Positive Selection and Horizontal Gene Transfer in the Genome of a Male-Killing *Wolbachia*

Tom Hill,[1,2] Robert L. Unckless [3], and Jessamyn I. Perlmutter [*,3]

[1]NIAID Collaborative Bioinformatics Resource, National Institutes of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

[2]Advanced Biomedical Computational Science, Frederick National Laboratory for Cancer Research, Frederick, MD, USA

[3]Department of Molecular Biosciences, University of Kansas, Lawrence, KS, USA

***Corresponding author:** E-mail: jessamyn.perlmutter@ku.edu.

Associate editor: Rebekah Rogers

Abstract

***Wolbachia* are a genus of widespread bacterial endosymbionts in which some strains can hijack or manipulate arthropod host reproduction. Male killing is one such manipulation in which these maternally transmitted bacteria benefit surviving daughters in part by removing competition with the sons for scarce resources. Despite previous findings of interesting genome features of microbial sex ratio distorters, the population genomics of male-killers remain largely uncharacterized. Here, we uncover several unique features of the genome and population genomics of four Arizonan populations of a male-killing *Wolbachia* strain, *w*Inn, that infects mushroom-feeding *Drosophila innubila*. We first compared the *w*Inn genome with other closely related *Wolbachia* genomes of *Drosophila* hosts in terms of genome content and confirm that the *w*Inn genome is largely similar in overall gene content to the *w*Mel strain infecting *D. melanogaster*. However, it also contains many unique genes and repetitive genetic elements that indicate lateral gene transfers between *w*Inn and non-*Drosophila* eukaryotes. We also find that, in line with literature precedent, genes in the *Wolbachia* prophage and Octomom regions are under positive selection. Of all the genes under positive selection, many also show evidence of recent horizontal transfer among *Wolbachia* symbiont genomes. These dynamics of selection and horizontal gene transfer across the genomes of several *Wolbachia* strains and diverse host species may be important underlying factors in *Wolbachia*'s success as a male-killer of divergent host species.**

***Key words:* *Wolbachia*, male killing, evolutionary genomics.**

Introduction

Wolbachia are the most widespread endosymbionts on the planet, infecting an estimated 40–52% of all insect species (Kondo et al. 2002; Zug and Hammerstein 2012; Weinert et al. 2015). These obligate intracellular Gram-negative α-proteobacteria of the order *Rickettsiales* primarily infect the gonads of their hosts and are predominantly transmitted vertically via the cytoplasm from mother to offspring (Hertig and Wolbach 1924; Serbus and Sullivan 2007). *Wolbachia* of insects and other arthropods have adopted cunning techniques to facilitate their matrilineal spread by manipulating host reproduction to increase the proportion of infected, transmitting females in the population (Werren et al. 2008; Hurst and Frost 2015; Kaur et al. 2021). The most common form of this reproductive parasitism is cytoplasmic incompatibility (CI), where crosses between uninfected females and infected males result in death of offspring. If the mother is also infected with a compatible strain, offspring are rescued from death, giving infected females a relative fitness advantage in the population over uninfected females (Yen and Barr 1971; Turelli and Hoffman 1991; Sinkins et al. 1995). Three other, less common forms of reproductive parasitism rely on sex ratio distortion to increase the proportion of transmitting females each generation. These phenotypes are known as parthenogenesis (asexual reproduction of females; Russell and Stouthamer 2011), feminization (genetic males physically develop and reproduce as females; Bouchon et al. 1998; Kageyama et al. 2002), and male killing (infected males die early in development; Hurst et al. 1999; Fujii et al. 2001; Dyson et al. 2002). In addition, despite being comparatively rare, there are reported cases of horizontal *Wolbachia* host switching. These include cases of switches either between host species where *Wolbachia* propagate in new taxa through mechanisms such as predation or transfer between individual hosts within a species (O'Neill et al. 1992; Vavre et al. 1999; Haine et al. 2005; Riegler et al. 2005; Werren et al. 2008; Ilinsky 2013; Turelli et al. 2018). These transfers likely influence the evolution of *Wolbachia* in relation to their hosts and play a key role in *Wolbachia*'s ubiquitous spread around the world (Sanaei et al. 2021).

The incredible success of *Wolbachia* in becoming one of the world's most widespread infections (Werren et al. 2008; LePage and Bordenstein 2013) is in part due to its diverse genetic toolkit and unique genome features. Indeed, to date,

Article

17 *Wolbachia* supergroups have been described, labeled supergroups A-F, H-Q, and S (Taylor et al. 2018; Laidoudi et al. 2020; Lefoulon et al. 2020). Studies on strains in supergroups A and B are well represented in the literature, and include many reproductive parasite strains of hosts such as mosquitoes and a large number of *Drosophila* species (Gerth et al. 2014). The first sequenced *Wolbachia* genome, strain *w*Mel of *Drosophila melanogaster*, demonstrated that this genus deviates from the canonical model of streamlined genome content typical of other endosymbionts. It instead contains unusually high levels of repetitive DNA content and mobile elements such as prophage insertions, insertion sequence (IS) elements, and transposons along with large numbers of gene duplications and genetic rearrangements (Wu et al. 2004). Further, there is evidence for frequent activity and transmission of these transposable elements (TEs) in the genome (Cordaux et al. 2008).

Many of the differences in content between *Wolbachia* genomes are contributed by prophage WO, the genome of phage WO of *Wolbachia*, that has inserted itself into the bacterial chromosome and replicates along with core *Wolbachia* genes (Ishmael et al. 2009; Kent and Bordenstein 2010; Bordenstein and Bordenstein 2016). This prophage is found in at least five different supergroups within nonmutualist symbioses (Kaur et al. 2021). Prophage WO sometimes retains the potential to form phage particles (Masui et al. 2001) and sometimes degrades over time into a relic prophage, losing the potential to form new viral particles due to loss of key virus structure genes, but with other genes remaining (Metcalf et al. 2014; Bordenstein and Bordenstein 2016). Prophages and phages are highly mobile and dynamic elements in the genome, often picking up new genes via horizontal gene transfer. Phage WO can transfer itself between *Wolbachia* strains coinfecting the same host, and nearby genes or gene fragments may hitchhike in the process, making the phage a vehicle of gene movement and a creator of genetic novelty (Bordenstein and Wernegreen 2004; Kent et al. 2011). Many unique prophage WO genes have the potential to confer important functions in interactions with the eukaryotic host (Bordenstein and Bordenstein 2016). Indeed, functional and evolutionary analyses of the genetic loci that underlie CI have shown that they are in fact prophage WO associated genes that interact directly with the eukaryotic host to manipulate reproductive and developmental processes (Beckmann et al. 2017, 2019; LePage et al. 2017; Lindsey et al. 2018). Related to phage WO is a cassette of eight genes known as Octomom, which are associated with a greater host protective benefit against RNA viruses such as *Drosophila* C virus and Flock House virus (Chrostek et al. 2013). This cassette contains paralogs of phage WO genes but replicates separately, and variance in copy number has functional consequences for the bacteria and host. Specifically, in a lab-derived strain of *w*Mel, copy number of Octomom is correlated with regulation of *Wolbachia* titer making strains pathogenic to the host (Chrostek and Teixeira 2015; Duarte et al. 2021).

Generally, specific genes and regions of genomes can horizontally transfer to and from bacteria of the same strain allowing for a recombination-like process which may facilitate adaptation. Or, these regions can transfer to different strains which allows for the acquisition of new genes, allowing for adaptation to better propagate within their hosts (Lawrence 1999; Dutta and Pan 2002). Regarding *Wolbachia* specifically, studies have demonstrated that lateral transfer with eukaryotes has likely occurred many times with *Wolbachia*, as it is common for entire *Wolbachia* genes or domains within genes to have homology with eukaryotic DNA (Wu et al. 2004; Bordenstein and Bordenstein 2016). Indeed, the transfers occur both ways, as many arthropod and nematode genomes also contain *Wolbachia* DNA (Kondo et al. 2002; Hotopp et al. 2007; Funkhouser-Jones et al. 2015; Leclercq et al. 2016; Wang et al. 2016). Genetic transfers may be aided by the abundant mobile elements within the genome, including phage WO and transposons. Indeed, the high and variable number of TEs in *Wolbachia* genomes has long been recognized as a useful method for differentiating strains due to their frequent movement (Wu et al. 2004; Duron et al. 2005), and they may be behind the movement of specific genes (Cooper et al. 2019) or disruption of others (Iturbe-Ormaetxe et al. 2005). In addition, the first plasmid of *Wolbachia* was recently described, pWCP of several populations of *Culex pipiens* mosquitoes, which itself contains a TE (Reveillaud et al. 2019). Thus, the growing number of described cases of genetic transfer correlates with the unusually high number of mobile genetic elements within *Wolbachia*, as well as its ability to switch hosts.

Despite the great diversity and interest in a variety of *Wolbachia* infections, most research attention has focused on CI, largely due to its use in vector control strategies (Zabalou et al. 2004). These programs take advantage of the natural abilities of *Wolbachia* to both block viral transmission and spread itself via reproductive parasitism (Hedges et al. 2008; Teixeira et al. 2008; O'Neill et al. 2018; Mains et al. 2019; Ross et al. 2019). Comparatively fewer analyses have been done on *Wolbachia* genomes of strains that induce male killing (Dyer and Jaenike 2004; Ishmael et al. 2009; Duplouy et al. 2013; Metcalf et al. 2014). However, male killing merits additional analysis due to its potential in vector control (Berec et al. 2016), role in shaping arthropod evolution (Jiggins et al. 2000), and the close relationship between CI and male killing (Dyer et al. 2005). Indeed, the CI genetic loci are located only a few genes away from the male-killing candidate gene, *wmk* (*WO-mediated killing*) in the *Wolbachia* strain of *Drosophila melanogaster* (*w*Mel) (Perlmutter et al. 2019). In addition, many male-killing and CI strains are closely related (Sheeley and McAllister 2009), and several strains are multipotent in that they can switch between the two phenotypes either within the same host or between different hosts (Hurst et al. 2000; Sasaki et al. 2002; Jaenike 2007). In addition, the strength of the phenotypes can be quite variable even within strains, suggesting there is a complex relationship between symbiont genotypes and host phenotypes that requires further study (Cooper et al. 2017, 2019). The close genetic relationship between, and likely prophage origins of, the two phenotypes indicate that studies on male killing may inform CI and vice versa. In addition, their overall similarities may

help narrow down evolutionary dynamics that are unique to each phenotype or shared between them.

Several studies have set the stage for male-killing genetics and genomics. Beyond *Wolbachia*, genomes of *Arsenophonus nasoniae* and *Spiroplasma poulsonii* have been sequenced. The *Arsenophonus* strain was the first full genome of a male-killer and revealed that it has a relatively large genome for an insect symbiont, with evidence of gene transfer with *Wolbachia* and other bacteria (Darby et al. 2010). The *Spiroplasma* genome has reduced metabolic capabilities despite its extracellular lifecycle, and also contains several lethal RIP toxins and a male-killing gene on a plasmid (Paredes et al. 2015; Harumoto and Lemaitre 2018; Garcia-Arraez et al. 2019). The first sequenced genome of a *Wolbachia* male-killer was that of the *w*Bol1b strain of *Hypolimnas bolina* butterflies (Duplouy et al. 2013). Among sequenced genomes at the time, it was most like *w*Pip of *C. pipiens* and contains type IV secretion system genes, many ankyrin repeat domains, and several prophage regions. In addition, *w*Bol1b has many genes with homology to non-*Wolbachia* bacteria, and a few genes homologous to genes in mosquitoes, suggesting gene transfer has occurred between this strain and both other prokaryotes and eukaryotes (Duplouy et al. 2013). The genome of the *w*Rec strain that causes CI in its native host, *D. recens*, and male killing in sister species *D. subquinaria* shows degradation in the prophage region (relic phage), with only a few dozen phage WO genes remaining and loss of essential phage particle genes, and no genes unique to the strain (Jaenike 2007; Metcalf et al. 2014).

Among the few known *Wolbachia* male-killers of flies is the strain infecting *Drosophila innubila* mushroom-feeders, *w*Inn (Dyer and Jaenike 2005). This strain is particularly interesting as it is closely related to *Wolbachia* found in the main *Drosophila* model species, *w*Mel of *D. melanogaster*, which causes CI (Sheeley and McAllister 2009). In addition, the symbiosis between *w*Inn and its host has been maintained for thousands of years, and despite this, there is no evidence of host resistance to male killing in modern populations (Jaenike and Dyer 2008; Unckless and Jaenike 2012). A previous microarray analysis compared the content of the *w*Inn genome to that of several other strains, including several CI strains and nonparasitic strains. The findings were that phage and ankyrin repeat genes were amongst the most divergent in the genome, that the ankyrins were particularly variable within *w*Inn, and that the strain had several unique but uncharacterized hypothetical proteins (Ishmael et al. 2009). Thus, several previous studies examined genomic content of male-killers, but little is known about their genomics on a population level. Due to the close relationship between *w*Mel and *w*Inn, the longstanding symbiosis of *w*Inn with its host, and building on previous findings suggesting unique *w*Inn genome content, we chose to conduct population genomics analysis on this strain. We aimed to identify both new genetic content and population genomic trends that may be important for *Wolbachia* or male-killers more specifically.

Here, we sequence the genome of the *Wolbachia* strain infecting *D. innubila*, *w*Inn, and conduct population genomic analyses using sequences from 48 *Wolbachia*-infected individual wild females from four populations in Arizona. We compare the genome content to that of similar supergroup A *Wolbachia* of *Drosophila* hosts to reduce variables when narrowing down unique genomic regions or population genetic trends of this male-killer. We demonstrate overall similarity of the genome content with *w*Mel, as previously shown, and newly identify several dozen unique genes and repetitive elements implying lateral gene transfer with divergent hosts. We determine that genes from prophage and Octomom regions show more evidence of positive selection than background genes, consistent with other strains and *Wolbachia*'s general ability to adapt to diverse hosts. Finally, we examine population structure and co-inheritance of *Wolbachia* with mitochondria to show that *w*Inn largely exhibits patterns of strict vertical inheritance within the population and mobile elements are likely responsible for any discordant phylogenies. Variance in sequencing coverage also putatively suggests that Octomom copy number may differ considerably across individuals and that prophage WO may form active particles in this strain.

Results

*w*Inn Genome Assembly Reveals a Genome Similar to *w*Mel and Evidence of Lateral Gene Transfer from Multiple Host Genera

Drosophila innubila is a mycophagous species in the *Drosophila* subgenus, found throughout the southwestern USA and northwestern Mexico on mountain-top forests known as "Sky Islands," separated by large expanses of desert (Jaenike et al. 2003; Dyer and Jaenike 2005; Dyer et al. 2005; Jaenike and Dyer 2008). Here, we examined the genome and population genomic variation of *w*Inn. In a previous survey we collected wild *D. innubila* from four isolated mountain locations and tested strains for *Wolbachia* using PCR to amplify the *wsp* locus and found 48 females infected with *Wolbachia* (supplementary table 1, Supplementary Material online, 13 from the Chiricahua mountains, 27 from Prescott, two from the Huachucas, and six from the Santa Ritas) (Hill and Unckless 2020a, 2020b).

We sequenced and assembled the genome using a combination of short and long reads for one strain. The *w*Inn genome is a single circular chromosome 1,290,587 base pairs long, with 35.1% GC content (fig. 1A). We found 1,341 genes, 1,301 found previously in other *Wolbachia*: 1,232 genes are shared with *w*Mel, 1,145 shared with *w*Rec, and 1,010 shared with *w*Ri. The *w*Inn genome had a BUSCO score of 81.9% (181 complete single-copy orthologs and two fragmented orthologs, from a total of 221) compared with 81.3% in *w*Mel (180 complete single-copy orthologs and two fragmented orthologs, from a total of 221). Of the 1,301 previously identified genes, 924 are conserved across all four genomes (fig. 1C and supplementary table 2, Supplementary Material online), including 12 prophage WO-A and 54 prophage WO-B genes in all genomes (supplementary table 3, Supplementary Material online), and nine Octomom genes (five orthologs to *w*Mel and four paralogs of these, genes linked to *Wolbachia* pathogenicity) (Chrostek and Teixeira 2015). Interestingly, these

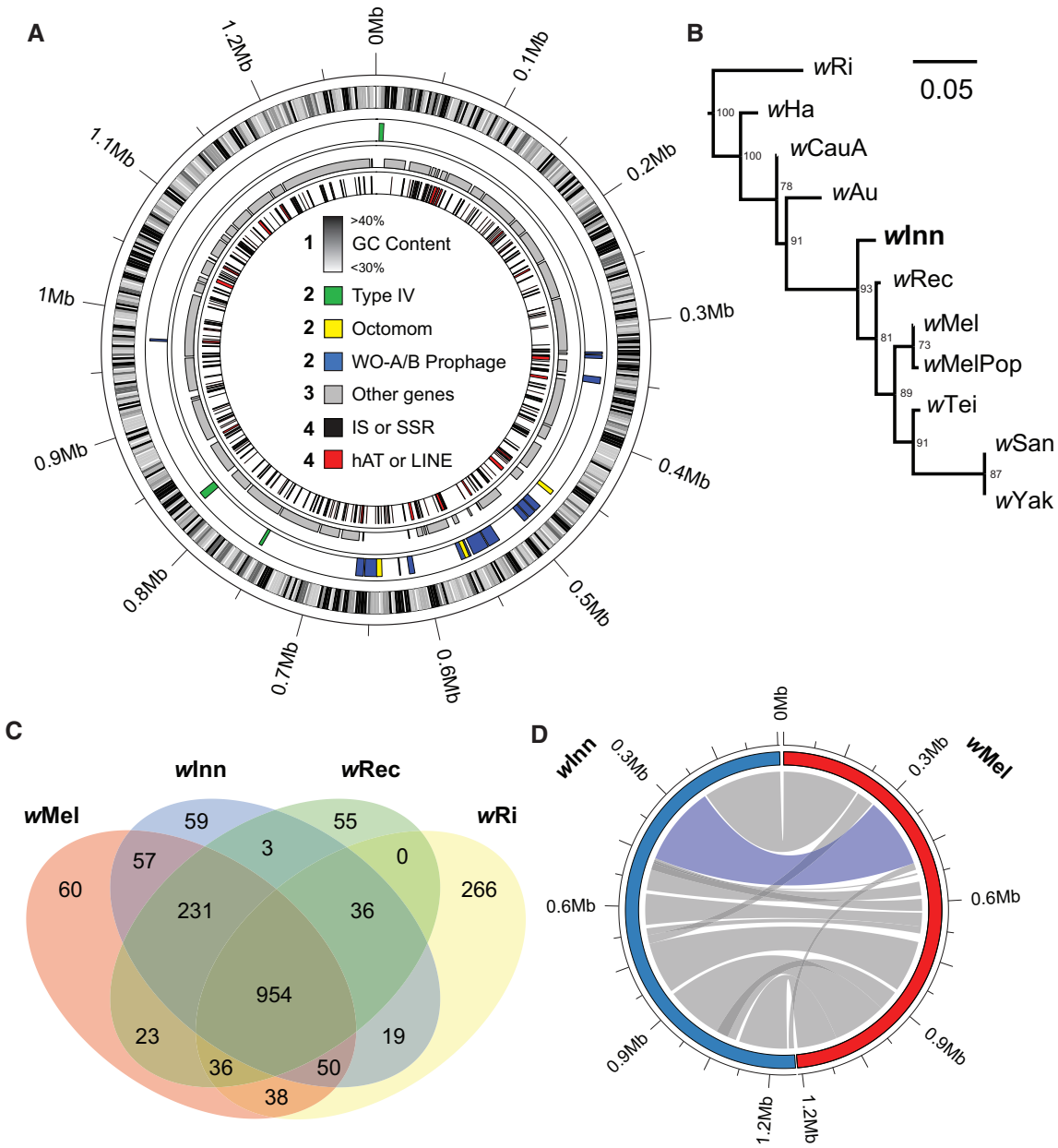

Fig. 1. (*A*) Schematic of the *w*Inn genome. Circles correspond to the following: 1) GC content of the *w*Inn genome in 10 kb windows, between 30% and 40%. Darker colors have higher GC content. 2) Locations of genes thought to interact with hosts, specifically prophage orthologous to WO-A and WO-B genes in *w*Mel (blue), Type IV secretion pumps (green), and Octomom genes (yellow). 3) Loci of nonphage genes. 4) Loci of repetitive content, with short simple repeats and interspersed satellites (SSR and IS, black), and hAT or LINE TE insertions (red). (*B*) Phylogeny of *Wolbachia* genomes closely related to *w*Inn for reference, as a subset of supplementary figure 1, Supplementary Material online. Bootstrap support of each branch is shown on the nodes (of 100 bootstraps). A description of each *Wolbachia* genome and the species they infect is given in supplementary table 1, Supplementary Material online. (*C*) The overlap of genes between *w*Inn, *w*Mel, and *w*Ri. (*D*) Synteny between *w*Mel and *w*Inn, with single large inversion shown in blue, whereas consistent synteny groups are shown in gray.

Octomom genes are not found in a single cassette like in *w*Mel but are instead spread throughout the genome (fig. 1A). The genes orthologous to WO-B genes of *w*Mel are found in three groups (fig. 1A, called WOInn-B1, WOInn-B2, and WOInn-B3). Despite the fragmentation of the prophage regions, the genes are syntenic to the prophage WO-B region in *w*Mel. We also found ten type IV secretion system genes, found in two cassettes, as in *w*Mel (fig. 1A). Consistent with previous results, *w*Inn is closely related to *w*Mel within supergroup A, clustering with other supergroup A *Wolbachia* genomes (fig. 1B and supplementary fig. 1, Supplementary Material online).

Three genes are shared between the male-killing *w*Inn genome and *w*Rec, but absent in the *w*Mel and *w*Ri genomes, both strains that induce CI in their native hosts (fig. 1C). Recall that *w*Rec reportedly kills males when introgressed into a sister species, but causes CI in its native host (Jaenike 2007). All three genes are hypothetical proteins found in other *Wolbachia* supergroup A genomes that do not cause male killing in their native hosts (including other varieties of *w*Mel).

In addition, *w*Inn does not appear to have a reduced, relic prophage genome-like *w*Rec, and instead shares most prophage genes with *w*Mel despite being more diverged from *w*Mel than *w*Rec (fig. 1C). The 57 genes absent in *w*Rec but present in *w*Inn and *w*Mel consists of 21 prophage genes, four transcription genes, nine metabolism genes, and 21 genes of unknown function. We attempted to further confirm the differences in genomic content by mapping short reads from *w*Rec, *w*Mel, *w*Inn, and *w*Ha to each of the genomes pairwise and find the exact same number of genes shared in each case, supporting the assembled genomes are not missing any shared genes. The *w*Rec genome appears to be missing portions of the regions orthologous to 0.35–0.55 and 1.24–1.28 Mb in the *w*Inn genome, which also includes a large portion of the prophage WO-B genome.

Of the 40 *w*Inn coding sequences not found in other *Wolbachia*, 23 of these have high similarity to genomic mRNA in *Formica* wood ants that may have an overlapping range with *D. innubila* (nonredundant megablast e-value <0.00005) (Francoeur 1973; Altschul et al. 1990). These 40 sequences largely contain predicted domains that are typically found in *Wolbachia* genomes, such as ankyrin repeats, PD-(D/E)XK nucleases/transposases, membrane transporters, and genes involved in amino acid and nucleotide metabolism (Bordenstein and Bordenstein 2016; Lindsey et al. 2018; Massey and Newton 2021). Similarly, the *D. innubila* host also contains multiple TE sequences shared with *Camponotus* (a genus within *Formica*) (Hill et al. 2019). Further, we find five unique *w*Inn genes have a high similarity to *Varroa destructor* mite transcriptome sequences (nonredundant megablast e-value <0.00005, supplementary table 3, Supplementary Material online). The 12 remaining sequences have no known orthologs. Among the 157 genes absent in *w*Inn but present in *w*Mel, we found no functional categories enriched ($P > 0.12$).

Like other *Wolbachia* (Foster et al. 2005; Woolfit et al. 2013) a large proportion of the *w*Inn genome is repetitive: 12.38% of the genome consists of repetitive content (fig. 1A). Most of these sequences are short simple repeats, satellites, and insertions from 16 bacterial ISs (selfish elements found in bacteria). About 1.49% of the genome consists of insertions of a single hAT family element (*hobo*-like DNA transposon found in *Drosophila*, rnd-1_family-6) inserted in 14 loci across the genome and 3.74% consists of three LINE elements inserted in 37 loci across the genome (long-interspersed nuclear elements, an RNA transposon order found in *Drosophila*), primarily in clusters (fig. 1A) (Wicker et al. 2007). Consistent with several of the above-described unique *w*Inn genes, one LINE element (rnd-1_family-12 with 14 insertions) is homologous to a LINE found previously in *V. destructor* (or a close relative), whereas another (rnd-1_family-165 with 12 insertions) is homologous to a LINE found in *Formica* wood ants (nonredundant megablast e-value <0.00005) (Altschul et al. 1990). No homologous sequence can be identified for the hAT element (nonredundant megablast e-value = 1 yielded no hits), which is also the only TE found with a complete sequence, as opposed to most of the LINE element insertions that are degraded (supplementary fig. 2, Supplementary Material online).

Octomom and Prophage Genes Are under Positive Selection in *w*Inn and Other *Wolbachia*

We next used codeML to determine genes with signatures of selection in *w*Inn compared with the closely related *Wolbachia* genomes (Yang 2007). For each ortholog set, we identified the proportion of synonymous (d*S*) substitutions and amino acid changes to nonsynonymous substitutions (d*N*) (per possible synonymous or nonsynonymous substitution, respectively). We did this for substitutions occurring on each branch of the phylogeny to identify changes between the gene sequence of *w*Inn, *w*Ha, and *w*Ri. We expect d*N*/d*S* to be higher when genes are under positive selection, due to more nonsynonymous fixations (Yang 2007). We chose *w*Ha (*w*Inn–*w*Ha d*S* ∼ 0.068) and *w*Ri (*w*Inn–*w*Ri d*S* ∼ 0.386) over *w*Mel or other genomes as these genomes are diverged enough from *w*Inn to provide some signal in synonymous divergence (unlike *w*Mel or *w*Rec, where d*S* ∼ 0.001 between genomes), while not being diverged enough to have too little similarity or saturated rates of d*S*. Octomom and prophage genes show values consistent with positive selection on all branches (fig. 2, GLM $t = 2.750$, $P = 0.0061$). The genes with the next closest values include DNA metabolism genes, which showed no significant difference in rates of positive selection on the *w*Inn branch compared with background (fig. 2, GLM $t = 1.868$, $P = 0.0622$). The DNA metabolism genes with the highest d*N*/d*S* values in *w*Inn are the nonphage genes WD1095 (*radC*, a DNA repair protein), WD0065 (a DNA-binding protein), WD0057 (a host integration factor), and WD0752 (*xerC*, a recombinase). Next, we looked at d*N* or d*S* alone to determine how selection individually differed among categories to ask if signals were driven by an increase in d*N* alone, or with an accompanying change in d*S*, which would tell us about overall genome evolution rates in *w*Inn compared with other strains. Elevated d*N* alone would suggest positive selection only, whereas elevated d*N* and d*S* would indicate not only selection, but also rapid evolution of DNA more generally. Consistent with the elevated divergence in Octomom and phage genes being driven by positive selection, we find these two categories have elevated d*N* across the total phylogeny compared with genes in other categories (Octomom GLM $t = 4.1$, $P = 6.61e\text{-}5$, phage GLM $t = 3.735$, $P = 1.1e\text{-}4$), whereas no other categories have elevated d*N* compared with the background (GLM $P > 0.05$). We do not find significantly elevated d*S* in Octomom or phage genes on across the other branches (Octomom GLM $t = -0.518$, $P = 0.6045$, phage GLM $t = -0.643$, $P = 0.5209$). Interestingly, we find both elevated d*N* (Octomom GLM $t = 3.699$, $P = 0.00028$, phage GLM $t = 3.926$, $P = 9.25e\text{-}5$) and elevated d*S* (Octomom GLM $t = 5.053$, $P = 5.2e\text{-}7$, phage GLM $t = 7.216$, $P = 1.1e\text{-}12$) on the *w*Inn branch in Octomom and phage genes.

Several genes have been implicated in reproductive parasitism in *Wolbachia* in *Drosophila*, so we specifically examined the evolution of these genes in *w*Inn and the other genomes (*wmk*: WD0626, *cifA*: WD0631, *cifB*: WD0632). These genes

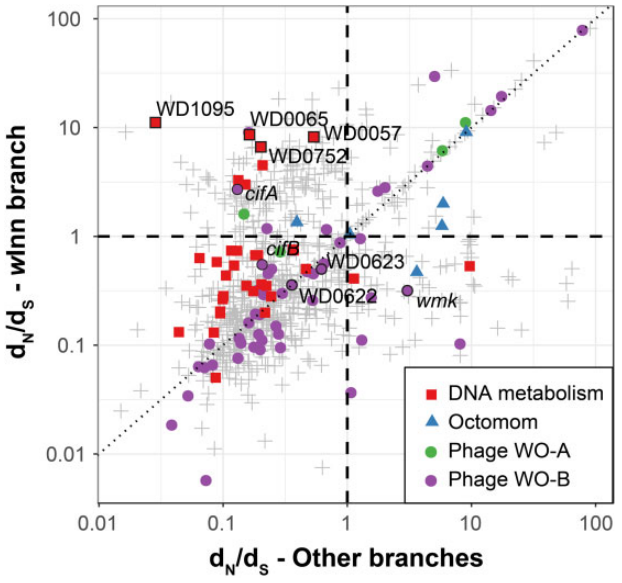

FIG. 2. Selection on genes of the *w*Inn branch versus evolution on the *w*Ri/*w*Ha branches. Functional categories of interest (DNA metabolism genes, prophage genes, and Octomom genes) are highlighted by different shapes and colors. Dashed lines show dN/dS=1 for both axes, whereas the dotted line shows where dN/dS is equal on the axes. Genes of interest, either due to putative involvement in *Wolbachia* pathogenicity, or due to high dN/dS in *w*Inn exclusively are named and labeled with a black outline to distinguish them.

show no difference from background rates of positive selection in *w*Inn (dN/dS = 0.27–1.56 in *w*Inn, versus dN/dS background median = 0.225, GLM $t = 0.224$, $P = 0.642$). The *wmk* male-killing candidate showed no significant difference on the *w*Ha and *w*Ri branches compared with *w*Inn (dN/dS = 1.56, vs. dN/dS background median = 0.225). Similarly, *cifA and cifB* (genes involved with CI) (Beckmann et al. 2017; LePage et al. 2017), had higher values (though not significantly different) in *w*Inn than the other two strains (fig. 2, WD0631 and WD0632, dN/dS = 1.10 and 0.27 respectively, $P = 0.845$). We also examined selection of specific codons in these genes but find no specific sites that are driving positive selection in these genes ($P > 0.05$). The Type IV secretion genes also do not show significantly different rates of positive selection than the background across the total phylogeny (GLM $t = 1.427$, $P = 0.154$).

Some Prophage Genes Show Evidence of Recent Horizontal Transfer in *w*Inn and across the *Wolbachia* Phylogeny

We suspected that many symbiont genes potentially involved in unique *w*Inn host–microbe interactions may be more likely to experience horizontal transfer between *Wolbachia* strains than other genes because of their association with the prophage. We looked for evidence of horizontal gene transfer since the divergence of *w*Inn from *w*Ha and *w*Ri. We used VHICA (Wallau et al. 2016) to compare synonymous divergence (dS) for the pairwise comparisons of *w*Inn–*w*Ha, *w*Inn–*w*Ri, and *w*Ri–*w*Ha for genes with orthologs in all three species. We used dS as a measure of the neutral mutation rate, as we expect that positive selection on synonymous mutations rarely leads to their fixation. If a gene has horizontally transferred into *w*Inn from another *Wolbachia* (but not *w*Ha or *w*Ri), we expect the dS to be higher in these comparisons than expected based on the distribution of dS across the rest of the genome (using the mean dS plus the variance in dS as the cut off for elevated dS) (Wallau et al. 2016). When a gene has transferred into the *w*Inn genome, we expect there will be elevated dS in both comparisons involving *w*Inn, but not the *w*Ha–*w*Ri comparison, as it would be significantly higher compared with these close relatives if it came from a more distantly related source. Overall, 13 genes have elevated dS in just the *w*Inn comparisons (fig. 3 and supplementary table 4, Supplementary Material online). These genes are enriched for prophage WO-A and WO-B genes (χ^2 test, $\chi^2 = 60.476$, df = 1, $P = 7.448e\text{-}15$), and Octomom genes (χ^2 test, $\chi^2 = 181.64$, df = 1, $P = 2.2e\text{-}16$) that have elevated divergence in *w*Inn. The putatively horizontally transferred genes do not significantly overlap with genes under positive selection in the *w*Inn genome ($\chi^2 = 0.009$, df = 1, $P = 0.9212$), though genes under positive selection across the total phylogeny do significantly overlap with genes with elevated dS in all comparisons ($\chi^2 = 9.8389$, df = 1, $P = 0.001709$). Domain analysis of the 13 genes demonstrates that most have weakly predicted domain functions, and those that do include predicted annotations such as *radC* DNA repair, patatin phospholipase, and *burrH*-like transcription factors with ankyrin repeats, all of which are typically found within phage WO (Bordenstein and Bordenstein 2016). The six putatively horizontally transferred genes not associated with the prophage or Octomom regions are all genes of unknown function. Horizontal transfer of the *Wolbachia* organism is unlikely to play a role in the elevated divergence seen here due to the lack of a significant difference in dS for the *w*Inn–*w*Ri and *w*Ha–*w*Ri comparisons (Wilcoxon rank sum test $W = 462{,}970$, $P = 0.7687$).

To examine if these gene categories frequently experience horizontal transfer, or if these transfers are unique to *w*Inn, we downloaded 54 *Wolbachia* genomes (all genomes available for download on NCBI genomes, described in supplementary table 1, Supplementary Material online) and made gene alignments for all orthologs and attempted to identify gene tree/species tree discordance. We assumed that excessive gene tree/species tree discordance would be due to large amounts of horizontal gene transfer. We attempted to find functional categories which show more tree discordance than expected and across 847 orthologous genes, and found excessive amounts of discordance for prophage WO genes (table 1, 36 of 47 genes, df = 1, $\chi^2 = 111.1$, $P = 5.62e\text{-}26$) and Octomom genes (table 1, seven of seven genes, df = 1, $\chi^2 = 71.27$, $P = 3.395e\text{-}17$) across large evolutionary distances, whereas no other categories have significantly more discordance than expected. We find a significant overlap in the genes which have horizontally transferred in *w*Inn and across the whole phylogeny, when put into a 2×2 contingency table (discordant vs. nondiscordant in VHICA and/or Ancient HT. $\chi^2 = 49.003$, $P = 2.556e\text{-}12$).

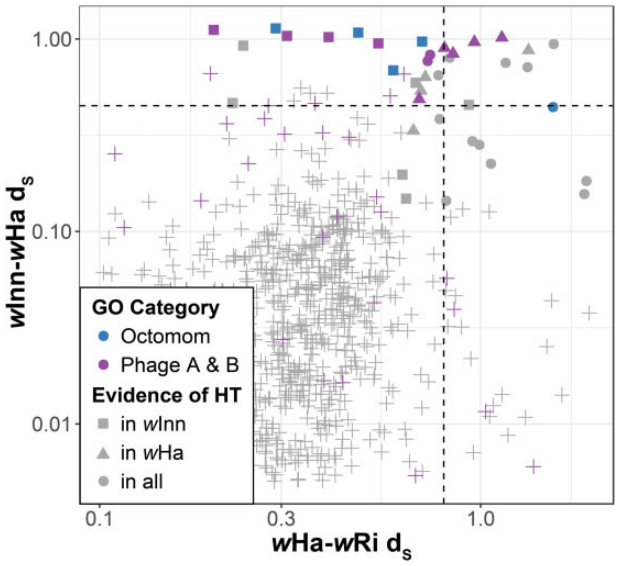

Fig. 3. Comparison of dS between pairs of *Wolbachia* suggesting horizontal transfer of genes. Point colors show the gene ontology categories (GO category) for Octomom genes and Prophage WO-B genes. Point shape indicates evidence of excessive divergence (and possible horizontal transfer) in either *w*Inn, *w*Ha, or both. Dashed lines show the mean plus the variance of dS for each axis as a rough cutoff for elevated synonymous divergence.

*w*Inn Is Highly Structured between Locations, and Is Not Always Co-inherited with the Mitochondria

We next mapped the short-read data for 48 samples of wild *D. innubila* infected with *w*Inn, collected from four locations in Arizona (fig. 4), to the repeat masked *w*Inn genome and called polymorphism. From this, we identified 30 SNPs as coding synonymous, 69 SNPs as coding nonsynonymous, and 235 SNPs as noncoding across all individuals. The *w*Inn samples are highly structured based on both the total and synonymous variation (fig. 4). Using a principal component analysis, we find three clear clusters, separating the Chiricahua and Prescott populations, and grouping the Santa Rita and the Huachuca populations together (fig. 4), as seen with the mitochondrial genome and consistent with previous findings (Jaenike et al. 2003; Dyer and Jaenike 2005; Jaenike and Dyer 2008; Hill and Unckless 2020a).

However, when building a maximum-likelihood tree of the *w*Inn samples using all polymorphisms in the core *Wolbachia* genes, we find some evidence of potential migration between populations (supplementary fig. 3*A*, Supplementary Material online). Specifically, we find two samples from PR cluster within CH and share the CH mitochondrial haplotype (supplementary fig. 3, Supplementary Material online). These two PR samples are also closer to CH than other PR samples in the principal component analysis (fig. 4*B* and *C*). This signature is not seen in the host, likely due to the recent establishment of *D. innubila* (Hill and Unckless 2020a), particularly in Prescott.

To identify if specific genomic regions are contributing to the population structure, we calculated the fixation index (F_{ST}), a measure of pairwise divergence between a subpopulation and the total population, between the three clustered groups. We expect F_{ST} to be elevated in cases where SNPs are found at high frequencies in a single population but not the remaining samples. As expected with the nonrecombining bacterial genome, we found signatures of F_{ST} are uniform genome wide, with no specific windows of elevated F_{ST} compared with the rest of the genome (1 kb windows, GLM $P > 0.432$) and no functional categories are enriched for high or low F_{ST} (supplementary table 6, Supplementary Material online, GLM $P > 0.611$).

When comparing the co-inheritance of the maternally transmitted *w*Inn and *D. innubila* mitochondria, we found little evidence of discordant inheritance, consistent with a previous study (Dyer and Jaenike 2005). We do however find evidence of 49 total SNPs spread across four clusters in the *w*Inn genome which show evidence of recombination-like events due to the presence of four allele combinations between the *Wolbachia* site and the mitochondria site across

Table 1. Summary of Species Tree/Gene Tree Discordance Analysis.

Functional Category	Discordant Genes	Genes in Category	χ^2	*P* value
Biosynthesis	**1**	**50**	**2.374**	**0.876**
Cell envelope	**1**	**31**	**0.952**	**0.671**
Cellular processes	**1**	**35**	**1.238**	**0.732**
DNA metabolism	**1**	**42**	**1.759**	**0.815**
Energy metabolism	**0**	**90**	**7.438**	**0.006**
Octomom	**7**	**7**	**71.278**	**3.395e-17**
Phage WO-A	**12**	**13**	**111.105**	**5.626e-26**
Phage WO-B	**24**	**33**	**165.927**	**5.817e-38**
Protein synthesis	**1**	**144**	**9.985**	**0.002**
Regulatory functions	**0**	**11**	**0.909**	**0.340**
Transcription	**0**	**52**	**4.298**	**0.038**
Transport and binding proteins	**0**	**44**	**3.636**	**0.057**
Type IV	**0**	**10**	**0.826**	**0.363**
Unknown function	**22**	**285**	**0.102**	**0.749**

NOTE.—Using 847 orthologous genes across 54 genomes (supplementary table 5, Supplementary Material online), we identified genes which showed significant discordance from the species tree. We calculated the expected number of discordant genes in each category based on the total number of discordant genes and the proportion of total genes in this category. We then used this value and the observed number of discordant genes in each category to perform a χ^2 test. The table shows the number of genes in each category showing discordance, and if this discordance is significant using a χ^2 test, using an expected number of discordant genes per category based on the size of the category. Note that these *P* values are not corrected for multiple tests.

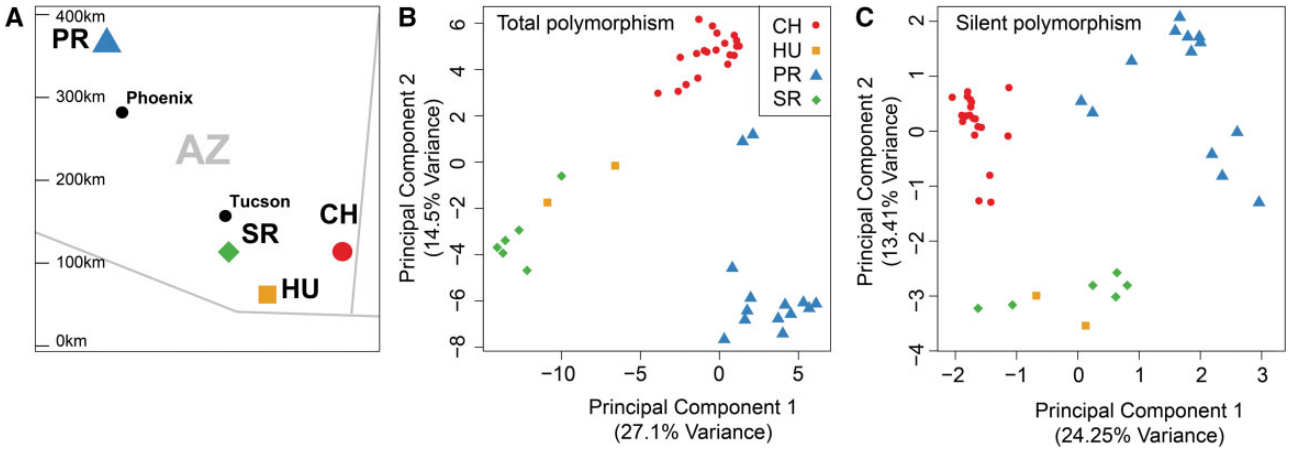

FIG. 4. Principal component analysis of genetic variation in *w*Inn. (*A*) Map of locations samples were taken from in this survey, adapted from Hill and Unckless (2020a). Phoenix and Tucson are shown as points of reference. (*B*) Total polymorphism and (*C*) silent polymorphism in *w*Inn samples, colored and shaped by location of collection, Chiricahuas (CH, red circles), Huachucas (HU, orange squares), Prescott (PR, blue triangles), and Santa Ritas (SR, green diamonds).

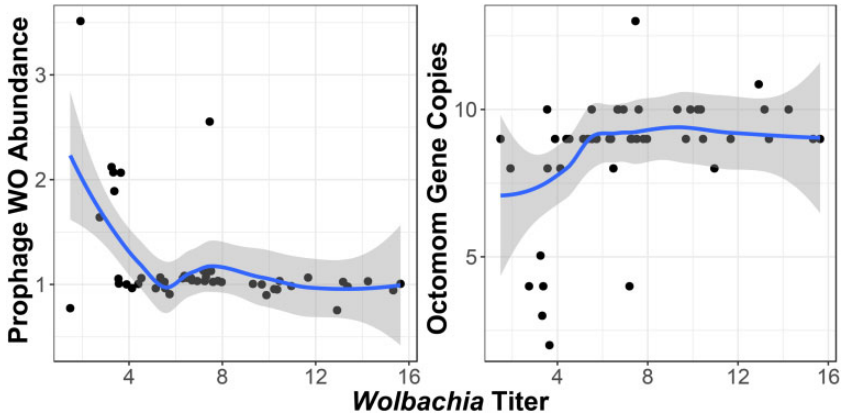

FIG. 5. Prophage abundance genes (prophage coverage/total *Wolbachia* coverage) compared with the *Wolbachia* titer (total *Wolbachia* coverage/host autosomal coverage) and Octomom copies compared with *Wolbachia* titer (total *Wolbachia* coverage/host autosomal coverage). A LOESS smoothed regression was added to the data for each plot to assist in visualizing the change in gene copy number with titer, showing the fitted curve of the average Octomom copy per 2-fold *Wolbachia* titer.

individuals (supplementary fig. 4, Supplementary Material online). In other words, if a mitochondrial site is polymorphic for A/C and a *Wolbachia* site is polymorphic for G/A, we find all pairwise combinations in the population (AA, AG, CA, CG). The two larger clusters of SNPs are contained within the mobile prophage WO-A and WO-B portions of the genome.

We also find the sequence coverage of the prophage regions differs between *w*Inn lines: the prophage region has significantly higher sequence coverage in eight of 48 lines, varying from 1× to 4× the average coverage of the *Wolbachia* genome (Wilcoxon rank sum $W > 443{,}221$, $P<0.04$, fig. 5). In addition, phage copy number is negatively correlated with *Wolbachia* titer (fig. 5, Spearman's $\rho = -0.584$ $t = -7.831$, $P = 2.35\text{e-}10$). Indeed, higher phage coverage correlates with lower *Wolbachia* titer up to a point, consistent with potentially active phage lysing symbiont cells.

Given the signals of selection on Octomom genes found here, and the previously identified association between increased pathogenicity and Octomom copy number in lab strains (Chrostek and Teixeira 2015), we also examined the correlation between Octomom sequencing coverage and *Wolbachia* titer within lines. We found all strains contain Octomom genes, some of which are in a cassette, and some not. The sequence coverage of the Octomom genes appears to also differ between lines (between 0× and 16× coverage, fig. 5). We note that while *Wolbachia* titer is not significantly different as Octomom coverage changes (a proxy for copy number, fig. 5, Spearman's $\rho = 0.115$, GLM $t = 1.930$, $P = 0.0597$), it is difficult to infer copy number when sequencing coverage is low (as it is in lower *Wolbachia* titer individuals) (Hill and Unckless 2020a). The lack of resolution precludes any conclusions on their relationship, as it is possible to falsely infer Octomom copy number in some samples. Variance in Octomom coverage does not significantly relate to coverage (GLM $t = -0.311$, $P = 0.0637$). Conservatively, we cannot confirm a relationship between Octomom copy number and *Wolbachia* titer in this data set, and any potential relationships will need to be confirmed in the future. The

differences in titer are not significantly associated with any SNPs or gene duplications outside of the Octomom or prophage WO regions ($P > 0.186$).

Discussion

We sequenced and assembled the genome of the *Wolbachia* strain infecting *Drosophila innubila*, *w*Inn. *w*Inn is one of the few strains of *Wolbachia* to cause male killing in *Drosophila* (Jaenike et al. 2003) and so we sought to examine its population genome dynamics, with particular focus on prophage and Octomom regions that have genes putatively or empirically involved in *Wolbachia* pathogenicity (Metcalf et al. 2014; Chrostek and Teixeira 2015; Bordenstein and Bordenstein 2016; Beckmann et al. 2017; LePage et al. 2017; Perlmutter et al. 2019). The genome content and dynamics of *w*Inn are largely like other closely related strains (fig. 1), sharing most of its genome with similar supergroup A *Wolbachia* strains, as previously reported (Ishmael et al. 2009). When comparing the genomes of *w*Inn and *w*Rec (capable of male killing) to the other closely related strains in this analysis (all CI-inducing), we identify only three unique genes. However, these genes are found in other *Wolbachia* strains not used in our initial analyses, based on the nonredundant BLAST database (Altschul et al. 1990; Pruitt et al. 2005), suggesting that these genes are more broadly found in nonmale-killing *Wolbachia* and were likely lost in the *w*Mel and *w*Ri strains used here. The lack of unique male-killer genes is consistent with the idea that male killing is often hidden. Indeed, many strains like *w*Rec do not cause male killing in their native hosts but do kill males when transferred to other hosts or vice versa (Fujii et al. 2001; Sasaki et al. 2002; Jaenike 2007; Hughes and Rasgon 2014). In addition, there is evidence of host resistance that suppresses the phenotype in many other systems, where the arms race between host and bacteria leads to loss of phenotype (Hornett et al. 2006; Jaenike 2007; Majerus and Majerus 2010). These factors indicate that absence of phenotype does not necessarily correlate with absence of symbiont genotype, and male killing is instead also heavily dependent on host background (i.e., male killing is not a simple matter of presence/absence of a male-killing toxin gene). The fact that the *wmk* male-killing gene candidate is found in many nonmale-killer genomes also supports the idea that male-killers do not necessarily have unique genetic content involved in the phenotype and a combination of host- and microbe-dependent factors are necessary for the phenotype to occur (Perlmutter et al. 2019).

We also find that while the overall genetic content of *w*Inn is like *w*Mel, it has key differences compared with the more similar *w*Rec strain. The *w*Inn and *w*Mel strains share 57 total genes (of which 21 were prophage genes) that are unique among those strains analyzed (including *w*Rec). These differences are largely due to the loss of phage regions in *w*Rec (Metcalf et al. 2014). It is intriguing to note that *w*Rec contains relic phage regions that have lost many genes compared with those in *w*Mel and *w*Inn, and is likely unable to produce viral particles because of the absence of key virus structural genes (Metcalf et al. 2014). Meanwhile, *w*Inn and *w*Mel have not (fig. 1). Both *w*Inn and *w*Rec are closely related, both are capable of male killing (Jaenike et al. 2003; Metcalf et al. 2014), and both infect mycophagous species, yet one has an eroded prophage region whereas the other does not, which may potentially be due to differences in phenotypes in their native hosts (CI vs. male killing, fig. 1). In addition, evidence here based on higher sequence coverage and the inverse relationship between phage titer and *Wolbachia* titer, further suggests that there may cautiously be active phage WO particle production in *w*Inn (supplementary fig. 5, Supplementary Material online). This remains to be confirmed via methods such as phage purification or EM imaging. If accurate, the variation in prophage coverage across samples may indicate that putative active phage particles are more abundant or in a different part of their replication cycle in each sample (the samples were not controlled for age or other factors that may affect the viral titer), or it could indicate variance in the number of phage insertions across the genomes of each sample. Similarly, the variation in Octomom sequence coverage across samples may also indicate different copy numbers across lines. It is unclear why *w*Inn and *w*Mel would putatively maintain viable phage particle production whereas *w*Rec would not. Future research will be needed to determine any functional consequences of phage particles that may be playing a role in their retainment or loss across different systems, and what role they may play in parasitic phenotypes.

We also identified 40 genes unique to the *w*Inn genome, and most intriguingly, 28 of these are homologous to *Formica* wood ant genes, an additional five share homology with *V. destructor* mites, and there are several TEs with homology to those in *Camponotus* ants. Along with evidence of horizontal gene transfer and rapid evolution, this homology with divergent hosts would suggest some possibilities for the genetic transfer routes in this system. Indeed, mites are known to transfer *Wolbachia* infections among *Drosophila* populations (Brown and Lloyd 2015), and *V. destructor* and its relatives are common parasitic species found throughout the United States and the rest of the world (Rosenkranz et al. 2010). Therefore, it is possible that this or a similar species has vectored either the entire *Wolbachia* symbiont or some genes among various arthropods, contributing to horizontal gene transfer in this strain. *Formica* wood ants and *Camponotus* ants are also common across the United States (Bolton et al. 2006), indicating that there is a possibility of interaction with the mites and/or *D. innubila*, although the transfers may have occurred before this *Wolbachia* strain established itself in its current host. The homology of *w*Inn genes with ant genes may indicate that there has been an exchange of genes or symbionts among these hosts, possibly via mites, and the ants and mites in the area may contain or have contained similar strains. In one possible model, the mites would vector either genes or symbionts among interacting hosts (Houck et al. 1993; Brown and Lloyd 2015), and in other models, the *w*Inn strain may have originated in these hosts. The potentially active phage particles could also play a role by primarily or excessively moving the prophage region among hosts, which would likely be much easier and more common than symbiont exchange. Given that the coverage of the

phage region differs between samples, it would suggest the possibility of phage actively forming and lysing bacterial cells.

Horizontal transfer of genes between *Wolbachia* strains and hosts is in line with existing literature demonstrating *Wolbachia*'s proclivity for genetic exchange. Indeed, the horizontal transfer of individual genes or the entire phage region among *Wolbachia* strains that is supported here reflects previously reported cases (Wang et al. 2016; Cooper et al. 2019). The presence of 13 genes with elevated dS in *w*Inn comparisons suggests horizontal acquisition from another *Wolbachia*, as they are divergent from recent ancestors but found in other strains. However, the genes that appear to have been horizontally acquired, are like other genes in various *Wolbachia* strains. The enrichment for specific phage WO genes identified here suggests they horizontally transfer between *Wolbachia* genomes more often than the rest of the genome, possibly due to the activity of putative phage particles (table 1 and fig. 3), or the activity of surrounding mobile elements, as may have happened in the *D. yakuba* clade with horizontal transfer of the CI loci and nearby transposons (Cooper et al. 2019).

Further, entire symbiont transfer may also potentially occur in this system, although it is rare if it does occur. Vertically transmitted symbionts, such as *Wolbachia*, are subject to strong selection within the host, and unlike frequently horizontally transferring microorganisms, can experience various degrees of genome reduction and other genetic adaptations that lead to essential ties with a specific host (Moran 2002; Dyer and Jaenike 2005; Jaenike and Dyer 2008; McCutcheon and Moran 2012). However, bypassing this phenomenon, there are *Wolbachia* strains that can transfer to new hosts via various modes of transmission and then sweep rapidly across a new and sometimes divergent host population (Riegler et al. 2005; Baldo et al. 2008; Turelli et al. 2018; Sanaei et al. 2021). Whole symbiont transmission could be aided by frequent horizontal transfer of genes or entire regions, such as the prophage, just as we report here in *w*Inn. More broadly, acquisition of new genetic variants that the eukaryotic host is unfamiliar with may confer an advantage that could allow the *Wolbachia* to be transferred to a new host or maintained in an existing host. Indeed, some known cases of horizontal phage WO gene transfer among symbionts have functional effects on the symbiont's ability to parasitize the host (Wang et al. 2016; Cooper et al. 2019). Most crucially, horizontal gene transfer in *Wolbachia* is not restricted to exchange among phages or bacteria, but also with the eukaryotic host. Phage WO stands unique among other described phages with its regions of large genes containing eukaryotic-like domains that imply both lateral transfer between animal and phage WO genomes and potential interaction with the eukaryotic host (Bordenstein and Bordenstein 2016). Thus, the acquisition of genes homologous to those in ants and mites in *w*Inn reflects *Wolbachia*'s unique genetic exchanges more broadly as well as its tripartite interactions among phages, bacteria, and animals. Elucidating the function and fitness impacts of these genes, if any, will be an interesting area of future research. In addition, if there is indeed frequent horizontal exchange in this system, this may then be a driving force for maintenance of phage particles, as they may confer selective advantages over time. Further supporting the idea of recurrent genetic exchanges in the *w*Inn ecosystem is evidence of both a complete hAT element, suggesting recent transmission of this element, and more degraded LINE elements homologous to those in *Formica* and *Varroa*, suggesting more ancient acquisition.

We find repeat content in the *w*Inn genome that is similar other *Wolbachia* (fig. 1A; supplementary table 2, Supplementary Material online; Wu et al. 2004; Ishmael et al. 2009). This contrasts with other obligate intracellular parasites that have relatively small, repeat free genomes (Woolfit et al. 2013). Most of the elements found in the *w*Inn genome are also shared with the host, *D. innubila* (Hill et al. 2019), suggesting that inefficient selection has not allowed these TEs to be maintained for extensive periods of time, but instead these elements are recent acquisitions (Yoshiyama et al. 2001). It is possible that eventually these TEs will be shed from the *w*Inn genome, as similar repeat families have been acquired in the past and eventually gone extinct and removed from the genome, in a cyclical fashion (Maruyama and Hartl 1991; Lohe et al. 1995). Previous work in *w*MelPop suggests that reduced selection allowed repeats to accumulate in the *Wolbachia* genomes (Woolfit et al. 2013). This could also be the case for *w*Inn, allowing these families to be maintained in the genome for extended periods of time, as opposed to removed immediately. The lack of similarity between *w*Inn and *w*MelPop repeats suggests that mobile elements have not been maintained since the common ancestor of these two *Wolbachia* strains, and that turnover is much more frequent.

Beyond genome content, we analyzed the population genomics of the four populations of *w*Inn. The overall finding, similar to previous findings on prophage biology and the *w*Inn genome (Ishmael et al. 2009; Duplouy et al. 2013), is that some prophage and Octomom genes appear to be under positive selection in all *Wolbachia* branches analyzed, whereas others are not. Additionally, no functional categories were significantly enriched for positive selection, and those closest to the cutoff were involved in DNA repair, DNA binding, host integration, and recombination (supplementary table 3, Supplementary Material online). Looking more specifically at the *cifA*/*B* CI genes and the *wmk* male-killing gene candidate, we find that they are not more rapidly evolving than the rest of the prophage region (fig. 2).

Regarding the Octomom regions, we find that these genes are under positive selection across all the *Wolbachia* genomes in the clade examined, as expected, not just in the male-killing *Wolbachia* (fig. 2). The selection on these genes may indicate they are frequently involved in host–pathogen interactions as suggested elsewhere (Chrostek and Teixeira 2015), and as is seen with immune genes and other genes involved in interspecies arms races (Dawkins and Krebs 1979; Sackton et al. 2007; Obbard et al. 2009; Palmer et al. 2018). Indeed, previous studies have found an association between Octomom copy number and increased titer (Chrostek and Teixeira 2015), as we see here (supplementary fig. 5, Supplementary Material online). Additionally, although prophage may be able to

excise themselves for transfer, Octomom may utilize TEs for horizontal transfer (Chrostek and Teixeira 2015). In line with this, the Octomom genes are fragmented across the genome rather than remaining in full cassettes, and the horizontal transfer of TEs is more and more frequently being found to occur between closely species with overlapping ranges (Peccoud et al. 2017; Hill and Betancourt 2018; Wallau et al. 2018). When looking for evidence of recent horizontal transfer to *w*Inn more generally (after divergence from *w*Ha and *w*Ri), 13 genes were identified as potentially horizontally transferred to *w*Inn. Of these, seven were prophage or Octomom genes (fig. 3), supporting all findings so far indicating particularly frequent transfer of prophage and Octomom genes.

Finally, we tested if the samples grouped geographically and compared the inheritance of mitochondria and *w*Inn in *D. innubila* populations (fig. 4) (Hill and Unckless 2020a, 2020b). Indeed, geographically, there is structure to the samples, which mostly group by capture location. This would suggest that there little to no movement from population to population, or that at least individuals do not reproduce after migrating. Low F_{ST} levels across the genome indicate that the structure found among the population groupings is not driven by any region. However, we also find some phylogenetic discordance in inheritance in the *Wolbachia* prophage region compared with host mitochondria, which could be due to imperfect co-inheritance, recurrent mutation, or a horizontal transfer event. Because the discordance is primarily associated with the prophage regions, the pattern is most likely due to horizontal prophage movement, rather than truly incongruous co-inheritance of symbiont and mitochondria in the population (fig. 5). This suggests that the prophages are likely transferring between symbiont genomes within their *D. innubila* populations, causing recombination-like signatures and resulting in the particularly high levels of horizontal transfer of the prophage region seen here (fig. 4, table 1, supplementary fig. 4, Supplementary Material online). Indeed, if there is active phage as suggested by the high variance in prophage coverage, the active particles would likely aid in this process of transfer between individuals in the population. When examining the phylogeny of *w*Inn genomes, we find some Prescott lines are grouped within Chiricahua lines (supplementary fig. 3, Supplementary Material online), potentially driven by this horizontal gene transfer. Alternatively, since the populations are recently established (Hill and Unckless 2020a, 2020b), it may simply indicate that these lines and mitotypes differentially segregated into these populations upon establishment and that the types dominant in Chiricahuas are rarer in Prescott.

In summary, we provide an updated genome description of *w*Inn of *D. innubila* and use various analyses to understand the population genomic trends of this symbiont. We find that the genome has many unique genes suggestive of lateral gene transfer with divergent eukaryotic hosts, along with likely recent and ancient repetitive elements that also indicate lateral transfer with diverse species has occurred on multiple occasions. It remains to be determined when these transfers occurred, if they occur in both directions, and what unique functional consequences they may have for this strain. We also find no evidence of genomic content unique to strains that can cause male killing, in line with predictions that the phenotype is not determined by microbial factors alone. Future work will need to focus on identifying host factors that may be responsible for the expression or absence of male killing in any given system. Reflecting literature precedent, parts of the prophage and Octomom regions are under positive selection, and we find that several genes have likely horizontally transferred into the symbiont from other *Wolbachia*. It is unknown which strains may have been a part of any genetic exchanges with *w*Inn, when, or in what context they may have been in contact. We also found evidence of structured populations that suggest little to no host migration between populations, high fidelity in mitochondria–*w*Inn co-inheritance, and a high likelihood of frequent prophage exchange among individuals within populations. It is unknown if most or all other male-killers exhibit such clear demographic structure among populations, and if this plays a role in the long-term maintenance of male killing. In light of the fact that not all strains that can kill males form phage particles (*w*Rec has a relic phage), it remains unknown if putative particles and common exchange of prophages among individuals plays an important role in parasitic phenotypes generally or among male-killers specifically. Finally, we provide preliminary evidence that there may be active phage particles among individuals that potentially aid in the dynamics described here. The presence or absence of phage particles and exact numbers of Octomom copies per sample will need to be confirmed in later work. Moving forward, it will be important to compare the population genomics of other male-killing strains to determine if they reflect the same principles uncovered here that may contribute to the success of *Wolbachia* symbionts in sex-biased populations or may reflect a broader strategy for survival and adaptation in diverse *Wolbachia* around the globe.

Materials and Methods

Genome Sequence of *w*Inn

For a single *Wolbachia*-positive strain described previously (Unckless and Jaenike 2012), we extracted DNA following the protocol described in (Chakraborty et al. 2018). We prepared the DNA as a sequencing library using the Oxford Nanopore Technologies Rapid 48-h (SQK-RAD002) protocol, which was then sequenced using a MinION (Oxford Nanopore Technologies, Oxford, UK; NCBI SRA: TBD) (Jain et al. 2016). The same DNA was also used to construct a fragment library with insert sizes of ~180 bp, and we sequenced this library on an Illumina HiSeq 4000 (150 bp paired-end, Illumina, San Diego, CA, NCBI SRA: TBD).

Oxford Nanopore sequencing read bases were called post hoc using the built in read_fast5_basecaller.exe program with options: –f FLO-MIN106 –k SQK-RAD002 –r–t 4. We assembled the raw Oxford Nanopore sequencing reads alongside the Illumina paired-end short sequencing reads using SPAdes version 3.13.0 (Bankevich et al. 2012), which generated an initial assembly of 83 contigs with homology to *Wolbachia*.

We then attempted to improve this initial assembly using the 83 assembled fragments, along with Nanopore sequencing reads and Illumina paired-end short sequencing reads in MaSuRCA version 3.4.1 (Zimin et al. 2013), defining the expected genome size as 1.5 million bp. This produced a single contig that was 1,286,799 bp long. We confirmed this contig was circular through aligning long-reads to this assembly and validating by eye. We then used Racon version 1.4.3 to polish the genome with minion fragments for three iterations (Walker et al. 2014) and further polished with Pilon version 1.23 for three iterations using the short read data (Vaser et al. 2017). We then verified the contiguity of the assembly using BUSCO version 3.0 (Simão et al. 2015). From a search for 221 proteobacteria orthologs, we found 181 complete single-copy orthologs and two fragmented orthologs (compared with 180 complete and two fragmented for the published *w*Mel genome: NC_002978.6).

Fly Collections and *Wolbachia* Infection Confirmation

We collected wild *Drosophila* at four mountainous locations across Arizona between August 22, 2017 and September 11, 2017 (Hill and Unckless 2020a, 2020b). Specifically, we collected at the Southwest research station in the Chiricahua mountains (31.871 latitude, −109.237 longitude), Prescott National Forest (34.540 latitude, −112.469 longitude), Madera Canyon in the Santa Rita mountains (31.729 latitude, −110.881 longitude), and Miller Peak in the Huachuca mountains (31.632 latitude, −110.340 longitude). Baits consisted of store-bought white button mushrooms (*Agaricus bisporus*) placed in large piles about 30 cm in diameter, at least five baits per location. A sweep net was used to collect flies over the baits in either the early morning or late afternoon between one and three days after the bait was left. Flies were sorted by sex and species based on morphology at the University of Arizona and were flash frozen at −80 °C before being shipped on dry ice to Lawrence, KS. Specifically, we separated individuals likely to be *Drosophila innubila* from the rest of the collections for further processing and genetic confirmation of species identification.

We further analyzed the 343 *D. innubila* flies which we homogenized and extracted DNA from using the Qiagen Gentra Puregene Tissue kit (USA Qiagen Inc., Germantown, MD) (Hill and Unckless 2020a, 2020b). We tested these samples for infection using *Wolbachia* primers specific to the *Wolbachia surface protein* (*wsp*) gene alongside a positive and negative control (Zhou et al. 1998).

The reaction mixture for the *wsp* PCR consisted of 1uL DNA, 1 μ 10× buffer (Solis Biodyne), 1.0 µl of 20 mM $MgCl_2$ (Solis Biodyne), 1 µl of dNTPs (20 µM each), 0.5 µl of forward (F) primer (81F 5′-TGGTCCAATAAGTGATGAAGAAAC-3′, 20 µM), 0.5 µl of reverse (R) primer (691R 5′-AAAAATTAAACGCTACTCCA-3′, 20 µM), 0.5 µl of Taq DNA polymerase (5 U/µl) (Solis Biodyne), and water to make up the final volume of 10 µl. The amplification reaction consisted of one cycle of 1 min at 94 °C, 1 min at 58 °C, and 2 min at 72 °C, followed by 35 cycles of 15 s at 94 °C, 1 min at 58 °C, and 2 min at 72 °C, and one cycle of 15 s at 94 °C, 1 min at 58 °C, and 7 min at 72 °C. These conditions yielded 610 bp PCR products, which we observed running out the product on a 1% agarose TAE gel. This survey yielded 48 *Wolbachia*-positive lines.

For the 48 *Wolbachia*-infected lines we previously extracted DNA and sequenced the host and *Wolbachia* genomes on two runs of an Illumina HiSeq 4000 (150 bp paired-end; Hill and Unckless 2020a, 2020b; Illumina, San Diego, CA), producing an average of 20,618,752 reads per sample, of which an average of 436,527 (∼85× coverage) mapped to *Wolbachia* per sample, as summarized in supplementary table 1, Supplementary Material online.

Genome Annotation

We annotated the *w*Inn genome using Prokka version 1.15.4 (Seemann 2014), detecting 2,686 total genes, of which 1,341 were retained following size and quality filtering ($>$ 50 bp, quality score $>$ 20, supplementary table 2, Supplementary Material online). Using this annotation of the genome, we extracted coding sequences and generated amino acid sequences using GFFread version 0.12.1 (Pertea and Pertea 2020). We also downloaded the coding sequence and amino acid sequences for open reading frames in the *Wolbachia* of *Drosophila melanogaster* (Canton S strain) (*w*Mel-CS, SAMN02604000), the *Wolbachia* of *Drosophila simulans* (Riverside strain) (*w*Ri, SAMN02603205), the *Wolbachia* of *Drosophila simulans* (Hawaii strain) (*w*Ha, SAMN02604273), and the *Wolbachia* of *C. pipiens* (*w*Pip, SAMN02296948). We used BlastP version 2.9.0 (Altschul et al. 1990) to identify orthologs for these genes in *w*Inn (parameters: hsp = 1, num_alignments = 1, e-value$<$0.00001). For each set of orthologs we created a gene alignment using MAFFT version 7.409 (parameters: –auto) and for 100 randomly chosen genes made a visual inspection of amino acid sequences to confirm similarity of putative orthologous sequences. We then verified the completeness of the extracted amino acids sequences using BUSCO version 3.0 (Simão et al. 2015). We then compared the orthologs with the published *w*Mel genome: NC_002978.6.

To annotate the repetitive content of the *w*Inn genome, we used RepeatModeler version 2.0.1 (Smit and Hubley 2008–2015) and RepeatMasker version 4.0.9 (parameters: -gff -gccalc -s -norna) (Smit et al. 2013–2015).

Genomic Variation in *w*Inn

For all 48 *Wolbachia*-positive lines collected in 2017, we mapped short reads to the *D. innubila* genome (Hill et al. 2019), masked using RepeatMasker version 4.0.9 (parameters: -gff -gccalc -pa 4 -s) (Smit et al. 2013–2015), a custom library of *D. innubila* repeats (Hill et al. 2019), and the masked *w*Inn genome using BWA MEM version 0.7.17-r1188 (Li and Durbin 2009) and SAMtools version 1.9 (Li et al. 2009). We then extracted aligned reads mapping to *w*Inn and used GATK version 4.0.0 to remove optical and PCR duplicates and realign around indels (McKenna et al. 2010; DePristo et al. 2011). We then called variants in the *w*Inn genome of each *Wolbachia*-positive lines using GATK HaplotypeCaller version 4.0.0 (McKenna et al. 2010; DePristo et al. 2011), considering only variants with a quality score greater than 500. Finally, we

combined VCFs using BCFtools version 1.7 (Narasimhan et al. 2016) to create a multiple sample VCF.

Using the generated BAM files, we calculated the coverage of the whole wInn genome relative to the host genome using BEDTools (Quinlan and Hall 2010). We also calculated the number of copies of prophage and Octomom per *Wolbachia* genome again using BEDTools. Following this, we fit a GLM to identify correlations between prophage and Octomom copy number and *Wolbachia* titer. We also fit a LOESS regression for visualization of these correlations in R (R Core Team 2020).

Detection of Selection on *Wolbachia* Genes

For each *w*Inn gene with an ortholog in *w*Ha and *w*Ri, we generated an alignment of the coding sequence of each gene using MAFFT version 7.409 (parameters: –auto). Following this alignment, we reformatted the alignment into a PAML version 1.3.1 usable format and generated a gene tree using PRANK version 0.170427 (parameters: +F -showtree -d = paml) (Löytynoja 2014). We next used codeML (Yang 2007) to calculate the nonsynonymous divergence (d*N*) and synonymous divergence (d*S*) across the entire gene tree and find the best fitting branches model (Model 7 or 8), as well as calculate d*N* and d*S* on each branch of the tree (Model 1), specifically looking at the estimates of d*N*/d*S* on the *w*Inn branch versus all other branches. For both the total tree and specifically the wInn branch, we looked for gene functional categories with higher d*N*/d*S* than all other genes, after controlling for gene length. We fit a GLM, comparing the d*N*/d*S* on the *w*Inn branch versus all other branches for all genes, including the genes functional group as a cofactor, Reporting the *t* and *P* values when relevant or significant. We also fit a GLM comparing the d*N*/d*S* distribution for specific genes of interest (functional groups, or suspected male-killing associated genes) to the remaining genes in the genome, again reporting the *t* and *P* values, which were two-tailed (R Core Team 2020).

Population Structure across *w*Inn Populations

For synonymous sites in the VCF, we used VCFtools version 0.1.16 (Danecek et al. 2011) to calculate the F_{st} between each population and the other populations (Brown 1970). We also performed a principal component analysis on the variation found across the samples in R version 3.5.1 (R Core Team 2013), using the VCF input as a presence/absence matrix.

Ancient and Recent Horizontal Transfer

We reasoned that in the absence of horizontal gene transfer, then *Wolbachia* variation would be perfectly linked to mitochondrial variation, whereas horizontal transfer would break that pattern. We wanted to test to determine if the horizontal transfer is occurring exclusively in specific genes and genomic regions, or is occurring across the whole *Wolbachia* genome, supporting movement of *Wolbachia* between organisms, instead of horizontal transfer of specific genes. To assess this, we looked at all pairwise combinations of mitochondrial and *Wolbachia* alleles and recorded sites with all four allele sets across the two loci across the 48 samples (e.g., GT, AT, GC, and AC), giving a recombination-like signature (suggesting nonvertical inheritance). We then counted the number of discordant and nondiscordant SNPs in 10 kb windows across the *w*Inn genome to identify specific sections enriched for discordant SNPs. We used a χ^2 test to identify specific functional categories enriched for discordant SNPs.

For longer term horizontal gene transfer, we used the VHICA R package version 0.2.7 to calculate synonymous divergence (d*S*) for all pairwise for all shared genes for pairwise combinations of *w*Inn, *w*Ha, and *w*Ri (Wallau et al. 2016). We reasoned that horizontal transfer of a gene from a highly divergent *Wolbachia* would produce a signal of increased d*S* between *w*Ha and *w*Inn for that gene and could polarize which species had the horizontal transfer event based on the d*S* of that gene in the pairs *w*Inn–*w*Ha, *w*Ha–*w*Ri, and *w*Inn–*w*Ri. We considered d*S* to be excessively high in a gene if it was greater than the mean d*S* plus the variance for all genes with a similar number of effective codons, as d*S* is on an average higher when the effective number of codons is higher in a gene (five codons window size, sliding five codons) (Wallau et al. 2016). This cutoff was based on the publication first describing VHICA and was used in subsequent publications (Wallau et al. 2016, 2018; Hill and Betancourt 2018). We considered a gene to be a putative horizontal acquisition in *w*Inn (or a recent ancestor) if d*S* of the *w*Inn gene compared with homologs in *w*Ha and *w*Ri is excessively high, but d*S* between the *w*Ha and *w*Ri homologs is not. We then performed a χ^2 test to look for functional categories that are enriched for putatively horizontally acquired genes. For this test, we calculated the expected number of putatively horizontally acquired genes in each gene category based on the total number of putatively horizontal genes and the proportion of the genes which make up this category. If all things being equal and genes are transferring randomly, we would expect the number of putatively horizontal genes in each category to be proportional to the number of genes in the category. We used the expected number and the observed number of putatively horizontally acquired genes in each category to calculate an enrichment and perform the χ^2 test.

Finally, we assessed the extent of ancient horizontal transfer across the *Wolbachia* phylogeny. We downloaded all *Wolbachia* genomes and their annotations from the NCBI genome database (summarized in supplementary table 1, Supplementary Material online). Then, based on the known NCBI annotations, we found groups of orthologous genes. We generated codon alignments for these orthologous genes using MAFFT (parameters: –auto) (Katoh et al. 2002), and generated a gene tree for each gene using PhyML (model = GTR, gamma = 8, bootstraps = 100) (Guindon et al. 2010). We also generated a whole species phylogeny for these genomes and to place *w*Inn on the phylogeny. For all genes found in all species with high confidence (231 genes, totaling 208,911 bp of the genome), we generated a multigene phylogeny with 100 bootstraps using PhyML (model = GTR, gamma = 8, bootstraps = 100). We then attempted to assess the extent of discordance between species and gene trees using CADM.global in APE to test for consistency between phylogenies, with the null hypothesis that the phylogenies are

different across 100,000 permutations per species/gene tree comparison (so a significant *P* value will suggest little discordance between phylogenies). Again, we performed a χ^2 test to look for functional categories that are enriched for putatively horizontally transferred genes. If horizontal transfer is random, we expect the number of horizontally transferring genes in a functional group to depend on the proportion of genes in this category. We therefore calculated the number of expected horizontal transfers per group by multiplying the number of events observed by the proportion of genes in a category. We then used this expected number of horizontal transfers and the observed number to calculate a χ^2 value per group and perform a χ^2 test.

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Supplementary Material

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

Acknowledgments

This work was supported by the Kansas Idea Network of Biomedical Research Excellence (K-INBRE, National Institutes of Health, Grant No. P20 GM103418, one each to T.H. and J.I.P.), the National Science Foundation (NSF Postdoctoral Research Fellowship in Biology, Grant No. DBI 2109772 to J.I.P.), and the National Institutes of Health (Grant Nos. R00 GM114714 and R01 AI139154 to R.L.U., Grant No. P20 GM103638 to KU CMADP Genome Sequencing Core). This manuscript was conceived and written thanks to suggestions and feedback from the Unckless lab at the University of Kansas. Collections were completed with assistance from Todd Schlenke at the University of Arizona, and the Southwest Research Station. The initial *w*Inn sample used for genome sequencing was graciously provided by John Jaenike. We thank Brittny Smith and Jenny Hackett at the KU CMADP Genome Sequencing Core and K-INBRE Bioinformatics Core for assistance in genome isolation, library preparation, sequencing, and computational resources. We thank Brandon Cooper for helpful comments on the manuscript.

Author Contributions

T.H. devised analyses, performed genome assembly, performed analyses, and wrote the manuscript. J.I.P. and R.L.U. devised analyses and wrote the manuscript.

Data Availability

The *w*Inn genome and annotation are available on the NCBI genome database under the accession SAMN18306550. All sequencing data used in this study are available on the NCBI SRA under the project accession PRJNA524688. Additional data regarding *Drosophila innubila* population genomics are available in the following FigShare folders: https://doi.org/10.6084/m9.figshare.13256102.v1, https://doi.org/10.6084/m9.figshare.13256099.v1, https://doi.org/10.6084/m9.figshare.13256105.v1, https://doi.org/10.6084/m9.figshare.13256123.v1, and https://doi.org/10.6084/m9.figshare.13256153.v1. Additional data is in the following DataDryad submission: https://datadryad.org/stash/share/wvfmDL39pdYrVUcgDFAfl33BOJu3KCJWuJyj-0M-qgA.

References

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

Baldo L, Ayoub NA, Hayashi CY, Russell JA, Stahlhut JK, Werren JH. 2008. Insight into the routes of *Wolbachia* invasion: high levels of horizontal transfer in the spider genus *Agelenopsis* revealed by *Wolbachia* strain and mitochondrial DNA diversity. *Mol Ecol*. 17:557–569.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham SON, Prjibelski AD, Pyshkin AV. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comp Biol*. 19:455–477.

Beckmann JF, Ronau JA, Hochstrasser M. 2017. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nat Microbiol*. 2:17007.

Beckmann JF, Sharma GD, Mendez L, Chen H, Hochstrasser M. 2019. The *Wolbachia* cytoplasmic incompatibility enzyme CidB targets nuclear import and protamine-histone exchange factors. *eLife* 8:e50026.

Berec L, Maxin D, Bernhauerova V. 2016. Male-killing bacteria as agents of insect pest control. *J Appl Ecol*. 53:1270–1279.

Bolton B, Alpert G, Ward PS, Naskrecki P. 2006. Bolton's catalogue of ants of the world. Cambridge: Harvard Press.

Bordenstein SR, Bordenstein SR. 2016. Eukaryotic association module in phage WO genomes from *Wolbachia*. *Nat Commun*. 7:13155.

Bordenstein SR, Wernegreen JJ. 2004. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol*. 21(10):1981–1991.

Bouchon D, Rigaud T, Juchault P. 1998. Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. *Proc Biol Sci*. 265(1401):1081–1090.

Brown AHD. 1970. The estimation of Wright's fixation index from genotypic frequencies. *Genetica* 41(3):399–406.

Brown AN, Lloyd VK. 2015. Evidence for horizontal transfer of *Wolbachia* by a *Drosophila* mite. *Exp Appl Acarol*. 66(3):301–311.

Chakraborty M, VanKuren NW, Zhao R, Zhang X, Kalsow S, Emerson J. 2018. Hidden genetic variation shapes the structure of functional elements in *Drosophila*. *Nat Genet*. 50(1):20–25.

Chrostek E, Marialva MS, Esteves SS, Weinert LA, Martinez J, Jiggins FM, Teixeira L. 2013. *Wolbachia* variants induce differential protection to viruses in *Drosophila melanogaster*: a phenotypic and phylogenomic analysis. *PLoS Genet*. 9:

Chrostek E, Teixeira L. 2015. Mutualism breakdown by amplification of *Wolbachia* genes. *PLoS Biol*. 13:e1002065.

Cooper BS, Ginsberg PS, Turelli M, Matute DR. 2017. *Wolbachia* in the *Drosophila yakuba* complex: pervasive frequency variation and weak cytoplasmic incompatibility, but no apparent effect on reproductive isolation. *Genetics* 205(1):333–351.

Cooper BS, Vanderpool D, Conner WR, Matute DR, Turelli M. 2019. *Wolbachia* acquisition by *Drosophila yakuba*-clade hosts and transfer of incompatibility loci between distantly related *Wolbachia*. *Genetics* 212(4):1399–1419.

Cordaux R, Pichon S, Ling A, Pérez P, Delaunay C, Vavre F, Bouchon D, Greve P. 2008. Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. *Mol Biol Evol*. 25(9):1889–1896.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.

Darby AC, Choi JH, Wilkes T, Hughes MA, Werren JH, Hurst GD, Colbourne JK. 2010. Characteristics of the genome of *Arsenophonus nasoniae*, son-killer bacterium of the wasp *Nasonia*. *Insect Mol Biol*. 19:75–89.

Dawkins R, Krebs JR. 1979. Arms race between and within species. *Proc R Soc B*. 205:489–511.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 43(5):491–498.

Duarte EH, Carvalho A, López-Madrigal S, Costa J, Teixeira L. 2021. Forward genetics in *Wolbachia*: regulation of *Wolbachia* proliferation by the amplification and deletion of an addictive genomic island. *PLoS Genet*. 17(6):e1009612.

Duplouy A, Iturbe-Ormaetxe I, Beatson SA, Szubert JM, Brownlie JC, McMeniman CJ, McGraw EA, Hurst GDD, Charlat S, O'Neill SL, et al. 2013. Draft genome sequence of the male-killing *Wolbachia* strain *w*Bol1 reveals recent horizontal gene transfers from diverse sources. *BMC Genomics* 14:1–13.

Duron O, Lagnel J, Raymond M, Bourtzis K, Fort P, Weill M. 2005. Transposable element polymorphism of *Wolbachia* in the mosquito *Culex pipiens*: evidence of genetic diversity, superinfection and recombination. *Mol Ecol*. 14(5):1561–1573.

Dutta C, Pan A. 2002. Horizontal gene transfer and bacterial diversity. *J Biosci*. 27(1 Suppl 1):27–33.

Dyer K, Jaenike J. 2005. Evolutionary dynamics of a spatially structured host-parasite association: *Drosophila innubila* and male-killing *Wolbachia*. *Evolution* 59(7):1518–1528.

Dyer KA, Jaenike J. 2004. Evolutionarily stable infection by a male-killing endosymbiont in *Drosophila innubila*: molecular evidence from the host and parasite genomes. *Genetics* 168(3):1443–1455.

Dyer KA, Minhas MS, Jaenike J. 2005. Expression and modulation of embryonic male-killing in *Drosophila innubila*: opportunities for multilevel selection. *Evolution* 59(4):838–848.

Dyson E, Kamath M, Hurst G. 2002. *Wolbachia* infection associated with all-female broods in *Hypolimnas bolina* (Lepidoptera: Nymphalidae): evidence for horizontal transmission of a butterfly male killer. *Heredity* 88(3):166–171.

Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, et al. 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol*. 3(4):e121.

Francoeur A. 1973. Révision taxonomique des éspèces nearctique du group fusca, genre *Formica* (Formicidae: Hymenoptera). *Mém Soc Entomol Qué*. 3:1–316.

Fujii Y, Kageyama D, Hoshizaki S, Ishikawa H, Sasaki T. 2001. Transfection of *Wolbachia* in Lepidoptera: the feminizer of the adzuki bean borer *Ostrinia scapulalis* causes male killing in the Mediterranean flour moth *Ephestia kuehniella*. *Proc R Soc B*. 268:855–859.

Funkhouser-Jones LJ, Sehnert SR, Martínez-Rodríguez P, Toribio-Fernández R, Pita M, Bella JL, Bordenstein SR. 2015. *Wolbachia* co-infection in a hybrid zone: discovery of horizontal gene transfers from two *Wolbachia* supergroups into an animal genome. *PeerJ* 3:e1479.

Garcia-Arraez MG, Masson F, Escobar JCP, Lemaitre B. 2019. Functional analysis of RIP toxins from the *Drosophila* endosymbiont *Spiroplasma poulsonii*. *BMC Microbiol*. 19:1–10.

Gerth M, Gansauge M-T, Weigert A, Bleidorn C. 2014. Phylogenomic analyses uncover origin and spread of the *Wolbachia* pandemic. *Nat Commun*. 5:1–7.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59:307–321.

Haine ER, Pickup NJ, Cook JM. 2005. Horizontal transmission of *Wolbachia* in a *Drosophila* community. *Ecol Entomol*. 30:464–472.

Harumoto T, Lemaitre B. 2018. Male-killing toxin in a bacterial symbiont of *Drosophila*. *Nature* 557:252.

Hedges LM, Brownlie JC, Neill SL, Johnson KN. 2008. *Wolbachia* and virus protection in insects. *Science* 322(5902):702.

Hertig M, Wolbach SB. 1924. Studies on *Rickettsia*-like micro-organisms in insects. *J Med Res*. 44(3):329–374.7.

Hill T, Betancourt A. 2018. Extensive horizontal exchange of transposable elements in the *Drosophila pseudoobscura* group. *Mob DNA*. 20:1–14.

Hill T, Koseva B, Unckless RL. 2019. The genome of *Drosophila innubila* reveals lineage-specific patterns of selection in immune genes. *Mol Biol Evol*. 36(7):1405–1417.

Hill T, Unckless R. 2020a. Adaptation, ancestral variation and gene flow in a 'Sky Island' *Drosophila* species. *Mol Ecol*. 30:83–99.

Hill T, Unckless R. 2020b. Recurrent evolution of high virulence in isolated populations of a DNA virus. *eLife* 9:e58931.

Hornett EA, Charlat S, Duplouy AM, Davies N, Roderick GK, Wedell N, Hurst GD. 2006. Evolution of male-killer suppression in a natural population. *PLoS Biol*. 4:e283.

Hotopp JCD, Clark ME, Oliveira DC, Foster JM, Fischer P, Torres MCM, Giebel JD, Kumar N, Ishmael N, Wang S. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317:1753–1756.

Houck MA, Clark JB, Peterson KR, Kidwell MG. 1993. Possible horizontal transfer of *Drosophila* genes by the mite *Protolaelaps regalis*. *Science* 253:1125–1129.

Hughes GL, Rasgon JL. 2014. Transinfection: a method to investigate *Wolbachia*-host interactions and control arthropod-borne disease. *Insect Mol Biol*. 23:141–151.

Hurst GD, Frost CL. 2015. Reproductive parasitism: maternally inherited symbionts in a biparental world. *Cold Spring Harb Perspect Biol*. 7:a017699.

Hurst GDD, Jiggins FM, Hinrich Graf von der Schulenburg J, Bertrand D, West SA, Goriacheva II, Zakharov IA, Werren JH, Stouthamer R, Majerus MEN. 1999. Male–killing *Wolbachia* in two species of insect. *Proc R Soc B*. 266:735–740.

Hurst GDD, Johnson AP, Schulenburg J, Fuyama Y. 2000. Male-Killing *Wolbachia* in *Drosophila*: a temperature-sensitive trait with a threshold bacterial density. *Genetics* 156(2):699–709.

Ilinsky Y. 2013. Coevolution of *Drosophila melanogaster* mtDNA and *Wolbachia* genotypes. *PLoS One* 8(1):e54373.

Ishmael N, Hotopp JCD, Ioannidis P, Biber S, Sakamoto J, Siozios S, Nene V, Werren J, Bourtzis K, Bordenstein SR, et al. 2009. Extensive genomic diversity of closely related *Wolbachia* strains. *Microbiology* 155(Pt 7):2211–2222.

Iturbe-Ormaetxe I, Burke GR, Riegler M, O'Neill SL. 2005. Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. *J Bacteriol*. 187(15):5136–5145.

Jaenike J. 2007. Spontaneous emergence of a new *Wolbachia* phenotype. *Evolution* 61(9):2244–2252.

Jaenike J, Dyer KA. 2008. No resistance to male-killing *Wolbachia* after thousands of years of infection. *J Evol Biol*. 21:1570–1577.

Jaenike J, Dyer KA, Reed LK. 2003. Within-population structure of competition and the dynamics of male-killing *Wolbachia*. *Evol Ecol Res*. 5:1023–1036.

Jain M, Olsen HE, Paten B, Akeson M. 2016. Erratum to: the Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*. 17(1):11.

Jiggins FM, Hurst GDD, Majerus MEN. 2000. Sex-ratio-distorting *Wolbachia* causes sex-role reversal in its butterfly host. *Proc R Soc B*. 267:69–73.

Kageyama D, Nishimura G, Hoshizaki S, Ishikawa Y. 2002. Feminizing *Wolbachia* in an insect, *Ostrinia furnacalis* (Lepidoptera: Crambidae). *Heredity* 88(6):444–449.

Katoh K, Misawa K, Kuma K-I, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30(14):3059–3066.

Kaur R, Shropshire JD, Cross KL, Leigh B, Mansueto AJ, Stewart V, Bordenstein SR, Bordenstein SR. 2021. Living in the endosymbiotic world of *Wolbachia*: a centennial review. *Cell Host Microbe*. 6:879–893.

Kent BN, Bordenstein SR. 2010. Phage WO of *Wolbachia*: lambda of the endosymbiont world. *Trends Microbiol*. 18(4):173–181.
Kent BN, Salichos L, Gibbons JG, Rokas A, Newton IL, Clark ME, Bordenstein SR. 2011. Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. *Genome Biol Evol*. 3:209–218.
Kondo N, Nikoh N, Ijichi N, Shimada M, Fukatsu T. 2002. Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *Proc Natl Acad Sci U S A*. 99(22):14280–14285.
Laidoudi Y, Levasseur A, Medkour H, Maaloum M, Ben Khedher M, Sambou M, Bassene H, Davoust B, Fenollar F, Raoult D, et al. 2020. An earliest endosymbiont, *Wolbachia massiliensis* sp. nov., strain PL13 from the bed bug (*Cimex hemipterus*), type strain of a new supergroup T. *Int J Mol Sci*. 21:8064.
Lawrence J. 1999. Gene transfer, speciation, and the evolution of bacterial genomes. *Curr Opin Microbiol*. 2(5):519–523.
Leclercq S, Thézé J, Chebbi MA, Giraud I, Moumen B, Ernenwein L, Grève P, Gilbert C, Cordaux R. 2016. Birth of a W sex chromosome by horizontal transfer of *Wolbachia* bacterial symbiont genome. *Proc Natl Acad Sci U S A*. 113(52):15036–15041.
Lefoulon E, Clark T, Borveto F, Perriat-Sanguinet M, Moulia C, Slatko B, Gavotte L. 2020. Pseudoscorpion *Wolbachia* symbionts: diversity and evidence for a new supergroup S. *BMC Microbiol*. 20:1–15.
LePage D, Bordenstein SR. 2013. *Wolbachia*: can we save lives with a great pandemic? *Trends Parasitol*. 29:385–393.
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.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis GR, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
Lindsey AR, Rice DW, Bordenstein SR, Brooks AW, Bordenstein SR, Newton IL. 2018. Evolutionary genetics of cytoplasmic incompatibility genes *cifA* and *cifB* in prophage WO of *Wolbachia*. *Genome Biol Evol*. 10(2):434–451.
Lohe AR, Moriyama EN, Lidholm DA, Hartl DL. 1995. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. *Mol Biol Evol*. 12(1):62–72.
Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. In: Russell DJ, editor. Multiple sequence alignment methods. Totowa (NJ): Humana Press. p. 155–170.
Mains JW, Kelly PH, Dobson KL, Petrie WD, Dobson SL. 2019. Localized control of *Aedes aegypti* (Diptera: Culicidae) in Miami, FL, via inundative releases of *Wolbachia*-infected male mosquitoes. *J Med Entomol*. 56(5):1296–1303.
Majerus TM, Majerus ME. 2010. Intergenomic arms races: detection of a nuclear rescue gene of male-killing in a ladybird. *PLoS Pathog*. 6(7):e1000987.
Maruyama K, Hartl DL. 1991. Evolution of the transposable element Mariner in *Drosophila* species. *Genetics* 128(2):319–329.
Massey JH, Newton IL. 2021. Diversity and function of arthropod endosymbiont toxins. *Trends Microbiol*. Advance Access published July 9, 2021, https://doi.org/10.10.1016/j.tim.2021.06.008.
Masui S, Kuroiwa H, Sasaki T, Inui M, Kuroiwa T, Ishikawa H. 2001. Bacteriophage WO and virus-like particles in *Wolbachia*, an endosymbiont of arthropods. *Biochem Biophys Res Commun*. 283(5):1099–1104.
McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*. 10(1):13–26.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky AM, Garimella KV, Altshuler D, Gabriel SB, Daly MJ, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Proc Int Conf Intell Capit Know Manag Org Learn*. 20:1297–1303.
Metcalf JA, Jo M, Bordenstein SR, Jaenike J, Bordenstein SR. 2014. Recent genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism. *PeerJ* 2:e529.
Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108(5):583–586.
Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 32(11):1749–1751.
O'Neill SL, Giordano R, Colbert A, Karr TL, Robertson HM. 1992. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. *Proc Natl Acad Sci U S A*. 89(7):2699–2702.
O'Neill SL, Ryan PA, Turley AP, Wilson G, Retzki K, Iturbe-Ormaetxe I, Dong Y, Kenny N, Paton CJ, Ritchie SA, et al. 2018. Scaled deployment of *Wolbachia* to protect the community from dengue and other Aedes transmitted arboviruses. *Gates Open Res*. 2:36.
Obbard DJ, Welch JJ, Kim KW, Jiggins FM. 2009. Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet*. 5:e1000698.
Palmer WH, Hadfield JD, Obbard DJ. 2018. RNA interference pathways display high rates of adaptive protein evolution in multiple invertebrates. *Genetics* 4:1585–1599.
Paredes JC, Herren JK, Schupfer F, Marin R, Claverol S, Kuo CH, Lemaitre B, Beven L. 2015. Genome sequence of the *Drosophila melanogaster* male-killing *Spiroplasma* strain MSRO endosymbiont. *mBio* 6:e02437–14.
Peccoud J, Loiseau V, Cordaux R, Gilbert C. 2017. Massive horizontal transfer of transposable elements in insects. *Proc Natl Acad Sci U S A*. 114:4721–4726.
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.
Pertea G, Pertea M. 2020. GFF Utilities: GffRead and GFFCompare. F1000Res. 9:304. Available from: https://github.com/gpertea/gffread.
Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 33:D501–D504.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
R Core Team. 2013. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
R Core Team. 2020. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
Reveillaud J, Bordenstein SR, Cruaud C, Shaiber A, Esen ÖC, Weill M, Makoundou P, Lolans K, Watson AR, Rakotoarivony I. 2019. The *Wolbachia* mobilome in *Culex pipiens* includes a putative plasmid. *Nat Commun*. 10:1–11.
Riegler M, Sidhu M, Miller WJ, O'Neill SL. 2005. Evidence for a global *Wolbachia* replacement in *Drosophila melanogaster*. *Curr Biol*. 15(15):1428–1433.
Rosenkranz P, Aumeier P, Ziegelmann B. 2010. Biology and control of *Varroa destructor*. *J Invert Pathol*. 103:S96–S119.
Ross PA, Turelli M, Hoffmann AA. 2019. Evolutionary ecology of *Wolbachia* releases for disease control. *Annu Rev Genet*. 53:93–116.
Russell JE, Stouthamer R. 2011. The genetics and evolution of obligate reproductive parasitism in *Trichogramma pretiosum* infected with parthenogenesis-inducing *Wolbachia*. *Heredity* 106(1):58–67.
Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. 2007. Dynamic evolution of the innate immune system in *Drosophila*. *Nat Genet*. 39(12):1461–1468.
Sanaei E, Charlat S, Engelstädter J. 2021. *Wolbachia* host shifts: routes, mechanisms, constraints and evolutionary consequences. *Biol Rev Camb Philos Soc*. 96(2):433–453.
Sasaki T, Kubo T, Ishikawa H. 2002. Interspecific transfer of *Wolbachia* between two Lepidopteran insects expressing cytoplasmic incompatibility: a *Wolbachia* variant naturally infecting *Cadra cautella* causes male killing in *Ephestia kuehniella*. *Genetics* 162(3):1313–1319.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069.
Serbus LR, Sullivan W. 2007. A cellular basis for *Wolbachia* recruitment to the host germline. *PLoS Pathog.* 3:e190.
Sheeley SL, McAllister BF. 2009. Mobile male-killer: similar *Wolbachia* strains kill males of divergent *Drosophila* hosts. *Heredity* 102:286–292.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
Sinkins SP, Braig HR, O'Neill SL. 1995. *Wolbachia* superinfections and the expression of cytoplasmic incompatibility. *Proc R Soc B.* 261:325–330.
Smit AFA, Hubley R. 2008–2015. RepeatModeler Open-1.0. Available from: http://www.repeatmasker.org.
Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. Available from: http:/www.repeatmasker.org
Taylor MJ, Bordenstein SR, Slatko B. 2018. Microbe Profile: *Wolbachia*: a sex selector, a viral protector and a target to treat filarial nematodes. *Microbiology* 164(11):1345–1347.
Teixeira L, Ferreira A, Ashburner M. 2008. The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biol.* 6(12):e2.
Turelli M, Cooper BS, Richardson KM, Ginsberg PS, Peckenpaugh B, Antelope CX, Kim KJ, May MR, Abrieux A, Wilson DA, et al. 2018. Rapid global spread of *w*Ri-like *Wolbachia* across multiple *Drosophila*. *Curr Biol.* 28(6):963–971.
Turelli M, Hoffman AA. 1991. Rapid spread of an inherited incompatibility factor in California *Drosophila*. *Nature* 353(6343):440–442.
Unckless RL, Jaenike J. 2012. Maintenance of a male-killing *Wolbachia* in *Drosophila innubila* by male-killing dependent and male-killing independent mechanisms. *Evolution* 66(3):678–689.
Vaser R, Sovic I, Nagarajan N, Sikic M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27(5):737–746.
Vavre F, Fleury F, Lepetit D, Fouillet P, Boulétreau M. 1999. Phylogenetic evidence for horizontal transmission of *Wolbachia* in host-parasitoid associations. *Mol Biol Evol.* 16(12):1711–1723.
Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
Wallau GL, Capy P, Loreto E, Rouzic AL, Hua-Van A. 2016. VHICA, a New method to discriminate between vertical and horizontal transposon transfer: application to the mariner family with in *Drosophila*. *Mol Biol Evol.* 33(4):1094–1109.
Wallau GL, Vieira C, Loreto ÉLS. 2018. Genetic exchange in eukaryotes through horizontal transfer: connected by the mobilome. *Mob DNA.* 9:1–16.
Wang GH, Sun BF, Xiong TL, Wang YK, Murfin KE, Xiao JH, Huang DW. 2016. Bacteriophage WO can mediate horizontal gene transfer in endosymbiotic *Wolbachia* genomes. *Front Microbiol.* 7:1867.
Weinert LA, Araujo-Jnr EV, Ahmed MZ, Welch JJ. 2015. The incidence of bacterial endosymbionts in terrestrial arthropods. *Proc Biol Sci.* 282(1807):20150249.
Werren JH, Baldo L, Clark ME. 2008. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol.* 6(10):741–751.
Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet.* 8(12):973–982.
Woolfit M, Iturbe-Ormaetxe I, Brownlie JC, Walker T, Riegler M, Seleznev A, Popovici J, Rances E, Wee BA, Pavlides J, et al. 2013. Genomic evolution of the pathogenic *Wolbachia* strain, *w*MelPop. *Genome Biol Evol.* 5(11):2189–2204.
Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, et al. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* 2(3):e69.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
Yen JH, Barr AR. 1971. New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens* L. *Nature* 232(5313):657–658.
Yoshiyama M, Tu Z, Kainoh Y, Honda H, Shono T, Kimura K. 2001. Possible horizontal transfer of a transposable element from host to parasitoid. *Mol Biol Evol.* 18(10):1952–1958.
Zabalou S, Riegler M, Theodorakopoulou M, Stauffer C, Savakis C, Bourtzis K. 2004. *Wolbachia*-induced cytoplasmic incompatibility as a means for insect pest population control. *Proc Natl Acad Sci U S A.* 101:15042–15045.
Zhou W, Rousset F, O'Neill S. 1998. Phylogeny and PCR-based classication of *Wolbachia* strains using *wsp* gene sequences. *Proc Biol Sci.* 265:509–515.
Zimin AV, Marcais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677.
Zug R, Hammerstein P. 2012. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS One* 7:e38544.