Toxin and Genome Evolution in a *Drosophila* Defensive Symbiosis

Matthew J. Ballinger[1,]*, Ryan M.R. Gawryluk, and Steve J. Perlman

Department of Biology, University of Victoria, British Columbia, Canada

[1]Present address: Department of Biological Sciences, Mississippi State University, Mississippi State, MS

*Corresponding author: E-mail: ballinger@biology.msstate.edu.

Accepted: December 11, 2018

Data deposition: This project has been deposited in GenBank under BioProject numbers PRNJA492288 and PRNJA495838 and nucleotide accession numbers MK251237-MK251240.

Abstract

Defenses conferred by microbial symbionts play a vital role in the health and fitness of their animal hosts. An important outstanding question in the study of defensive symbiosis is what determines long term stability and effectiveness against diverse natural enemies. In this study, we combine genome and transcriptome sequencing, symbiont transfection and parasite protection experiments, and toxin activity assays to examine the evolution of the defensive symbiosis between *Drosophila* flies and their vertically transmitted *Spiroplasma* bacterial symbionts, focusing in particular on ribosome-inactivating proteins (RIPs), symbiont-encoded toxins that have been implicated in protection against both parasitic wasps and nematodes. Although many strains of *Spiroplasma*, including the male-killing symbiont (sMel) of *Drosophila melanogaster*, protect against parasitic wasps, only the strain (sNeo) that infects the mycophagous fly *Drosophila neotestacea* appears to protect against parasitic nematodes. We find that RIP repertoire is a major differentiating factor between strains that do and do not offer nematode protection, and that sMel RIPs do not show activity against nematode ribosomes in vivo. We also discovered a strain of *Spiroplasma* infecting a mycophagous phorid fly, *Megaselia nigra*. Although both the host and its *Spiroplasma* are distantly related to *D. neotestacea* and its symbiont, genome sequencing revealed that the *M. nigra* symbiont encodes abundant and diverse RIPs, including plasmid-encoded toxins that are closely related to the RIPs in sNeo. Our results suggest that distantly related *Spiroplasma* RIP toxins may perform specialized functions with regard to parasite specificity and suggest an important role for horizontal gene transfer in the emergence of novel defensive phenotypes.

Key words: *Spiroplasma*, ribosome-inactivating proteins, horizontal gene transfer, insect–microbe symbiosis, parasites, nematodes.

Introduction

Multicellular organisms commonly host microbial symbionts that provide protection against natural enemies, and it is now clear that these defensive symbionts play an important role in host ecology and evolution (Haine 2008; Vorburger and Perlman 2018). Two important and related questions in the field of defensive symbiosis are as follows: How stable is the defensive symbiosis over ecological and evolutionary time, and what is the specificity of protection provided by the symbiont? Both of these answers will depend to a great extent on the mechanism of symbiont-mediated protection, and one of the most common mechanisms is through symbiont-produced antimicrobials and toxins. Genomic and metabolomic studies have revealed a broad spectrum of toxin evolution patterns across defensive partnerships. For example, the antibiotic mixture produced by defensive streptomycetes of digger wasps has remained relatively stable over the lifetime of this ancient symbiosis, nearly 70 Myr (Kaltenpoth et al. 2005, 2014; Engl et al. 2018), where it provides broad-scale protection against pathogenic fungi in the soil. In contrast, the toxin that kills parasitic wasps in aphid hosts is exclusively phage-encoded, and neither the phage nor its bacterial host *Hamiltonella defensa* are fixed in aphid populations (Oliver et al. 2003, 2009; Brandt et al. 2017). In addition, phage and their highly variable toxin cassettes are routinely exchanged between symbionts (Degnan and Moran 2008), which is likely due to variation in toxin effectiveness against wasps.

In the genus *Drosophila,* multiple species harbor heritable strains of the bacterial symbiont, *Spiroplasma.* Symbioses between *Drosophila* and *Spiroplasma* are persistent and intimate at ecological timescales, but across evolutionary time they are characterized by horizontal transmission among unrelated hosts (Haselkorn et al. 2009). Interestingly, some closely related strains of *Spiroplasma* exhibit differences in their defensive capabilities. For example, the *Spiroplasma poulsonii* strains that infect *Drosophila melanogaster* (hereafter sMel), *Drosophila hydei*, and *Drosophila neotestacea* (sNeo) kill larval parasitoid wasps as they develop inside hosts (Xie et al. 2010, 2014; Haselkorn and Jaenike 2015); however, only sNeo can also protect hosts from sterilization by the parasitic nematode, *Howardula aoronymphium* (Haselkorn and Jaenike 2015). Despite this difference, a similar mechanism has been implicated in both defenses, with both nematodes and wasps showing evidence of attack by ribosome-inactivating proteins (RIPs), secreted toxins that include the well-known poisons, ricin and Shiga toxin. In both sMel and sNeo, RIPs comprise multicopy gene families (Ballinger and Perlman 2017). Whether there are important functional differences among the *Spiroplasma* RIP genes is unknown but copy number variation and sequence diversity between strains suggests dynamic evolutionary histories for each gene family that may relate to their biological roles.

Here, we use genome and transcriptome sequencing, symbiont transfection and parasite infection experiments, and wild fly screening to investigate the evolution of toxin-mediated defense and parasite specificity in the *Drosophila-Spiroplasma* system. When transferred into *D. neotestacea*, sMel is ineffective against the parasitic nematode, *H. aoronymphium* and, importantly, does not depurinate nematode ribosomes. We find that one of the few differences in the genomes of sNeo and sMel is their RIP toxin repertoire, with sNeo encoding RIPs that are absent from sMel's genome. One of these sNeo-specific RIPs is encoded on a plasmid. We also discovered a *Spiroplasma* strain infecting a species of phorid fly, *Megaselia nigra*, that shares mushroom breeding sites with *D. neotestacea.* We sequenced its genome, and found that although distantly related to *S. poulsonii*, it encodes an exceptionally diverse family of RIP genes, including plasmid-encoded RIPs that are closely related to those of sNeo. Our study highlights the importance of toxin gene repertoire in protection against multiple parasites and suggests horizontal gene transfer plays a pivotal part in expanding the defensive reach of heritable symbionts.

Materials and Methods

Symbiont, Insect, and Nematode Lines and Collections

The *Spiroplasma* symbionts in *D. neotestacea* and *D. melanogaster* are different strains of *S. poulsonii* and are referred to as sNeo and sMel (Ballinger and Perlman 2017), respectively. *Spiroplasma*-infected *D. neotestacea* were collected from mushroom baits in West Hartford, CT, USA in 2006 and maintained in the laboratory in vials containing mushrooms (*Agaricus bisporus*) and Carolina instant fly food. *Spiroplasma*-free flies were obtained by treating with tetracycline (Jaenike et al. 2010). Both of these lines were also previously cured of a *Wolbachia* symbiont infection via rifampicin treatment. Because of its male-killing phenotype, *s*Mel is also known as MSRO (Melanogaster Sex Ratio Organism; Montenegro et al. 2005; Harumoto and Lemaitre 2018). This strain was originally collected in *D. melanogaster* in Uganda, Africa (Pool et al. 2006) and was introduced into the Oregon-R strain of *D. melanogaster* by microinjection and provided to us by Bruno Lemaitre.

A line of *D. neotestacea* stably infected with the male-killing *Spiroplasma* from *D. melanogaster* was established via intrathoracic injection. Hemolymph was collected from *D. melanogaster* harboring sMel and 50.6 nl was injected intrathoracically into 3- to 5-day-old adult female *D. neotestacea.* Isolines were established from injected mothers and offspring were monitored. Because sMel is also a highly penetrant male-killer, even in *D. neotestacea* (Haselkorn and Jaenike 2015), lines were selected for sex ratio distortion, that is, vials that produced males were discarded. We therefore mated sMel-infected females to uninfected *D. neotestacea* males every generation to maintain the culture. All experiments performed with sMel in *D. neotestacea* involved flies infected with sMel for at least eight generations. *Howardula aoronymphium* was collected from West Hartford, CT, USA in 2006 and maintained in the laboratory in *Drosophila falleni*, as previously described (Hamilton et al. 2016).

Megaselia nigra was collected from mushroom baits in Victoria, British Columbia, Canada, in August 2016 and 2017. Species identification was made by Dr Emily Hartop at the Los Angeles Natural History Museum. Mitochondrial and nuclear DNA markers were amplified using PCR (see supplementary table S1, Supplementary Material online for PCR primers and reaction conditions) and Sanger sequenced (Sequetech, CA). The mitochondrial genes cytochrome oxidase subunit 1 and NADH dehydrogenase subunit 1 shared $\geq$99% nucleotide identity with *M. nigra* collected in Sweden (NCBI GenBank accession numbers for Victoria: MK251237 and MK251238; for Sweden: KX774871, KX774872, KX775048, KX775049). The nuclear marker, the 28S ribosomal RNA gene, was 99.8% identical (NCBI accession numbers: MK251239, KX529393, KX529394). *Megaselia nigra* individuals from Victoria were screened for *Spiroplasma* by PCR with primers targeting citri, poulsonii, and ixodetis clade taxa at several loci, including dnaA, ITS, and rpoB (supplementary table S1, Supplementary Material online). An 818-bp amplicon produced from citri/poulsonii dnaA primed reactions was Sanger sequenced and used for preliminary strain identification (GenBank accession number MK251240). This locus shares 99.2% nucleotide identity with *Spiroplasma phoeniceum* and 98.9% with *Spiroplasma kunkelii*. We subsequently

used the rpoB locus to compare this strain, which we refer to as sNigra, with others across the genus, as this marker is sequenced for most strains where genomes are unavailable. In particular, we wanted to examine its relationship to several strains of *Drosophila* symbionts for which limited sequences are available. We used PhyML with the GTR + I + G substitution model to perform a maximum-likelihood phylogenetic analysis from a nucleotide sequence alignment of rpoB gene sequences of 17 different *Spiroplasma* strains within the citri, poulsonii, and chrysopicola clades.

Transcriptional Response of sNeo to *Howardula* Infection

To examine differential sNeo gene expression during nematode parasitism, we performed nematode infection experiments coupled with short-read sequencing and analysis. *Drosophila neotestacea* flies harboring sNeo were incubated overnight on mushroom agar at 21 °C. Eggs were picked the subsequent morning in sets of 20 and placed onto mushroom (*A. bisporus*) wedges that were treated with either *D. falleni* homogenate containing ~400 *Howardula* infective juveniles in Ringer's solution (S^+HA^+ condition), or an equivalent volume of homogenate from uninfected *D. falleni* (S^+HA^- condition). A single mushroom wedge was placed onto moistened cheesecloth in each vial. Additional mushroom wedges (without nematodes) were added after depletion of the original mushroom by larval feeding.

Each adult fly was reared alone in a mushroom vial after eclosion. Female flies were aged for two days and homogenized in 100 µl of TRIzol (Invitrogen) with 10–15 1-mm silica-zirconium beads in a Mini Beadbeater 8 for 10 s. Trizol homogenates were snap-frozen in liquid nitrogen and stored at –80 °C until RNA extractions were performed. Total RNA was dissolved to 250 ng/µl in RNAse-free water and treated with DNAse I (Invitrogen) for 10 min at 37 °C, with 2 units/µl RNAseOut (Invitrogen). RNA was extracted again with 100 µl of Trizol and precipitated with ethanol. RNA was dissolved in RNAse-free water, and 500 ng of RNA was used for reverse transcription primed by random hexamers. The resulting cDNA was used to assess *Howardula* and sNeo infection status in all flies via qPCR (*D. neotestacea* positive control = rpl32; *Howardula* marker = 28S rRNA control region; *Spiroplasma* = RpoB). Total RNA from either three S^+HA^- or S^+HA^+ flies was pooled into a single replicate; four biological replicates were made for each condition (i.e., 24 flies total; 12 per condition). Flies were chosen from a wide variety of collection days and vials in order to reduce day and vial effects.

RNA was depleted of eukaryotic and bacterial rRNA with Ribo-Zero (Illumina) and used to construct strand-specific, paired end cDNA libraries (Genome Québec Innovation Centre). The four S^+HA^+ biological replicates were multiplexed onto a single Illumina lane (HiSeq 2500; 125-bp reads); S^+HA^- replicates were similarly multiplexed onto a single HiSeq2500 lane. Raw sequence reads were trimmed for quality and adaptor sequence with bbduk (Bushnell 2014) and reads ≥35 bp were retained. Trimmed reads were mapped to our sNeo genome assembly with bbmap, and featureCounts (Liao et al. 2014) was used to assign reads to genes predicted from our genome assembly with RASTtk (Brettin et al. 2015). Differential expression analyses were performed using DESeq2 (Love et al. 2014).

Howardula Protection Assays

sMel and sNeo-infected and uninfected *D. neotestacea* flies were exposed to *H. aoronymphium* as described above. Adult females were separated upon emergence and were aged as virgins for one week before their ovaries were dissected and all their mature eggs (at stage 10B or later) were counted.

Howardula Ribosome Depurination Assays

RNA was extracted from *Howardula*-infected and uninfected *D. neotestacea* larvae, pupae, and adults using TRIzol reagent (Invitrogen). RIPs are N-glycosidases that target an adenine residue in the α-sarcin/ricin loop of eukaryotic 28S rRNA. We used an established RT-qPCR based assay to quantify RIP activity (Melchior and Tolleson 2010; Pierce et al. 2011; Hamilton et al. 2016). In brief, the N-glycosidase activity of RIPs leaves an abasic site at the position of the cleaved adenine, and when reverse transcriptase encounters this position during cDNA synthesis, it incorporates a deoxyadenosine monophosphate (dAMP) into the nascent strand, while complementary DNA (cDNA) constructed from intact templates will receive a thymidine monophosphate (TMP). For detection of these targets by qRT-PCR, forward primers were designed such that the 3′-most primer position complements intact or depurinated variants of the RIP-targeted position, and a secondary mismatch was introduced at an adjacent position for both primer sets to improve their ability to distinguish between their similar sequence targets. The reverse primer targets a nearby region of 28S rRNA exhibiting high sequence divergence between *Howardula* and *Drosophila*. Primer sequences and validation details are presented in supplementary table S1, Supplementary Material online. For timecourse experiments, third instar larvae and pupae at day two and six were collected and RNA was extracted ($n = 6$ per time point). A second independent timecourse experiment was carried out to confirm the pilot results and examine additional time points, pupal days two, five, and eight ($n = 6$ per time point). A third independent exposure was carried out to collect coupled protection and depurination data; RNA was extracted from $n = 8$ two-day old adult flies per treatment whereas the rest were aged one week for ovary dissection as described above. Because sMel kills males, we used only female flies from the S-, sMel, and sNeo lines in this experiment. *Spiroplasma* rpoB was used as a marker to confirm *Spiroplasma* infection for all flies used in RIP assays. None of the flies from the sNeo-infected *D. neotestacea* line was

negative for rpoB. Two of the eight flies from the sMel-transinfected line were negative for *Spiroplasma* rpoB in the third experiment and were treated as *Spiroplasma*-free during data analysis. Fold changes in ribosomal RNA abundance were calculated according to the Pfaffl ratio method (Pfaffl 2001). Rather than comparing treated sample cycle threshold values (C_T) to those of controls, that is, *Spiroplasma*-infected to *Spiroplasma*-free, we subtracted the C_T of each sample from that of the global mean to plot data from S– (control) and S+ (treated) samples side-by-side. Primer efficiency (E) was incorporated into the calculations as $1 + E^{\Delta CT}$. The reference gene used to normalize depurinated ribosome abundance was a nearby region of the 28S rRNA that is not targeted by RIPs. Data were plotted in R version 3.3.3 using ggplot2.

sNeo and sNigra *Spiroplasma* DNA Extraction, Sequencing, and Genome Assembly

DNA extraction and sequencing methods used to generate the sNeo genome have been described previously (Ballinger and Perlman 2017). Briefly, sNeo-infected *D. neotestacea* adults were aged for three weeks before ovaries were dissected from 40 individuals and pooled for DNA extraction. Tissue was homogenized with a pestle in a 1.5 ml microfuge tube containing lysis buffer. DNA was isolated using the phenol-chloroform method. Total nucleic acids were treated with RNase A (Qiagen) and repurified by phenol-chloroform and ethanol precipitation. Pacific Biosciences long-read libraries were prepared by Genome Quebec and sequenced on three Pacific Biosciences RS II SMRT cells. Paired-end Illumina 125 bp libraries were prepared by Genome Quebec using the NEBNext Ultra II kit (New England Biolabs) and sequenced on 0.4 of a HiSeq 2500 sequencing lane. All Pacific Biosciences subreads were mapped to the *S. poulsonii* strain MSRO (sMel) genome assembly (v2) (Masson et al. 2018) and the previous version of the sNeo assembly (Ballinger and Perlman 2017) using bwa mem (Li and Durbin 2010). Short reads were mapped to the same references using bbmap (Bushnell 2014). Unicycler was used to assemble the genome using short and long read sets (Wick et al. 2017). The genome has been submitted to NCBI under BioProject PRNJA492288.

Genomic DNA was extracted from *Spiroplasma*-infected *M. nigra* using similar methods. Fifteen individual *M. nigra* flies collected in August of 2017 were screened for sNigra by qPCR using primers targeting the internal transcribed spacer region of ribosomal RNA (supplementary table S1, Supplementary Material online). Eleven of the fifteen were positive and were pooled. Library preparation for the sNigra genome was done by Genome Quebec and 125 bp paired-end reads were sequenced on 0.33 of a HiSeq 2500 sequencing lane. To assemble sNigra genomic DNA from 114 million mixed genomic reads of *M. nigra* and its associated microbes, we performed two cycles of genome assembly. Low-quality and adaptor sequences were trimmed from the complete read set, and reads with GC content above 31% were removed with bbduk (Bushnell 2014). The remaining low-GC reads were assembled with SPAdes 3.10.1 (Bankevich et al. 2012), resulting in a large assembly of ~484,000 contigs. Contigs 500 nucleotides or greater in length were submitted to the BusyBee web server for binning analysis with taxonomic annotation (Laczny et al. 2017). BusyBee estimated boundaries for ten bins, one of which was composed of *Spiroplasma*, *Mycoplasma*, and *Plectrovirus* sequences (supplementary fig. S1, Supplementary Material online). A second round of read processing was carried out using these contigs, as well as the genomic and plasmid sequences of *S. kunkelii* and *Spiroplasma citri*, to retain all mapping reads from the adapter-trimmed read set. Assembly was carried out with SPAdes. The resulting assembly was further curated by predicting and translating ORFs, then searching protein sequences against the nr protein database available on NCBI using BLAST. Nucleotide contigs that encoded a protein-coding gene for which the best hit by BLAST was to *Spiroplasma* were retained as the draft genome of sNigra. The genome assembly has been submitted to NCBI under BioProject PRJNA495838.

Comparative Gene Content and Genome Completeness

The draft genomes of sNeo and sNigra were uploaded to the KEGG Automatic Annotation Server v 2.1 (Moriya et al. 2007) and KEGG orthologs were identified by bidirectional best BLAST hit to the representative gene data set for prokaryotes. To generate comparable data sets for sMel and *S. kunkelii*, we also uploaded the complete genome sequences for these strains and performed the analysis using the same parameters. KO numbers among strains were compared and visualized using Venny 2.1 (available online at http://bioinfogp.cnb.csic.es/tools/venny/; last accessed September 24, 2018).

Prior to BUSCO analysis of the sNeo and sNigra genomes, genes were annotated using RAST in RASTtk mode (Brettin et al. 2015). All protein coding genes were extracted and translated, then assessed by BUSCO v3 (Simão et al. 2015) in protein mode. This program identifies single-copy orthologs between an input data set and a taxon-specific database. To determine how many of the conserved orthologs in the Tenericutes database should be expected in a *Spiroplasma* genome, we analyzed the complete genomes of sMel, *S. citri*, and *Spiroplasma apis*, which encoded 161, 160, and 160 of the 166 single-copy orthologs genes in the database, respectively. Therefore, we used 161 as the denominator to calculate estimated genome completeness for sNeo and sNigra.

Estimating Copy Number of Putative Plasmid Genes

Abundance of sNigra RIP3 and RIP6 was measured in seven of the individual extractions that had been pooled for genome sequencing by qPCR using primers designed for each RIP gene and *rpoB* as a normalizer. Primer sequences and validation

details are found in supplementary table S1, Supplementary Material online. As ΔC_T, we used the C_T difference between each RIP and the *rpoB* locus for two technical replicates per sample and calculated copy number as $1 + E^{\Delta CT}$.

Results

Transcriptome of the Nematode-defensive *Spiroplasma*, sNeo

We sequenced transcripts of sNeo in *H. aoronymphium*-infected (HA+) and uninfected (HA–) adult *D. neotestacea* as a potential avenue toward identification of previously undescribed genes involved in defense. In sum, only 0.7% of ~420 million paired-end reads from HA+ and HA– mapped to sNeo genes, reflective of a low overall sNeo titer. Although we recovered reads derived from 1,658 sNeo genes (with an average of ≥5 transcripts per million), there were only eight transcripts that were significantly more abundant in HA+ hosts. Four of these are hypothetical proteins, three encode conserved proteins involved in RNA-binding, sugar transport, or oxidative stress response, and one is a RIP (RIP3). Three of the hypothetical proteins are encoded adjacent to each other in the sNeo genome and are among the most highly expressed sNeo genes (supplementary table S2, Supplementary Material online); aside from RIP3, these three transcripts represent the only secreted proteins more abundant in the HA+ condition. Homologs could be identified for two of the hypothetical proteins in diverse *Spiroplasma* species, but two others—both secreted—lack clear homologs in other species. Although only one RIP was found to be more highly transcribed in response to HA infection, all four of the RIPs encoded by sNeo were transcribed at high levels, on par with abundant housekeeping genes. Other previously described toxin-encoding genes, including two putative chitinases and two putative epsilon toxins, were also robustly transcribed but at reduced levels in comparison with RIPs. Given that this analysis did not strongly implicate any novel *Spiroplasma* effectors, we focused the remaining experiments on examining the role of RIPs in parasite-specific defense.

Nematode Defense and Depurination Is *Spiroplasma* Strain Specific

As had been previously shown (Haselkorn and Jaenike 2015), sMel does not provide protection again nematodes. *Drosophila neotestacea* that were stably infected with sMel are sterilized by *H. aoronymphium* nematodes, whereas flies infected with their native sNeo symbiont are not (fig. 1*A*).

To ask whether RIP activity also differs between strains in accordance with the pattern of protection, we used a highly sensitive qRT-PCR assay to measure the hallmark of RIP-mediated inactivation, cleavage of a single adenine base from the α-sarcin/ricin loop (SRL) of *Howardula* 28S rRNA. Although there was an abundance of depurinated parasite ribosomes in flies harboring sNeo (fig. 1*B*; Tukey test, $P < .001$), we found no signal of parasite ribosome depurination in flies harboring sMel relative to S-flies (fig. 1*B*; Tukey test, $P = 0.14$). We previously reported that RIP activity on host ribosomes localized in the hemolymph creates a signal of background host depurination (Ballinger and Perlman 2017), and in sMel-infected *D. neotestacea*, we were able to detect this, indicating that sMel RIPs are expressed and active (supplementary fig. S2, Supplementary Material online). The lack of any *Howardula* ribosome depurination by sMel suggests that their RIPs do not gain entry into *Howardula* cells.

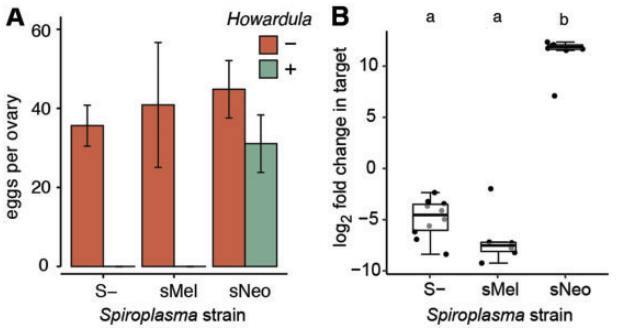

Fig. 1.—Nematode protection and RIP activity is *Spiroplasma* strain-specific in *Drosophila neotestacea*. (*A*) Development of eggs in ovaries of *Howardula*-parasitized and unparasitized *D. neotestacea* was examined as a proxy for *Spiroplasma* defense. In the absence of nematodes, egg development was normal for all unparasitized females (Tukey test). Among flies harboring symbionts, only those with sNeo exhibited rescued fecundity (though not completely restored relative to sNeo HA-females; Welch's *t*-test $t_{17.91} = 4.148$, $P < 0.001$) whereas those harboring sMel were no more protected than *Spiroplasma*-free controls. (*B*) Evidence of in vivo RIP activity on *Howardula* 28S ribosomal RNA by each *Spiroplasma* strain during *Howardula* infection. The RIPs of sMel do not depurinate *Howardula* ribosomes at the sarcin-ricin loop.

We also examined the timing of sNeo RIP attack during nematode infection by quantifying nematode depurination in infected fly larvae, pupae, and adults. Infective juvenile nematodes infect second to third instar larval *D. neotestacea*, but interestingly, RIPs do not depurinate nematode ribosomes at these early stages, immediately following their exposure to S+ host hemolymph. Rather, depurination is delayed until the host pupal stage (fig. 2). We repeated this experiment a second time to confirm the onset of depurination is delayed until the second day of host pupation (fig. 2). This is markedly different to RIP activity in wasps, where depurination is detected upon wasp hatching.

Comparative Genomics of sNeo and sMel

We assembled the genome of the *Spiroplasma* symbiont of *D. neotestacea* (sNeo) from Illumina and Pacific Biosciences sequencing reads using Unicycler and compared it with the *D. melanogaster*-infecting strain (sMel). The sNeo genome assembly is 1.78 Mb with an N50 of 97.8 kb (table 1) and

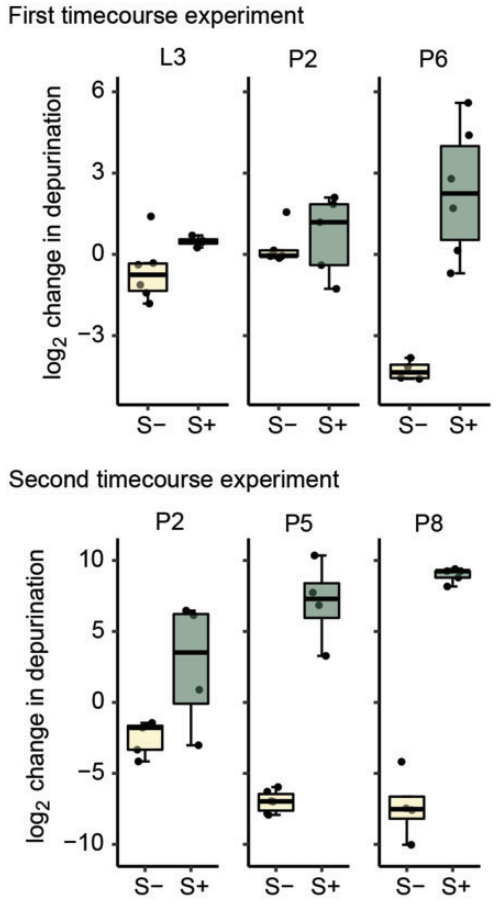

Fig. 2.—Delayed onset of ribosome depurination in nematodes. Levels of depurinated nematode ribosomes during infection of sNeo-free and infected *Drosophila neotestacea* across three developmental timepoints in a late larval stage and throughout pupation. In each of two experiments, modest levels of depurinated ribosomes are detected on and before approximately the second day of host pupation, followed by a significant increase in abundance over the subsequent 4–6 days.

BUSCO estimates its gene content completeness at 98%. The vast majority of the genome (1.67 Mb) is encoded on 24 large contigs (>10 kb); most of the remaining 263 contigs are short (≪10 kb). Four contigs were small, closed circles—potential plasmids or phage genomes—but only one of them encodes genes with strong sequence similarity to plasmid genes, and we refer to this putative plasmid as psNeo. As has been described previously for citri clade *Spiroplasma*, insertion sequence (IS) family transposases and phage comprise sizeable fractions of the chromosomal genome (Carle et al. 2010; Lo et al. 2013; Paredes et al. 2015). Approximately 10% of the genome was IS family transposase (0.88% IS3, 9.21% IS5) and 2.47% could be attributed to *Spiroplasma* phages SVGII3 and 1-R8A2B. Gene prediction by RAST yielded 2,142 putative protein-coding genes, 31 tRNAs and 3 ribosomal RNA genes. Of the protein-coding genes, 101 had no detectable homolog encoded by sMel, but 90 of those did have homologs in other strains of *Spiroplasma* and are thus unlikely to be genes involved in defense specificity. General biological functions could be assigned to 40 of them (supplementary fig. S3, Supplementary Material online). The most common functional categories for these genes were metabolism, membrane transport, and conjugation. Of the 11 remaining, no significant matches to any organism or domain was found by BlastP or HMMER. As these are among the shortest of the predicted annotations (8 of the 11 are <45 amino acids in length), it is also possible that some do not encode functional proteins. A list of all 101 genes found in this analysis is available in supplementary table S3, Supplementary Material online.

Divergent sNeo RIPs Were Recently Acquired by Plasmid Exchange

We examined the draft genome assembly of sNeo to gain insight into the evolutionary history of the RIPs. Both of the two RIPs private to the sNeo strain are flanked by genes that encode type IV secretion system component proteins involved in conjugative plasmid transfer, whereas the other RIPs in sNeo and sMel are not. In particular, sNeo-RIP2, encoded on the predicted plasmid psNeo, is flanked by sequence identical in gene content and coding structure to the *S. citri* plasmid, pSci2, excepting that psNeo also encodes a RIP and a hypothetical protein that displays low domain similarity to the epsilon pore-forming toxin of *Clostridium perfringens*. A homologous epsilon toxin is also encoded in the genome of sMel, but it contains a stop codon in the N-terminus; therefore, it is unclear if a functional protein is produced from the truncated reading frame. Beyond these two strains, the epsilon-like toxin has not been found in any other sequenced *Spiroplasma* genomes. The RIP2-encoding plasmid is present as a complete, closed circle with perfect nucleotide identity in the overlapping termini of the linear contig. The previous version of the genome contained three insertion/deletion mismatches in this overlap which we confirmed were errors by Sanger sequencing. The absence of orthologous RIPs in sMel and their association with plasmid sequences similar to pSci2 strongly suggests that the two divergent RIPs of sNeo were recently acquired by horizontal transfer from an unknown strain of *Spiroplasma*.

Spiroplasma in a Mushroom-feeding Phorid Fly, *M. nigra*

We found that *M. nigra*, a species of phorid fly that commonly visits the same mushroom baits as *D. neotestacea* in Victoria, harbors a *Spiroplasma* symbiont at high frequency. We found that 77 of 96 and 11 of 15 *M. nigra* adults collected at mushroom baits in August 2015 and August 2017, respectively, were positive for *Spiroplasma*. This strain of *Spiroplasma* (sNigra) is not closely related to sNeo and sMel; phylogenetic analysis places it in the citri clade, closely related to the insect-vectored plant pathogens *S. kunkelii* and *S. phoeniceum* (fig. 3*A*).

Table 1
Genome Assembly Statistics and Gene Content of *Spiroplasma*

Genome	Size (bp)	Contigs	GC%	Contig N50	CDS	Plasmids	Reference
sNeo	1,783,629	263	26.37	97,866	2,142	1	This study
Spiroplasma poulsonii	1,883,005	1	26.4	—	2,868	1	Paredes et al. (2015), Masson et al. (2018)
Spiroplasma kunkelii	1,463,926	1	24.97	—	1,646	4	Davis et al. (2015)
sNigra	1,397,726	329	26.4	7,570	1,775	N.D.	This study

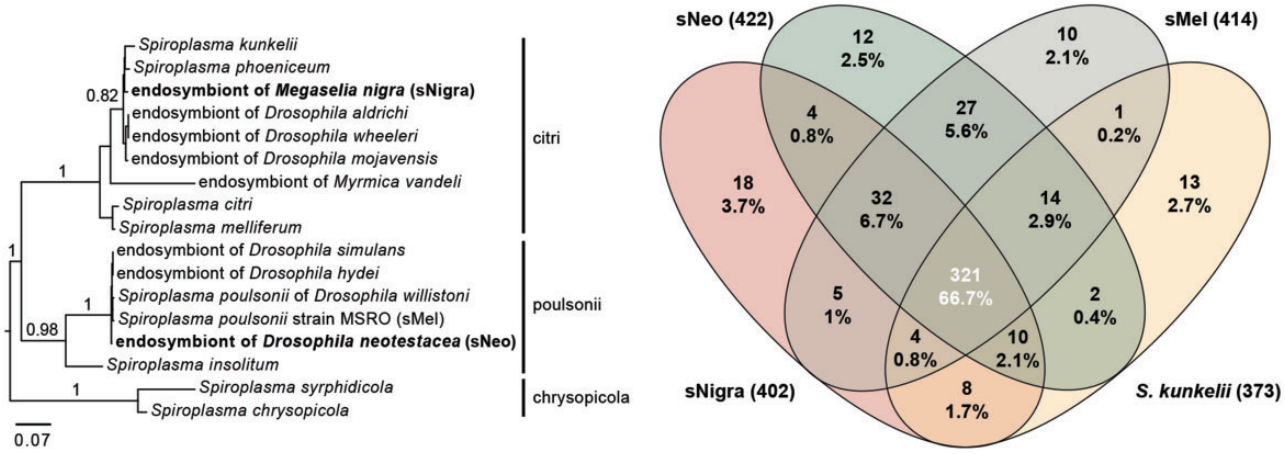

Fig. 3.—Evolutionary relationships and comparative gene content of a *Spiroplasma* in the phorid *Megaselia nigra*. (*A*) A phylogram constructed from aligned nucleotide sequences of the gene rpoB shows the *Spiroplasma* strain discovered in *M. nigra* (sNigra) is most closely related to citri clade taxa, whereas the defensive symbionts of *Drosophila melanogaster* (sMel) and *Drosophila neotestacea* (sNeo) are in the sister clade, poulsonii. Branches are labeled with SH-like approximate likelihood ratio scores > 0.75. The three main clades of *Spiroplasma* shown on the phylogeny are designated above branches and the tree is rooted with the apis clade taxa. (*B*) A Venn diagram compares conserved orthologous genes present in sMel, sNeo, sNigra, and *Spiroplasma kunkelii*. Upper numbers display the number of genes that overlap in each comparison, lower numbers indicate the percentage of the data set represented by the upper number. Numbers in parentheses next to strains display the number of KEGG orthologous genes from each that were identified and included in the analysis.

The Genome of sNigra Encodes Diverse RIPs

Our draft genome of sNigra assembled as 329 contigs totaling 1.4 Mb, with an N50 of 7,570 bp and 26.4% GC-content (table 1). RAST predicted 1,775 protein coding genes, 3 ribosomal RNA genes, and 18 tRNAs. BUSCO estimated its gene content completeness at 98%. The longest contig, 39,039 bases in length, could be aligned to a 34 kb region of synteny with the *S. kunkelii* strain CR2–3x genome, confirming the similarity we reported using the *rpoB* locus (94.8% identity across 34 kb of protein-coding and noncoding sequence). Assignment and comparison of KEGG orthology numbers for sNigra, sNeo, sMel, and *S. kunkelii* returned the greatest number of conserved bacterial orthologs in sNeo (422) and the fewest in *S. kunkelii* (373). About sixty-seven percent of genes in the data set were identified in all four strains (fig. 3*B*). Despite high sequence similarity between sNigra and *S. kunkelii*, there are more than twice as many conserved orthologs shared between sNigra and the poulsonii strains (sMel and sNeo) than between *S. kunkelii* and those strains, probably due to host and lifestyle shifts.

Using the TBlastN algorithm, we identified 8 putative RIPs in the genome of sNigra. Phylogenetic analysis revealed that sNigra encodes at least one RIP in each of the three major clades of sNeo and sMel RIPs, with notable enrichment of diversity in RIPs that are closely related to sNeo RIPs 1 and 2; we refer to the clade encompassing these RIPs as the mycophagous fly clade RIPs, or McRIPs (fig. 4*A*). We characterized the genomic architecture surrounding the McRIPs in sNigra and found that the two copies that show greatest sequence similarity to the McRIPs in sNeo, RIPs 5 and 6, are flanked by genes involved in conjugative transfer and plasmid maintenance, similar to sNeo RIPs 1 and 2. sNigra RIP6, which is closely related to sNeo RIPs 1 and 2, is flanked by plasmid-derived sequences identical in gene content and coding structure to the plasmid sequence flanking sNeo RIP2 (fig. 4*B*). In addition to the RIP, it encodes a homolog of the putative epsilon toxin found directly upstream of RIP1 and RIP2 in sNeo. We used qPCR to determine whether the abundance of sNigra RIP6 DNA among individual hosts displays variation consistent with its location on an extrachromosomal plasmid.

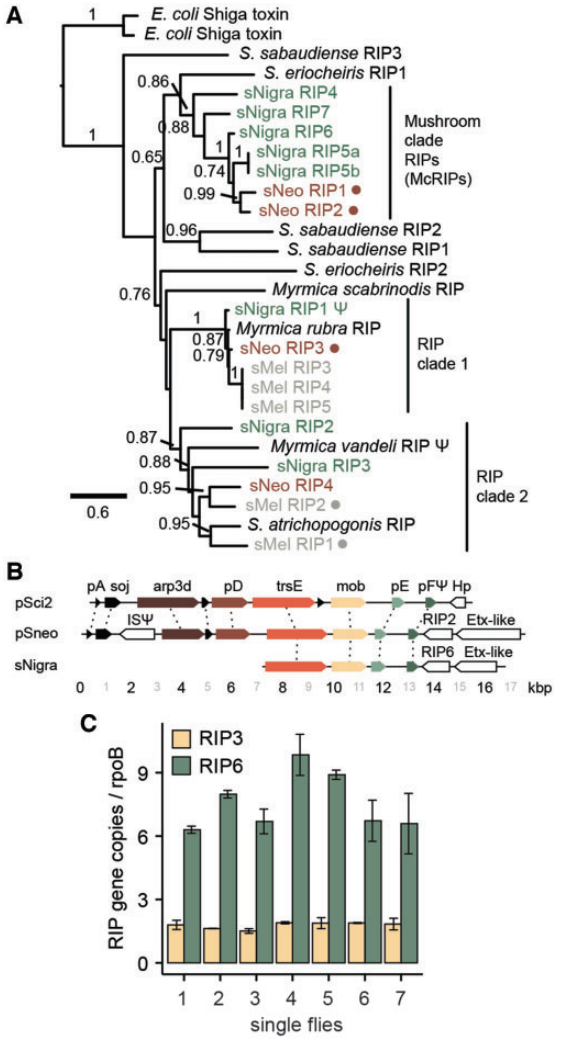

Fig. 4.—Diversity, genomic synteny, and copy number variation of *Spiroplasma* RIPs. (*A*) A midpoint-rooted phylogram constructed from aligned amino acid sequences of *Spiroplasma* RIPs shows evidence of duplications, losses, and horizontal transfer. Branches are labeled with SH-like approximate likelihood ratio scores > 0.65. Clade designations are shown to the right of taxon labels. Filled circles are used to label RIPs shown to have the highest transcript abundance in larvae and pupae (Ballinger and Perlman 2017). (*B*) Diagram showing the coding structure of plasmids from *Spiroplasma citri* (pSci2), the *Spiroplasma* endosymbiont of *Megaselia nigra* (sNigra), and the defensive *Spiroplasma* symbiont of *Drosophila neotestacea* (sNeo). Genes involved in conjugation and plasmid maintenance that are present in all compared sequences are labeled and marked by color-filled arrows on the linear diagrams, whereas genes that are not present on all three sequences are outlined arrows. RIPs and epsilon toxins (Etx-like) are present only in the insect symbiont strains and are also labeled. Hp, hypothetical protein. Numbers below indicate sequence length in kilobases. (*C*) In sNigra, RIP6 (green) DNA copy number is elevated and exhibits variable abundance relative to the *Spiroplasma* housekeeping gene *rpoB* among individual flies, consistent with plasmid localization, whereas a second RIP (yellow), exhibits similar DNA copy number and no interhost variability relative to *rpoB*, consistent with a chromosomal location.

This gene is 6–10 times more abundant than the single copy *Spiroplasma* gene *rpoB*; its abundance relative to *rpoB* is highly variable among individuals compared with that of a putatively chromosomally encoded single copy sNigra RIP, RIP3 (fig. 4*C*).

Discussion

In this study, we provide evidence that the ability of defensive *Spiroplasma* symbionts to protect against different classes of natural enemy is largely driven by different toxin repertoires acquired via horizontal transfer. Although they are very closely related, the inherited *S. poulsonii* symbionts that infect *D. neotestacea* and *D. melanogaster* exhibit very different ecologies. The neotestacea strain (sNeo) does not manipulate reproduction but is the only *Spiroplasma* that has been shown to defend against parasitic nematodes (Haselkorn and Jaenike 2015); its high prevalence in *D. neotestacea* in nature can be explained by the strong fitness benefit it provides to its host against the common virulent parasite *H. aoronymphium* (Jaenike et al. 2010; Cockburn et al. 2013). Although both strains can protect against parasitic wasps (Xie et al. 2010; Haselkorn and Jaenike 2015; Ballinger and Perlman 2017), there is no data yet showing that this occurs in nature. The fact that the melanogaster strain (sMel) is found at very low prevalence in the wild (Pool et al. 2006) suggests that its frequency is largely driven by its male-killing phenotype.

The genomes of sNeo and sMel revealed few differences in gene content. Interestingly one of the notable differences was the presence of two divergent ribosome RIPs, sNeo RIPs 1 and 2, that are recently acquired from or encoded on plasmids, respectively. We previously implicated RIPs in protection (Hamilton et al. 2014, 2016; Ballinger and Perlman 2017), with both nematode and wasp ribosomes exhibiting the hallmark of RIP attack—depurination at a specific residue in the α-sarcin-ricin loop of 28S rRNA. In agreement with previous work (Haselkorn and Jaenike 2015), we too found that sMel is unable to protect against nematodes and our experiments add that *Howardula* ribosomes are not depurinated in *D. neotestacea* stably infected with sMel. Also, the timecourse of depurination is very different in wasps versus nematodes. Wasp depurination occurs soon after wasp eggs hatch inside the fly maggot (Ballinger and Perlman 2017), whereas nematode depurination occurs much later, when the nematode has molted to the motherworm stage, inside the fly pupa. This is when we would expect to see RIP activity against nematodes, as motherworms in *Spiroplasma*-infected flies are small, sickly, and produce few juveniles (Jaenike et al. 2010).

We speculate that the RIPs in sNeo but not sMel can successfully access nematode ribosomes. However, much is unknown about the mechanism of *Spiroplasma* RIP action. For example, we do not know how *Spiroplasma* RIPs enter target cells, and whether this requires specific cell surface receptors or additional *Spiroplasma* partner proteins. *Spiroplasma*

encodes type 1 RIPs consisting of an enzymatic A chain but not a carbohydrate-binding B chain. Surprisingly little is known of general strategies used by type 1 RIPs to access the cytoplasm. The lack of a B chain is thought to result in a drastic reduction in receptor binding and thus cell entry, which could explain the reduced toxicity relative to type 2 RIPs. However, receptor binding has been reported in some cases, for example, the type 1 RIP saporin and the ricin A chain have both been reported to use the vertebrate α2-macroglobulin receptor (Cavallaro et al. 1995). Thus, it is possible that the McRIPs specifically bind nematode cell surface receptors that are not recognized by the other *Spiroplasma* type 1 RIPs. It is also interesting that sNeo RIP2 and sNigra RIP6 are adjacent to a gene with distant homology to a pore-forming epsilon toxin in *Clostridium* (see also Hamilton et al. 2014); perhaps this toxin plays a role in cell entry. A very recently published paper (Masson et al. 2018) reported the ability to culture sMel. This breakthrough will hopefully soon lead to the possibility to genetically transform both sMel and sNeo, which will allow us to directly test whether specific RIPs, as well as other *Spiroplasma* factors, are necessary for specific defense.

That some *Spiroplasma* RIPs are found on plasmids strongly suggests that they were recently acquired via horizontal transfer. Interestingly, the genetic basis of male-killing in sMel was recently discovered, and was found to be due to a novel toxin, called Spaid, that also lies on a plasmid (Harumoto and Lemaitre 2018), suggesting that this is a common pattern for ecologically specialized symbiont functions. We also recently documented dynamic gains and losses of RIP toxins, including plasmid-associated ones, in the genomes of ant-associated *Spiroplasma* (Ballinger et al. 2018).

So where did the sNeo RIP toxins come from? We thought that other mycophagous flies might be an appropriate place to look, as this group of insects is commonly infected with diverse parasitic nematodes (Jaenike and Perlman 2002). Indeed, we found that the phorid fly *M. nigra*, a visitor to mushroom baits in Victoria (along with *D. neotestacea*) is commonly infected with a strain of *Spiroplasma* in the citri clade. Although distantly related to sNeo, this symbiont's genome encodes closely related RIPs, including one on a plasmid similar to psNeo. It will be interesting to test whether sNigra protects against nematodes. Although we have not yet found nematodes infecting wild *M. nigra*, its congener and fellow mushroom-feeder *Megaselia halterata* is commonly infected with the sterilizing nematode *Howardula husseyi* (Riding and Hague 1974). Although primarily vertically transmitted over ecological timescales, the phylogenetic distribution of inherited *Spiroplasma* shows rampant colonization of new hosts via horizontal transmission (Haselkorn et al. 2009), and one experimentally validated mode of transfer is via mites (Jaenike et al. 2007), which are common associates of both *Drosophila* and *Megaselia*, as both parasites and phoretic passengers.

Dynamic gains and losses via horizontal gene transfer through plasmids, phage, and other mobile genetic elements, appear to be very common in *Spiroplasma* (Lo and Kuo 2017; Tsai et al. 2018) and in defensive symbionts in general. Some of the best examples include toxins in *Hamiltonella* symbionts of aphids (Degnan and Moran 2008; Brandt et al. 2017) and biosynthetic gene clusters that encode toxic metabolites *Pseudonocardia* symbionts of fungus-growing ants (Sit et al. 2015; VanArnam et al. 2018) that are encoded on phage and plasmids, respectively. Unlike obligate nutritional symbionts, facultative symbionts frequently colonize novel hosts and encounter many other microbes that they can potentially exchange genes with, providing a rapid way to respond to dynamic biotic challenges.

Supplementary Material

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

Acknowledgments

We thank Emily Hartop for assisting with *Megaselia* species identification and Bruno Lemaitre for providing sMel in *Drosophila melanogaster*. This work was funded by a grant from the Swiss National Science Foundation awarded to S.J.P.

Literature Cited

Ballinger MJ, Moore LD, Perlman SJ. 2018. Evolution and diversity of inherited Spiroplasma symbionts in Myrmica ants. Appl Environ Microbiol. 84(4):1–14.

Ballinger MJ, Perlman SJ. 2017. Generality of toxins in defensive symbiosis: ribosome-inactivating proteins and defense against parasitic wasps in Drosophila. PLoS Pathog. 13:1–19.

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Brandt JW, et al. 2017. Culture of an aphid heritable symbiont demonstrates its direct role in defence against parasitoids. Proc R Soc B Biol Sci. 284(1866):20171925.

Brettin T, et al. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep. 5:1–6.

Bushnell B. 2014. BBMap short read aligner.

Carle P, et al. 2010. Partial chromosome sequence of *Spiroplasma citri* reveals extensive viral invasion and important gene decay. Appl Environ Microbiol. 76(11):3420–3426.

Cavallaro U, et al. 1995. α-Macroglobulin receptor mediates binding and cytotoxicity of plant ribosome-inactivating proteins. Eur J Biochem. 232(1):165–171.

Cockburn SN, et al. 2013. Dynamics of the continent-wide spread of a Drosophila defensive symbiont. Ecol Lett. 16(5):609–616.

Davis RE, et al. 2015. Complete genome sequence of *Spiroplasma kunkelii* strain CR2–3x, causal agent of corn stunt disease in *Zea mays* L. Genome Announc. 3:6–7.

Degnan PH, Moran NA. 2008. Diverse phage-encoded toxins in a protective insect endosymbiont. Appl Environ Microbiol. 74(21):6782–6791.

Engl T, et al. 2018. Evolutionary stability of antibiotic protection in a defensive symbiosis. Proc Natl Acad Sci U S A. 115(9):E2020–E2029.
Haine ER. 2008. Symbiont-mediated protection. Proc R Soc B Biol Sci. 275:353–361.
Hamilton PT, et al. 2014. Transcriptional responses in a Drosophila defensive symbiosis. Mol Ecol. 23(6):1558–1570.
Hamilton PT, et al. 2016. A ribosome-inactivating protein in a Drosophila defensive symbiont. Proc Natl Acad Sci U S A. 113:1518648113.
Harumoto T, Lemaitre B. 2018. Male-killing toxin in a bacterial symbiont of Drosophila. Nature. 557:252–255.
Haselkorn TS, Jaenike J. 2015. Macroevolutionary persistence of heritable endosymbionts: acquisition, retention and expression of adaptive phenotypes in Spiroplasma. Mol Ecol. 24(14):3752–3765.
Haselkorn TS, Markow TA, Moran NA. 2009. Multiple introductions of the Spiroplasma bacterial endosymbiont into Drosophila. Mol Ecol. 18(6):1294–1305.
Jaenike J, Perlman J. 2002. Ecology and evolution of host-parasite associations: mycophagous Drosophila and their parasitic nematodes. Am Nat. 160:S23–S39.
Jaenike J, et al. 2007. Interspecific transmission of endosymbiotic Spiroplasma by mites. Biol Lett. 3(1):23–25.
Jaenike J, et al. 2010. Adaptation via symbiosis: recent spread of a Drosophila defensive symbiont. Science 329(5988):212–215.
Kaltenpoth M, et al. 2005. Symbiotic bacteria protect wasp larvae from fungal infestation. Curr Biol. 15(5):475–479.
Kaltenpoth M, et al. 2014. Partner choice and fidelity stabilize coevolution in a Cretaceous-age defensive symbiosis. Proc Natl Acad Sci U S A. 111(17):6359–6364.
Laczny CC, et al. 2017. BusyBee Web: Metagenomic data analysis by bootstrapped supervised binning and annotation. Nucleic Acids Res. 45:W171–W179.
Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26(5):589–595.
Liao Y, Smyth GK, Shi W. 2014. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30(7):923–930.
Lo WS, et al. 2013. Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. BMC Genomics 14, 1–13.
Lo WS, Kuo CH. 2017. Horizontal acquisition and transcriptional integration of novel genes in mosquito-associated Spiroplasma. Genome Biol Evol. doi:10.1093/gbe/evx244.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:1–21.
Masson F, et al. 2018. In vitro culture of the insect endosymbiont *Spiroplasma poulsonii* highlights bacterial genes involved in host symbiont interaction. MBio 9:1–11.
Melchior WB, Tolleson WH. 2010. A functional quantitative polymerase chain reaction assay for ricin, Shiga toxin, and related ribosome-inactivating proteins. Anal Biochem. 396(2):204–211.
Montenegro H, et al. 2005. Male-killing Spiroplasma naturally infecting *Drosophila melanogaster*. Insect Mol Biol. 14(3):281–287.
Moriya Y, et al. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182–W185.
Oliver K, et al. 2009. Bacteriophages encode factors required for protection in a symbiotic mutualism. Science (80-) 325(5943): 992–994.
Oliver KM, et al. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proc Natl Acad Sci U S A. 100(4):1803–1807.
Paredes JC, et al. 2015. Genome sequence of the *Drosophila melanogaster* male-killing. MBio 6:1–12.
Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29(9):e45.
Pierce M, et al. 2011. Development of a quantitative RT-PCR assay to examine the kinetics of ribosome depurination by ribosome inactivating proteins using *Saccharomyces cerevisiae* as a model. RNA 17(1):201–210.
Pool J, Wong A, Aquadro C. 2006. Finding of male-killing Spiroplasma infecting *Drosophila melanogaster* in Africa implies transatlantic migration of this endosymbiont. Heredity (Edinb) 97(1):27–32.
Riding IL, Hague NG. 1974. Some observations on a *Tylenchid nematode* Howardula sp. parasitizing the mushroom phorid *Megaselia halterata* (Phoridae, Diptera). Ann Appl Biol. 78(3):205–2011.
Simão FA, et al. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.
Sit CS, et al. 2015. Variable genetic architectures produce virtually identical molecules in bacterial symbionts of fungus-growing ants. Proc Natl Acad Sci U S A. 112(43):13150–13154.
Tsai YM, Chang A, Kuo CH. 2018. Horizontal gene acquisitions contributed to genome expansion in insect–symbiotic *Spiroplasma clarkii*. Genome Biol Evol. doi:10.1093/gbe/evy113.
VanArnam EB, Currie CR, Clardy J. 2018. Defense contracts: molecular protection in insect–microbe symbioses. Chem Soc Rev. 47(5):1638–1651.
Vorburger C, Perlman SJ. 2018. The role of defensive symbionts in host–parasite coevolution. Biol Rev. 93(4):1747–1764.
Wick RR, et al. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 13:1–22.
Xie J, et al. 2014. Male killing Spiroplasma protects *Drosophila melanogaster* against two parasitoid wasps. Heredity (Edinb) 112(4):399–408.
Xie J, Vilchez I, Mateos M. 2010. Spiroplasma bacteria enhance survival of *Drosophila hydei* attacked by the parasitic wasp *Leptopilina heterotoma*. PLoS One 5:1–7.

Associate editor: Daniel Sloan