) Infectious speciation revisited: impact of symbiont-depletion on female fitness and mating behavior of *Drosophila paulistorum*. PLoS Pathog 6:e1001214. https://doi.org/10.1371/journal.ppat.1001214
Montenegro H, Petherwick A, Hurst G, Klaczko L (2006) Fitness effects of *Wolbachia* and *Spiroplasma* in *Drosophila melanogaster*. Genetica 127:207–215. https://doi.org/10.1007/s10709-005-3766-4
Montenegro H, Solferini VN, Klaczko LB, Hurst GDD (2005) Male-killing *Spiroplasma* naturally infecting *Drosophila melanogaster*. Insect Mol Biol 14:281–288. https://doi.org/10.1111/j.1365-2583.2005.00558.x
Montllor CB, Maxmen A, Purcell AH (2002) Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress. Ecol Entomol 27:189-195
Moran NA, Plague GR (2004) Genomic changes following host restriction in bacteria. Curr Opin Genet Dev 14:627-633. https://doi.org/10.1016/j.gde.2004.09.003
Nielsen H (2017) Predicting Secretory Proteins with SignalP. Methods Mol Biol 1611:59-73. https://doi.org/10.1007/978-1-4939-7015-5_6
Oishi K (1971) Spirochaete-mediated abnormal sex-ratio (SR) condition in *Drosophila*: a second virus associated with spirochaetes and its use in the study of the SR condition. Genetics Research 18:45-56. https://doi.org/10.1017/S0016672300012404
Oishi K, Poulson D, Williamson D (1984) Virus mediated change in clumping properties of *Drosophila* SR *Spiroplasma*. Curr Microbiol 10:153-158. https://doi.org/10.1007/BF01576777
Oishi K, Poulson DF (1970) A virus associated with SR-spirochetes of *Drosophila nebulosa*. Proc Natl Acad Sci U S A 67:1565-1572. https://doi.org/10.1073/pnas.67.3.1565
Owen SV, Perez-Sepulveda BM, Adriaenssens EM (2021) Detection of bacteriophages: sequence-based systems. Bacteriophages: Biology, Technology, Therapy:621-644. https://doi.org/10.1007/978-3-319-41986-2_19

Paredes JC, Herren JK, Schüpfer F, Marin R, Claverol S, Kuo C-H, Lemaitre B, Béven L (2015) Genome sequence of the *Drosophila melanogaster* male-killing *Spiroplasma* strain MSRO endosymbiont. mBio 6:e02437-02414. https://doi.org/10.1128/mBio.02437-14

Perlmutter JI, Bordenstein SR, Unckless RL, LePage DP, Metcalf JA, Hill T, Martinez J, Jiggins FM, Bordenstein SR (2019) The phage gene *wmk* is a candidate for male killing by a bacterial endosymbiont. PLoS Pathog 15:e1007936. https://doi.org/10.1371/journal.ppat.1007936

Pool JE, Wong A, Aquadro CF (2006) Finding of male-killing *Spiroplasma* infecting *Drosophila melanogaster* in Africa implies transatlantic migration of this endosymbiont. Heredity 97:27-32. https://doi.org/10.1038/sj.hdy.6800830

Regassa LB, Gasparich GE (2006) Spiroplasmas: evolutionary relationships and biodiversity. Front Biosci 11:2983-3002

Roossinck MJ, Bazán ER (2017) Symbiosis: viruses as intimate partners. Annual Review of Virology 4:123–139. https://doi.org/10.1146/annurev-virology-110615-042323

Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-2069. https://doi.org/10.1093/bioinformatics/btu153

Shropshire JD, Kalra M, Bordenstein SR (2020a) Evolution-guided mutagenesis of the cytoplasmic incompatibility proteins: Identifying CifA's complex functional repertoire and new essential regions in CifB. PLoS Pathog 16:e1008794. https://doi.org/10.1371/journal.ppat.1008794

Shropshire JD, Leigh B, Bordenstein SR (2020b) Symbiont-mediated cytoplasmic incompatibility: what have we learned in 50 years? Elife 9:e61989. https://doi.org/10.7554/eLife.61989

Shropshire JD, On J, Layton EM, Zhou H, Bordenstein SR (2018) One prophage WO gene rescues cytoplasmic incompatibility in *Drosophila melanogaster*. Proc Natl Acad Sci U S A 115:4987-4991. https://doi.org/10.1073/pnas.1800650115

Shropshire JD, Rosenberg R, Bordenstein SR (2021) The impacts of cytoplasmic incompatibility factor (cifA and cifB) genetic variation on phenotypes. Genetics 217:1-13. https://doi.org/10.1093/genetics/iyaa007

Siguier P, Gourbeyre E, Varani A, Ton-Hoang B, Chandler M (2015) Everyman's Guide to Bacterial Insertion Sequences. Microbiol Spectr 3:MDNA3-0030-2014. https://doi.org/10.1128/microbiolspec.MDNA3-0030-2014

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32-36. https://doi.org/10.1093/nar/gkj014

Smit AFA, Hubley R, Green P (2013-2015) RepeatMasker Open-4.0.

Stone E, Campbell K, Grant I, McAuliffe O (2019) Understanding and Exploiting Phage-Host Interactions. Viruses 11. https://doi.org/10.3390/v11060567

Takemura-Uchiyama I, Uchiyama J, Kato S, Inoue T, Ujihara T, Ohara N, Daibata M, Matsuzaki S (2013) Evaluating efficacy of bacteriophage therapy against Staphylococcus aureus infections using a silkworm larval infection model. FEMS Microbiol Lett 347:52-60. https://doi.org/10.1111/1574-6968.12220

Valentine RC, Shapiro BM, Stadtman ER (1968) Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from Escherichia coli. Biochemistry 7:2143-2152. https://doi.org/10.1021/bi00846a017

van der Wilk F, Dullemans AM, Verbeek M, van den Heuvel JF (1999) Isolation and characterization of APSE-1, a bacteriophage infecting the secondary endosymbiont of Acyrthosiphon pisum. Virology 262:104-113. https://doi.org/10.1006/viro.1999.9902

Waldor MK, Mekalanos JJ (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. Science 272:1910-1914. https://doi.org/10.1126/science.272.5270.1910

Weldon SR, Strand MR, Oliver KM (2013) Phage loss and the breakdown of a defensive symbiosis in aphids. Proc R Soc Lond B Biol Sci 280:20122103. https://doi.org/10.1098/rspb.2012.2103

Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology. Nat Rev Microbiol 6:741-751. https://doi.org/10.1038/nrmicro1969

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

Yang M, Derbyshire MK, Yamashita RA, Marchler-Bauer A (2020) NCBI's Conserved Domain Database and Tools for Protein Domain Analysis. Curr Protoc Bioinformatics 69:e90. https://doi.org/10.1002/cpbi.90

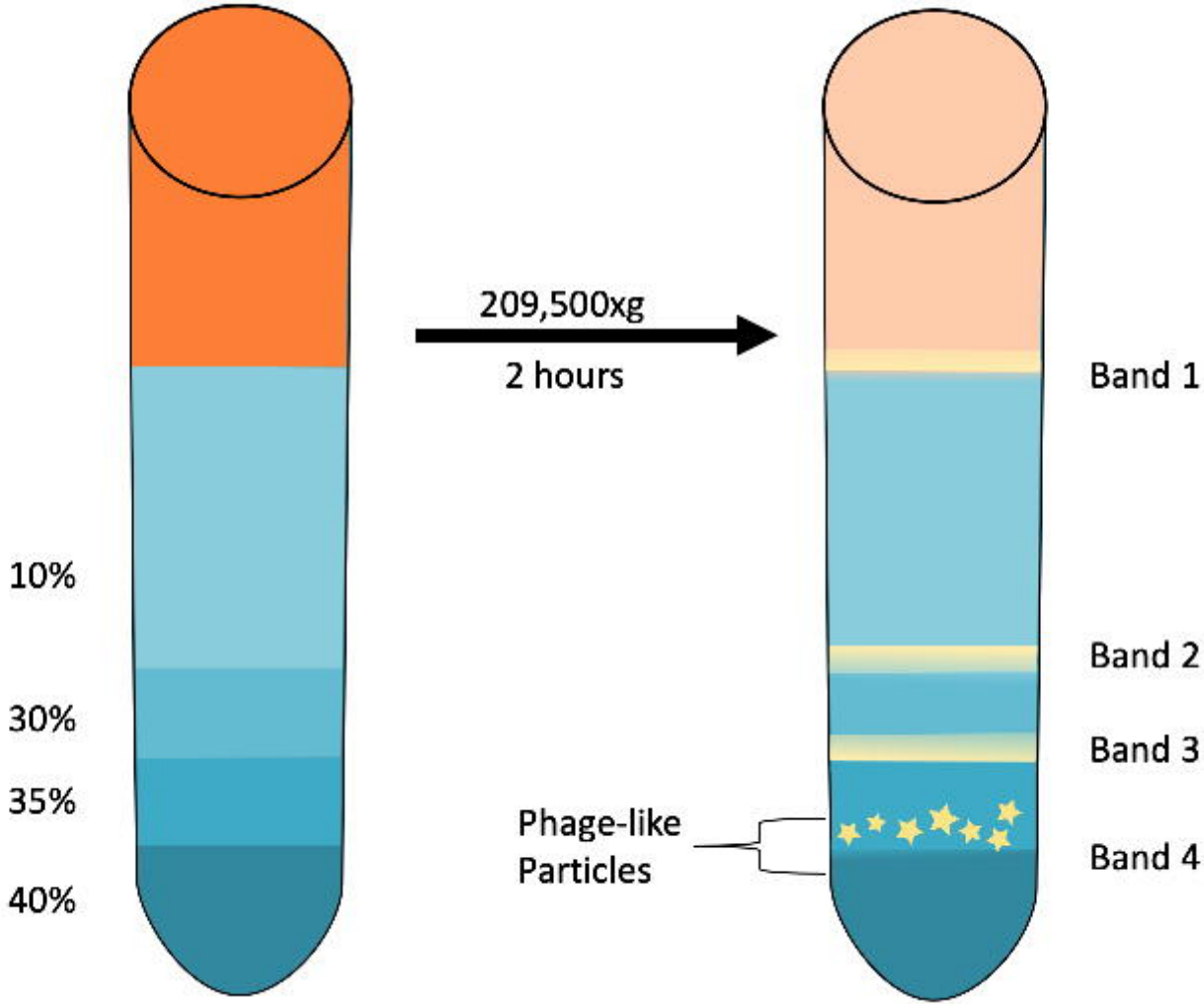
Yen HC, Hu NT, Marrs BL (1979) Characterization of the gene transfer agent made by an overproducer mutant of Rhodopseudomonas capsulata. J Mol Biol 131:157-168. https://doi.org/10.1016/0022-2836(79)90071-8

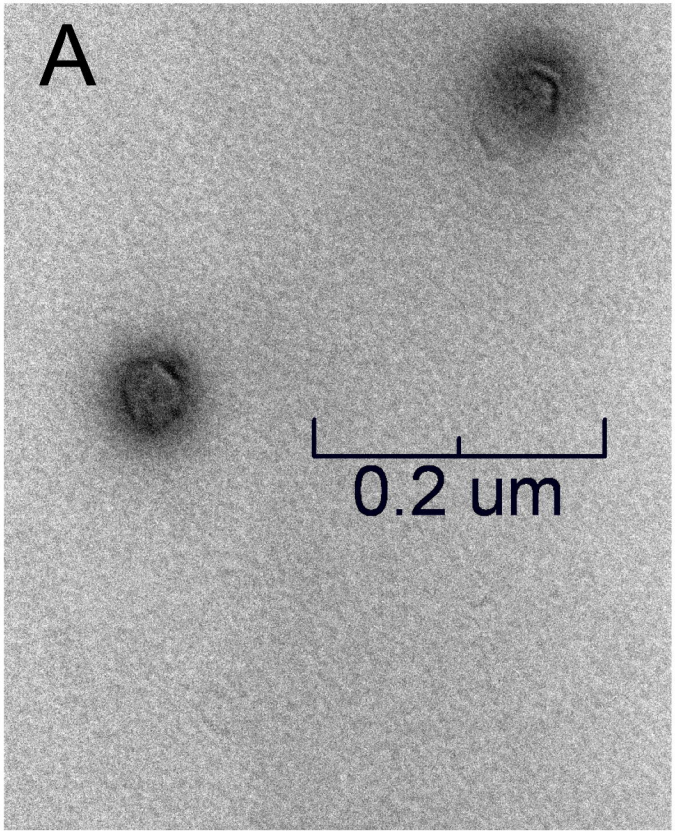
Yen JH, Barr AR (1971) New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens* L. Nature 232:657-658. https://doi.org/10.1038/232657a0

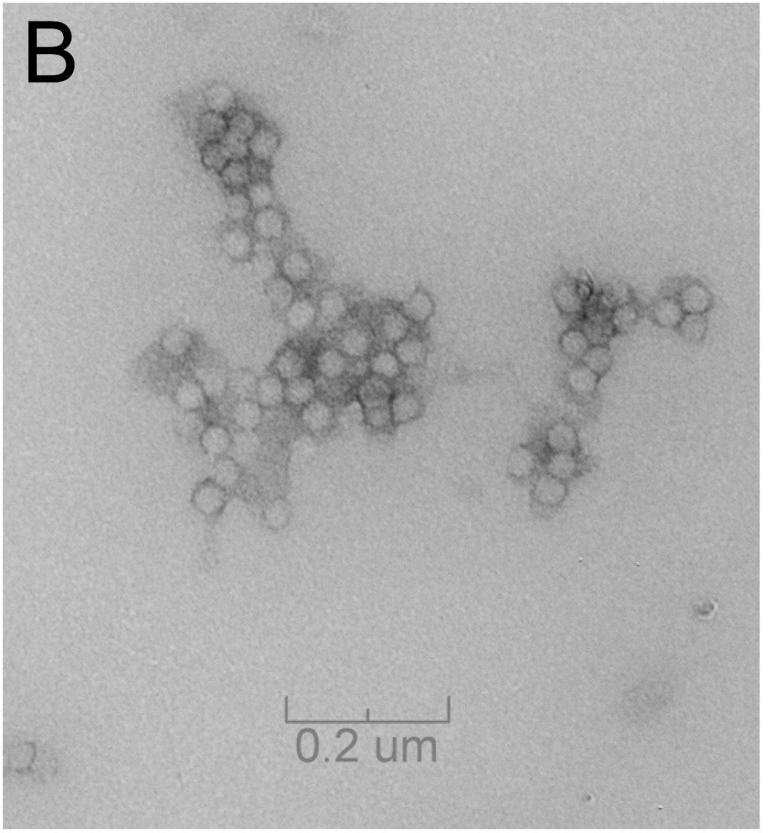
Zhou WG, Rousset F, O'Neill S (1998) Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. Proc R Soc Lond B Biol Sci 265:509–515. https://doi.org/10.1098/rspb.1998.0324

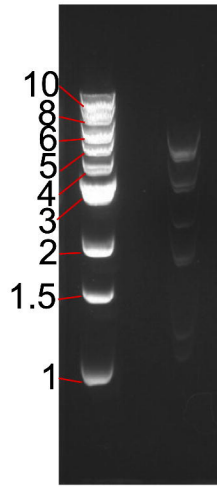
Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, Alva V (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol 430:2237-2243. https://doi.org/10.1016/j.jmb.2017.12.007

Zivanovic Y, Goulet I, Prunell A (1986) Properties of supercoiled DNA in gel electrophoresis: The V-like dependence of mobility on topological constraint. DNA-matrix interactions. J Mol Biol 192:645-660. https://doi.org/10.1016/0022-2836(86)90282-2


209,500xg
2 hours
10%
30%
35%
40%
Band 1
Band 2
Band 3
Band 4
Phage-like
Particles


A
0.2 um


B
0.2 um


10
8
6
5
4
3
2
1.5
1

A NSRO-P1

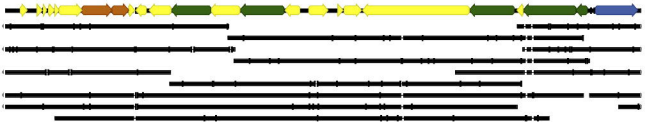
NSRO-P1

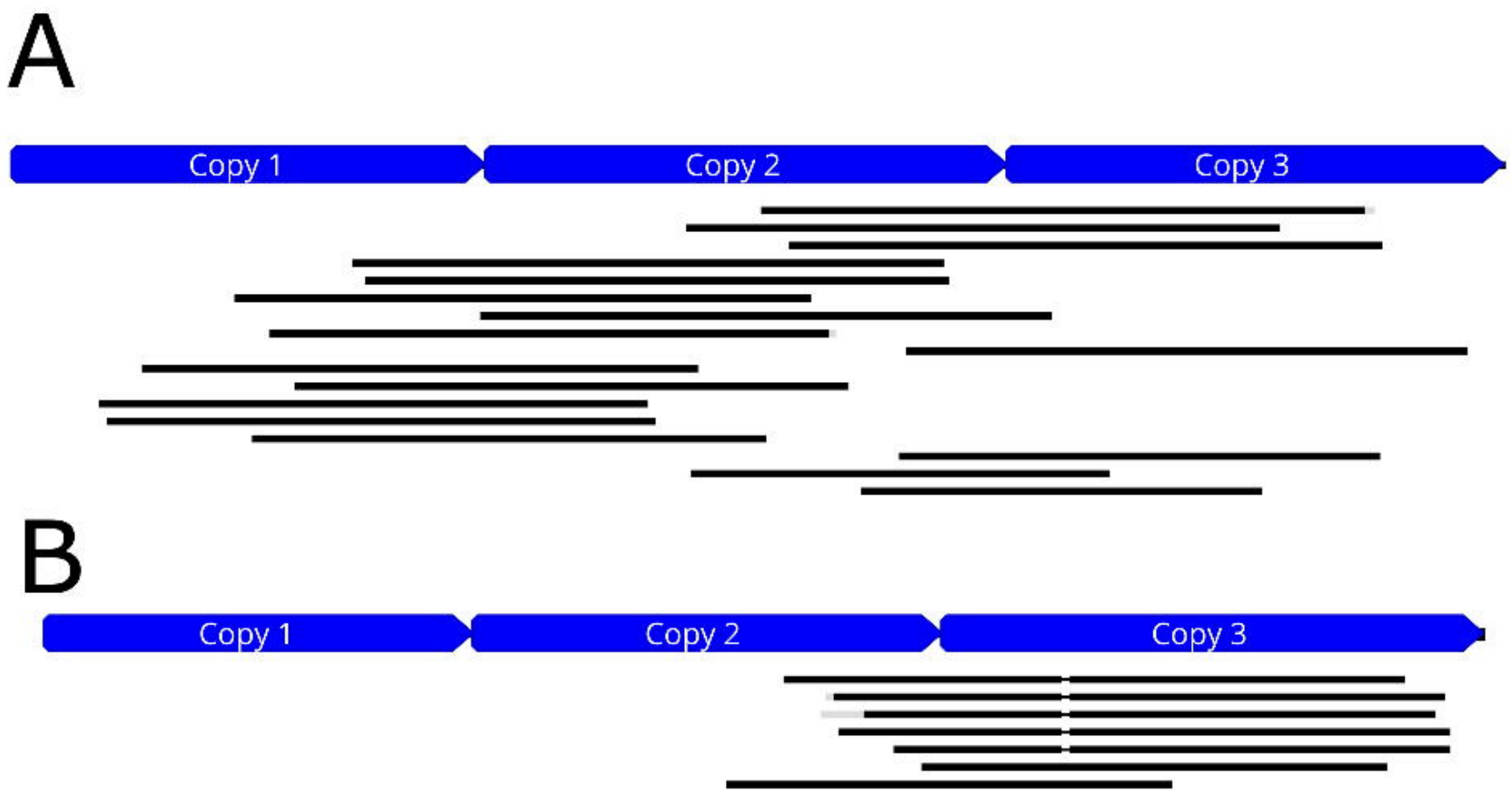
FWD 0ce8328d-9be8-459c-a79c-f3c5e3ef9ce0 (fragment 10506 -> 21028)
FWD 0ce8328d-9be8-459c-a79c-f3c5e3ef9ce0 (fragment 1 -> 10505)
FWD 1b052564-1761-41f6-92bd-5525a106c054 (fragment 1 -> 10534)
FWD 1b052564-1761-41f6-92bd-5525a106c054 (fragment 10535 -> 21081)
FWD 4c34ce76-9338-4a72-b955-6ea51c7605e3_1 (reversed) (fragment 1 -> 10319)
FWD 4c34ce76-9338-4a72-b955-6ea51c7605e3_1 (reversed) (fragment 10320 -> 207...
FWD 7f43a28e-88f7-49c7-9609-aef56840c607 (reversed) (fragment 1 -> 10363)
FWD 7f43a28e-88f7-49c7-9609-aef56840c607 (reversed) (fragment 10364 -> 20588)
FWD 8a7f9359-4764-43e0-a1e2-7fc2be68765a (fragment 1 -> 10123)
FWD 8a7f9359-4764-43e0-a1e2-7fc2be68765a (fragment 10124 -> 20296)
FWD 313b8580-946f-4246-9df4-1dcef2de3b05 (fragment 1 -> 10316)
FWD 313b8580-946f-4246-9df4-1dcef2de3b05 (fragment 10317 -> 20685)
FWD 4371caea-e3e3-4a13-afbb-1d80bbda274f (fragment 1 -> 10630)
FWD 4371caea-e3e3-4a13-afbb-1d80bbda274f (fragment 10631 -> 21273)
FWD 44338461-e843-41a9-9118-96671f977c03 (reversed) (fragment 1 -> 9531)
FWD 44338461-e843-41a9-9118-96671f977c03 (reversed) (fragment 9532 -> 19087)
FWD 67466420-9cfa-45f0-891e-a7ba308e990f (fragment 1 -> 10949)
FWD 67466420-9cfa-45f0-891e-a7ba308e990f (fragment 10950 -> 21964)
FWD a4be8860-6c94-4984-85ea-6cc79365fbd3 (fragment 1 -> 10806)
FWD a4be8860-6c94-4984-85ea-6cc79365fbd3 (fragment 10807 -> 21547)
FWD abc09d85-0146-4b6f-94f9-500e34166bae (reversed) (fragment 1 -> 10795)
FWD abc09d85-0146-4b6f-94f9-500e34166bae (reversed) (fragment 10796 -> 21565)
FWD dafd7268-6b2a-4ccd-8f41-9086655c3a08_1 (reversed) (fragment 1 -> 10492)
FWD dafd7268-6b2a-4ccd-8f41-9086655c3a08_1 (reversed) (fragment 10493 -> 21049)
FWD e28a928c-7993-42bd-a89f-5b0a4226dd18 (fragment 1 -> 10569)
FWD e28a928c-7993-42bd-a89f-5b0a4226dd18 (fragment 10570 -> 21052)
REV caa57e90-7799-4f5e-90aa-51934bb9faae

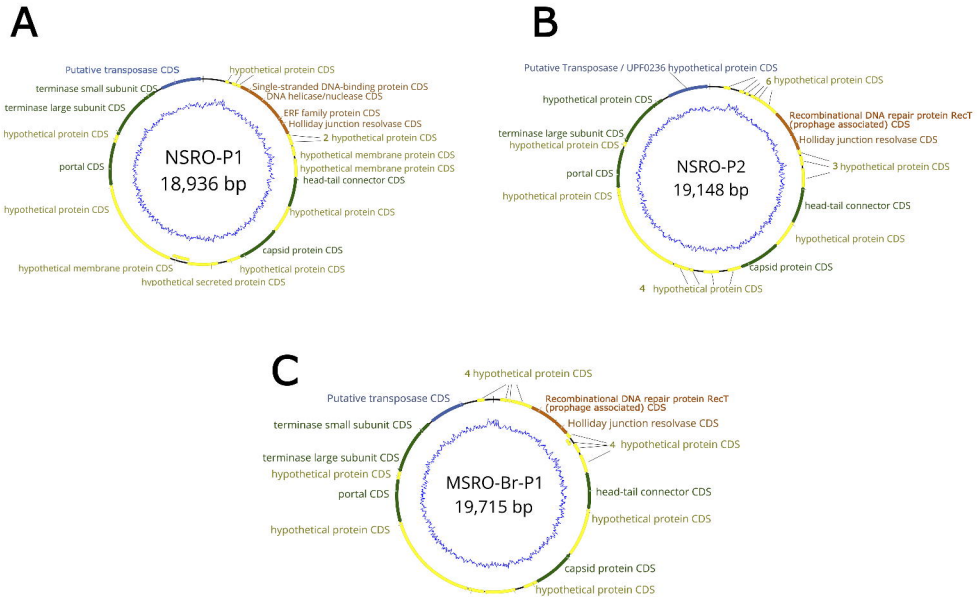
B NSRO-P2

NSRO-P2

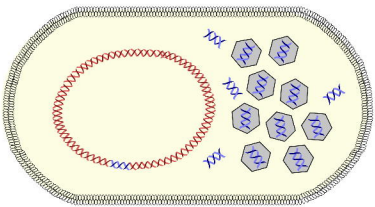
FWD 05b54b55-7f01-4cab-8403-13ba3f9aab8f (fragment 1 -> 10714)
FWD 05b54b55-7f01-4cab-8403-13ba3f9aab8f (fragment 10715 -> 21334)
FWD 8a057449-2e54-4f98-8f77-3f647516272e (fragment 1 -> 10900)
FWD 8a057449-2e54-4f98-8f77-3f647516272e (fragment 10901 -> 21926)
FWD d8d46348-880a-4aac-b49d-f364d2a0e66a (fragment 1 -> 10685)
FWD d8d46348-880a-4aac-b49d-f364d2a0e66a (fragment 10686 -> 21577)
REV dac7cf50-1f43-4735-9f83-22b5838cc23d
REV bdc8edaf-5074-4e75-9b16-89416aedaa3c
REV edad25ca-d1e3-4b84-af58-ce4e5a211309



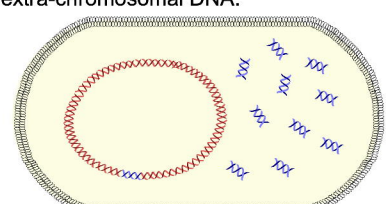
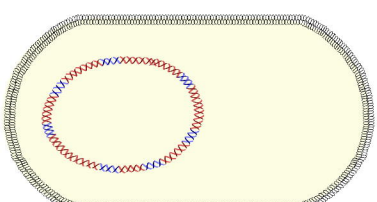

A
Copy 1
Copy 2
Copy 3
B
Copy 1
Copy 2
Copy 3


A
NSRO-P1
18,936 bp
Putative transposase CDS
hypothetical protein CDS
terminase small subunit CDS
Single-stranded DNA-binding protein CDS
DNA helicase/nuclease CDS
terminase large subunit CDS
ERF family protein CDS
Holliday junction resolvase CDS
hypothetical protein CDS
2 hypothetical protein CDS
hypothetical membrane protein CDS
portal CDS
hypothetical membrane protein CDS
head-tail connector CDS
hypothetical protein CDS
hypothetical protein CDS
capsid protein CDS
hypothetical membrane protein CDS
hypothetical protein CDS
hypothetical secreted protein CDS
B
NSRO-P2
19,148 bp
Putative Transposase / UPF0236 hypothetical protein CDS
6 hypothetical protein CDS
hypothetical protein CDS
Recombinational DNA repair protein RecT (prophage associated) CDS
terminase large subunit CDS
Holliday junction resolvase CDS
hypothetical protein CDS
3 hypothetical protein CDS
portal CDS
hypothetical protein CDS
head-tail connector CDS
hypothetical protein CDS
capsid protein CDS
4 hypothetical protein CDS
C
MSRO-Br-P1
19,715 bp
4 hypothetical protein CDS
Putative transposase CDS
Recombinational DNA repair protein RecT (prophage associated) CDS
terminase small subunit CDS
Holliday junction resolvase CDS
4 hypothetical protein CDS
terminase large subunit CDS
hypothetical protein CDS
portal CDS
head-tail connector CDS
hypothetical protein CDS
hypothetical protein CDS
capsid protein CDS
hypothetical protein CDS
hypothetical protein CDS
hypothetical protein CDS

Scenario 1. is abundant as extra-chromosomal elements, and is packaged into particles. Thus, its recovery reflects its presence in particles (as phage or Gene Transfer Agents)



Scenario 2. is abundant as an extra-chromosomal element, but not necessarily in particles. Thus, its recovery in the particle extraction reflects contamination from extra-chromosomal DNA.



Scenario 3. is abundant in the bacterial chromosome. Thus, its recovery in the particle extraction reflects contamination from chromosomal DNA.