A single synonymous nucleotide change impacts the male-killing phenotype of prophage WO gene *wmk*

Jessamyn I. Perlmutter[a,b,c#], Jane E. Meyers[a,c], & Seth R. Bordenstein[a,c,d,e#]

[a]Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA

[b]Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas, USA

[c]Vanderbilt Microbiome Innovation Center, Vanderbilt University, Nashville, Tennessee, USA

[d]Department of Pathology, Microbiology, and Immunology, Vanderbilt University, Nashville, Tennessee, USA

[e]Vanderbilt Institute for Infection, Immunology, and Inflammation, Vanderbilt University, Nashville, Tennessee, USA

#Address correspondence to jessamyn.perlmutter@outlook.com and s.bordenstein@vanderbilt.edu

Abstract

Wolbachia are the most widespread bacterial endosymbionts in animals. Within arthropods, these maternally-transmitted bacteria can selfishly hijack host reproductive processes to increase the relative fitness of their transmitting females. One such form of reproductive parasitism called male killing, or the selective killing of infected males, is recapitulated to degrees by transgenic expression of the *WO-mediated killing* (*wmk*) gene. Here, we characterize the genotype-phenotype landscape of *wmk*-induced male killing in *D. melanogaster* using transgenic expression. While phylogenetically distant *wmk* homologs induce no sex-ratio bias, closely-related homologs exhibit complex phenotypes spanning no death, male death, or death of all hosts. We demonstrate that alternative start codons, synonymous codons, and notably a single synonymous nucleotide in *wmk* can ablate killing. These findings reveal previously unrecognized features of transgenic *wmk*-induced killing and establish new hypotheses for the impacts of post-transcriptional processes in male killing variation. We conclude that synonymous sequence changes are not necessarily silent in nested endosymbiotic interactions with life-or-death consequences.

Introduction

Wolbachia are maternally-transmitted, obligate intracellular bacteria primarily residing in the cells of germline tissues in many arthropod species worldwide[1,2]. To facilitate their spread through the host matriline, these bacteria hijack sex ratios, sex determination systems, and embryonic viability to cause various reproductive parasitism phenotypes[3-6]. One such phenotype is male killing, whereby sons of infected females are selectively killed[4,7,8]. Male killing selfishly drives the spread of the bacteria when, for instance, brothers and sisters compete for limited resources, and male lethality yields females (the transmitters) a relative fitness benefit due to reduced competition with fewer siblings[9-12]. This form of reproductive parasitism can have profound impacts on host behavior and evolution, including a sex role reversal[13], host nuclear genome changes to resist the phenotype[14,15], and potentially host population extinction[16]. Population modelling also suggests male killing may be deployed as a control method to crash the population size of arthropod pests and disease vectors[17].

Many lines of evidence suggest a complex relationship between male killing genotype and phenotype. For example, male killing can be suppressed by host background[14,15,18,19], where male hosts exhibit resistance to the phenotype even in the presence *Wolbachia*. In addition, *Wolbachia* that do not cause male killing in one host species can cause male killing in a second host upon introgression or transfer to a naïve host[18,20]. Furthermore, some strains induce death at different host developmental stages (embryo vs. larvae) or across different sex determination systems (ZW lepidopterans, XY dipterans, XO arachnids), and the number of surviving males may vary widely[21-23]. Such findings specify that the expression of male killing, and particularly the genotype-phenotype relationship in these symbioses, is more complex than simply the presence or absence of a male-killing gene.

We recently identified a male-killing candidate gene, *wmk* (*WO-mediated killing*), in the *w*Mel *Wolbachia* strain of *Drosophila melanogaster* based on comparative genomic, transgenic, and cytological approaches[24,25]. It is a putative transcription factor with two predicted helix-turn-helix (HTH) DNA-binding domains, and transgenic expression in *D. melanogaster* embryos recapitulates many cytological and molecular aspects of male killing including accumulation of DNA damage in male embryos that overlaps with sites of dosage compensation activity[21,25,26]. The *wmk* gene and two cytoplasmic incompatibility factor (*cif*) genes that underlie cytoplasmic incompatibility (a parasitism phenotype whereby offspring die in crosses between infected males and uninfected females) occur in the eukaryotic association module (EAM) of prophage WO[24,25,27], which refers to the phage WO genome that is inserted into the bacterial chromosome. The EAM is common in WO phages across several *Wolbachia* strains[28,29] and is rich in genes that are homologous to eukaryotic genes or annotated with eukaryotic functions[24]. As such, the expression of reproductive parasitism genes from the EAM and tripartite interactions between phage WO, *Wolbachia*, and eukaryotic hosts are central to *Wolbachia*'s ability to interact with and modify host reproduction.

The discovery of *wmk* has now enabled investigation of the impacts of genetic variation on the transgenic male-killing phenotype. Although the role of *wmk* in male killing has not yet been assessed in nature, evidence based on native *wmk* genotypes and transgenic *wmk* expression in flies suggest refined intricacy to the interactions between genotype and phenotype. For example, the native *Wolbachia wmk* homologs in three closely-related strains (*w*Mel of *D. melanogaster*, *w*Rec of *D. recens*, and *w*Suzi of *D. suzukii*) have nearly identical nucleotide sequences that vary by one to three SNPs[25], and yet these strains induce different forms of reproductive parasitism. *w*Mel and *w*Rec induce cytoplasmic incompatibility, *w*Rec also causes male killing in some strains

of a sister species, and *w*Suzi is not known to induce any reproductive phenotype[18,30,31]. Previous transgenic testing also demonstrated that addition of nine additional amino acids to the N-terminal region of the Wmk protein at the site of a putative alternative start codon ablates the phenotype[32]. Therefore, the Wmk N-terminus is particularly sensitive to genetic alterations, which is notable since small modifications such as protein tags in this location are often common additions to proteins that do not typically interfere with function. Furthermore, *D. melanogaster* does not harbor *Wolbachia* known to cause male killing, though transgenic *wmk* expression recapitulates the phenotype in this host[25], and the *w*Mel strain is most closely related to the *w*Rec male-killing strain[33]. In addition, some *D. melanogaster* contain male-killing *Spiroplasma* symbionts whose plasmid gene *Spaid* can also transgenically recapitulate male killing[34]. In summary, the evidence for *Wolbachia* and *wmk* transgenic male killing suggests an intricate and multifaceted genotype-phenotype landscape.

Building on this background, several key questions emerge: Do closely-related or distantly-related homologs induce male killing? How sensitive is the *wmk* phenotype to small or large genetic changes? And how adapted are *wmk* homologs to their arthropod hosts? Here, we evaluate codon-optimized *wmk* homologs that span a spectrum of genetic divergence (Fig. 1), including homologs (i) from distantly-related hosts, (ii) with putative upstream alternative start codons, (iii) with a single amino acid change, or (iv) with one or more synonymous codon changes, all in the *D. melanogaster* host. In this way, we aimed to investigate how a variety of genetic alterations to *wmk*, from a single synonymous nucleotide to the gene level, affect the phenotype. In particular, we hypothesized that divergent strains would not induce the phenotype due to co-adaptation with a distantly-related host. We also anticipated that the alternative start codons would inhibit function based on previous results[32]. In contrast, we hypothesized that neither a single amino acid change

nor synonymous codon changes would alter the male-killing phenotype. We report that while distant homologs do not cause a male-killing phenotype in this host, single amino acid and synonymous nucleotide changes remarkably do alter the phenotype. Thus, *wmk* male killing is sensitive to the full spectrum of genetic alterations at a fine-scale level not previously recognized. Notably, synonymous sequence changes and post-transcriptional processes appear to play a role in controlling the genotype-phenotype relationships that underpin *wmk* male killing.

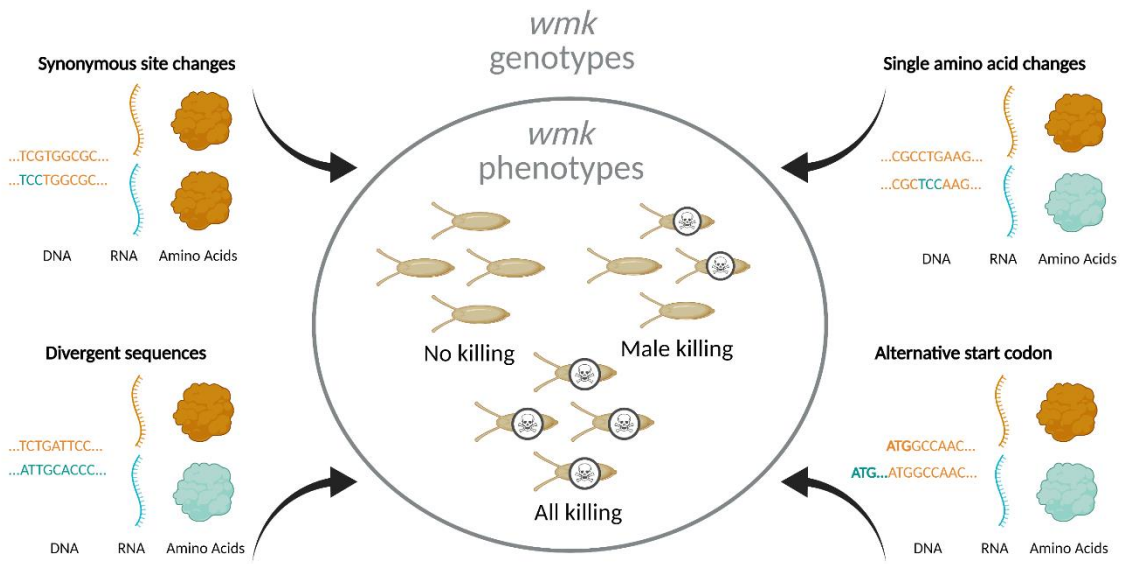
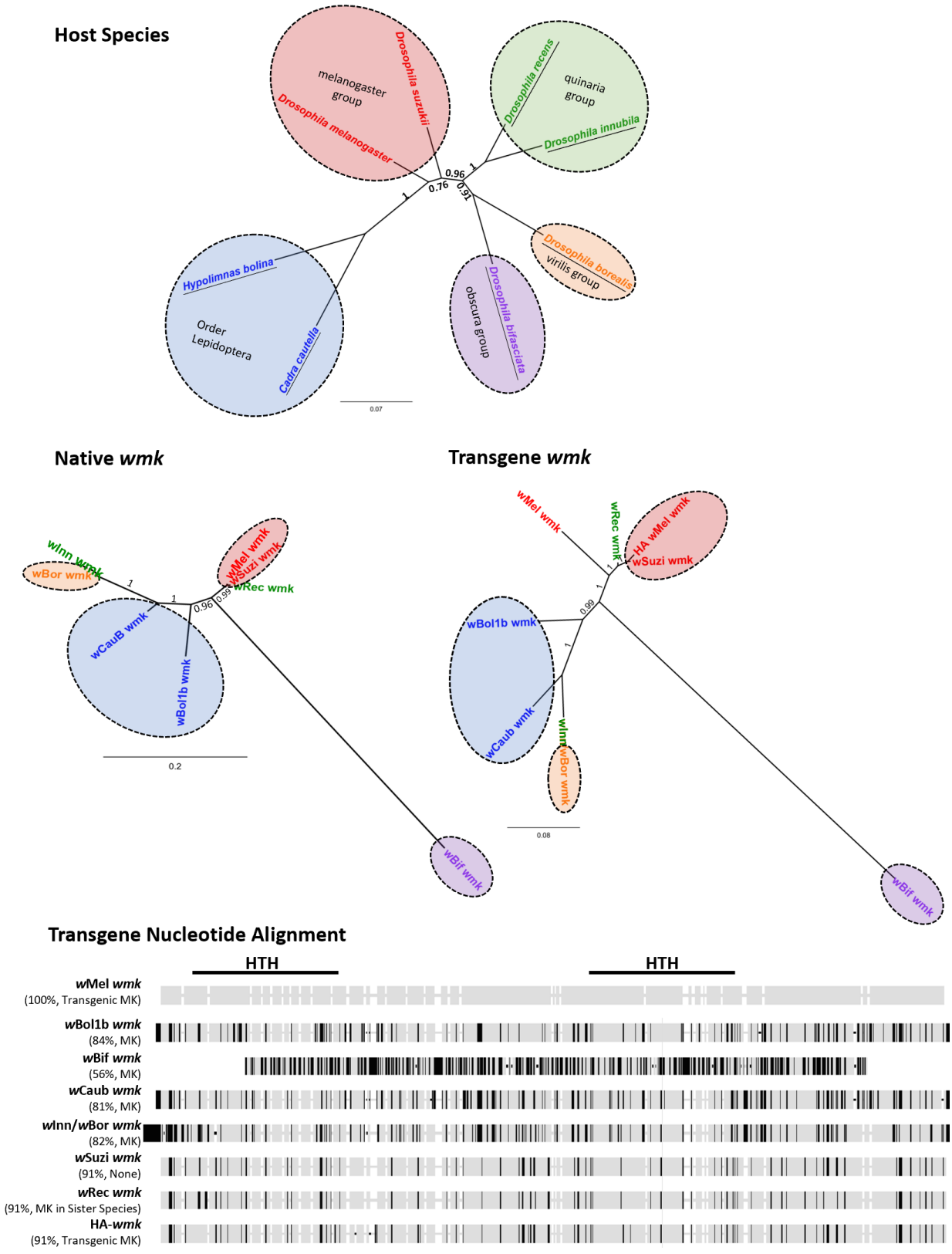


Figure 1. Overview of experimental design. To investigate the genotype-phenotype landscape, we transgenically expressed *wmk* homologs with varying degrees of genetic changes. These sequences are codon-optimized based on different codon biases due to different tRNA abundances in the divergent bacterial source and eukaryotic destination species[35,36]. Transgenic *wmk* in *Drosophila melanogaster* embryos results in three different phenotypes: no killing, male killing, and killing of males and females. Compared to *w*Mel *wmk*, these transgenes were either divergent homologs from other *Wolbachia* strains, a homolog with a single amino acid change, homologs with an additional 9 codons at the 5' ends of the genes starting at an alternative upstream start codon, or variants with a single synonymous codon or nucleotide difference. These genotypes resulted in varying degrees of RNA sequence- and amino acid-level changes. Created with BioRender.com.

Results

***Closely-related homologs of* wmk.** We tested homologs from strains related to *w*Mel that occur in several groups of species in the *Drosophila* genus, along with two from Order Lepidoptera (Fig. 1- figure supplement 1). The phylogenetic clades of both the native *Wolbachia* genes and their transgenes are similar to those of the host species. Roughly divided, the transgenes group into two clusters that offer a range of genetic divergence for evaluating genotype-phenotype relationships (Fig. 1- figure supplement 1) - those distantly related to the *w*Mel *wmk* transgene (less than 90% codon-optimized nucleotide sequence similarity to transgenic *w*Mel *wmk*) and those closely related (greater than 90% identity, from *w*Rec, *w*Suzi, and an HA-tagged *w*Mel homolog).

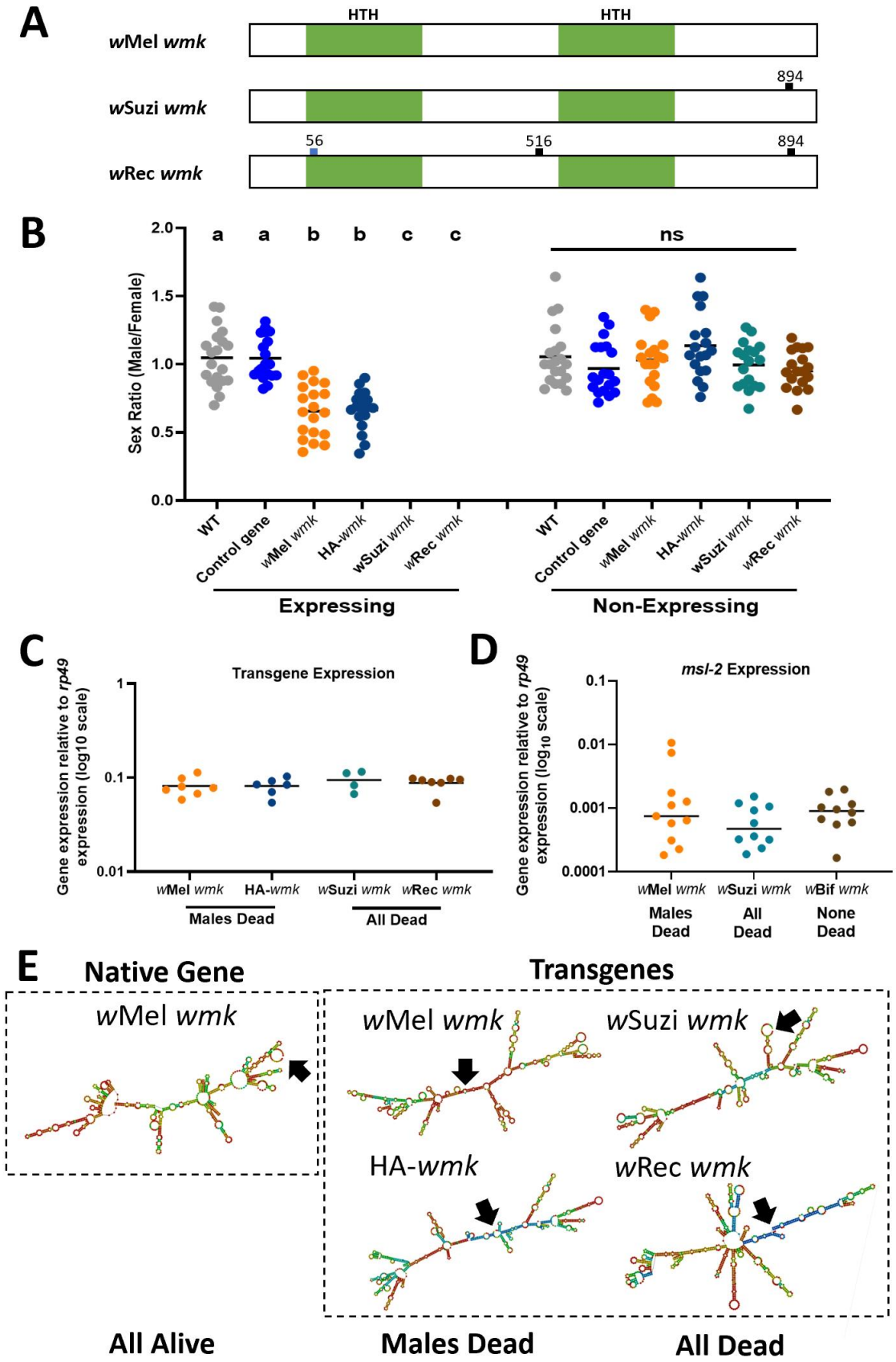

Host Species
melanogaster group
Drosophila suzukii
Drosophila melanogaster
Drosophila recens
quinaria group
Drosophila innubila
Drosophila borealis
virilis group
Drosophila bifasciata
obscura group
Hypolimnas bolina
Order Lepidoptera
Cadra cautella
0.96
0.76
0.91
1
0.07
Native wmk
wInn wmk
wBor wmk
wMel wmk
wSuzi wmk
wRec wmk
wCauB wmk
wBol1b wmk
wBif wmk
0.96
0.93
0.2
Transgene wmk
wMel wmk
wRec wmk
HA wMel wmk
wSuzi wmk
wBol1b wmk
wCaub wmk
wInn
wBor wmk
wBif wmk
0.99
0.08
Transgene Nucleotide Alignment
HTH
HTH
wMel wmk
(100%, Transgenic MK)
wBol1b wmk
(84%, MK)
wBif wmk
(56%, MK)
wCaub wmk
(81%, MK)
wInn/wBor wmk
(82%, MK)
wSuzi wmk
(91%, None)
wRec wmk
(91%, MK in Sister Species)
HA-wmk
(91%, Transgenic MK)

133

Figure 1- figure supplement 1. Homologs of *wmk* tested in this study include variation in native gene and transgene sequence identity as well as host species. (Top) Bayesian nucleotide phylogeny of insect hosts based on 652 bp of the cytochrome oxidase subunit 1 (COI) gene from *D. melanogaster*, *H. bolina*, *D. bifasciata*, *C. cautella*, *D. innubila*, *D. borealis*, *D. suzukii*, and *D. recens*. Branch labels and scale bar indicate posterior probability. Species names are colored either by group within the *Drosophila* genus or by Order Lepidoptera. Underlines indicate hosts in which male-killing strains have been reported. Colored circles indicate host species group. (Middle) Bayesian nucleotide phylogenies of native (middle left, non-transgenic) or transgene (middle right) *wmk* sequences based on 690 or 686 bp, respectively. Label colors reflect groups from host phylogeny for comparison. Colored circles indicate host species group. Branch labels and scale bar indicate posterior probability. *w*Bif *wmk* branches distantly due to a highly divergent sequence. *w*Bor and *w*Inn *wmk* share a branch because they share the same transgenic sequence. (Bottom) Nucleotide alignment of transgenes tested in this study as compared to the previously tested *w*Mel *wmk*, with the regions encompassing the Helix-turn-helix (HTH) protein domains marked with black lines. Black ticks indicate sequence differences with the *w*Mel *wmk* reference strain, light grey indicates sequence matches to the reference sequence, and white indicates an indel in at least one strain. In parentheses under strain names, percentages refer to nucleotide similarity compared to the *w*Mel *wmk* transgene. MK indicates a male killing strain in its native host, transgenic MK indicates the ability to induce a biased sex ratio transgenically only, and MK in sister species indicates the ability to kill males in a sister host species but not in the native host species. Alignment excludes the HA tag in the tagged strain, which is located between the two HTH domains.

The closely-related *wmk* homologs to *w*Mel include those from *Wolbachia* strains (i) *w*Suzi of the fruit pest species *D. suzukii* that has no confirmed reproductive phenotype but notably occurs in populations with female-biased sex ratios[30,37] and (ii) *w*Rec of *D. recens* that kills males when introgressed into its sister species, *D. subquinaria*[18]. The natural *w*Suzi *wmk* homolog has one synonymous single nucleotide polymorphism (SNP) compared to the *w*Mel *wmk* reference, thus yielding the same amino acid sequence. The natural *w*Rec *wmk* homolog has three SNPs, one of which is non-synonymous relative to *w*Mel *wmk*, located in the first HTH DNA-binding domain (Fig. 2A). These transgenes were codon optimized for expression in *D. melanogaster*. We hypothesized that both transgenes would induce a biased sex ratio comparable to *wmk* when expressed in *D. melanogaster*. However, transgenic expression of both homologs unexpectedly resulted in death of all expressing flies, both male and female (Fig. 2B). In addition, we simultaneously tested a transgene of *w*Mel *wmk* with an internal 3X HA tag epitope in the linker

region between the two HTH DNA-binding domains. This transgene, HA-*wmk*, exhibited a sex ratio bias comparable to *w*Mel *wmk*, as expected (Fig. 2B). Therefore, two transgenes of *w*Mel *wmk* (one with a tag) resulted in the biased sex ratio previously reported[25], while two closely-related transgenes from strains that at least associate with female-biased host sex ratios yielded an all-killing phenotype.


A
HTH
HTH
wMel wmk
894
wSuzi wmk
56
516
894
wRec wmk
B
a a b b c c
ns
Sex Ratio (Male/Female)
2.0
1.5
1.0
0.5
0.0
WT
Control gene
wMel wmk
HA-wmk
wSuzi wmk
wRec wmk
Expressing
Non-Expressing
C
Transgene Expression
Gene expression relative to rp49 expression (log10 scale)
1
0.1
0.01
wMel wmk
HA-wmk
wSuzi wmk
wRec wmk
Males Dead
All Dead
D
msl-2 Expression
Gene expression relative to rp49 expression (log10 scale)
0.1
0.01
0.001
0.0001
wMel wmk
wSuzi wmk
wBif wmk
Males Dead
All Dead
None Dead
E
Native Gene
Transgenes
wMel wmk
wMel wmk
wSuzi wmk
HA-wmk
wRec wmk
All Alive
Males Dead
All Dead

174

Figure 2. Transgenic expression of closely-related *wmk* homologs causes male killing and all killing phenotypes in *D. melanogaster*. (A) Schematic of *w*Mel, *w*Suzi, and *w*Rec *wmk* native nucleotide sequences. The blue tick mark indicates a non-synonymous nucleotide difference. Black tick marks indicate synonymous nucleotide changes. Numbers indicate nucleotide position across the entire 912 nucleotide sequence. (B) Sex ratios of adult flies are shown for expressing (*Act5c*-Gal4) and non-expressing (CyO) embryonic offspring. Each sample point represents the adult offspring (N=50-157, mean 86) produced by a replicate family of ten mothers and two fathers, with expressing and non-expressing flies of a given genotype being siblings. Bars represent the mean sex ratio. Statistics are based on a Kruskal-Wallis, one-way ANOVA followed by Dunn's correction across either expressing or non-expressing flies. *w*Rec and *w*Suzi *wmk* have no points in the expressing category due to death of most or all males and females. HA-*wmk* contains a 3X HA tag in the linker region between the two helix-turn-helix domains. This experiment was performed twice. Data and statistical outputs are available in Figure 2 - Source Data 1 and Figure 2- Source Data 2, respectively. (C) Gene expression in embryos 4-5 h AED of each indicated *wmk* transgene from (B), relative to *Drosophila* housekeeping gene, *rp49*. There is no significant difference in expression based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. Data and statistical outputs are available in Figure 2 - Source Data 3 and Figure 2- Source Data 4, respectively. (D) Gene expression in embryos 4-5 h AED of the host *msl-2* dosage compensation gene relative to *rp49* under simultaneous expression of the indicated transgene. There is no significant difference in expression based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. Data and statistical outputs are available in Figure 2 - Source Data 5 and Figure 2- Source Data 6, respectively. (E) Predicted RNA secondary structures of native *w*Mel *wmk* and several transgene strains. Black arrows point to the location of the start codon within each structure.

Gene expression similarities. To assess if variation in gene expression underpins the phenotypic variation, we measured the transcript levels of the transgenes in embryos 4-5 h AED (after egg deposition) when *wmk* kills males[25]. Gene expression levels are not significantly different across *w*Suzi, *w*Mel, *w*Rec, or HA tag *w*Mel *wmk* transgenes, indicating that transcript levels do not account for the phenotypic differences (Fig. 2C). An alternative hypothesis is that transgenic expression of the *w*Suzi and *w*Rec homologs impacts native expression of a host gene in *D. melanogaster* that causes male and female lethality. For instance, transgenic expression of the DNA-binding dosage compensation gene, *male-specific lethal 2* (*msl-2*), can induce total lethality with male-killing *Spiroplasma*[38]. Based on this, we quantified *msl-2* transcript levels in embryos expressing *w*Mel *wmk* (sex ratio bias), *w*Suzi *wmk* (all expressing hosts die), and *w*Bif

wmk (no killing or sex ratio bias) and found that *msl-2* levels were comparable across all genotypes and phenotypes (Fig. 2D).

RNA structural model variation. We next considered mRNA secondary structural differences of the transgene transcripts as a factor explaining the observed phenotypic variation. Modelling RNA structures showed they were substantially different, even in the case of few to no amino acid level changes from one sequence to another. Notably, some RNA structural features grouped by phenotype, such as the location of the start codon (Fig. 2E). Two transgenic transcripts, *w*Mel *wmk* and HA-*wmk*, cause a sex ratio bias and have start codons in the middle of the structure, while the other transgenes, *w*Suzi *wmk* and *w*Rec *wmk*, that kill all flies have start codons on outer loops or hairpins. Although caution is warranted with this predicted structural analysis, mRNA secondary structure could explain some phenotypic outcomes of closely-related *wmk* homologs. Native *w*Mel *wmk* also has a start codon on an outer loop but does not induce an "all killing" phenotype like the *w*Suzi and *w*Rec *wmk* transgenes, which may be due to lower expression of native genes[25], and/or other structural differences.

***Distantly-related homologs of* wmk.** To determine if *wmk* homologs from distantly-related strains induce a biased sex ratio in *D. melanogaster*, we transgenically expressed four codon-optimized homologs from known male-killing strains of *Wolbachia*: the *w*Bol1b strain from *Hypolimnas bolina* butterflies[39], the *w*Bif strain from *Drosophila bifasciata* flies[21], the *w*Caub strain from *Cadra cautella* moths[22], and the *w*Inn strain from *D. innubila* flies[40] (same gene sequence as the *w*Bor male-killing strain from *D. borealis* flies[41]) (Fig. 1 – figure supplement 1). While transgene expression of *w*Mel *wmk* induces a biased sex ratio (~ one third of expressing males die), none of the more distantly-related *w*Bol1b, *w*Bif, *w*Caub, or *w*Inn/*w*Bor *wmk* homologs

yield a biased sex ratio, demonstrating that they do not recapitulate male killing when transgenically expressed in *D. melanogaster* under the conditions tested (Fig. 3).

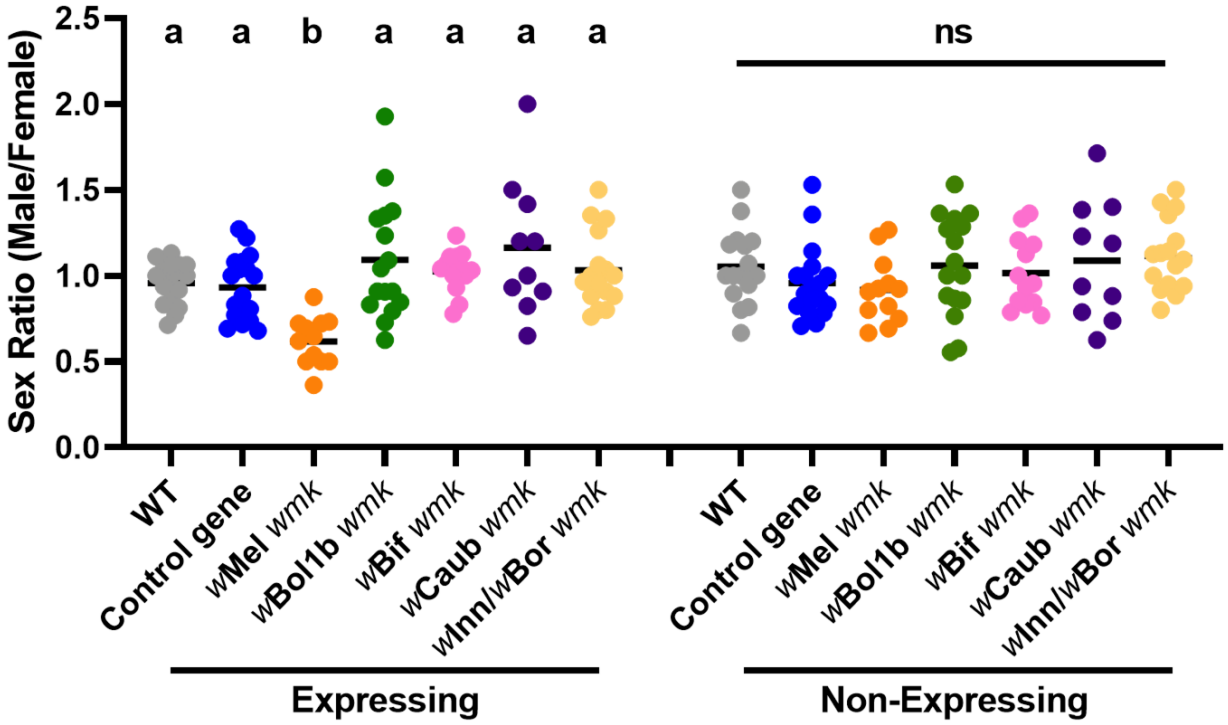


Figure 3. Divergent homologs of *wmk* from male-killing strains do not induce a biased sex ratio in *D. melanogaster*. Sex ratios of adult flies are shown from either expressing (*Act5c*-Gal4) or non-expressing (CyO) offspring. WT refers to the background insertion line and Control gene refers to the WD0034 control transgene that induces no sex ratio bias. Each sample point represents the adult offspring (N=50-132, mean 69) produced by a replicate family of ten mothers and two fathers, with expressing and non-expressing flies of a given genotype being siblings. Bars represent the mean sex ratio. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction across either expressing or non-expressing flies. This experiment was performed twice. Data and statistical outputs are available in Figure 3 - Source Data 1 and Figure 3- Source Data 2, respectively.

Alternative start codon variation and male killing. Relevant to the studies here, we previously provided evidence that some strains contain alternative start codons upstream of the annotated start for *wmk* and these upstream regions are expressed in the *w*Mel strain[32]. When *w*Mel *wmk* was transgenically expressed with the most likely upstream start codon, the phenotype was lost, and no biased sex ratio resulted. We also showed that some non-male-killing strains tended to have more of the alternative start codons[32]. To determine if *w*Rec and *w*Suzi *wmk* transgene

phenotypes are similarly sensitive to transcript changes, we expressed them with upstream codons that are native to each of their genetic sequences. As previously observed with other homologs, they lost their killing phenotype with only nine amino acids added to the 5’ end of the gene, despite being smaller than many commonly used protein tags (Fig. 2- figure supplement 1). All expressing flies survived with no sex ratio bias. Returning to the RNA structure models, we find that simply adding the corresponding nucleotides at the 5’ end of each homolog resulted in several predicted differences in RNA secondary structure for each transgene compared to the structures without the additional 5’ nucleotides (Fig. 2- figure supplement 1). This includes additional loops and different predicted placement of the start codon.

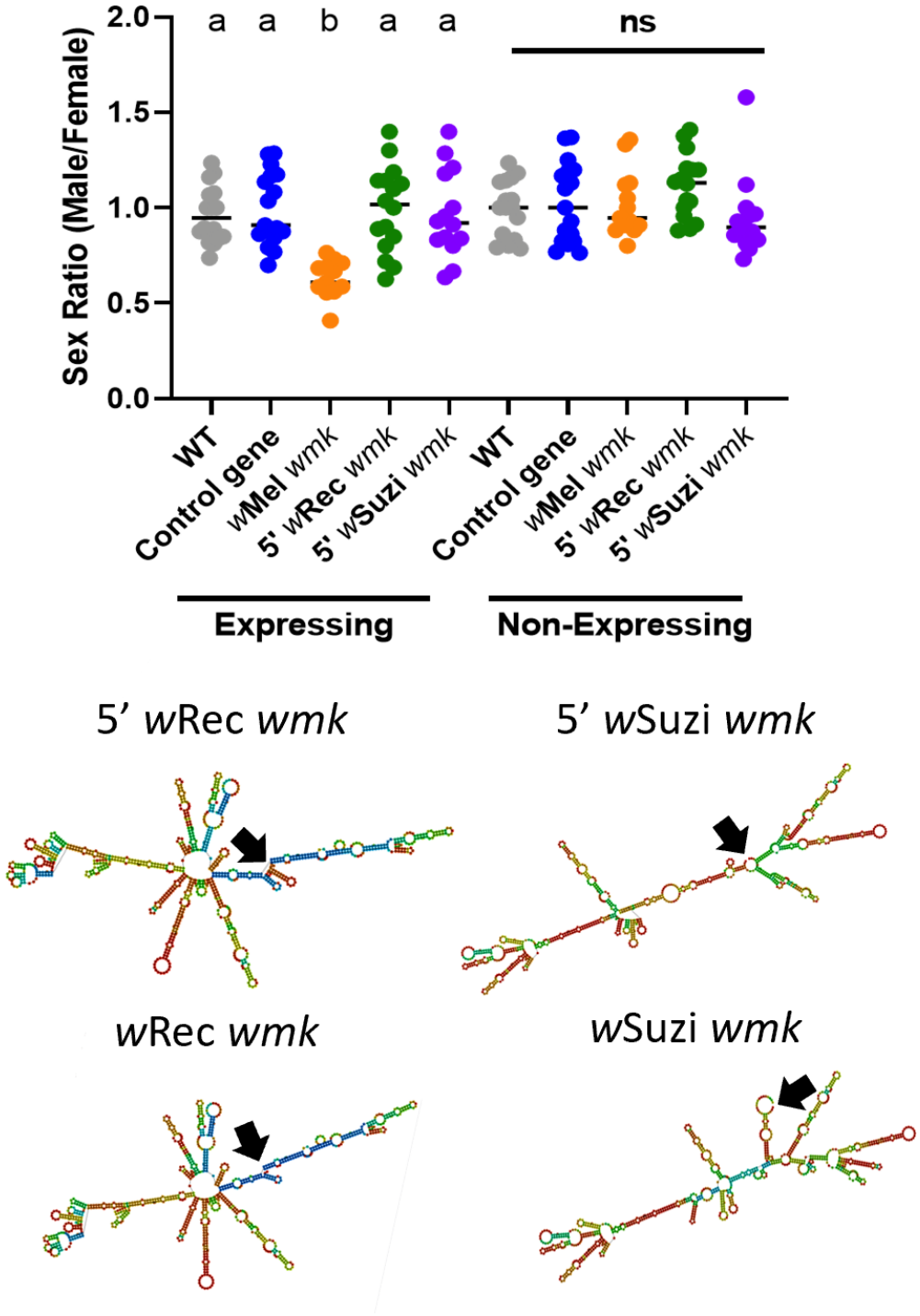


Figure 2- figure supplement 1. *w*Rec and *w*Suzi transgenes expressed with an alternative start codon lose their transgenic phenotypes. (Top) Sex ratios of adult flies are shown for expressing (*Act5c*-Gal4) and non-expressing (CyO) offspring. Each sample point represents the adult offspring (N=50-120, mean 84) produced by a replicate family of ten mothers and two fathers, with expressing and non-expressing flies of a given genotype being siblings. Bars represent the mean sex ratio. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's

correction across either expressing or non-expressing flies. This experiment was performed twice. Data and statistical outputs are available in Figure 2 - Source Data 7 and Figure 2- Source Data 8, respectively. (Bottom) Predicted RNA secondary structures of the *w*Rec and *w*Suzi *wmk* transgenes with the additional 5' sequence exhibit slight structural changes compared to their non-lengthened counterparts (included again from Fig. 2 for ease of comparison). Black arrows point to the location of the start codon within each structure.

Silent site variation and male-killing. Finally, to identify particular nucleotides that may account for phenotypic variation among the homologs, we aligned the sequences of the four transgenes in Fig. 2 and investigated codons that clustered by phenotype (sex ratio bias for *w*Mel *wmk* and HA-*wmk*, or all killing for *w*Rec and *w*Suzi *wmk*). Across the length of the genes (and excluding the HA tag), there were only two codon differences: one at the sixteenth amino acid position and another near the end of the gene. As previous work demonstrated that changes at the 5' end of this gene affect phenotype[32] and since approximately the first ten codons in model prokaryotic genes are known to substantially affect mRNA structure and resulting translation rate[42], we focused on the earlier codon at site 16. This codon, which codes for Serine in all homologs, segregates among sequences by phenotype. The HA-*wmk* and *w*Mel *wmk* transgenes, which recapitulate male killing, have a TCG codon, while the all-killing transgenes *w*Suzi and *w*Rec *wmk* have TCC and AGC, respectively (Fig. 4A).

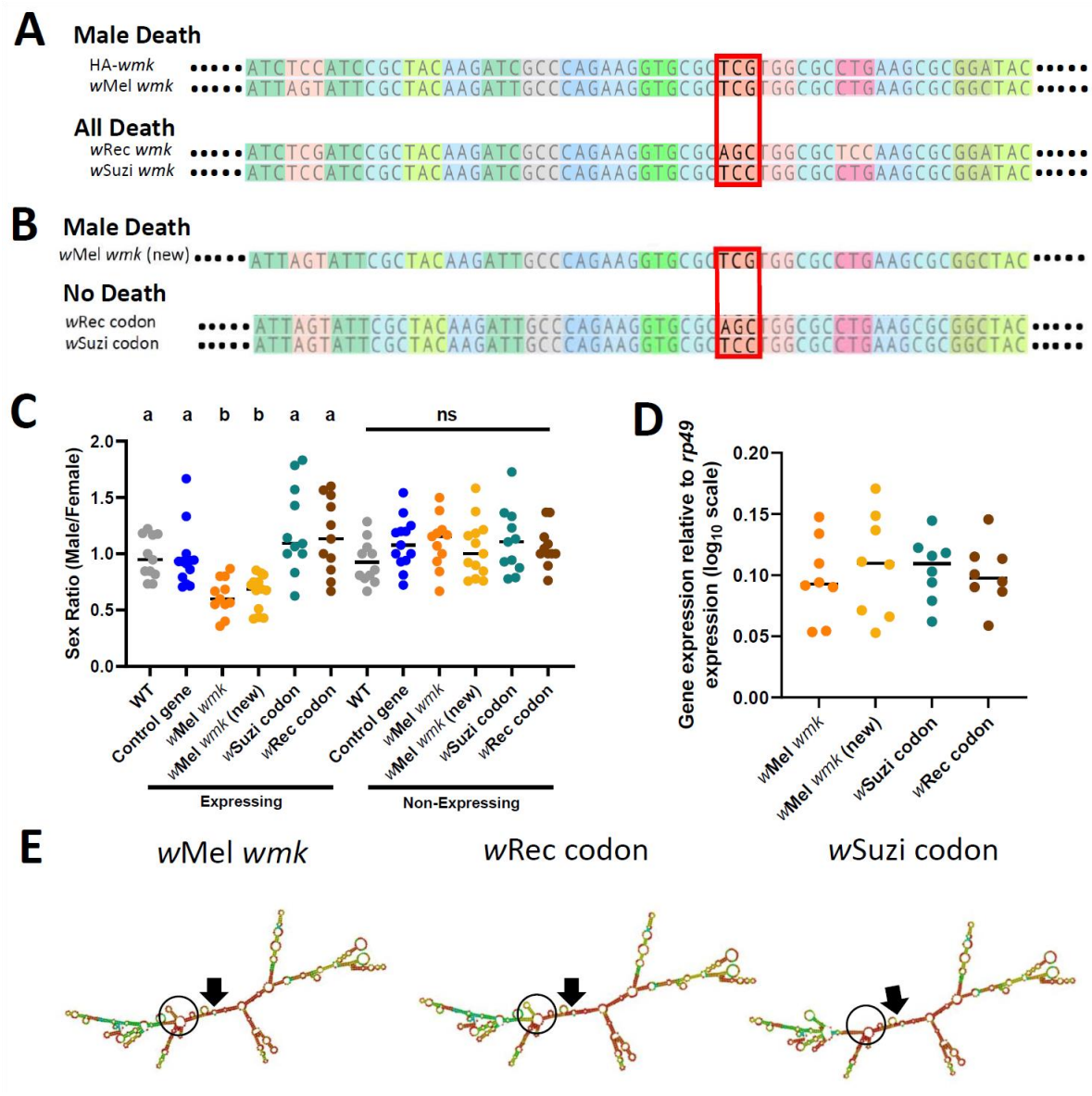


Figure 4. Synonymous nucleotide changes in the 16th codon position of *wmk* alters resulting phenotype. (A) Sequence alignment of transgenic *wmk* homologs. The codon farthest on the left is the fourth codon in the sequence, and the highlighted codon is the 16th, with the farthest right representing the 23rd codon, and ellipses indicating codons continuing on either side. The red box outlines where the genotypes cluster by phenotype. The "HA-*wmk*" and "*w*Mel *wmk*" genotypes share the same codon in this position, and both induce male-specific death. The "*w*Rec *wmk*" and "*w*Suzi *wmk*" genotypes both exhibit different codons from the previous two and exhibit an all-killing phenotype. Colors correlate with amino acid identity. (B) Sequence alignment of transgenes with either the *w*Mel *wmk* sequence made anew (*w*Mel *wmk* new), or with the 16th codon (red box) replaced with the synonymous codons from the *w*Rec and *w*Suzi *wmk* transgenes. The colors and symbols reflect those in (A). (C) Sex ratios of adult transgenic flies are shown for expressing (*Act5c*-Gal4) and non-expressing (CyO) offspring that include the original transgene *w*Mel *wmk* strain used in previous figures, along with the newly created identical *w*Mel *wmk* (new) transgene

and the additional transgenes with the single codon swapped out for the indicated codons noted in (A). *w*Suzi codon and *w*Rec codon refer to the strains that have the same sequence as the *w*Mel *wmk*, but with one or three silent sites changed in the single codon at the 16th amino acid position. Each sample point represents the adult offspring (N=50-161, mean 73) produced by a replicate family of ten mothers and two fathers, with expressing and non-expressing flies of a given genotype being siblings. Bars represent the mean sex ratio. Statistics are based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction across either expressing or non-expressing flies. This experiment was performed twice. Data and statistical outputs are available in Figure 4 - Source Data 1 and Figure 4- Source Data 2, respectively. (D) Gene expression in embryos 4-5 h AED denotes expression of each transgene relative to that of *rp49*. There is no significant difference in expression based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. Data and statistical outputs are available in Figure 4 - Source Data 3 and Figure 4- Source Data 4, respectively. (E) Predicted RNA secondary structures of the *w*Mel *wmk* transcript compared to both of the *w*Rec or *w*Suzi codon transgenes exhibiting slight structural differences. The structure for transgene *w*Mel *wmk* is included again from Fig. 3 for ease of comparison. Black arrows point to the location of the start codon within each structure. Black circles highlight a key area of difference between the structures, with a stem absent in the *w*Suzi codon strain, and different base pair match probabilities calculated for each as indicated by color (blue to red, low to high probability). Within the black circle, the *w*Suzi codon transgene structure is missing a predicted stem that the others have. The stem in the *w*Rec codon line, while present, has a weaker prediction as noted by the cooler colors, so there may be structural differences compared to the *w*Mel *wmk* model.

To functionally test if the silent changes in this Serine codon accounted for the phenotype differences, we generated three new transgenes (Fig. 4B): (i) *w*Mel *wmk* control with no changes compared to the previously tested transgene (TCG codon, labelled "*w*Mel *wmk* (new)"), (ii) *w*Mel *wmk* with three nucleotide changes in the codon that reconstitutes the AGC present in the *w*Rec transgene line (labelled "*w*Rec codon"), and (iii) *w*Mel *wmk* with one nucleotide change in the codon that reconstitutes the TCC present in the *w*Suzi transgene line (labelled "*w*Suzi codon"). When these three otherwise identical genes were transgenically expressed, the wMel *wmk* (new) transgene with no changes caused a biased sex ratio as expected; however, and remarkably, expression of transgenes with two different Serine codons ablated the phenotype and resulted in a non-biased sex ratio with normal numbers of expressing flies (Fig. 4C). This ablation occurs even though transcript levels remain similar across all transgenes and despite sequencing confirmation of the single codon differences (Fig. 4D). Thus, a minimum of one single synonymous site change

in the 5' region was sufficient to alter the sex ratio phenotype. However, while the nucleotide changes in the codon changed the phenotype, they did not recapitulate the all-killing phenotypes of their corresponding homologs. The predicted RNA secondary structures from the transgenes with the single codon changes are similar to the original *w*Mel *wmk* transgene, but they differ in some aspects such as presence or absence of stems and loops and the probability score of the base pair match as indicated by color (scale of red to blue, warmer colors indicate high probability, cooler colors indicate low probability). (Fig. 4E).

Discussion

Linking mutations to function is crucial in resolving the evolutionary dynamics of adaptations, especially when the functions emerge in a nested system of tripartite phage-bacteria-animal interactions. In a simple case, the direct impact of genetic divergence is functional divergence with increasing numbers of mutations leading to increased functional divergence. Here, we investigated evolutionary and molecular hypotheses related to the genotype-phenotype relationships between the prophage *WO-mediated killing* gene, *wmk*[25,32], and male killing in *D. melanogaster*. Sequences among *wmk* homologs in nature can differ across divergent hosts, leading to the hypothesis that highly divergent alleles are functionally fine-tuned. Using a variety of sequences, we report three key results: (i) mutations even at the synonymous codon and single nucleotide levels can alter the male-killing phenotype, (ii) phenotype, genotype, and RNA structural variation exhibit some correlation with each other, and (iii) distantly-related homologs do not induce a male-killing phenotype in transgenic *D. melanogaster*, while some closely-related homologs from *Drosophila* species/populations with female-biased sex ratios induce all-killing phenotypes. We discuss how these findings expand and support mechanistic models of *wmk* male

killing, emphasize important implications for transgenic assays in endosymbiont studies, and relate these findings to aspects of male killing such as phenotype switching and host resistance.

The most surprising and unanticipated result was that transgene expression of highly similar homologs and even single synonymous site changes alter phenotype, and they make the difference between life and death for some males. We tested *wmk* homologs with high native sequence identity compared to native *w*Mel *wmk*: *w*Rec *wmk* from the mushroom-feeding *D. recens* (native sequence has 2 synonymous and 1 non-synonymous nucleotide differences) and *w*Suzi *wmk* from the fruit crop pest *D. suzukii* (native sequence has 1 synonymous nucleotide difference). Although we anticipated similar results to *w*Mel *wmk* expression, we found that transgenic expression of these genes killed all flies (Fig. 2), even though the *w*Suzi *wmk* transgene produces an identical protein to transgenic *w*Mel Wmk. Of particular note, transgenic *w*Suzi *wmk* (all killing) and the original transgenic *w*Mel *wmk* sequence (male killing) produce proteins of the exact same amino acid sequence, but they shared only a 91% sequence identity in the codon-optimized transgene, largely due to an updated codon optimization algorithm. Codon optimization is the norm in transgenic symbiosis research under the common assumption that synonymous codons are functionally redundant. However, codon optimization algorithms often choose a codon based on factors including codon adaptation, mRNA folding, regulatory motifs, nucleotide bias, or codon correlations and biases[35]. With even a few different codons input into the algorithm as well as algorithm updates over the years, the tested transgenes had different nucleotide sequences.

Based on the aforementioned results, we next sought to assess the sensitivity of male killing to *wmk* transcriptional and post-transcriptional changes and asked if one synonymous codon or site change was sufficient to alter phenotype. We created three transgenic lines, each with different DNA sequences coding for the same Serine at position 16. With only these minor changes that

encode an identical protein sequence to *w*Mel Wmk, the transgenic male-killing phenotype was ablated (Fig. 4). Thus, one single codon and even a single nucleotide, remarkably determined male viability. Importantly, this key result implies at the molecular level that endosymbiont phenotypes of reproductive parasites may not simply be governed by DNA sequence alone.

A key question related to our findings is how synonymous changes cause vastly different phenotypes. It is often assumed in endosymbiont transgenic experiments that synonymous changes do not affect function, and codons for the same amino acid are functionally redundant. However, several decades of research have uncovered mounting evidence that the functional redundancy hypothesis is not always accurate[35]. Indeed, codon bias varies across species (hence the need for codon optimization)[44,45], and there are several mechanisms through which single or few codons can influence transcription or translation. When there is a rare codon, there are fewer tRNAs available, and the translation rate may be correspondingly lowered[46]. Notably, there is a bias for rare codons in N-terminal regions of bacterial genes, which is likely due to their influence on mRNA structure[47]. Lower 5' RNA stability or weaker structure can also be correlated with ease of initiation and faster translation[48], and stem loop structures may affect translation rate in varying ways. Indeed, strong 5' RNA secondary structure may inhibit ribosomal initiation or translational efficiency[49,50], and there can be selection against SNPs that alter RNA secondary structure[51]. There is also evidence for selection across the domains of life on codon usage in the first 30-60 nucleotides of a gene, likely due to their impact on mRNA structure near the site of initiation[48]. Additionally, early codons with low frequencies can be beneficial by slowing elongation, thus preventing ribosome collisions, or potentially helping to recruit chaperones to the emerging peptide for proper folding[52,53]. In addition, some specific codons across the sequence are favored

for their lower likelihood of mistranslation[54]. Indeed, synonymous codon differences in genes can result, for instance, in altered gene expression[55], lower organism fitness[56], or disease[57].

Beyond codon sequence, gene expression levels across representatives of all phenotypes are similar (Fig. 2), supporting post-transcriptional differences as the source of functional variation and refuting the functional redundancy hypothesis in this context. Examining the three synonymous transgenes more closely, we find that codon usage indeed correlates with phenotype. According to the codon usage table for *D. melanogaster* on the Genscript website, the original *w*Mel *wmk* codon at this position (TCG) has a frequency 16.7 per 1000 codons. In contrast, the *w*Suzi codon (TCC) has a frequency of 19.5/1000, and the *w*Rec codon (AGC) has a frequency of 20.5/1000[43]. Therefore, there is a pattern where the more common codons are present in the all-killing transgenes, and at similar frequencies, while the strain expressing partial male killing has a codon at a lower frequency. Thus, some of the phenomena described in studies on synonymous codons and function likely underlie the phenotypic variation observed here.

Given that transgenic *wmk* expression with silent site changes results in different phenotypes, we used software that estimated RNA secondary structures for the tested transgenes. Indeed, we found that the predicted structures correlate somewhat with phenotype, so synonymous codons may change transcript structure or other post-transcriptional features of the transcript or protein (Fig. 2). Other small changes including adding a 9-codon N-terminal sequence ablated the phenotype and reversed it to no death, despite the additional sequence being smaller than many common gene tags that typically do not interfere with function. This small change also alters predicted RNA structure (Fig. 2- supplement 1). We were unable to assess protein levels to determine translational differences due to lack of an antibody. However, the data demonstrates

that the *wmk* transgene is functionally sensitive to some post-transcriptional changes, at least in the N-terminal region.

What, then, do these results illuminate about transgenic *wmk* function and the utility of transgenic research in light of the potential for marked influences of synonymous changes on phenotype? First, as discussed, the *wmk* transgenic phenotypes are likely sensitive to post-transcriptional processes. This has important implications for understanding *wmk* in a natural male-killing context as soon as feasible techniques are available, since heterologous expression and its reliance on codon optimization may obscure our understanding of the gene's biology. Second, since three different transgenic phenotypes (sex ratio bias, all killing, no killing) have been found so far with only a few sequences analyzed, it is possible that further testing of new codons may increase the partial male killing to a full male-killing phenotype. Therefore, it would be fruitful and may be possible to continue to uncover a transgenic sequence that fully recapitulates the phenotype to refine this system as a study tool for *Wolbachia* male killing. Third and critically, although the results show that there is some relationship between synonymous codons and phenotype, several points remain for further testing. For example, we cannot conclude that the particular codon tested here is responsible for phenotype alterations in other host genetic backgrounds or species. It is possible that this codon plays a functional role only in a singular host genetic context. Here, we changed *wmk* sequences while holding the host genetic background fixed, but the reverse is required to conclude whether or not the particular codon plays a general role in other genotypes or natural contexts. Second, due to possible coevolution, various codons may or may not yield similar functional effects across different host backgrounds, and additional synonymous sites may contribute to the male killing phenotype. Thus, the results here illuminate a previously unrecognized need for future research on the functional impacts of synonymous

substitutions in endosymbionts. Future work may focus on determining if there is one specific synonymous codon that affects the male killing function in all cases, if a more general feature exists where alteration of any or a subset of N-terminal or other *wmk* codons affects function, or if the effect of synonymous changes is specific to this background.

In addition, these findings are informative with regards to the more general study of phenotypes induced by endosymbiont transgenes. It is standard practice to codon optimize genes for maximizing host expression when testing endosymbiont gene function[25,27,34], with the assumption that synonymous codons will not alter function. However, if codon optimization potentially changes the interpretation of transgenic findings, then phenotypes should be corroborated in natural contexts once tools such as genetic editing are available in the relevant organism. Specifically, *wmk* should be knocked out in native contexts once it is more technically feasible. In addition, codon optimization algorithms are updated with new information periodically with the assumption that they yield improved results, although it is unclear in practice whether an algorithm is better optimized to produce results that reflect the true biology of a transgene. Future work is necessary to explore these concepts further. For example, comparisons of alleles may need to be performed with alleles identical in sequence except for any engineered differences, and the algorithm should remain constant across all transgenes that are compared to each other. Further, careful analysis and comparison of transgenic phenotypes produced by different algorithms may be necessary in some cases where the phenotype is known or expected. This approach could ensure the algorithm produces a transgenic phenotype that most closely resembles the natural phenotype. In addition, the reliability of certain molecular evolutionary signatures, such as dN/dS, may be compromised since synonymous mutations are assumed to be neutral in these analyses. These principles are particularly important for research on endosymbionts that increasingly relies on

heterologous gene expression for functional studies. The smaller genomes of endosymbionts tend to have low GC content overall[58], and GC content is the strongest driver of codon usage bias[59,60] and contributes to the strength of mRNA secondary structure[35], so careful attention to the effects of synonymous changes may be of acute interest to the endosymbiont research community.

We also find interesting support for the hypothesis that there is co-adaptation between *wmk* homologs and their hosts since male-killing genes may be evolutionary matched for the host sex determination and molecular machinery that they manipulate. We first analyzed *wmk* homologs that are more distantly related to *w*Mel *wmk*, including from male-killing strains *w*Bol1b, *w*Bif, *w*Caub, and *w*Inn/*w*Bor[22,39,40,61,62]. They did not recapitulate male killing when transgenically expressed in *D. melanogaster*. The lack of a sex ratio bias is expected if either *wmk* does not underpin male killing in these systems or divergent homologs are required to be closely co-adapted to their hosts. This latter hypothesis is based on observations that resistance to male killing is common (evidence of a potential host-microbe arms race)[8,14,18,19,63], and some strains can cause male killing anew when transferred to another, usually closely-related host[18,20,22] while not causing male killing in all recipient host species[64,65]. It is possible, then, that the *wmk* homologs tested induce a sex ratio bias in their natural hosts, but not in *D. melanogaster*.

If we make the assumption that *wmk* is involved in male killing in nature, which requires confirmation beyond transgenic recapitulation of the phenotype, then the results here give the basis for additional hypotheses that require further testing. First, the killing of both males and females by the closely-related *w*Rec and *w*Suzi *wmk* homologs could indicate that the target is something shared in both males and females, but functions differently within each sex (assuming no off-target effects). Indeed, previous studies on *wmk* and *Wolbachia* male killing demonstrate a positive correlation with between dosage compensation complex (DCC) activity and DNA defects in male

embryos[21,25,26]. Notably, four of the five protein components of the DCC are expressed in both males and females, and it is only *msl-2* that is male-specific and catalyzes formation of the complex[66]. Thus, there are several non-sex-specific *wmk* target candidates that the gene product may interact with to cause lethality in males and females. Second, protein divergence and resultant conformational changes may impact the specificity between host target and the Wmk toxin, and could underlie development of host resistance. Male-killing genes are expected to evolve rapidly within hosts in order to counteradapt host resistance mechanisms and keep evolutionary pace with the rapid evolution of sex-related genes that may be the target of a male-killing toxin[67,68]. As such, major or minor divergence in protein or transcript sequence of either the host target or the microbial toxin may underpin changes that lead to common host resistance, *wmk*-host coadaptation, and functional transgene differences in the foreign *D. melanogaster* host. Third, these findings leave open the possibility of a variety of functionally-relevant *wmk* protein or transcript conformations in nature, which could contribute to the marked diversity of *Wolbachia* male killing in terms of host species and sex determination systems[4,8,22,23,40,61,69,70].

Taken together, his work reports previously unrecognized relationships for *wmk*-induced killing and establishes new hypotheses for the impacts of RNA structure and post-transcriptional processes in *wmk*-induced male killing. It also highlights several critical features for the research community regarding assumptions about the broad use of transgenes and the role of synonymous mutations in gene function. If *wmk* is involved in natural male killing, then this work could indicate how silent sequence changes may relate to known male-killing phenomena such as frequent host resistance or male-killing function in diverse hosts.

Materials and Methods

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
gene (*Wolbachia pipientis*)	WD0626	NCBI	NCBI:WD_RS02815	Also known as *wmk* (*WO-mediated killing*)
genetic reagent (*D. melanogaster*)	Act5c-Gal4/CyO	Bloomington Drosophila Stock Center	BDSC:3953; FlyBase FBti0012290	P{AyGAL4}25
genetic reagent (*D. melanogaster*)	WT (y^1w^{67c23}; P[CaryP]P2)	Bloomington Drosophila Stock Center	BDSC:8622; FlyBase FBti0040535	WT strain used in this study; P{CaryP}attP2
genetic reagent (*D. melanogaster*)	*w*Mel *wmk*	This paper; Perlmutter *et al.*, 2019 PMID 31504075		Expresses codon-optimized transgene; UAS promoter
genetic reagent (*D. melanogaster*)	*w*Bol1b *wmk*	This paper		Expresses codon-optimized transgene; UAS promoter
genetic reagent (*D. melanogaster*)	*w*Bif *wmk*	This paper		Expresses codon-optimized transgene; UAS promoter

genetic reagent (*D. melanogaster*)	*w*Caub *wmk*	This paper		Expresses codon-optimized transgene; UAS promoter
genetic reagent (*D. melanogaster*)	*w*Inn/*w*Bor *wmk*	This paper		Expresses codon-optimized transgene; UAS promoter; *w*Inn and *w*Bor *wmk* have same exact sequence
genetic reagent (*D. melanogaster*)	*w*Suzi *wmk*	This paper		Expresses codon-optimized transgene; UAS promoter
genetic reagent (*D. melanogaster*)	*w*Rec *wmk*	This paper		Expresses codon-optimized transgene; UAS promoter
genetic reagent (*D. melanogaster*)	HA-*wmk*	This paper		Expresses codon-optimized transgene; UAS promoter; 3X HA tag epitope in linker between HTH domains of *w*Mel *wmk*

genetic reagent (*D. melanogaster*)	5' *w*Rec *wmk*	This paper		Expresses codon-optimized transgene; UAS promoter; Sequence has additional 9 amino acids starting at upstream alternative start codon
genetic reagent (*D. melanogaster*)	5' *w*Suzi *wmk*	This paper		Expresses codon-optimized transgene; UAS promoter; Sequence has additional 9 amino acids starting at upstream alternative start codon
genetic reagent (*D. melanogaster*)	*w*Mel *wmk* (new)	This paper		Expresses codon-optimized transgene; UAS promoter; Same exact sequence as *w*Mel *wmk*, newly transformed strain
genetic reagent (*D. melanogaster*)	*w*Suzi codon	This paper		Expresses codon-optimized transgene; UAS

				promoter; Same as *w*Mel *wmk*, but with 16th amino acid position using TCC Serine codon from *w*Suzi *wmk* strain
genetic reagent (*D. melanogaster*)	*w*Rec codon	This paper		Expresses codon-optimized transgene; UAS promoter; Same as *w*Mel *wmk*, but with 16th amino acid position using AGC Serine codon from *w*Rec *wmk* strain
recombinant DNA reagent	pTIGER (plasmid)	Ferguson *et al.*, 2012; PMID 22328499		Modified pUASp plasmid for enhanced germline expression under Gal4/UAS control
sequence-based reagent	Rp49_F	This paper	PCR primers	CGGTTACG GATCGAA CAAGC
sequenced-based reagent	Rp49_R	This paper	PCR primers	CTTGCGCT TCTTGGAG GAGA

sequenced-based reagent	wmk_homologs_opt_F	This paper	PCR primers	CTGTATGCCATTGCCGAGACCCT
sequenced-based reagent	wmk_homologs_opt_R	This paper	PCR primers	TCACCAGATCCTTGGCGATCTTCATC
sequenced-based reagent	Msl-2_F	This paper	PCR primers	GGATTAACGCGGTCTAAGCATGTGTAACTG
sequenced-based reagent	Msl-2_R	This paper	PCR primers	GTATGCCGTCTGGGCCATGATG
commercial assay or kit	Direct-zol RNA MiniPrep Kit	Zymo	R2051	
commercial assay or kit	Superscript VILO cDNA Synthesis Kit	ThermoFisher	11754050	
chemical compound, drug	DNase, RNase-free	Ambion, Life Technologies	AM2222	
chemical compound, drug	iTaq Universal SYBR Green Mix	Bio-Rad	1725120	
software, algorithm	GraphPad Prism 8	GraphPad Prism 8	RRID:SCR_002798	
software, algorithm	Geneious Pro v.2019.2; Geneious Pro v.2020.2.4	Geneious	RRID:SCR_010519	

software, algorithm	jModelTest	jModelTest	RRID:SCR_015244	
software, algorithm	RNAfold WebServer	University of Vienna	Gruber *et al.*, 2008, PMID 18424795; Lorenz *et al.*, 2011, PMID 22115189	http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi

**Reagents source from this paper may be obtained from Bordenstein lab

Drosophila strains and maintenance

D. melanogaster strains used in this study include *Act5c*-Gal4/CyO (BDSC 3953, ubiquitously-expressing zygotic driver), the WT background line of genotype y^1w^{67c23}; P[CaryP]P2 (BDSC 8622), the WD0626 (*wmk*) and WD0034 (control gene) transgene constructs previously described[25], and several new transgene constructs. Briefly, each *Wolbachia* gene of interest was codon-optimized for optimal *D. melanogaster* expression using algorithms developed by GenScript Biotech (Piscataway, NJ). These sequences were then synthesized as DNA nucleotides and cloned by GenScript into the pTIGER plasmid[71]. The pTIGER vector is a UASp-based plasmid optimized for germline expression and uses PhiC31 integrase[72] for targeted integration into the *D. melanogaster* genome. It also includes UAS promoters for inducible expression with a Gal4 driver[73] as well as a w^+ red eye marker for transformant screening. The plasmids were then sent to BestGene (Chino Hills, CA), which performed injections of the vectors into embryos of the BDSC 8622 background line with PhiC31 integrase to integrate the vector into the attP2 insertion site in the genome. Successful transformants were selected based on the red eye marker, and each transgene line is descended from the offspring of a single transformant (isofemale).

D. melanogaster were reared on standard cornmeal, molasses, and yeast (CMY) media (5% w/v cornmeal (Quaker, Chicago IL), 1.875% w/v yeast (Red Star Yeast, Milwaulkee WI), 7.8% v/v molasses (Sweet Harvest Foods, Rosemount MN), 0.5% w/v type II *Drosophila* agar (Genesee Scientific, San Diego CA), 0.056% w/v tegosept (Genesee Scientific, San Diego CA), and 0.39% v/v propionic acid (Sigma Aldrich)). Stocks were maintained at 25°C with virgin flies stored at room temperature. During virgin collections, stocks were kept at 18°C overnight and 25°C during the day. All flies were kept on a 12-hour light/dark cycle.

Sex ratio assays

To assess the effect of transgene expression on adult sex ratios (measurement of male killing), sex ratio assays were performed as previously described[25]. Briefly, twenty biological replicates of 10 uninfected, 4- to 7-day-old virgin, female *Act5c*-Gal4/CyO driver flies and 2 uninfected, 1- to 2-day-old virgin, male transgene flies were each set up in vials with CMY media. Individuals were randomly allocated to each vial after all females or males of a given genotype were mixed together. They were left on the media to lay eggs for 4 days at 25°C with a 12 h light/dark cycle, at which point adults were discarded. The vials are then left at 25°C until the offspring are counted. After 9 days of adult offspring emergence, they were scored for both sex and expression (red eye color from *Act5c*-Gal4 chromosome) or non-expression (curly wings from CyO balancer chromosome). The number of adult offspring per vial across all experiments ranges from 50-170, with a mean of 120 (ranges and means per experiment are included in figure captions). Any vials with fewer than 50 adult offspring were removed from the analysis, as this indicates either abnormally poor egg laying or hatching (typically 0-2 vials per group). In addition, vials with no adult emergence, while others of the same genotype had typical levels of offspring,

were also excluded (typically 2-3 vials per group). Results were graphed in GraphPad Prism 8.4.0, which applies a "Standard Plot Appearance" correction for visibility of data distribution where the width of distribution of points is proportional to the number of points at that y-value.

RNA secondary structures

RNA secondary structures were generated by uploading the nucleotide sequences of the indicated gene to the RNA fold web server[74,75]. The structures shown are the graphical outputs of the MFE (minimum free energy) secondary structures. Colors indicate base pair probabilities, from blue to red, with blue indicating a probability of 0 and red indicating a probability of 1.

Gene expression

Gene expression was measured in *Drosophila* embryos aged 4-5 h AED. Each point represents a biological replicate with the RNA of 30 pooled embryos from crosses between a unique set of 60 uninfected, 4- to 7-day-old virgin, female *Act5c*-Gal4/CyO driver flies and 12 uninfected, 1- to 2-day-old virgin, male transgene flies of the indicated genotype. Each point represents a biological replicate from different bottles. Individuals were randomly allocated to each bottle after all females or males of a given genotype were mixed together. Each collection chamber consisted of a grape juice agar plate with yeast in an 8 oz round bottom bottle, and flies. These were placed in a 25°C incubator overnight (16 h). Then, the plates were swapped with fresh ones. The flies were allowed to lay eggs for 1 h. The plates were then left at 25°C for an additional 4 h to age them to be 4–5 h old (the estimated time of male death in *wmk* crosses). Embryos were then gathered in groups of 30 (each group from a unique bottle/biological replicate) and flash frozen in liquid nitrogen. RNA was extracted using the Direct-zol RNA MiniPrep Kit (Zymo), RNase-free

DNase (Ambion, Life Technologies), cDNA was generated with SuperScript VILO (Invitrogen), and RT-qPCR was run using iTaq Universal SYBR Green Mix (Bio-Rad). qPCR was performed on a Bio-Rad CFX-96 Real-Time System. Primers are listed in the Key Resources Table. Conditions were as follows: 50°C 10 min, 95°C 5 min, 40x (95°C 10 s, 55°C 30 s), 95°C 30 s. Differences in gene expression were done by calculating $2^{-\Delta ct}$ (difference in ct values of two genes of interest). Data points were excluded if a sample had low quality cDNA that did not amplify in qPCR. Data points of each biological replicate are measured as the mean of two technical replicates from each sample.

Phylogenetic trees

The nucleotide phylogenetic trees of host COI genes and *wmk* native gene or transgene sequences were inferred based on a MUSCLE alignment in Geneious Prime 2020.2.4 followed by stripping all sites with gaps. The resulting 652, 690, 686 nucleotide base pair alignments (respectively) were then analyzed via jModelTest 2.1.10 v20160303[76,77]. The AICc-corrected best model, JC, was predicted for all three alignments and was used to build the trees using the MrBayes[78,79] Geneious plugin with the JC69 model[80] and equal rate variation.

Transgene sequence alignments

The gene sequence alignments of different *wmk* transgenes in Figs. 2 and 6 were conducted in Geneious Pro v.2019.2 using a MUSCLE alignment. Black bars in Fig. 2 indicate sequence mismatches compared to the *w*Mel *wmk* transgene reference sequence with any gaps stripped. Codons in Fig. 6 are colored by amino acid.

Statistical analyses

Sample sizes for experiments were based on previous publications, which demonstrated repeatability in relative differences between treatment groups. Each experiment was completed twice, and statistical tests were applied to both to confirm repetition in differences between treatment groups. In all plots, the first experiment is the representative one shown. For sex ratios, we tested different sample sizes for reliability in sex ratio measurements, and found that 20 biological samples per group, with 10 females and 2 males per sample, resulted in consistently replicable data, which is the standard we apply here. qPCR data was approached similarly, by previous work in the lab demonstrating that with embryos 4-5 h AED, we are able to get consistent, high quality, replicable data with 30 embryos per sample, and at least 8 samples per group.

All statistical analyses for sex ratios and were performed using GraphPad Prism 8 software. For sex ratios, a non-parametric Kruskal-Wallis one-way ANOVA followed by Dunn's multiple comparisons test was applied to all gene-expressing categories, followed by the same test but on all non-expressing categories. For gene expression, groups were compared using the same Kruskal-Wallis one-way ANOVA with Dunn's multiple comparisons test. Significant results are indicated with * symbols in figures, with accompanying p values in captions. Any comparisons with no symbol are nonsignificant. Full statistical information and outputs for all sex ratio and qPCR data is available in the Source Data file.

Acknowledgements

This work was supported by National Institutes of Health (NIH) grant R21 AI133522 to S.R.B, the Vanderbilt Microbiome Innovation Center, and NIH grants F31 AI143152, K-INBRE P20 GM103418, and National Science Foundation (NSF) Postdoctoral Research Fellowship

(PRFB) DBI 2109772 to J.I.P. We would also like to thank R. L. Unckless for helpful comments on phylogenies.

Competing Interests

Both J.I.P. and S.R.B. are inventors on a patent related to the transgenic use of *wmk* in vector control: US Patent 20210000092 16/982708.

References

1 Hurst, G. D. & Frost, C. L. Reproductive parasitism: maternally inherited symbionts in a biparental world. *CSH Perspect Biol* **7**, a017699 (2015).

2 Taylor, M., Bordenstein, S. & Slatko, B. Microbe Profile: *Wolbachia*: a sex selector, a viral protector and a target to treat filarial nematodes. *Microbiology* **164**, 1345-1347 (2018).

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

4 Hurst, G. D. D., Jiggins, F. M., von der Schulenburg, J. H. G., Bertrand, D., West, S. A., Goriacheva, I. I., Zakharov, I. A., Werren, J. H. Stouthamer, R. & Majerus, M. E. Male–killing *Wolbachia* in two species of insect. *Proc Royal Soc B* **266**, 735-740 (1999).

5 Bouchon, D., Rigaud, T. & Juchault, P. Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. *Proc Royal Soc B* **265**, 1081-1090 (1998).

6 Hunter, M. S. The influence of parthenogenesis-inducing *Wolbachia* on the oviposition behaviour and sex-specific developmental requirements of autoparasitoid wasps. *J Evol Biol* **12**, 735-741 (1999).

7 Hurst, G. D. & Jiggins, F. M. Male-killing bacteria in insects: mechanisms, incidence, and implications. *Emerg Infect Dis* **6**, 329-336 (2000).

8 Charlat, S., Hornett, E. A., Dyson, E. A., Ho, P. P., Loc, N. T., Schilthuizen, M., Davies, N., Roderick, G. K. & Hurst G. D. Prevalence and penetrance variation of male-killing *Wolbachia* across Indo-Pacific populations of the butterfly *Hypolimnas bolina*. *Mol Ecol* **14**, 3525-3530 (2005).

9 Jaenike, J., Dyer, K. A. & Reed, L. K. Within-population structure of competition and the dynamics of male-killing *Wolbachia*. *Evol Ecol Res* **5**, 1023-1036 (2003).

10 Unckless, R. L. & Jaenike, J. Maintenance of a male-killing *Wolbachia* in *Drosophila innubila* by male-killing dependent and male-killing independent mechanisms. *Evolution* **66**, 678-689 (2012).

11 Skinner, S. W. Son-killer: a third extrachromosomal factor affecting the sex ratio in the parasitoid wasp, *Nasonia* (=Mormoniella) *vitripennis*. *Genetics* **109**, 745-759 (1985).

12 Hurst, L. D. The incidences and evolution of cytoplasmic male killers. *Proc Royal Soc B* **244**, 91-99 (1991).

13 Jiggins, F. M., Hurst, G. D. & Majerus, M. E. Sex-ratio-distorting *Wolbachia* causes sex-role reversal in its butterfly host. *Proc Royal Soc B* **267**, 69-73 (2000).

14 Majerus, T. M. & Majerus, M. E. Intergenomic arms races: detection of a nuclear rescue gene of male-killing in a ladybird. *PLoS Pathog* **6**, e1000987 (2010).

15 Hornett, E. A., Charlat, S., Duplouy, A. M. R., Davies, N., Roderick, G. K., Wedell, N., & Hurst, G. D. D. Evolution of male-killer suppression in a natural population. *PLoS Biol* **4**, e283 (2006).

16 Groenenboom, M. A. & Hogeweg, P. Space and the persistence of male–killing endosymbionts in insect populations. *Proc Royal Soc B* **269**, 2509-2518 (2002).

17 Berec, L., Maxin, D. & Bernhauerova, V. Male-killing bacteria as agents of insect pest control. *J Appl Ecol* (2016).

677 18 Jaenike, J. Spontaneous emergence of a new *Wolbachia* phenotype. *Evolution* **61**, 2244-
678 2252 (2007).
679 19 Mitsuhashi, W., Ikeda, H. & Muraji, M. Fifty-year trend towards suppression of
680 *Wolbachia*-induced male-killing by its butterfly host, *Hypolimnas bolina*. *J Insect Sci* **11**,
681 92 (2011).
682 20 Fujii, Y., Kageyama, D., Hoshizaki, S., Ishikawa, H. & Sasaki, T. Transfection of *Wolbachia*
683 in Lepidoptera: the feminizer of the adzuki bean borer *Ostrinia scapulalis* causes male
684 killing in the Mediterranean flour moth *Ephestia kuehniella*. *Proc Royal Soc B* **268**, 855-
685 859 (2001).
686 21 Riparbelli, M. G., Giordano, R., Ueyama, M. & Callaini, G. *Wolbachia*-mediated male
687 killing is associated with defective chromatin remodeling. *PLoS One* **7**, e30045 (2012).
688 22 Sasaki, T., Kubo, T. & Ishikawa, H. Interspecific transfer of *Wolbachia* between two
689 lepidopteran insects expressing cytoplasmic incompatibility: a *Wolbachia* variant
690 naturally infecting *Cadra cautella* causes male killing in *Ephestia kuehniella*. *Genetics*
691 **162**, 1313-1319 (2002).
692 23 Zeh, D. W., Zeh, J. A. & Bonilla, M. M. *Wolbachia*, sex ratio bias and apparent male killing
693 in the harlequin beetle riding pseudoscorpion. *Heredity* **95**, 41-49 (2005).
694 24 Bordenstein, S. R. & Bordenstein, S. R. Eukaryotic association module in phage WO
695 genomes from *Wolbachia*. *Nat Commun* **7**, 13155 (2016).
696 25 Perlmutter, J. I., Bordenstein, S. R., Unckless, R. L., LePage, D. P., Metcalf, J. A., Hill, T.,
697 Martinez, J., Jiggins, F. M. & Bordenstein, S. R. The phage gene *wmk* is a candidate for
698 male killing by a bacterial endosymbiont. *PLoS Pathog* **15**, e1007936 (2019).
699 26 Harumoto, T., Fukatsu, T. & Lemaitre, B. Common and unique strategies of male killing
700 evolved in two distinct *Drosophila* symbionts. *Proc Royal Soc B* **285**, 20172167 (2018).
701 27 LePage, D. P., Metcalf, J. A., Bordenstein, S. R., On, J., Perlmutter, J. I., Shropshire, J. D.,
702 Layton, E. M., Funkhouser-Jones, L. J., Beckmann, J. F. & Bordenstein, S. R. Prophage WO
703 genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature*
704 **543**, 243-247 (2017).
705 28 Masui, S., Kuroiwa, H., Sasaki, T., Inui, M., Kuroiwa, T. & Ishikawa H. Bacteriophage WO
706 and virus-like particles in *Wolbachia*, an endosymbiont of arthropods. *Biochem Biophys*
707 *Res Commun* **283**, 1099-1104 (2001).
708 29 Wu, M., Sun, L. V., Vamathevan J., Riegler, M., Deboy, R., Brownlie, J. C., McGraw, E. A.,
709 Martin, W., Esser, C., Ahmadinejad, N. & Wiegand C. Phylogenomics of the reproductive
710 parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic
711 elements. *PLoS Biol* **2**, E69 (2004).
712 30 Hamm, C. A. *et al.* Wolbachia do not live by reproductive manipulation alone: infection
713 polymorphism in Drosophila suzukii and D. subpulchrella. *Molecular ecology* **23**, 4871-
714 4885 (2014).
715 31 Hoffmann, A. A. Partial cytoplasmic incompatibility between two Australian populations
716 of *Drosophila melanogaster*. *Entomol Exp Appl* **48**, 61-67 (1988).
717 32 Perlmutter, J. I., Meyers, J. E. & Bordenstein, S. R. Transgenic testing does not support a
718 role for additional candidate genes in *Wolbachia* male killing or cytoplasmic
719 incompatibility. *mSystems* **5**, e00658-00619 (2020).

720 33 Metcalf, J. A., Jo, M., Bordenstein, S. R., Jaenike, J. & Bordenstein, S. R. Recent genome
721 reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates
722 for reproductive parasitism. *PeerJ* **2**, e529 (2014).
723 34 Harumoto, T. & Lemaitre, B. Male-killing toxin in a bacterial symbiont of *Drosophila*.
724 *Nature* **557**, 252 (2018).
725 35 Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of
726 codon bias. *Nat Rev Genet* **12**, 32-42 (2011).
727 36 Gustafsson, C., Govindarajan, S. & Minshull, J. Codon bias and heterologous protein
728 expression. *Trends Biotechnol* **22**, 346-353 (2004).
729 37 Drummond, F., Ballman, E. & Collins, J. Population dynamics of spotted wing *Drosophila*
730 (*Drosophila suzukii* (Matsumura)) in Maine wild blueberry (*Vaccinium angustifolium*
731 Aiton). *Insects* **10**, 205 (2019).
732 38 Cheng, B., Kuppanda, N., Aldrich, J. C., Akbari, O. S. & Ferree, P. M. Male-killing
733 *Spiroplasma* alters behavior of the dosage compensation complex during *Drosophila*
734 *melanogaster* embryogenesis. *Current Biol* **26**, 1339-1345 (2016).
735 39 Dyson, E., Kamath, M. & Hurst, G. *Wolbachia* infection associated with all-female broods
736 in *Hypolimnas bolina* (Lepidoptera: Nymphalidae): evidence for horizontal transmission
737 of a butterfly male killer. *Heredity* **88**, 166 (2002).
738 40 Dyer, K. A. & Jaenike, J. Evolutionarily stable infection by a male-killing endosymbiont in
739 *Drosophila innubila*: molecular evidence from the host and parasite genomes. *Genetics*
740 **168**, 1443-1455 (2004).
741 41 Carson, H. A female producing strain of *D. borealis* Patterson. *Drosoph Inf Serv* **30**, 109-
742 110 (1956).
743 42 Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z. & Blüthgen, N. Efficient translation
744 initiation dictates codon usage at gene start. *Mol Syst Biol* **9** (2013).
745 43 GenScript. *GenScript Codon Usage Frequency Table (chart) Tool*,
746 <https://www.genscript.com/tools/codon-frequency-table> (2020).
747 44 Sharp, P. M., Averof, M., Lloyd, A. T., Matassi, G. & Peden, J. F. DNA sequence evolution:
748 the sounds of silence. *Proc Royal Soc B* **349**, 241-247 (1995).
749 45 Andersson, S. & Kurland, C. Codon preferences in free-living microorganisms. *Microbiol*
750 *Mol Biol Rev* **54**, 198-210 (1990).
751 46 Li, G.-W., Oh, E. & Weissman, J. S. The anti-Shine–Dalgarno sequence drives
752 translational pausing and codon choice in bacteria. *Nature* **484**, 538-541 (2012).
753 47 Goodman, D. B., Church, G. M. & Kosuri, S. Causes and effects of N-terminal codon bias
754 in bacterial genes. *Science* **342**, 475-479 (2013).
755 48 Gu, W., Zhou, T. & Wilke, C. O. A universal trend of reduced mRNA stability near the
756 translation-initiation site in prokaryotes and eukaryotes. *PLoS Comp Biol* **6** (2010).
757 49 de Smit, M. H. & Van Duin, J. Secondary structure of the ribosome binding site
758 determines translational efficiency: a quantitative analysis. *PNAS* **87**, 7668-7672 (1990).
759 50 Bettany, A. J., Moore, P. A., Cafferkey, R., Bell, L. D., Goodey, A. R., Carter, B. L. & Brown,
760 A. J. 5'-Secondary structure formation, in constrast to a short string of non-preferred
761 codons, inhibits the translation of the pyruvate kinase mRNA in yeast. *Yeast* **5**, 187-198
762 (1989).

763 51 Salari, R., Kimchi-Sarfaty, C., Gottesman, M. M. & Przytycka, T. M. Sensitive
764 measurement of single-nucleotide polymorphism-induced changes of RNA
765 conformation: application to disease studies. *Nucleic Acids Res* **41**, 44-53 (2013).
766 52 Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O.,
767 Furman, I. & Pilpel, Y. An evolutionarily conserved mechanism for controlling the
768 efficiency of protein translation. *Cell* **141**, 344-354 (2010).
769 53 Fredrick, K. & Ibba, M. How the sequence of a gene can tune its translation. *Cell* **141**,
770 227-229 (2010).
771 54 Qin, H., Wu, W. B., Comeron, J. M., Kreitman, M. & Li, W.-H. Intragenic spatial patterns
772 of codon usage bias in prokaryotic and eukaryotic genomes. *Genetics* **168**, 2245-2260
773 (2004).
774 55 Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of
775 gene expression in *Escherichia coli*. *Science* **324**, 255-258 (2009).
776 56 Agashe, D., Martinez-Gomez, N. C., Drummond, D. A. & Marx, C. J. Good codons, bad
777 transcript: large reductions in gene expression and fitness arising from synonymous
778 mutations in a key enzyme. *Mol Biol Evol* **30**, 549-560 (2013).
779 57 Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous
780 mutations to human disease. *Nat Rev Genetics* **12**, 683-691 (2011).
781 58 McCutcheon, J. P., McDonald, B. R. & Moran, N. A. Origin of an alternative genetic code
782 in the extremely small and GC–rich genome of a bacterial symbiont. *PLoS Genet* **5**
783 (2009).
784 59 Knight, R. D., Freeland, S. J. & Landweber, L. F. A simple model based on mutation and
785 selection explains trends in codon and amino-acid usage and GC composition within and
786 across genomes. *Genome Biol* **2**, 1-13 (2001).
787 60 Chen, S. L., Lee, W., Hottes, A. K., Shapiro, L. & McAdams, H. H. Codon usage between
788 genomes is constrained by genome-wide mutational processes. *PNAS* **101**, 3480-3485
789 (2004).
790 61 Hurst, G. D., Johnson, A. P., Schulenburg, J. H. & Fuyama, Y. Male-killing *Wolbachia* in
791 *Drosophila*: a temperature-sensitive trait with a threshold bacterial density. *Genetics*
792 **156**, 699-709 (2000).
793 62 Sheeley, S. & McAllister, B. Mobile male-killer: similar *Wolbachia* strains kill males of
794 divergent *Drosophila* hosts. *Heredity* **102**, 286 (2009).
795 63 Hornett, E. A., Duplouy, A. M., Davies, N., Roderick, G. K., Wedell, N., Hurst, G. D. &
796 Charlat, S. You can't keep a good parasite down: evolution of a male-killer suppressor
797 uncovers cytoplasmic incompatibility. *Evolution* **62**, 1258-1263 (2008).
798 64 Veneti, Z., Zabalou, S., Papafotiou, G., Paraskevopoulos, C., Pattas, S., Livadaras, I.,
799 Markakis, G., Herren, J. K., Jaenike, J. & Bourtzis, K. Loss of reproductive parasitism
800 following transfer of male-killing *Wolbachia* to *Drosophila melanogaster* and *Drosophila*
801 *simulans*. *Heredity* **109**, 306-312 (2012).
802 65 Hughes, G. L. & Rasgon, J. L. Transinfection: a method to investigate *Wolbachia*-host
803 interactions and control arthropod-borne disease. *Insect Mol Biol* **23**, 141-151 (2014).
804 66 Lucchesi, J. C. & Kuroda, M. I. Dosage compensation in *Drosophila*. *CSH Perspect Biol* **7**,
805 a019398 (2015).

806 67 Marın, I. & Baker, B. S. The evolutionary dynamics of sex determination. *Science* **281**,
807 1990-1994 (1998).
808 68 Rodriguez, M. A., Vermaak, D., Bayes, J. J. & Malik, H. S. Species-specific positive
809 selection of the male-specific lethal complex that participates in dosage compensation
810 in *Drosophila*. *PNAS* **104**, 15412-15417 (2007).
811 69 Fialho, R. F. & Stevens, L. Male-killing *Wolbachia* in a flour beetle. *Proc Royal Soc B* **267**,
812 1469-1473 (2000).
813 70 Van Borm, S., Wenseleers, T., Billen, J. & Boomsma, J. J. *Wolbachia* in leafcutter ants: a
814 widespread symbiont that may induce male killing or incompatible matings. *J Evol Biol*
815 **14**, 805-814 (2001).
816 71 Ferguson, S. B., Blundon, M. A., Klovstad, M. S. & Schupbach, T. Modulation of gurken
817 translation by insulin and TOR signaling in *Drosophila*. *J Cell Sci* **125**, 1407-1419 (2012).
818 72 Groth, A. C., Fish, M., Nusse, R. & Calos, M. P. Construction of transgenic *Drosophila* by
819 using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775-1782 (2004).
820 73 Southall, T. D., Elliott, D. A. & Brand, A. H. The GAL4 system: A versatile toolkit for gene
821 expression in *Drosophila*. *CSH Protoc* **7**, pdb-top49 (2008).
822 74 Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The vienna RNA
823 websuite. *Nucleic Acids Res* **36**, W70-W74 (2008).
824 75 Lorenz, R/. Bernhart, S. H., Zu Siederdissen, C. H., Tafer, H., Flamm, C., Stadler, P. F. &
825 Hofacker, I. L. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).
826 76 Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new
827 heuristics and parallel computing. *Nat Methods* **9**, 772 (2012).
828 77 Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large
829 phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704 (2003).
830 78 Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees.
831 *Bioinformatics* **17**, 754-755 (2001).
832 79 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under
833 mixed models. *Bioinformatics* **19**, 1572-1574 (2003).
834 80 Jukes, T. H. & Cantor, C. R. Evolution of protein molecules. *Mammal Prot Metabol* **3**, 21-
835 132 (1969).
836

837