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

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

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

[b]Vanderbilt Microbiome Initiative, Vanderbilt University, Nashville, Tennessee, USA

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

[d]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* gene *wmk*. Here, we characterize the genotype-phenotype landscape of *wmk*-induced male killing in *D. melanogaster*. 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 and, notably, one synonymous mutation in *wmk* can ablate killing. These findings reveal previously unrecognized relationships of *wmk*-induced killing and establish new hypotheses for the impacts of post-transcriptional processes in *wmk*-induced male killing. We conclude that single synonymous sequence changes are not necessarily silent in important nested symbiotic interactions.

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 the reduced competition from 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].

We recently identified a male-killing candidate gene, *wmk* (*WO-mediated killing*), in the eukaryotic association module of prophage WO in *Wolbachia* based on comparative genomic, transgenic, and cytological approaches[18,19]. *wmk* is a putative transcription factor with two predicted helix-turn-helix (HTH) DNA-binding domains. Transgenic expression recapitulates many aspects of male killing in *Drosophila melanogaster* embryos. This includes a female-biased sex ratio due to selective death of males, and chromatin defects such as DNA damage, chromatin bridging, and formation of pyknotic nuclei similar to wild type male killing in *D. bifasciata*[19-21]. Additionally, transgenic *wmk* expression in flies results in DNA damage at sites of dosage compensation activity involved in *D. melanogaster* male-specific gene expression[19,20].

Indeed, *wmk* is a top candidate for male killing based on functional and comparative genomic analyses, but several key questions remain. Importantly, what are the genotype-phenotype relationships that underpin *wmk* function? Do *wmk* homologs from different prophage WO variants and *Wolbachia* strains cause male-killing? To address these questions here, we transgenically expressed *wmk* homologs with varying degrees of sequence divergence, ultimately testing variants that differ by as little as one synonymous nucleotide change. We also tested homologs with predicted alternative start codons to determine their impact on phenotype. The homologs are derived from a set of male-killing and non-male-killing strains and are tested using the Gal4-UAS *Drosophila melanogaster* expression system and assays for reproductive parasitism[19,22,23].

We hypothesized that since male killing is often suppressed by host background[14,15,24] and transgenic phenotypes can be influenced by alternative start codons[22], adaptation between the host background and *wmk* homolog will be important for expression of male killing. Related, phylogenetically similar *wmk* sequences will cause male lethality in the same host, while distantly-related variants will not. Notably, *D. melanogaster* does not harbor *Wolbachia* known to cause male killing in contexts tested thus far, though it does naturally harbor a strain closely-related to a natural male killer[25], and some individuals contain male-killing *Spiroplasma* symbionts whose plasmid gene *Spaid* can transgenically recapitulate male killing[26]. The tripartite *D. melanogaster*-*Wolbachia*-phage WO transgene symbiosis is the most genetically tractable system to date for deciphering the fine-scale genetics and mechanism of *Wolbachia*-induced male killing.

To address the hypothesis that genetic distance negatively associates with male-killing, we transgenically and bioinformatically evaluated *wmk* homologs that span a range of genetic divergence (Fig. 1). In particular, we hypothesized that the Wmk protein from the male-killing

*w*Rec strain of *D. recens* would cause a sex ratio bias owing to its single amino acid difference relative to the previously-tested Wmk from prophage WO in *w*Mel of *D. melanogaster*[19]. The *w*Rec strain, which naturally causes cytoplasmic incompatibility (CI, death of offspring when infected males mate with uninfected females) in *D. recens*, causes male killing when transferred to a strain of its sister species, *D. subquinaria*[24]. We also tested variants with more divergent sequences, alternative start codons, and synonymous nucleotide changes that associate with male-killing expression. Here, we discover that not only does transgenic expression of *wmk* homologs from strains non-native to *D. melanogaster* result in complex outcomes, but single, synonymous, nucleotide changes remarkably ablate the phenotype. Thus, single to few synonymous changes and post-transcriptional processes play a previously-unrecognized role in controlling the genotype-phenotype relationships that underpin *wmk* male killing.

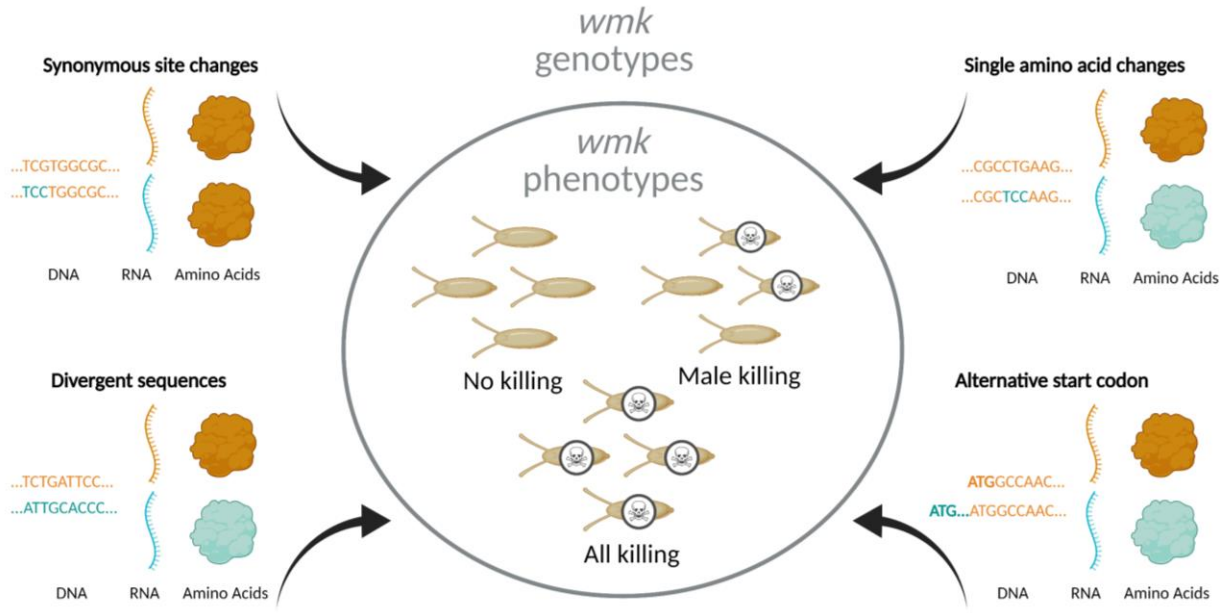

Figure 1. Overview of experiments in this paper. Expression of various *wmk* transgene sequences in *Drosophila melanogaster* embryos results in three different phenotypes. Compared to *w*Mel *wmk*, these transgenes were either divergent homologs from other *Wolbachia* strains, a homolog with a single amino acid change, homologs with an additional 9 codons at the 5' ends of the genes starting at an alternative upstream start codon, or a variant with a single synonymous codon difference. These genotypes resulted in varying degrees of RNA sequence- and amino acid-level changes. Created with BioRender.com.

Results

Strains and* wmk *genes. To determine if *wmk* homologs from other 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[27], the *w*Bif strain from *Drosophila bifasciata* flies[21], the *w*Caub strain from *Cadra cautella* moths[28], and the *w*Inn strain from *D. innubila* flies[29] (same gene sequence as the *w*Bor male-killing strain from *D. borealis* flies[30]). These hosts span several species groups in the *Drosophila* genus, along with two from Order Lepidoptera (Fig. 2A). In addition, most harbor natural male-killing *Wolbachia*. Further, some of the native *wmk* homologs present in prophage WO of the *Wolbachia* strains cluster into expected host species groups except for those in the quinaria group, supporting the hypothesis of *wmk*-host specialization (Fig. 2B). Moreover, the codon-optimized *wmk* transgenes also follow the phylogenetic structure of the native genes (Fig. 2C) and are roughly divided into two sequence groups of those distantly related to the *w*Mel *wmk* transgene (<90% codon-optimized nucleotide sequence similarity) and those closely-related (from *w*Rec, *w*Suzi, and an HA-tagged *w*Mel homolog), thus offering a range of allelic divergence for evaluating genotype-phenotype relationships (Fig. 2D).

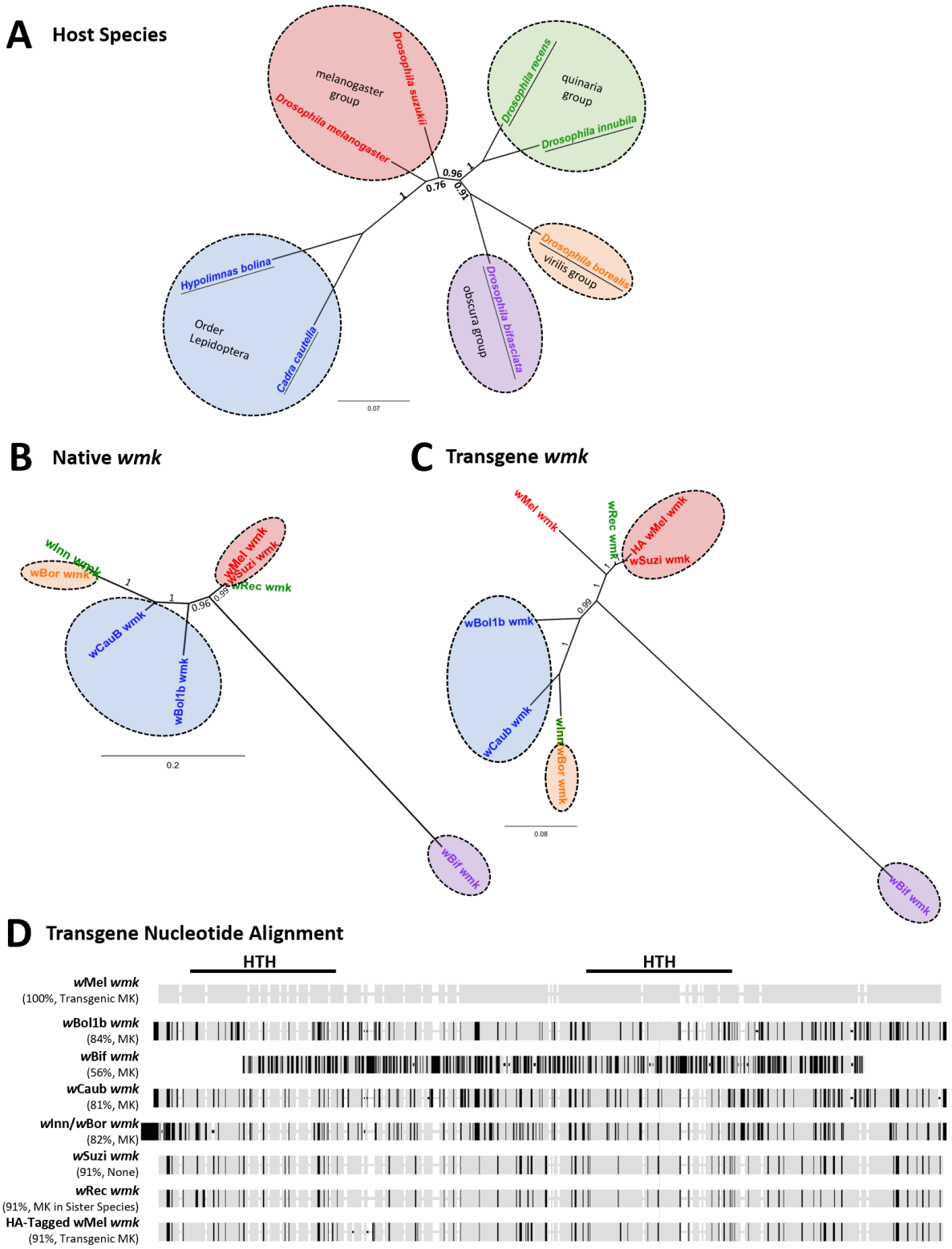

Figure 2. Homologs of *wmk* tested in this study include variation in native gene and transgene sequence identity as well as host species. (A) Bayesian nucleotide phylogeny of insect hosts based on 652 bp of the cytochrome oxidase subunit 1 (COI) gene from *D. melanogaster*, *H. bolina*,

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. (B) and (C) Bayesian nucleotide phylogenies of native (B, non-transgenic) or transgene (C) *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 share a branch because they share the same transgenic sequence. (D) 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.

Distantly-related homologs. 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 male-killing *wmk* homologs yielded a biased sex ratio, indicating that they do not recapitulate male killing when transgenically expressed in *D. melanogaster* under the conditions tested here (Fig. 3).

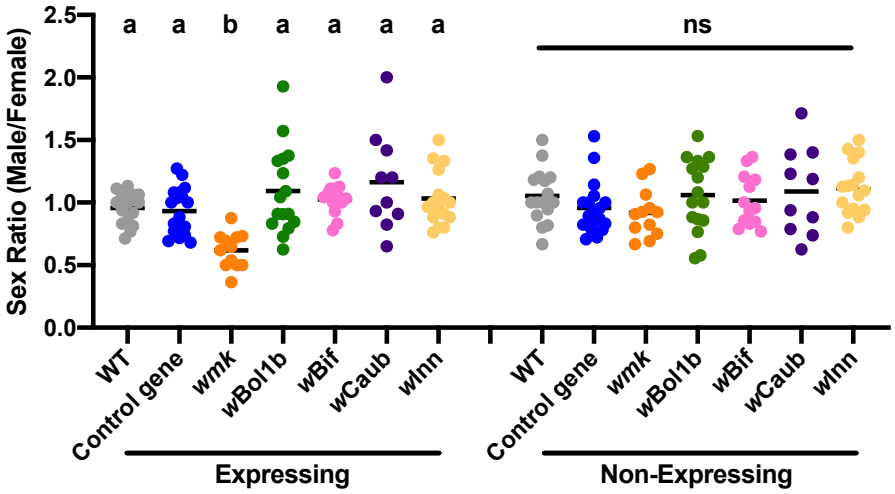

Figure 3. Divergent homologs of *wmk* from male-killing strains do not induce a biased sex ratio in *D. melanogaster*. Sex ratios of adult flies are shown from either expressing (*Act5c*-Gal4) or non-expressing (CyO) offspring. WT refers to the background insertion line and Control gene refers to the 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. Full statistical information is available in Figure 3 - Source Data 1.

Closely-related homologs. We tested *wmk* homologs from strains more similar to *w*Mel, including *w*Suzi of the pest species *D. suzukii* that has no confirmed reproductive phenotype, but does notably occur in populations with low male:female sex ratios[31,32], and *w*Rec of *D. recens* that kills males when introgressed into sister species, *D. subquinaria*[24]. The natural *w*Suzi homolog has one synonymous single nucleotide polymorphism (SNP) compared to the *w*Mel reference, thus yielding the same amino acid sequence. The natural *w*Rec homolog has three SNPs, one of which is non-synonymous relative to *w*Mel, located in the first HTH DNA-binding domain (Fig. 4A). All transgenes were codon optimized (as were those in Fig. 3), a process that produces the same amino acid sequences as the native genes, but also uses different sets of synonymous codons to optimize expression in the eukaryotic host (Fig. 2C). We hypothesized that both would induce a biased sex ratio comparable to *wmk* when transgenically expressed in *D. melanogaster*. However, transgenic expression of both homologs unexpectedly resulted in death of all expressing flies, both male and female (Fig. 4B). 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. It exhibited a sex ratio bias comparable to *w*Mel *wmk*, as expected (Fig. 4B). Therefore, two transgenes of *w*Mel *wmk* (one with a tag) resulted in the biased sex ratio previously reported[19], while two closely related transgenes yielded an all-killing phenotype.

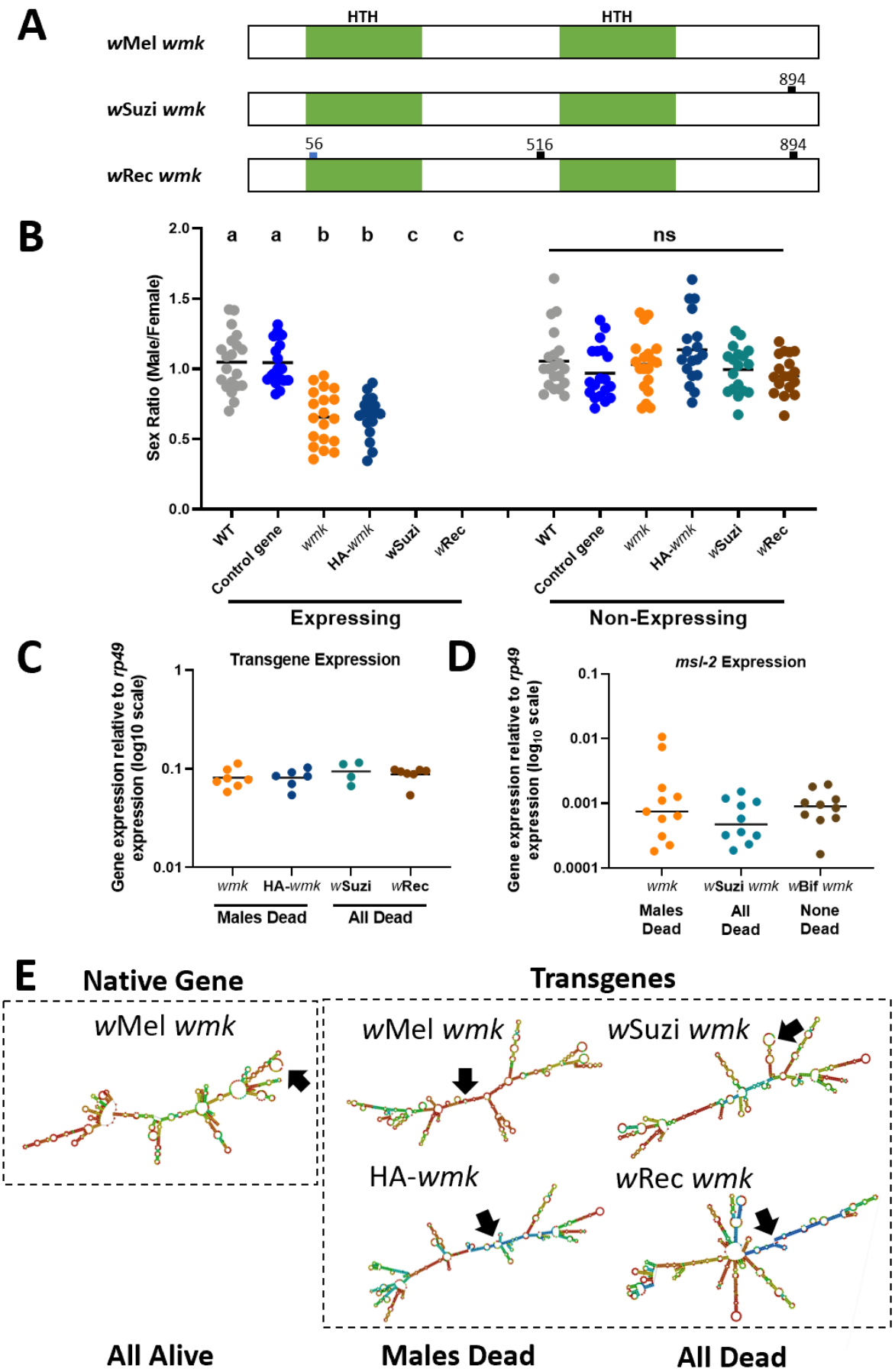
A
HTH
HTH
wMel wmk
wSuzi wmk
894
wRec wmk
56
516
894
B
Sex Ratio (Male/Female)
2.0
1.5
1.0
0.5
0.0
a
a
b
b
c
c
ns
WT
Control gene
wmk
HA-wmk
wSuzi
wRec
WT
Control gene
wmk
HA-wmk
wSuzi
wRec
Expressing
Non-Expressing
C
Transgene Expression
Gene expression relative to rp49 expression (log10 scale)
1
0.1
0.01
wmk
HA-wmk
wSuzi
wRec
Males Dead
All Dead
D
msl-2 Expression
Gene expression relative to rp49 expression (log10 scale)
0.1
0.01
0.001
0.0001
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

Figure 4. Closely-related *wmk* homologs exhibit several transgenic 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 have no points in the expressing category due to death of most or all males and females. HA-tagged *wmk* contains a 3X HA tag in the linker region between the two helix-turn-helix domains. This experiment was performed twice. Full statistical information is available in Figure 4 - Source Data 1. (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. Full statistical information is available in Figure 4 - Source Data 2. (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. Full statistical information is available in Figure 4 - Source Data 3. (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[19]. 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 phenotype differences (Fig. 4C). An alternative hypothesis is that expression of the *w*Suzi and *w*Rec homologs impact expression of a host gene in a way that is lethal to all hosts. Indeed, in *D. melanogaster* flies artificially expressing the *male-specific lethal 2* gene, *msl-2*, both males and females are susceptible to *Spiroplasma poulsonii*-induced killing (another male-killing symbiont)[33]. Based on this, we hypothesized that increased *msl-2* expression may underpin the *wmk*-induced non-sex-specific killing. We quantified *msl-2* transcript levels in embryos expressing *wmk* from *w*Mel (sex ratio bias), *w*Suzi (all expressing hosts die), and *w*Bif

(no killing or sex ratio bias). However, *msl-2* levels were comparable across groups of all genotypes and phenotypes, suggesting *msl-2* expression does not impact the phenotype of *w*Suzi *wmk*-expressing embryos (Fig. 4D).

RNA structural model variation. Next, we considered that secondary structural differences of the transgene transcripts that may contribute to the observed phenotypic variation. We modelled the mRNA structure of several transgenes and determined the RNA structures were substantially different, even in the case of few to no amino acid level changes from one protein to another. Moreover, some RNA structural features grouped by phenotype, such as location of the start codon (Fig. 4E). Although this is cautiously a predicted structural analysis, mRNA secondary structure may correlate with phenotype. Two transgenic transcripts that cause a sex ratio bias have start codons in the middle of the structure, while the other transgenes causing different phenotypes have start codons on outer loops or hairpins. However, native *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 transgenes. This difference may be related to native *wmk* expression by the bacteria instead of the transgenic eukaryotic host where we have tested the correlation, lower expression levels compared to the transgenes[19], and/or other structural differences. Thus, while speculative at this point, it is possible that transgene expression involving RNA structure variation and codon optimization contributes to the resulting differences in phenotype, which we test below.

Alternative start codon variation. Relevant to the work here, we previously provided evidence that transcriptional changes in *wmk* expression may contribute to differences in male-killing abilities between *Wolbachia* strains or hosts[22]. For instance, some strains contain alternative

start codons upstream of the annotated start for *wmk* and these upstream regions are expressed in the *w*Mel strain. When *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, establishing a hypothesis that differences in transcript length could significantly alter phenotype[22]. 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 seen 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. 5A). 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. 5B). This includes additional loops and different predicted placement of the start codon. Thus, we conclude that small or large *wmk* transcript sequence changes can lead to alterations in predicted RNA structure or possibly protein structure that may relate to altered phenotypes in *wmk* homologs.

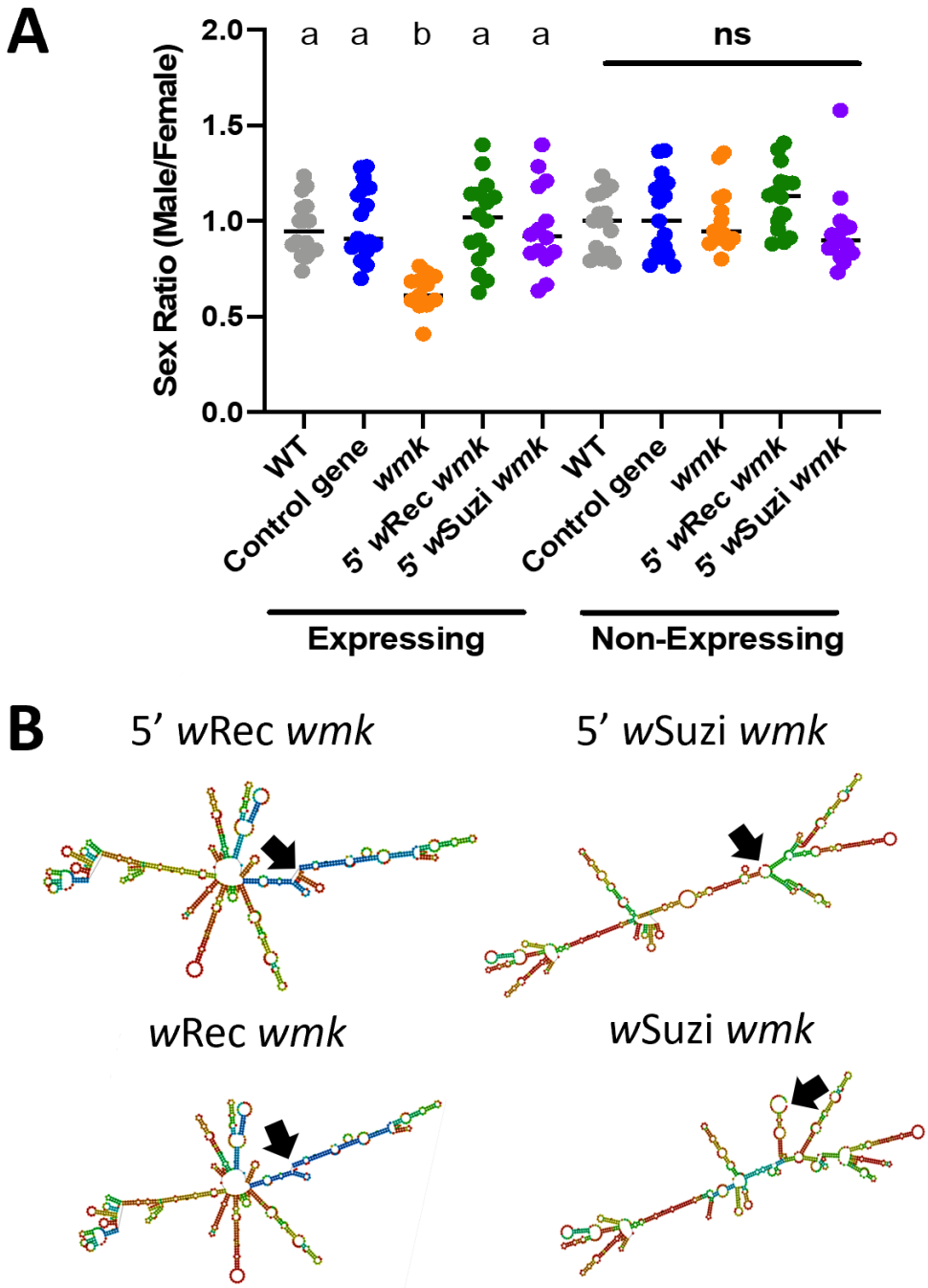

Figure 5. *w*Rec and *w*Suzi transgenes expressed with an alternative start codon lose their transgenic phenotypes. (A) Sex ratios of adult flies are shown for expressing (*Act5c*-Gal4) and non-expressing (CyO) offspring. Each sample point represents the adult offspring (N=50-120, mean 84) produced by 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. Full statistical information is available in Figure 5 - Source Data 1. (B) 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. 4 for ease of comparison). Black arrows point to the location of the start codon within each structure.

Silent site variation and phenotype. Finally, to identify particular nucleotides that may account for phenotypic variation among the homologs, we aligned the sequences of the four transgenes and looked for any codons that clustered by phenotype (sex ratio bias for *w*Mel *wmk* and HA tag *w*Mel *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[22] and since approximately the first ten codons in model prokaryotic genes are known to substantially affect mRNA structure and resulting translation rate[34], 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 tag and *wmk* transgene from *w*Mel, which recapitulate male killing, have a TCG codon, while the all-killing transgenes from *w*Suzi and *w*Rec have TCC and AGC, respectively (Fig. 6A).

To functionally test if the silent changes in this Serine codon accounted for the phenotype differences, we generated three new transgenes for phenotyping: (i) *w*Mel *wmk* control with no changes compared to the previously tested transgene (TCG codon, labelled "*w*Mel *wmk*"), (ii) *w*Mel *wmk* with three nucleotide changes in the codon that reconstitutes the AGC present in the *w*Rec transgene line (labelled "*w*Mel *wmk* with *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*Mel *wmk* with *w*Suzi codon"). When these three otherwise identical genes were transgenically expressed, the *wmk* 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. 6B). This ablation occurs despite the fact that transcript expression levels remain similar across all transgenes and despite sequencing confirmation of the single codon differences (Fig. 6C). 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 *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. 6D).

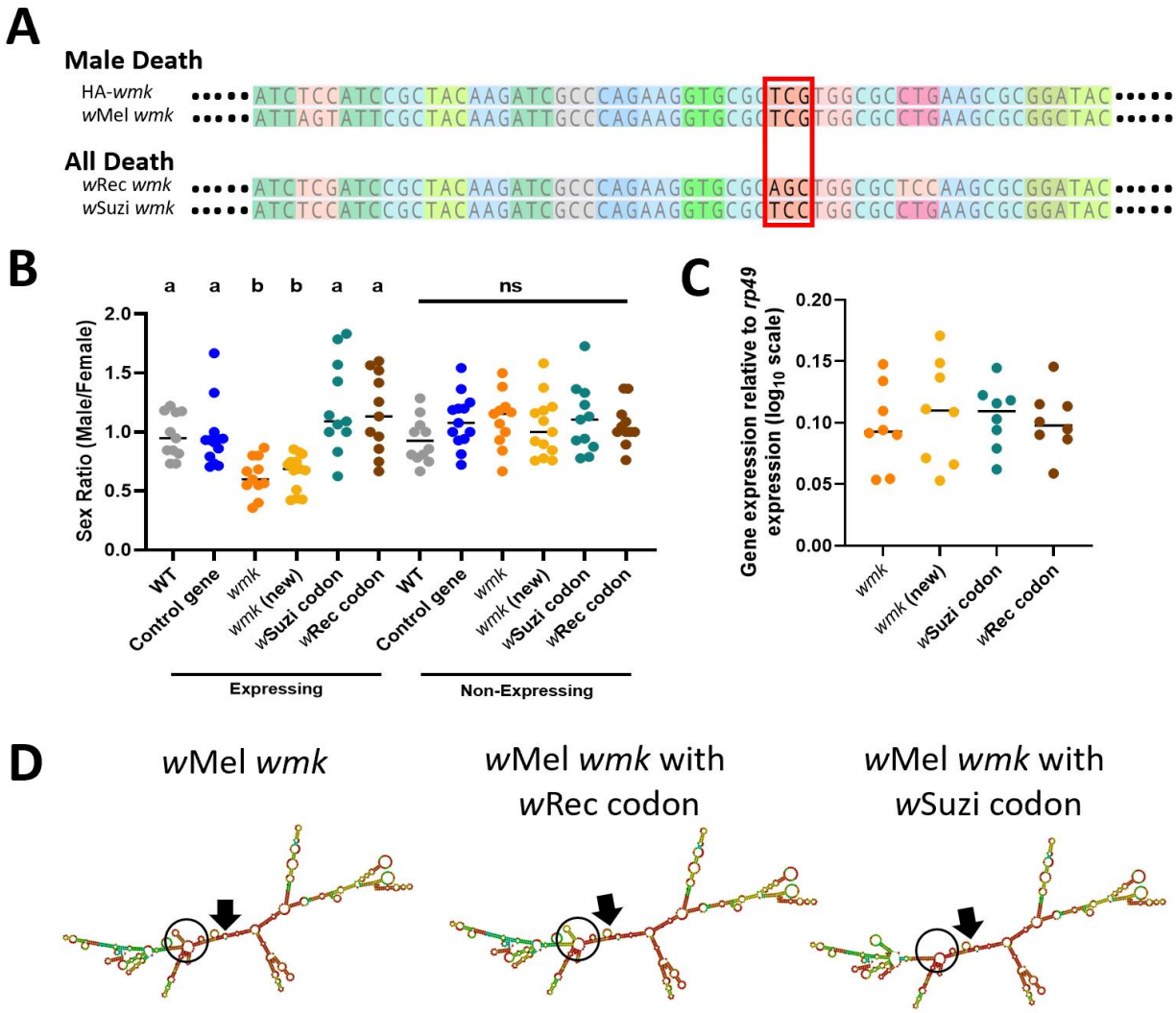

Figure 6. Synonymous nucleotide changes in the 16th codon position of *wmk* alters resulting phenotype. (A) Sequence alignment of genotypes used in the experiments below. The codon farthest on the left is the fourth codon in the sequence, and the highlighted codon is the 16th, with the farthest right 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. (B) Sex ratios of adult transgenic flies are shown for expressing (*Act5c*-Gal4) and non-expressing (CyO) offspring that include the original transgene *wmk* strain used in previous figures, along with the newly created identical *wmk* 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. Full statistical information is available in Figure 6 - Source Data 1. (C) Gene expression in embryos 4-5 h AED denotes expression of each transgene relative to that of *rp49*. Wmk (new) refers to the simultaneously created *w*Mel *wmk* control that has the exact same nucleotide sequence as the original. *w*Suzi codon and *w*Rec codon refer to the strains that have the same sequence as the *w*Mel *wmk*, but with the single codon at the 16th amino acid position changed. There is no significant difference in expression based on a Kruskal-Wallis one-way ANOVA followed by Dunn's correction. Full statistical information is available in Figure 6 - Source Data 2. (D) Predicted RNA secondary structures of the *w*Mel *wmk* transcript with either the *w*Rec or *w*Suzi codon exhibiting slight structural changes compared to the original structure (Fig. 4E). The structure for transgene *w*Mel *wmk* is included again from Fig. 4 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 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 structure is missing a predicted stem that the others have. The *w*Rec stem, while present, has a weaker prediction as noted by the cooler colors, so there may be structural differences compared to the *wMel* model.

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 sought to determine if and how genetic changes correlate with male killing expression using evolutionarily-related and distant homologs of the recently discovered phage *WO-mediated killing* gene, *wmk*[19,22]. Sequence divergence among *wmk* homologs can be minor or substantial across divergent hosts, leading to the hypothesis that highly divergent alleles are functionally fine-tuned to kill males in a variety of hosts. Moreover, some alleles of *wmk* have few mutations relative to the reference *w*Mel gene and offer an opportunity to investigate the impacts of small sequence changes on function that are less dependent on host divergence.

Using a variety of sequences from strains beyond the reference *w*Mel strain in *D. melanogaster*, we report three key results: (i) distantly-related homologs do not induce a male-

killing phenotype in *D. melanogaster*, while closely-related homologs from *w*Rec and *w*Suzi strains induce all-killing phenotypes, and *w*Mel and HA-tagged *w*Mel homologs kill only males, (ii) phenotypic variation associates with small changes in genotype that associate with RNA structure rather than amino acid identity, and (iii) single synonymous nucleotide changes presumed to be functionally silent in addition to alternative start codons ablate *wmk* male killing. These findings reveal that minor and silent nucleotide changes, rather than extensive amino acid divergence, markedly impact the genotype-phenotype relationships between *Wolbachia* and *D. melanogaster*. We discuss how these findings expand and support mechanistic models of *wmk* male killing, and how they relate to *Wolbachia* phenotype switching and host suppression of male killing that are common in this dynamic symbiosis.

Distantly-related homologs. We first analyzed *wmk* homologs that are more distantly related to *w*Mel, including from male-killing strains *w*Bol1b, *w*Bif, *w*Caub, and *w*Inn/*w*Bor. These homologs (all codon-optimized <90% nucleotide sequence identity compared to *w*Mel *wmk*) are from hosts that selectively kill males in their native host species[27-29,35,36]. 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 (and therefore the host and male-killing gene product are potentially in an arms race of adaptation and counter adaptation)[8,14,24,37,38], and some strains can cause male killing anew when transferred to another, usually closely-related host[24,28,39] while not causing male killing in all recipient host species[40,41]. It is possible, then, that the *wmk* homologs tested induce a sex ratio bias in their natural hosts, but not in *D. melanogaster*. Notably, *w*Bol1b and *w*Caub are both from lepidopteran hosts of ZW sex determination systems, and *wmk* has never

been tested functionally in a ZW host. The *w*Bif homolog has the lowest sequence identity compared to *w*Mel *wmk* and its divergent host is in the obscura group of the *Sophophora* subgenus of *Drosophila* as opposed to the melanogaster group tested here[42]. Similarly, the *w*Inn/*w*Bor strains, while more closely related to *w*Mel than others among this group of male killers, are also considerably divergent to *w*Mel. Notably, previous trans-infection experiments demonstrated that *w*Inn is not capable of male killing in *D. melanogaster*[40]. Therefore, these homologs were not expected to cause a sex ratio bias in this host if the co-adaptation hypothesis is true. Their lack of a sex ratio bias lends support to this hypothesis.

Closely-related homologs. We tested two closely-related homologs to that of *w*Mel: one from *w*Rec (2 synonymous and 1 non-synonymous nucleotide differences), and one from *w*Suzi (1 synonymous nucleotide difference). The *w*Rec homolog occurs in a *Wolbachia* that induces male killing in a sister host species, whereas it does not cause male killing in its native host[24]. We also tested the *wmk* homolog in *w*Suzi, which is not confirmed as a male killer, but is a closely related strain in an important pest species. Interestingly, there is a low abundance of males relative to females in some wild populations, raising the possibility of male killing in the population[32,43]. We expected both to produce a sex ratio bias when transgenically expressed due to *w*Suzi having the same amino acid sequence and *w*Rec having only one amino acid difference. However, we unexpectedly found that they killed all expressing hosts, both male and female. At the same time that we tested the *w*Rec and *w*Suzi homologs, we also tested a HA-tagged *w*Mel *wmk* strain that produced the same biased sex ratio as the original *w*Mel *wmk* strain. This sequence has a 91% sequence identity to *w*Mel, as do the *w*Rec and *w*Suzi *wmk* transgene strains. However, the three do not share the same exact nucleotide differences despite all having 91% identity overall, Fig. 2.

Therefore, two different sequences of *wmk* cause a sex ratio bias, two homologs kill all hosts, and several more homologs do not kill any hosts, with all sequences confirmed via gene sequencing.

***RNA structure*.** All genes went through the codon optimization algorithm at Genscript Biotech (Piscataway, NJ) to optimize expression in *D. melanogaster*, however, they were made several years apart. The original *w*Mel *wmk* strain was created about 4 years before the others, and the codon optimization algorithm changed in that time. In addition, codon optimization algorithms choose a codon based on factors including codon adaptation, mRNA folding, regulatory motifs, nucleotide bias, or codon correlations and biases[44]. With different codons input into the algorithm, and an updated algorithm, the new transgene strains had different nucleotide sequences. It is typically assumed that these synonymous differences will not influence the function of the protein since they code for the same amino acid. However, this assumption is false in this case given the results of the sex ratio assays.

Of particular note, the *w*Suzi *wmk* homolog (no killing) and the original *w*Mel *wmk* sequence (male killing) produce proteins of the exact same amino acid sequence, but share only a 91% sequence identity in the codon-optimized transgene. The comparison between these two in particular led to the next hypothesis that *wmk* phenotype may not be completely determined by amino acid sequence, but may be partially influenced by the RNA structure or other differences. To gain preliminary insight into the role of *wmk* RNA structure in resulting phenotypes, we input the sequences into modeling software to establish models of RNA secondary structures. Indeed, there were model differences from the sequences used here that correlated with phenotype, thus supporting the prediction that RNA structure rather than total amino acid divergence can be crucial in phenotypic outcome. We also measured gene expression and demonstrated that transgene

expression levels are similar and do not associate with phenotype, further supporting post-transcriptional differences as the source of functional variation.

Alternative start codon variation. In addition, we previously showed that transcript length differences change the *wmk* phenotype by expressing transgenes with an additional nine amino acids starting at a putative alternative upstream start codon[22]. Here, we tested the *w*Suzi and *w*Rec transgenes with these upstream start codons and found that phenotype was also ablated. Although the genes without the upstream sequence kill all hosts, adding a small sequence smaller than many common tags that would not be expected to interfere with function resulted in no death. This means that not only do small changes in transcript sequence affect the phenotype, but so do short increases in transcript length. We were unable to assess protein levels to determine translational differences due to lack of an antibody. However, the data demonstrates that *wmk* is functionally sensitive to post-transcriptional changes.

Silent site variation. We sought to assess the degree of sensitivity of *wmk* to post-transcriptional changes and to determine if codon differences definitively underpin phenotypic differences. To do so, we compared the transgene sequences of the two male killer strains to the two all-killing strains and identified candidate base-pairs where genotype and phenotype correlated. Across the entire length of the transgenes, only two such codons were identified. We chose to focus on the upstream codon position due to our previous work showing that modifications to the 5' region of the gene affects phenotype[22], and literature demonstrating the earliest codons can substantially influence mRNA structure and translation rate in model prokaryotes[34]. We created three new transgene strains: one that had the exact same sequence as the original *w*Mel *wmk* strain, and two more that were the same except for changing only the single codon to either the codon from the *w*Suzi or the *w*Rec homolog. All codons corresponded to a

serine, so the changes were synonymous. Expression of these three transgenes remarkably demonstrated that although the newly created *w*Mel *wmk* transgene strain still caused a sex ratio bias similar to the original strain, single synonymous codon changes from each of the other two transgene strains resulted in ablation of phenotype. However, the single codon changes did not recapitulate the all-killing phenotype from their originator transgene strains, which have many additional codon differences. According to the codon usage table for *D. melanogaster* on the Genscript website, the original *wmk* codon at this position (TCG) has a frequency per 1000 codons of 16.7/1000. 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[45]. Therefore, there is a pattern where the more common codons are present in the all-killing transgenes, and at similar frequencies, while the partial sex ratio bias strains have a codon at a lower frequency. Thus, codon usage correlates with phenotype in these examples.

A key question related to our findings is how do synonymous changes cause vastly different phenotypes. It is often assumed in transgene experiments that synonymous changes do not affect function, and codons for the same amino acid are functionally redundant. Several decades of research, however, have uncovered mounting evidence that the functional redundancy hypothesis is not so simple[44]. Indeed, codon bias varies across species (hence the need for codon optimization)[46,47], and there are several mechanisms through which 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[48]. 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[49]. Lower 5' RNA stability or weaker structure can also be correlated with ease of initiation and faster translation[50], and stem loop structures may affect translation rate in varying ways. Indeed, strong

5’ RNA secondary structure may inhibit ribosomal initiation or translational efficiency[51,52], and there can be selection against SNPs that alter RNA secondary structure[53]. 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 impact on mRNA structure near the site of initiation[50]. Additional evidence shows that 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[54,55]. In addition, some codons in different regions of the gene are favored for their lower likelihood of mistranslation[56]. Many of the above factors may explain our results here, namely why the *wmk* allele with the least frequently used codon in the 5’ region of the gene maintains a sex ratio bias, while alleles with more common codons have a different phenotype. These factors may also explain why the *w*Rec and *w*Suzi homologs killed all hosts, as they may have been folded or translated differently. More broadly beyond *wmk*, and due to many of the factors just described, synonymous differences in genes can result, for instance, in altered gene expression[57], lower organism fitness[58], or cause disease[59].

What, then, do these results illuminate about *wmk* function and the utility of transgenic research in light of the powerful influence of synonymous changes on phenotype? First, as discussed, *wmk* transgene phenotype appears highly susceptible to post-transcriptional mechanisms. At least three different phenotypes (sex ratio bias, all killing, no killing) correlate with only synonymous changes to the gene sequences. Transcript levels remain the same despite phenotype, and changes to either the length or the sequence of the transcript results in altered phenotypes. Second, if we continued to explore the relationship between codon sequence and phenotype of this gene, it is possible that we may identify one that produces a full male-killing phenotype instead of partial male killing. Effects of synonymous codons from codon optimization

could be an underlying reason why the phenotype is not complete in the transgenic system. In previous work, we also suggested that varying transcript lengths can affect phenotype, and here we expand this model. Indeed, few sequence differences in *wmk* could result in different RNA secondary structure and/or protein conformations that cause no sex ratio bias, a weak functional ability to kill males, or an enhanced ability to kill all hosts. Further, resistance to male killing is commonly reported[14,15,24], and if *wmk* function is easily altered with either synonymous or nonsynonymous changes, then this could contribute to the ease of resistance, and also to the phenotype changes from trans-infections between hosts.

If we make the hypothetical assumption that *wmk* is involved in male killing, then the results give the basis for additional hypotheses. First, the killing of both males and females by the *w*Rec and *w*Suzi 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 implicate dosage compensation due to correlation with DCC activity or timing of DNA defects in male embryos[19-21]. 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[60], so there are several non-sex-specific *wmk* target candidates. With small changes, transgenic or natural gene products may be altered by post-transcriptional mechanisms in such a way that they have lethal effects in females as well as males. Second, if the post-transcriptional differences are related to factors such as protein folding, then protein conformation is one potential mechanism for specificity between host target and toxin. As stated earlier, resistance to male killing is common and notably, some proteins involved in sex determination and dosage compensation are rapidly evolving[61,62]. So, small toxin protein conformation alterations, coupled with rapid evolution of potential host targets, could

explain both the ease of host resistance and specificity between host and toxin. 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[4,8,28,29,35,63-65].

Finally, these findings are informative with regards to the use of transgenes. It is standard practice to codon-optimize genes for maximized host expression when testing endosymbiont gene function[19,23,26], with the assumption that synonymous codons will not alter function. However, codon optimization may change the interpretation of results from transgenic experiments. Negative data may not indicate absolute proof that a gene does not have the function that was measured, and positive results may not be a guarantee of natural function without additional corroborating evidence. Investigations of native infections that serendipitously occur in genetically-tractable hosts, such as *Spiroplasma poulsonii* male killing and *Wolbachia* cytoplasmic incompatibility in *D. melanogaster*[23,26,66] will be the simplest cases. With *wmk* or any other non-native host or phenotypic contexts, transgenic function should eventually be corroborated in natural cases once the genetic tools are available in these non-model organisms. Regarding other transgenic work, caution is warranted with experimental design using codon optimization. For example, comparisons across alleles should be done when alleles are identical in sequence except for any deliberate differences. If there are sequence differences beyond these, it may affect the interpretation of the results. The reliability of certain calculations may be reduced in some cases, such as dN/dS as a proxy for positive selection, where mainly non-synonymous changes are assumed to be functionally consequential. These principles are particularly important for research on endosymbionts that increasingly relies on heterologous gene expression for functional studies. Importantly, the smaller genomes of endosymbionts tend to have low GC content overall[67], and

GC content is the strongest driver of codon usage bias[68,69] and contributes to the strength of mRNA secondary structure[44], so careful attention to the effects of synonymous changes may be of acute interest to the endosymbiont research community.

This 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 emphasizes why it may be particularly sensitive to synonymous nucleotide changes that may be dependent on host cell codon biases.

Materials and Methods

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[19], and several new transgene constructs. Briefly, these lines were constructed with codon optimization of sequences by GenScript Biotech (Piscataway, NJ) into the pTIGER pUASp-based vector for germline expression that integrates using PhiC31 integrase. The plasmids were then sent to BestGene (Chino Hills, CA), which performed injections of the vectors into the BDSC 8622 background line with transformants selected based on w^+ eye color.

D. melanogaster were reared on a standard cornmeal, molasses, and yeast (CMY) media. 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[19]. 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).

RNA secondary structures

RNA secondary structures were generated by uploading the nucleotide sequences of the indicated gene to the RNA fold web server[70,71]. 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), DNase treated with DNA-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 Table S1. 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[72,73]. The AICc-corrected best model, JC, was predicted for all three alignments and was used to build the trees using the MrBayes[74,75] Geneious plugin with the JC69 model[76] 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. 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 Initiative, and NIH grant F31 AI143152 to J.I.P. We would also like to thank R. L. Unckless for helpful comments on the manuscript.

Competing Interests

Both J.I.P. and S.R.B. are inventors on a patent related to the use of *wmk* in vector control.

References

1 Hurst, G. D. & Frost, C. L. Reproductive parasitism: maternally inherited symbionts in a biparental world. *Cold Spring Harbor perspectives in biology* **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, doi:10.1038/232657a0 (1971).

4 Hurst, G. D. D., Jiggins, F. M., Hinrich Graf von der Schulenburg, J., Bertrand, D., West, S. A., Goriacheva, I. I., Zakharov, I. A., Werren, J. H., Stouthamer, R. & Majerus, M. E. N. Male–killing Wolbachia in two species of insect. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **266**, 735-740, doi:10.1098/rspb.1999.0698 (1999).

5 Bouchon, D., Rigaud, T. & Juchault, P. Evidence for widespread Wolbachia infection in isopod crustaceans: molecular identification and host feminization. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **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. *Journal of Evolutionary Biology* **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, doi:10.3201/eid0604.000402 (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, doi:10.1111/j.1365-294X.2005.02678.x (2005).

9 Jaenike, J., Dyer, K. A. & Reed, L. K. Within-population structure of competition and the dynamics of male-killing Wolbachia. *Evolutionary Ecology Research* **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, doi:10.1111/j.1558-5646.2011.01485.x (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. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **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. *Proceedings. Biological sciences* **267**, 69-73, doi:10.1098/rspb.2000.0968 (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, doi:10.1371/journal.ppat.1000987 (2010).

15 Hornett, E. A., Charlat, S., Duplouy, A. M., Davies, N., Roderick, G. K., Wedell, N. & Hurst, G. D. Evolution of male-killer suppression in a natural population. *PLoS Biol* **4**, e283, doi:10.1371/journal.pbio.0040283 (2006).

16 Groenenboom, M. A. & Hogeweg, P. Space and the persistence of male–killing endosymbionts in insect populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**, 2509-2518 (2002).
17 Berec, L., Maxin, D. & Bernhauerova, V. Male-killing bacteria as agents of insect pest control. *Journal of Applied Ecology* (2016).
18 Bordenstein, S. R. & Bordenstein, S. R. Eukaryotic association module in phage WO genomes from Wolbachia. *Nature communications* **7**, 13155, doi:10.1038/ncomms13155 (2016).
19 Perlmutter, J. I., Bordenstein, S. R., Unckless, R. L., LePage, D. P., Metcalf, J. A., Hill, T., Martinez, J., Jiggins, F. M. & Bordenstein, S. R. The phage gene wmk is a candidate for male killing by a bacterial endosymbiont. *PLoS pathogens* **15**, e1007936 (2019).
20 Harumoto, T., Fukatsu, T. & Lemaitre, B. Common and unique strategies of male killing evolved in two distinct Drosophila symbionts. *Proceedings. Biological sciences* **285**, doi:10.1098/rspb.2017.2167 (2018).
21 Riparbelli, M. G., Giordano, R., Ueyama, M. & Callaini, G. Wolbachia-mediated male killing is associated with defective chromatin remodeling. *PLoS One* **7**, e30045, doi:10.1371/journal.pone.0030045 (2012).
22 Perlmutter, J. I., Meyers, J. E. & Bordenstein, S. R. Transgenic testing does not support a role for additional candidate genes in Wolbachia male killing or cytoplasmic incompatibility. *mSystems* **5**, e00658-00619, doi:10.1128/mSystems.00658-19 (2020).
23 LePage, D. P., Metcalf, J. A., Bordenstein, S. R., On, J., Perlmutter, J. I., Shropshire, J. D., Layton, E. M., Funkhouser-Jones, L. J., Beckmann, J. F. & Bordenstein, S. R. Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility. *Nature* **543**, 243-247, doi:10.1038/nature21391 (2017).
24 Jaenike, J. Spontaneous emergence of a new Wolbachia phenotype. *Evolution* **61**, 2244-2252, doi:10.1111/j.1558-5646.2007.00180.x (2007).
25 Metcalf, J. A., Jo, M., Bordenstein, S. R., Jaenike, J. & Bordenstein, S. R. Recent genome reduction of Wolbachia in Drosophila recens targets phage WO and narrows candidates for reproductive parasitism. *PeerJ* **2**, e529, doi:10.7717/peerj.529 (2014).
26 Harumoto, T. & Lemaitre, B. Male-killing toxin in a bacterial symbiont of Drosophila. *Nature* **557**, 252 (2018).
27 Dyson, E., Kamath, M. & Hurst, G. Wolbachia infection associated with all-female broods in Hypolimnas bolina (Lepidoptera: Nymphalidae): evidence for horizontal transmission of a butterfly male killer. *Heredity* **88**, 166 (2002).
28 Sasaki, T., Kubo, T. & Ishikawa, H. 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**, 1313-1319 (2002).
29 Dyer, K. A. & Jaenike, J. Evolutionarily stable infection by a male-killing endosymbiont in Drosophila innubila: molecular evidence from the host and parasite genomes. *Genetics* **168**, 1443-1455, doi:10.1534/genetics.104.027854 (2004).
30 Carson, H. A female producing strain of D. borealis Patterson. *Drosophila Information Service* **30**, 109-110 (1956).
31 Hamm, C. A., Begun, D. J., Vo, A., Smith, C. C., Saelao, P., Shaver, A. O., Jaenike, J. & Turelli, M. Wolbachia do not live by reproductive manipulation alone: infection

polymorphism in Drosophila suzukii and D. subpulchrella. *Molecular ecology* **23**, 4871-4885 (2014).

32 Drummond, F., Ballman, E. & Collins, J. Population dynamics of spotted wing Drosophila (Drosophila suzukii (Matsumura)) in Maine wild blueberry (Vaccinium angustifolium Aiton). *Insects* **10**, 205 (2019).

33 Cheng, B., Kuppanda, N., Aldrich, J. C., Akbari, O. S. & Ferree, P. M. Male-Killing Spiroplasma Alters Behavior of the Dosage Compensation Complex during Drosophila melanogaster Embryogenesis. *Current biology : CB* **26**, 1339-1345, doi:10.1016/j.cub.2016.03.050 (2016).

34 Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z. & Blüthgen, N. Efficient translation initiation dictates codon usage at gene start. *Molecular systems biology* **9** (2013).

35 Hurst, G. D., Johnson, A. P., Schulenburg, J. H. & Fuyama, Y. Male-killing Wolbachia in Drosophila: a temperature-sensitive trait with a threshold bacterial density. *Genetics* **156**, 699-709 (2000).

36 Sheeley, S. & McAllister, B. Mobile male-killer: similar Wolbachia strains kill males of divergent Drosophila hosts. *Heredity* **102**, 286 (2009).

37 Mitsuhashi, W., Ikeda, H. & Muraji, M. Fifty-year trend towards suppression of Wolbachia-induced male-killing by its butterfly host, Hypolimnas bolina. *Journal of insect science (Online)* **11**, 92, doi:10.1673/031.011.9201 (2011).

38 Hornett, E. A., Duplouy, A. M., Davies, N., Roderick, G. K., Wedell, N., Hurst, G. D. & Charlat, S. You can't keep a good parasite down: evolution of a male-killer suppressor uncovers cytoplasmic incompatibility. *Evolution* **62**, 1258-1263, doi:10.1111/j.1558-5646.2008.00353.x (2008).

39 Fujii, Y., Kageyama, D., Hoshizaki, S., Ishikawa, H. & Sasaki, T. Transfection of Wolbachia in Lepidoptera: the feminizer of the adzuki bean borer Ostrinia scapulalis causes male killing in the Mediterranean flour moth Ephestia kuehniella. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**, 855-859 (2001).

40 Veneti, Z., Zabalou, S., Papafotiou, G., Paraskevopoulos, C., Pattas, S., Livadaras, I., Markakis, G., Herren, J. K., Jaenike, J. & Bourtzis, K. Loss of reproductive parasitism following transfer of male-killing Wolbachia to Drosophila melanogaster and Drosophila simulans. *Heredity* **109**, 306-312, doi:10.1038/hdy.2012.43 (2012).

41 Hughes, G. L. & Rasgon, J. L. Transinfection: a method to investigate Wolbachia-host interactions and control arthropod-borne disease. *Insect Mol Biol* **23**, 141-151, doi:10.1111/imb.12066 (2014).

42 Huey, R. B., Moreteau, B., Moreteau, J.-C., Gibert, P., Gilchrist, G. W., Ives, A. R., Garland Jr, T. & David, J. R. Sexual size dimorphism in a Drosophila clade, the D. obscura group. *Zoology* **109**, 318-330 (2006).

43 Walsh, D. B., Bolda, M. P., Goodhue, R. E., Dreves, A. J., Lee, J., Bruck, D. J., Walton, V. M., O'Neal, S. D. & Zalom, F. G. Drosophila suzukii (Diptera: Drosophilidae): invasive pest of ripening soft fruit expanding its geographic range and damage potential. *Journal of Integrated Pest Management* **2**, G1-G7 (2011).

44 Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics* **12**, 32-42 (2011).

45 GenScript. *GenScript Codon Usage Frequency Table (chart) Tool*, <https://www.genscript.com/tools/codon-frequency-table> (2020).

46 Sharp, P. M., Averof, M., Lloyd, A. T., Matassi, G. & Peden, J. F. DNA sequence evolution: the sounds of silence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **349**, 241-247 (1995).
47 Andersson, S. & Kurland, C. Codon preferences in free-living microorganisms. *Microbiology and Molecular Biology Reviews* **54**, 198-210 (1990).
48 Li, G.-W., Oh, E. & Weissman, J. S. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538-541 (2012).
49 Goodman, D. B., Church, G. M. & Kosuri, S. Causes and effects of N-terminal codon bias in bacterial genes. *Science (New York, N.Y.)* **342**, 475-479 (2013).
50 Gu, W., Zhou, T. & Wilke, C. O. A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS computational biology* **6** (2010).
51 de Smit, M. H. & Van Duin, J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proceedings of the National Academy of Sciences* **87**, 7668-7672 (1990).
52 Bettany, A. J., Moore, P. A., Cafferkey, R., Bell, L. D., Goodey, A. R., Carter, B. L. & Brown, A. J. 5′-Secondary structure formation, in constrast to a short string of non-preferred codons, inhibits the translation of the pyruvate kinase mRNA in yeast. *Yeast* **5**, 187-198 (1989).
53 Salari, R., Kimchi-Sarfaty, C., Gottesman, M. M. & Przytycka, T. M. Sensitive measurement of single-nucleotide polymorphism-induced changes of RNA conformation: application to disease studies. *Nucleic acids research* **41**, 44-53 (2013).
54 Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I. & Pilpel, Y. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344-354 (2010).
55 Fredrick, K. & Ibba, M. How the sequence of a gene can tune its translation. *Cell* **141**, 227-229 (2010).
56 Qin, H., Wu, W. B., Comeron, J. M., Kreitman, M. & Li, W.-H. Intragenic spatial patterns of codon usage bias in prokaryotic and eukaryotic genomes. *Genetics* **168**, 2245-2260 (2004).
57 Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in Escherichia coli. *Science (New York, N.Y.)* **324**, 255-258 (2009).
58 Agashe, D., Martinez-Gomez, N. C., Drummond, D. A. & Marx, C. J. Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. *Molecular biology and evolution* **30**, 549-560 (2013).
59 Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nature Reviews Genetics* **12**, 683-691 (2011).
60 Lucchesi, J. C. & Kuroda, M. I. Dosage compensation in Drosophila. *Cold Spring Harbor perspectives in biology* **7**, a019398 (2015).
61 Marın, I. & Baker, B. S. The evolutionary dynamics of sex determination. *Science (New York, N.Y.)* **281**, 1990-1994 (1998).
62 Rodriguez, M. A., Vermaak, D., Bayes, J. J. & Malik, H. S. Species-specific positive selection of the male-specific lethal complex that participates in dosage compensation in Drosophila. *Proceedings of the National Academy of Sciences* **104**, 15412-15417 (2007).
63 Fialho, R. F. & Stevens, L. Male-killing Wolbachia in a flour beetle. *Proceedings. Biological sciences* **267**, 1469-1473, doi:10.1098/rspb.2000.1166 (2000).

64 Zeh, D. W., Zeh, J. A. & Bonilla, M. M. Wolbachia, sex ratio bias and apparent male killing in the harlequin beetle riding pseudoscorpion. *Heredity* **95**, 41-49, doi:10.1038/sj.hdy.6800666 (2005).
65 Van Borm, S., Wenseleers, T., Billen, J. & Boomsma, J. J. Wolbachia in leafcutter ants: a widespread symbiont that may induce male killing or incompatible matings. *Journal of Evolutionary Biology* **14**, 805-814 (2001).
66 Beckmann, J. F., Ronau, J. A. & Hochstrasser, M. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nature microbiology* **2**, 17007 (2017).
67 McCutcheon, J. P., McDonald, B. R. & Moran, N. A. Origin of an alternative genetic code in the extremely small and GC–rich genome of a bacterial symbiont. *PLoS genetics* **5** (2009).
68 Knight, R. D., Freeland, S. J. & Landweber, L. F. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome biology* **2**, research0010. 0011 (2001).
69 Chen, S. L., Lee, W., Hottes, A. K., Shapiro, L. & McAdams, H. H. Codon usage between genomes is constrained by genome-wide mutational processes. *Proceedings of the National Academy of Sciences* **101**, 3480-3485 (2004).
70 Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The vienna RNA websuite. *Nucleic acids research* **36**, W70-W74 (2008).
71 Lorenz, R., Bernhart, S. H., Zu Siederdissen, C. H., Tafer, H., Flamm, C., Stadler, P. F. & Hofacker, I. L. ViennaRNA Package 2.0. *Algorithms for molecular biology* **6**, 26 (2011).
72 Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods* **9**, 772, doi:10.1038/nmeth.2109 (2012).
73 Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* **52**, 696-704 (2003).
74 Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* **17**, 754-755 (2001).
75 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)* **19**, 1572-1574 (2003).
76 Jukes, T. H. & Cantor, C. R. Evolution of protein molecules. *Mammalian protein metabolism* **3**, 21-132 (1969).

Supplementary Information

Table S1. Primers used in this study.

Primer Name	Primer Sequence
Rp49_F	CGGTTACGGATCGAACAAGC
Rp49_R	CTTGCGCTTCTTGGAGGAGA

wmk_homologs_opt_F	CTGTATGCCATTGCCGAGACCCT
wmk_homologs_opt_R	TCACCAGATCCTTGGCGATCTTCATC
Msl-2_F	GGATTAACGCGGTCTAAGCATGTGTAACTG
Msl-2_R	GTATGCCGTCTGGGCCATGATG